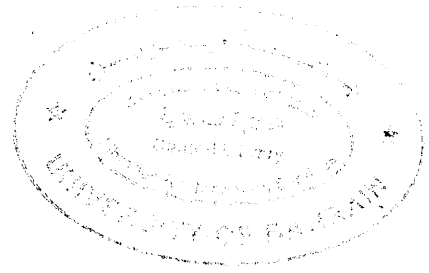


UNIVERSITY OF BAHRAIN



**IMPACTS OF MEASUREMENT ERRORS ON ESTIMATION
OF PARAMETERS IN LINEAR MODELS**

BY

HASSAN ABDULLA AL-BASRI

**A Thesis Submitted to the Department of Mathematics as
a Partial Fulfillment of the Requirements for the Degree
of
Master of Science**

Under the Supervision of

Dr. Moh'd Akkas Ali

April 1996



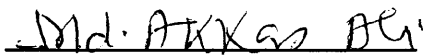
1 October 1996

The Thesis Defence Committee considers the Thesis titled:

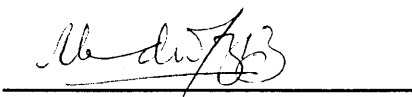
**"Impacts of Measurement Errors on Estimation of
Parameters in Linear Models"**

Submitted by *Hassan Al-Basri* to the Department of Mathematics
satisfactory and acceptable for the Master of Science Degree in
Mathematics.

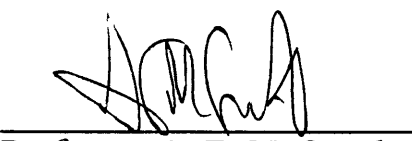
Thesis Defence Committee:


Dr. M. A. Akkas
Supervisor

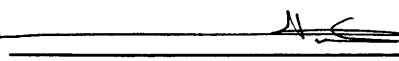
Department of Mathematics
University of Bahrain


Dr. A. A. Lahji
Member

Department of Mathematics
University of Bahrain


Professor A. F. M. Smith
Member

Department of Mathematics
Imperial College of Science
Technology & Medicine
London


Dr. A. S. A. Al-Hammadi
Acting Chairman

Department of Mathematics
University of Bahrain

Acknowledgment

By the grace of God, and kind instruction of Dr. Moh'd Akkas Ali, I have been successful in completing my thesis for the Degree of Master of Science.

My heartfelt thanks go to all concerned.

Contents

Acknowledgment	iii
Preface	vi
 Chapter I: General linear models, estimation and prediction	
1.1 Introduction	1
1.2 The basic paradigm	1
1.3 Model	3
1.4 Assumptions	4
1.5 Estimation	4
1.6 Roles of H in linear regression	6
1.7 Properties of the prediction matrix	8
 Chapter II: Analysis of residuals	
2.1 Introduction	16
2.2 Types of residuals	17
2.3 Correlation among residuals	20
2.4 Examination of residuals	24
 Chapter III: Influential observation	
3.1 Introduction	33
3.2 Definition of influential observations	34
3.3 The leverage	34
3.4 Measures of influence	36
3.5 Other measures of influence	39

Chapter IV: Errors of measurement in regression

4.1	Introduction	45
4.2	Errors in the independent variables of general linear model	47
4.3	Simultaneous estimation of linear functions of parameters when independent variables are subject to measurement errors	50
4.4	Errors in the dependent variable of a general linear model	53
4.5	Simultaneous estimation of linear functions of parameters when dependent variable is subject to measurement errors	55
4.6	Remarks	57

Chapter V: Errors of measurement and prior information in linear models

5.1	Introduction	59
5.2	Prior information and errors in the independent variables	59
5.3	Prior information and errors in the dependent variable	67

Chapter VI: Identification of location shift outliers

6.1	Introduction	72
6.2	Location shift outliers	74
6.3	Measures for identification of potential location shift outliers	77
6.4	Example	83

References	86
-------------------	----

Preface

Science is today becoming increasingly quantitative. Scientists are being confronted more and more with large amounts of numerical data gathered from their laboratories, field experiments and surveys. But the mere collecting and recording of data achieves nothing. Data that having been collected, must be analyzed and interpreted. Over the last several decades the linear regression model has surely become among the most widely employed quantitative tools of the applied social sciences and many of the physical sciences. In vast majority of regression analysis the least squares approach is used to fit a linear model for the data. The popularity of ordinary least squares is attributable to its low computational costs and its intuitive plausibility in a wide variety of circumstances. The coefficients of regression play important role in fitting a linear model for the data. The accuracy of the model depends on the accuracy of the estimation of the regression coefficients. The estimation of the parameters can be affected by several factors, such as influential observations, errors of measurement whether in the dependent variable or in independent variables, availability of prior information on the regression coefficients, and outliers. In this thesis an attempt is made to assess the magnitude of impacts of the above mentioned factors on the estimation of the coefficients of regression. This thesis consists of six chapters.

Chapter I deals with estimation of unknown regression coefficients and prediction of future observations. Also it studies the role and properties of the projection matrix.

Chapter II is devoted to the residuals. Residuals are used in many procedures designed to detect the disagreement between data and assumed model. By 1975 residuals analysis was widely regarded as an integral part of any regression problem. Types of residuals and the correlation between the residuals are discussed briefly. Also the graphical methods of examining the residuals in order to check the model are considered.

In chapter III we discuss the influential observations and describe the diagnostic measures used to identify the influential observations. Influential observations deserve further investigation, because they strongly influence the regression results. Influential data, of course, are not necessarily bad data points; they may contain some of the most interesting sample information. Only after such data points have been identified can their quality be assessed and appropriate action be taken.

Chapter IV studies the impacts of errors of measurement on regression analysis. The assumption that the dependent variable or the independent variables are measured without errors is very far from the fact, specially with recent rapid increase in data gathering projects in social and medical sciences. 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 regression coefficients are also considered.

In chapter V we discuss the impact of prior information on the errors of measurement. We show, in some sense, that the prior information reduce the effect of errors of measurement on the estimation of the regression coefficients.

Chapter VI is devoted to identification of location shift outliers. Three simple measures to identify potential location shift outliers have been proposed. An example is given to illustrate our procedures.

CHAPTER I

GENERAL LINEAR MODELS, ESTIMATION AND PREDICTION

1.1 Introduction

In any system in which variables change, it is of interest to examine the effects that some variables exert on others. There may in fact be a simple functional relationship between variables. However, often there exists a functional relationship which is too complicated to describe in simple terms. In this case we may wish to approximate to this functional relationship by some simple mathematical function, such as a polynomial, which contains the appropriate variables and which approximates to the true function over some limited ranges of the variables involved. By examining such a function we may be able to learn more about the underlying true relationship and the effect produced by changes in certain important variables.

Any method of fitting equations to data may be called regression. Such equations are valuable for at least two purposes; making predictions and judging the strength of relationships.

Because they provide a way of empirically identifying how a variable is affected by other variables, regression methods have become essential in a wide range of fields including social and biological sciences, engineering, medical and business.

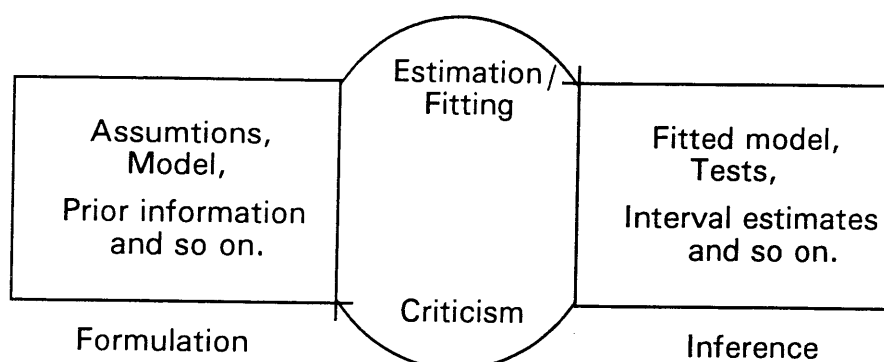
1.2 The Basic Paradigm

The paradigm that should be used in fitting regression models is illustrated in figure.

1.1. The paradigm is adopted from Cook and Weisberg (1982, P-7).

Figure 1.1

Schematic Outline of Model Fitting.



We begin with the left - hand square by formulating a problem of interest, deciding on an appropriate set of variables to study, choosing a functional form for a model and making certain assumptions. If Bayesian analysis is planned, we seek out prior information to aid in the process. The upper arrow pointing to the right corresponds to fitting the model to observed data, obtaining estimates, posterior distributions and so on. We call this the estimation step. In principle, the method actually used in the estimation step should depend on the assumptions in the left - hand square. More frequently, the method of fitting is determined by available software.

However, in vast majority of regression analysis, the least squares approach is used to carry out fitting. Whatever method of fitting is used, the right-hand square corresponds to the fitted model, including estimates, tests, confidence intervals and so forth.

The bottom arrow in figure. 1.1 is labeled criticism. It is meant to describe the act of critical assessment of the assumptions and the assumed model, given the fit in the right hand square and the actual data values. Criticism of a model may lead to modification of assumptions and thus further iteration through the system. The problems of the

estimation step, at least via least squares, are now easily and quickly solved. The analyst can consider inherently more complicated problems of criticism. Most of the methods for criticism (diagnostics) require computation of statistics that have values for each case in a data set, such as residuals and other related statistics.

1.3 Model

Regression analysis is a statistical tool for evaluating the relationship of one or more independent variables (or regressors) to a single continuous dependent variable. Usually, it is assumed that a certain type of relationship between the independent variables and the dependent variable holds. Further, it is assumed that ..

$$\text{Data} = \text{Model} + \text{Error}$$

The assumption of a particular error distribution helps us fit the specified model in optimum manner. We consider the general regression model

$$Y = X\beta + e \quad (1.3.1)$$

where;

Y is an $n \times 1$ vector of observations;

X is an $n \times k$ ($n > k$) matrix of predictors (independent variables or regressors) possibly included one constant predictor;

β is a $k \times 1$ vector of unknown coefficients (parameters) to be estimated; and e is an $n \times 1$ vector of random disturbances (errors).

1.4 Assumptions

The following assumptions are made to fit a postulated model to a given set of data :

(i) **Linearity assumption:** This assumption is implicit in the definition of model (1.3.1), which says that each observed response value Y_i , the i^{th} component of Y , can be written as a linear function of the i^{th} row of X, x'_i , that is,

$$Y_i = x'_i \beta + e_i \quad i = 1, 2, 3, \dots, n.$$

(ii) **Computational assumption:** In order to find a unique estimate of β it is necessary that $(X'X)^{-1}$ exists, or equivalently $\text{rank}(X) = k$.

(iii) **Distributional assumption:** The statistical analysis based on least squares (e.g. the t-test, the F-test, etc) assume that X is measured without errors, and e_i , the i^{th} component of e , does not depend on x'_i , $i = 1, 2, 3, \dots, n$ and

$$e \sim N(0, \sigma^2 I)$$

1.5 Estimation

The least squares estimator of β is obtained by minimizing

$e'e = (Y - X\beta)'(Y - X\beta)$. Taking the derivative of $e'e$ with respect to β yields the normal equations

$$(X'X)\beta = X'Y \quad (1.5.1)$$

The system of linear equations (1.5.1) has a unique solution if and only if $(X'X)^{-1}$ exists. In this case, premultiplying (1.5.1) by $(X'X)^{-1}$ gives the least squares estimator of β , that is

$$b = (X'X)^{-1} X'Y \quad (1.5.2)$$

If the assumptions in section (1.4) hold, the least squares theory provides us with the following well-known results (see, Graybill, 1976).

$$(a) \quad E(b) = \beta, E[(b-\beta)(b-\beta)'] = \sigma^2 (X'X)^{-1}$$

$$b \sim N[\beta, \sigma^2 (X'X)^{-1}]$$

(b) The $nx1$ vector of fitted (predicted) values,

$$\hat{Y} = Xb = X(X'X)^{-1} X'Y = HY \quad (1.5.3)$$

where $H = X (X'X)^{-1} X'$, has the following properties:

$$E(\hat{Y}) = X\beta, E[(\hat{Y} - X\beta)(\hat{Y} - X\beta)'] = \sigma^2 H,$$

$$\text{and } \hat{Y} \sim N(X\beta, \sigma^2 H) \quad (1.5.4)$$

(c) The $nx1$ vector of ordinary residuals,

$$\hat{e} = Y - \hat{Y} = Y - HY = (I - H) Y$$

has the following properties:

$$E(\hat{e}) = 0, E(\hat{e}\hat{e}') = \sigma^2 (I - H),$$

$$\hat{e} \sim N[0, \sigma^2 (I - H)], \quad (1.5.5)$$

and

$$(\hat{e}'\hat{e} / \sigma^2) \sim \chi^2_{(n-k)}$$

where $\chi^2_{(n-k)}$ denotes a χ^2 distribution with $(n-k)$ degrees of freedom and $\hat{e}'\hat{e}$ is the residual sum of squares.

(d) An unbiased estimator of σ^2 is given by

$$\hat{\sigma}^2 = \hat{\mathbf{e}}' \hat{\mathbf{e}} / (n-k) = \mathbf{Y}' (\mathbf{I} - \mathbf{H}) \mathbf{Y} / (n-k).$$

Note that the matrices $\mathbf{H}$ and $(\mathbf{I} - \mathbf{H})$ are singular and hence the normal distributions in (1.5.4) and (1.5.5) are singular.

1.6 Roles Of $\mathbf{H}$ In Linear Regression

The matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ (1.6.1)

determines many of the standard least squares results. This matrix plays an important role in linear regression analysis. It possesses several interesting properties. The matrix $\mathbf{H}$ is sometimes called the hat matrix because it maps $\mathbf{Y}$ into $\hat{\mathbf{Y}}$ or put a hat on $\mathbf{Y}$.

Chatterjee and Hadi (1986) call it the prediction matrix because it is the transformation matrix that, when applied to $\mathbf{Y}$, produces the predicted values.

Let h_{ij} be the ij th element of $\mathbf{H}$. then the i^{th} diagonal element of $\mathbf{H}$, is

$$h_{ii} = \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i, \quad i = 1, 2, 3, \dots, n. \quad (1.6.2)$$

From (1.5.4) and (1.5.5), we have $\text{var}(\hat{Y}_i) = \sigma^2 h_{ii}$,

and $\text{var}(\hat{e}_i) = \sigma^2 (1 - h_{ii})$. Thus, aside from a constant factor σ^2 , h_{ii} determines both

$\text{Var}(\hat{Y}_i)$ and $\text{Var}(\hat{e}_i)$. Here $\hat{Y}_i$ and $\hat{e}_i$ denote the i^{th} component of $\hat{\mathbf{Y}}$ and $\hat{\mathbf{e}}$ respectively.

The ij th element of $\mathbf{H}$ is

$$h_{ij} = \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_j, \quad i, j = 1, 2, 3, \dots, n. \quad (1.6.3)$$

Since
$$\hat{e} = (I - H) Y = (I - H) (X \beta + e)$$

$$= (I - H) e ,$$

hence the relationship between $\hat{e}$ and e depends only on H . Furthermore, from (1.5.3) it follows that the i^{th} predicted value $\hat{Y}_i$ can be written as

$$\hat{Y}_i = \sum_{j=1}^n h_{ij} Y_j = h_{ii} Y_i + \sum_{j \neq i} h_{ij} Y_j, \quad i = 1, 2, \dots, n. \quad (1.6.4)$$

Thus

$$\frac{\partial \hat{Y}_i}{\partial Y_i} = h_{ii}, \quad i = 1, 2, 3, \dots, n.$$

Therefore, h_{ii} can be interpreted as the amount of leverage each value Y_i has in determining $\hat{Y}_i$, regardless of the actual value of Y_i (Hoaglin and Welsch-1978).

Similarly, h_{ij} can be regarded as the amount of leverage each value Y_j has in determining $\hat{Y}_i$.

The reciprocal of h_{ii} can be thought of as the effective or equivalent number of observations that determine $\hat{Y}_i$ (Huber, 1981). That is, if $h_{ii} = 1$, $\hat{Y}_i$ is determined by Y_i alone. On the other hand, if $h_{ii} = 0$, Y_i has no influence on $\hat{Y}_i$. Behnken and Draper (1972) indicated that wide variation in h_{ii} reflects nonhomogeneous spacing of the rows of X .

1.7 Properties Of The Prediction Matrix

Property 1:

H is invariant under non-singular linear transformations of the form $X \rightarrow XE$, where E is any $k \times k$ non-singular matrix.

Proof: Let $Z = XE$, and

$$\begin{aligned} H_Z &= Z (Z'Z)^{-1} Z' \\ &= (XE) [E' X' X E]^{-1} E' X' \\ &= X E E^{-1} (X' X)^{-1} E^{-1} E' X' \\ &= X (X' X)^{-1} X' = H. \end{aligned}$$

Property 1 indicates that if E is non-singular then the prediction matrices for X and XE are the same.

Let E be any $k \times k$ non-singular matrix and let the prediction matrices for X and XE be denoted by H_X and H_{XE} respectively. By property 1 we have $H_{XE} = H_X$, and hence

$$\hat{e}_{Y:XE} = (I - H_{XE}) Y = (I - H_X) Y = \hat{e}_{Y:X}.$$

Thus if model (1.3.1) contains a constant term, the residuals and their estimated variances are invariant under scale and location transformation of X , while if (1.3.1) does not contain a constant, they are invariant only under scale transformation of X . Furthermore, property 1 allows for any nonsingular reparameterization of model (1.3.1). That is, if $\beta^* = E^{-1} \beta$, then the model

$Y = X\beta + e$ and $Y = XE\beta^* + e$ are equivalent in the sense of producing the same $\hat{Y}$.

Property 2:

(a) H and $(I - H)$ are symmetric and idempotent matrices.

(b) $\text{tr}(H) = k$, and $\text{tr}(I - H) = (n - k)$,

where $\text{tr}(G)$ denotes the trace of the square matrix G .

Proof: (a) Clearly $H = X(X'X)^{-1}X' = H'$, and

$$H^2 = H \cdot H = H.$$

Thus H is symmetric and idempotent. Similarly, it can be shown that $(I - H)$ is also symmetric and idempotent.

$$(b) \text{tr}(H) = \text{tr} [X (X'X)^{-1} X']$$

$$= \text{tr} [(X'X)^{-1} X' X]$$

$$= \text{tr} [I_k] = k$$

Here I_k denotes the k -dimensional identity matrix. the rest is obvious.

Property 3:

Let $X = (X_1|X_2)$, where X_1 is an $n \times r$ matrix of rank r and X_2 is an $n \times (k-r)$ matrix of rank $(k-r)$.

Let $H_1 = X_1(X_1'X_1)^{-1}X_1'$ be the projection matrix for X_1 , and $H_2 = W(W'W)^{-1}W'$ be the projection matrix for W , where $W = (I - H_1)X_2$. Then H can be expressed as

$$H = H_1 + H_2.$$

Proof: H can be written in the form

$$H = \begin{bmatrix} X_1 & X_2 \end{bmatrix} \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1} \begin{bmatrix} X_1' \\ X_2' \end{bmatrix}.$$

Now $\begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1} = \begin{bmatrix} (X_1'X_1)^{-1} + FMF' & -FM \\ -MF' & M \end{bmatrix},$

where $F = (X_1'X_1)^{-1} X_1'X_2$, and

$$\begin{aligned} M &= \left[X_2'X_2 - X_2'X_1(X_1'X_1)^{-1} X_1'X_2 \right]^{-1} \\ &= \left[X_2'(I - H_1)X_2 \right]^{-1}. \end{aligned}$$

Thus we have

$$\begin{aligned} H &= \begin{bmatrix} X_1 & X_2 \end{bmatrix} \begin{bmatrix} (X_1'X_1)^{-1} + FMF' & -FM \\ -MF' & M \end{bmatrix} \begin{bmatrix} X_1' \\ X_2' \end{bmatrix} \\ &= \begin{bmatrix} X_1 & X_2 \end{bmatrix} \begin{bmatrix} (X_1'X_1)^{-1} X_1' + FMF'X_1' - FMX_2' \\ -MF'X_1' + MX_2' \end{bmatrix} \\ &= H_1 + X_1 FMF'X_1' - X_1 FMX_2' - X_2 MF'X_1' + X_2 MX_2' \\ &= H_1 + X_1 FM(F'X_1' - X_2') - X_2 M(F'X_1' - X_2') \\ &= H_1 + (X_1 FM - X_2 M)(X_1 F - X_2)' \\ &= H_1 + (X_1 F - X_2) M (X_1 F - X_2)' \\ &= H_1 + (H_1 X_2 - X_2) M (H_1 X_2 - X_2)' \\ &= H_1 + (H_1 - I) X_2 M X_2' (H_1 - I)' \\ &= H_1 + (I - H_1) X_2 M X_2' (I - H_1)' \\ &= H_1 + W M W' = H_1 + W (W'W)^{-1} W' \end{aligned}$$

$= H_1 + H_2$. This completes the proof.

Property 4:

For $i = 1, 2, 3, \dots, n$ and $j = 1, 2, 3, \dots, n$,

we have

$$(a) \quad 0 \leq h_{ii} \leq 1 \quad \text{for all } i$$

$$(b) \quad -0.5 \leq h_{ij} \leq 0.5 \quad \text{for all } j \neq i$$

where h_{ij} denotes the ij th element of H .

Proof: Since $H = H^2$ and $H = (h_{ij})$,

$i = 1, 2, 3, \dots, n$ and $j = 1, 2, 3, \dots, n$. Therefore,

$$h_{ij} = \sum_{j=1}^n h_{ij}^2 = h_{ii}^2 + \sum_{j \neq i} h_{ij}^2 . \quad (1.7.1)$$

Thus, $h_{ii} \geq h_{ii}^2 \Rightarrow h_{ii} (1 - h_{ii}) \geq 0$.

Hence, $0 \leq h_{ii} \leq 1$, for all i . This proves part (a) . Identity (1.7.1) can also be expressed as

$$h_{ii} = h_{ii}^2 + h_{ij}^2 + \sum_{r \neq i, j} h_{ir}^2 .$$

Hence, $h_{ij}^2 \leq h_{ii} (1 - h_{ii})$, and since

$0 \leq h_{ii} \leq 1$, part (b) follows.

Now suppose that X contains a constant column, and let $X = (J \mid X_2)$, where J is n -vector of ones. From property 3 , we have

$$H_1 = J(J'J)^{-1}J' = \frac{1}{n}JJ' ,$$

and

$$W = (I - H_1)X_2 = \left(I - \frac{1}{n}JJ'\right)X_2 = \tilde{X} ;$$

then $H_2 = \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'$.

Thus

$$H = H_1 + H_2 = \frac{1}{n}JJ' + \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'.$$

Each of the diagonal elements of $\frac{1}{n}JJ'$ is equal to $\frac{1}{n}$ and since H_2 is a prediction matrix, by property 4 (a), its diagonal elements are non-negative, hence in this case

$$h_{ii} \geq \frac{1}{n} \quad \text{for all } i.$$

We observe that $\tilde{X}'J = 0$. Hence $H_2J = 0$.

Therefore, $HJ = H_1J = J$.

Property 5:

Let X be an $n \times k$ matrix of rank k . Then for fixed n , h_{ii} , $i = 1, 2, 3, \dots, n$ is nondecreasing in k .

Proof: From property 3 we have $H = H_1 + H_2$; where H_1 and H_2 are prediction matrices. By property 4 (a), the diagonal elements of H_1 and H_2 are non-negative. Therefore, for fixed n , h_{ii} , $i = 1, 2, 3, \dots, n$ is non-decreasing in k .

Property 6:

Let X be $n \times k$ matrix of rank k . Then, for fixed k , h_{ii} , $i = 1, 2, 3, \dots, n$ is nonincreasing in n .

Proof: Suppose we currently have $(n - 1)$ observations, denoted by $X_{(i)}$ for which the prediction matrix $H_{(i)}$ has been computed, and one additional observation x_i is now available.

Let

$$X = \begin{bmatrix} X_{(i)} \\ x_i' \end{bmatrix}, \text{ hence}$$

$$X'X = X_{(i)}'X_{(i)} + x_i x_i'.$$

Therefore

$$\begin{aligned} (X'X)^{-1} &= \left[X_{(i)}'X_{(i)} + x_i x_i' \right]^{-1} \\ &= \left[X_{(i)}'X_{(i)} \right]^{-1} - \frac{\left[X_{(i)}'X_{(i)} \right]^{-1} x_i x_i' \left[X_{(i)}'X_{(i)} \right]^{-1}}{1 + x_i' \left[X_{(i)}'X_{(i)} \right]^{-1} x_i}. \end{aligned}$$

See, Rao (1973, P. 33).

Hence

$$h_{ii} = h_{ii}(-i) - \frac{h_{ii}(-i)h_{ii}(-i)}{1 + h_{ii}(-i)}, \quad (1.7.2)$$

where

$$h_{ii} = x_i' (X'X)^{-1} x_i$$

$$h_{ii}(-i) = x_i' \left[X_{(i)}'X_{(i)} \right]^{-1} x_i.$$

From (1.7.2), we have

$$h_{ii} = \frac{h_{ii}(-i)}{1 + h_{ii}(-i)} \leq h_{ii}(-i)$$

This completes the proof.

Property 7:

Let λ_r and V_r , $r = 1, 2, 3, \dots, k$, be the eigenvalues and the corresponding normalized eigenvectors of $X'X$, respectively. If θ_{ir} is the angle between x_i and V_r , then

$$h_{ij} = \|x_i\| \cdot \|x_j\| \sum_{r=1}^k \lambda_r^{-1} \cos \theta_{ir} \cos \theta_{rj} ,$$

and hence

$$h_{ii} = (x_i' x_i) \sum_{r=1}^k \lambda_r^{-1} \cos^2 \theta_{ir}$$

Proof: Let Λ be the diagonal matrix containing the eigenvalues of $X'X$ on its diagonal. Further, let V be the corresponding matrix of normalized eigenvectors. Then by the spectral decomposition theorem (see Searle, 1982) we have

$$X'X = V\Lambda V', \text{ and } (X'X)^{-1} = V\Lambda^{-1}V'.$$

Therefore,

$$h_{ij} = x_i' V \Lambda^{-1} V' x_j = \sum_{r=1}^k \lambda_r^{-1} x_i' V_r V_r' x_j , \quad i = 1, 2, 3, \dots, n , \text{ and} \\ j = 1, 2, 3, \dots, n ,$$

where λ_r is an eigenvalue of $(X'X)$ and V_r is the corresponding normalized eigenvector.

Now, h_{ij} can be written as

$$h_{ij} = \|x_i\| \cdot \|x_j\| \sum_{r=1}^k \lambda_r^{-1} \frac{x_i' V_r}{\|x_i\| \cdot \|V_r\|} \cdot \frac{V_r' x_j}{\|x_j\| \cdot \|V_r\|} .$$

(Note that $\|V_r\| = 1$). Hence,

$$h_{ij} = \|x_i\| \cdot \|x_j\| \sum_{r=1}^k \lambda_r^{-1} \cos \theta_{ir} \cos \theta_{rj}.$$

Clearly, $h_{ii} = \|x_i\|^2 \sum_{r=1}^k \lambda_r^{-1} \cos^2 \theta_{ir}.$

If x_i and V_r have the same direction, then $\cos^2 \theta_{ir} = 1$ and if they are perpendicular, then $\cos^2 \theta_{ir} = 0$. Thus h_{ii} will be large if the following two conditions hold (Cook and Weisberg 1982, P.13).

- (i) $x_i'x_i$ is large relative to the other observations (i.e. x_i is far removed from the bulk of the other points in the data set), and
- (ii) x_i is substantially in the direction of an eigenvector corresponding to a small eigenvalue of $X'X$.

CHAPTER 2

ANALYSIS OF RESIDUALS

2.1 Introduction

The residuals are defined as the n differences $\hat{e}_i = Y_i - \hat{Y}_i$, $i = 1, 2, \dots, n$, where Y_i is an observation and $\hat{Y}_i$ is the corresponding fitted value obtained by use of the fitted regression equation. We can see from this definition that the residuals $\hat{e}_i$ 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. Residuals have, for a long time, helped statisticians with approaches to understanding data and models. Exploratory data analysis places a heavy emphasis on looking at and analyzing residuals. Many residual-using techniques are effective for exploration (Tukey, 1977). Usually, it is assumed that

$$\begin{aligned}\text{Data} &= \text{Model} + \text{Random Error} \\ &= \text{Fit} + \text{Residual}\end{aligned}$$

Moreover, in performing the regression analysis we make some additional assumptions about the errors. We assume that the errors are independent, have zero mean, and constant variance σ^2 , and follow a normal distribution. The last assumption is needed for tests and confidence intervals. Thus if our fitted model is correct, the residuals should exhibit tendencies that tend to confirm the assumptions we have made, or at least should not exhibit a denial of the assumptions. This latter idea is the

one that should be kept in mind when analysing the residuals. After our analysis of residuals we should be able to conclude either

- a) The assumptions appear to be violated (in a way that can be specified), or
- b) The assumptions do not appear to be violated.

2.2 Types Of Residuals

The ordinary residuals are scale dependent. Several transformation of the ordinary residuals have been suggested to overcome this shortcoming. We first consider a set of residuals that have distributions independent of the scale of measurement. These residuals are closely related to a location shift outlier model. Alternatively, the residuals can be transformed to have a selected covariance structure.

Internal Studentized Residuals

For many diagnostic procedure, it is useful to define a standardized version of the residuals that does not depend on σ^2 . In least squares regression, the internal studentized residuals are defined by

$$r_i = \frac{\hat{e}_i}{\{S^2(1-h_{ii})\}^{1/2}}, \quad i = 1, 2, 3, \dots, n, \quad (2.2.1)$$

where

$$S^2 = \frac{\sum_{i=1}^n \hat{e}_i^2}{(n-k)} \quad \text{is the residual mean square.}$$

This transformation of residuals has been studied by Srikantan (1961), Anscombe and Tukey (1963), Ellenberg (1973, 1976), Beckman and Trusell (1974), Prescott (1975) and many others.

Ellenberg (1973) obtained the distribution of standardized residuals assuming that $e \sim N(0, \sigma^2 I)$. He showed that the probability density function of r_i is

$$f(r_i) = \begin{cases} \frac{\Gamma(v + \frac{1}{2})}{\Gamma(v)\Gamma(\frac{1}{2})\sqrt{n-k}} (1 - \frac{r_i^2}{n-k})^{v-1} & \text{for } |r_i| \leq \sqrt{n-k} \\ 0 & \text{else where,} \end{cases} \quad (2.2.2)$$

where $v = (n - k - 1) / 2$. Hence $\frac{r_i^2}{(n-k)}$ follows a Beta distribution with parameters

$1/2$ and $(n - k - 1) / 2$. It follows from (2.2.2) that

$$E(r_i) = 0 \text{ and } \text{var}(r_i) = 1$$

The standardized residuals are used as replacements for the ordinary residuals in graphical procedures and test for outliers.

External Studentized Residuals

For studentized residuals, an estimator of σ^2 that is independent of $\hat{e}_i$ is required. Let

$$S^2(i) = \frac{\left[\sum_{j=1}^n \hat{e}_j^2 - \frac{\hat{e}_i^2}{(1 - h_{ii})} \right]}{(n - k - 1)}$$

$$= \frac{\left[\hat{\mathbf{e}}' \hat{\mathbf{e}} - \frac{\hat{e}_i^2}{(1 - h_{ii})} \right]}{(n - k - 1)}$$

where $\hat{\mathbf{e}} = (\hat{e}_1, \hat{e}_2, \dots, \hat{e}_n)'$ is the vector of residuals and h_{ii} is the i^{th} diagonal element of

$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$. It is well known that if $\mathbf{e} \sim N(0, \sigma^2 \mathbf{I})$, then

$$\hat{\mathbf{e}} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{Y} = (\mathbf{I} - \mathbf{H}) \mathbf{Y}$$

$$= (\mathbf{I} - \mathbf{H}) (\mathbf{X}\beta + \mathbf{e}) = (\mathbf{I} - \mathbf{H}) \mathbf{e}$$

is distributed as $N(0, (\mathbf{I} - \mathbf{H})\sigma^2)$ and $\hat{e}_i \sim N(0, (1 - h_{ii})\sigma^2)$

Moreover,

$$E(\hat{\mathbf{e}}' \hat{\mathbf{e}}) = \sigma^2 \text{Tr}(\mathbf{I} - \mathbf{H}) = \sigma^2 (n - k)$$

and

$$E(\hat{e}_i^2) = \sigma^2 (1 - h_{ii}).$$

$$\text{Therefore } (n - k - 1) E(S^2(i)) = E(\hat{\mathbf{e}}' \hat{\mathbf{e}}) - \frac{E(\hat{e}_i^2)}{(1 - h_{ii})}$$

$$= \sigma^2 (n - k) - \sigma^2 = (n - k - 1) \sigma^2.$$

Hence

$$\begin{aligned} S^2(i) &= \frac{\left[\sum_{j=1}^n \hat{e}_j^2 - \frac{\hat{e}_i^2}{(1 - h_{ii})} \right]}{(n - k - 1)} \\ &= \frac{\left\{ (n - k) S^2 - \frac{\hat{e}_i^2}{(1 - h_{ii})} \right\}}{(n - k - 1)} \end{aligned}$$

is unbiased estimate of σ^2 .

Under normality, $S^2(i)$ and $\hat{e}_i$ are independent, see Cook and Weisberg (1982, p - 120)

Cook and Weisberg (1982, p-20) also showed that

$$S^2(i) = S^2 \left(\frac{n - k - r_i^2}{n - k - 1} \right) .$$

The studentized residuals are defined by

$$\begin{aligned} t_i &= \frac{\hat{e}_i}{\{S^2(i)(1 - h_{ii})\}^{1/2}} , \quad i = 1, 2, 3, \dots, n \\ &= \frac{r_i \sqrt{S^2}}{(S^2(i))^{1/2}} = \frac{r_i \sqrt{S^2} (n - k - 1)^{1/2}}{\left\{ (n - k)S^2 - \frac{\hat{e}_i^2}{(1 - h_{ii})} \right\}^{1/2}} \\ &= \frac{r_i (n - k - 1)^{1/2}}{\left\{ n - k - \frac{\hat{e}_i^2}{[S^2(1 - h_{ii})]} \right\}^{1/2}} \\ &= \frac{r_i \sqrt{n - k - 1}}{\sqrt{n - k - r_i^2}} = r_i \left(\frac{n - k - 1}{n - k - r_i^2} \right)^{1/2} \end{aligned}$$

Cook and Weisberg (1982, p-20) pointed out that t_i has the student t - distribution with $(n - k - 1)$ degrees of freedom.

2.3 Correlations Among Residuals

In a general regression situation, when k parameters are estimated from n observations, the n residuals are associated with only $(n-k)$ degrees of freedom. Thus the residuals cannot be independent and correlations exist among them. If the model under consideration is of the form

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

and if $X'X$ is non singular, then the residuals can be written in matrix form as

$$\begin{aligned}\hat{e} &= Y - \hat{Y} = Y - Xb = Y - X(X'X)^{-1}X'Y \\ &= (I - H) Y ,\end{aligned}$$

where $b = (X'X)^{-1}X'Y$ is the least squares estimator of β , $\hat{Y} = Xb$, $H = X(X'X)^{-1}X'$ and

$Y = (Y_1, Y_2, \dots, Y_n)'$ is the vector of observations. Since $E(Y) = X\beta$, it follows that

$$\hat{e} - E(\hat{e}_i) = (I - H)(Y - X\beta) = (I - H)e.$$

The variance - covariance matrix of $\hat{e}$ is defined as

$$\begin{aligned}\text{Var}(\hat{e}) &= E\left\{\left[\hat{e} - E(\hat{e})\right]\left[\hat{e} - E(\hat{e})\right]'\right\} \\ &= (I - H)E(ee')(I - H) = \sigma^2(I - H).\end{aligned}$$

The i^{th} diagonal element of $\sigma^2(I - H)$ gives the variance of $\hat{e}_i$. So,

$$\text{Var}(\hat{e}_i) = \sigma^2(1 - h_{ii}),$$

where h_{ii} is the i^{th} diagonal element of H . The covariance between $\hat{e}_i$ and $\hat{e}_j$ is given

by the $(i, j)^{\text{th}}$ element of the matrix $\sigma^2(I - H)$.

The correlation between $\hat{e}_i$ and $\hat{e}_j$ is given by

$$\begin{aligned}\text{Cor}(\hat{e}_i, \hat{e}_j) &= \rho_{ij} = \frac{\text{Cov}(\hat{e}_i, \hat{e}_j)}{\sqrt{\text{Var}(\hat{e}_i)\text{Var}(\hat{e}_j)}} \\ &= \frac{-h_{ij}}{\sqrt{(1 - h_{ii})(1 - h_{jj})}}\end{aligned}$$

where h_{ij} denotes the $(i, j)^{\text{th}}$ element of H .

Since

$$-\frac{1}{2} \leq h_{ij} \leq \frac{1}{2}, \text{ therefore,}$$

$$-\frac{1}{2\sqrt{(1-h_{ii})(1-h_{jj})}} \leq \rho_{ij} \leq \frac{1}{2\sqrt{(1-h_{ii})(1-h_{jj})}}.$$

Clearly, the values of ρ_{ij} depend entirely on the elements of the matrix X .

Anscombe and Tukey (1963) remarked that although correlations and constraints affect distributions of the functions of the residuals the corresponding effects on the graphical procedures can usually be neglected.

Now we shall study the sample (observed)

(i) correlation between $\hat{e}$ and Y .

(ii) correlation between $\hat{e}$ and $\hat{Y}$.

Let $\bar{\hat{e}} = J'\hat{e}/n$ and $\bar{Y} = J'Y/n$, where $J = (1, 1, \dots, 1)'$ the n -dimensional vector of ones.

When we fit a simple linear regression by least squares, we choose $b_0, b_1, \dots, b_{k-1}$ to minimize

$$\sum_{i=1}^n \hat{e}_i^2 = \sum_{i=1}^n \left(Y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \dots - b_{k-1} x_{i(k-1)} \right)^2$$

Differentiation with respect to b_0 and b_j shows that

$$\sum_{i=1}^n \hat{e}_i = 0, \text{ and } \sum_{i=1}^n x_{ij} \hat{e}_i = 0, \quad j = 1, 2, 3, \dots, k-1.$$

Hence

$$n\bar{\hat{e}} = J'(\hat{e}_1, \hat{e}_2, \dots, \hat{e}_n)' = \sum_{i=1}^n \hat{e}_i = 0.$$

$$\begin{aligned} \sum_{i=1}^n (\hat{e}_i - \bar{\hat{e}})(Y_i - \bar{Y}) &= \sum_{i=1}^n \bar{e}_i (Y_i - \bar{Y}) = \sum_{i=1}^n \hat{e}_i Y_i \\ &= \hat{e}'Y, \end{aligned}$$

where

$$\hat{e}' = (\hat{e}_1, \hat{e}_2, \dots, \hat{e}_n), \text{ and } Y' = (Y_1, Y_2, \dots, Y_n).$$

We observe that

$$\begin{aligned} \hat{e}'\hat{e} &= Y'(I - H)Y = Y'\hat{e} \\ &= \hat{e}'Y. \end{aligned}$$

Therefore,

$$\sum_{i=1}^n (\hat{e}_i - \bar{\hat{e}})(Y_i - \bar{Y}) = \hat{e}'\hat{e} = \text{Residual sum of squares.}$$

The sample correlation between $\hat{e}$ and Y is defined by

$$\begin{aligned} r_{\hat{e}, Y} &= \frac{\sum_{i=1}^n (\hat{e}_i - \bar{\hat{e}})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (\hat{e}_i - \bar{\hat{e}})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} \\ &= \frac{\hat{e}'\hat{e}}{\sqrt{\hat{e}'\hat{e} \sum_{i=1}^n (Y_i - \bar{Y})^2}} \\ &= \frac{(\hat{e}'\hat{e})^{1/2}}{\sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \end{aligned}$$

$$\begin{aligned}
&= \sqrt{\frac{\text{Residual sum of squares}}{\text{Total corrected sum of squares}}} \\
&= (1 - R^2)^{1/2},
\end{aligned}$$

where

$$R^2 = \frac{\text{Regression sum of squares}}{\text{Total corrected sum of squares}}.$$

Clearly $r_{\hat{e}, Y}$ will be zero only when the model is a perfect fit. Again,

$$\begin{aligned}
\sum_{i=1}^n (\hat{e}_i - \bar{\hat{e}})(\hat{Y}_i - \bar{\hat{Y}}) &= \sum_{i=1}^n \hat{e}_i (\hat{Y}_i - \bar{\hat{Y}}) \\
&= \sum_{i=1}^n \hat{e}_i \hat{Y}_i = \hat{e}'\hat{Y}. \\
&= Y' (I - H) H Y = 0.
\end{aligned}$$

Therefore, under normality assumption $\hat{e}$ and $\hat{Y}$ are independently distributed.

The sample correlation between $\hat{e}$ and $\hat{Y}$ is zero.

2.4 Examination Of Residuals

We now give ways of examining the residuals in order to check the model. These ways are all graphical, are easy to do, and are usually very revealing when the assumptions are violated. The principal ways of plotting the residuals are

- (i) Overall.
- (ii) Against the fitted values $\hat{Y}_i$, $i = 1, 2, \dots, n$.
- (iii) Against the values of one or more independent variables.

In addition to these basic plots, the residuals should also be plotted

- (iv) in any way that is sensible for the particular problem under consideration.

We now explain these plots in more detail. The following simple example will be used for illustrative purposes.

Example 1: A study was made on the effect of temperature on the yield of a chemical process. The following data were collected.

Temperature (X) : -5 , -4 , -3 , -2 , -1 , 0 , 1 , 2 , 3 , 4 , 5

Yield (Y) : 1 , 5 , 4 , 7 , 10 , 8 , 9 , 13 , 14 , 13 , 18

The data give the following results

$$\sum X = 0, \quad \sum Y = 102, \quad \sum X^2 = 110,$$

$$\sum Y^2 = 1194, \quad \sum XY = 158,$$

$$\sum (Y - \bar{Y})^2 = \sum Y^2 - (\sum Y)^2 / n = 248.1818,$$

$$b = \frac{\sum XY}{\sum X^2} = 1.4364, \text{ and}$$

$$a = \bar{Y} - b\bar{X} = \frac{102}{11} - (1.4364)(0) = 9.2727.$$

The straight line model $\hat{Y} = 9.2727 + 1.4364X$ is fitted to the above data . The fitted values and the residuals are given in the following table:

Table. 1 : Fitted values and Residuals

X (Temp.)	Y (Yield)	$\hat{Y}$ (Fitted values)	$\hat{e}$ (Residual)= Y - $\hat{Y}$
-5	1	2.09070	-1.0907
-4	5	3.52710	1.4729
-3	4	4.96350	-0.9635
-2	7	6.39990	0.6001
-1	10	7.83630	2.1637
0	8	9.27270	-1.2727
1	9	10.7091	-1.7091
2	13	12.1455	0.8545
3	14	13.5819	0.4181
4	13	15.0183	-2.0183
5	18	16.4547	1.5453

Residuals sum of squares , $\sum \hat{e}_i^2 = 21.2364$

$$1-R^2 = \frac{\text{Residual sum of squares}}{\text{Total corrected sum of squares}} = \frac{21.2364}{248.1818} = 0.0856$$

Overall Plot

When the residuals of the above example are plotted overall we obtain the diagram shown in figure 2.1. If our model is correct these residuals should approximately resemble eleven observations from normal distribution with zero mean. Our plot exhibits slight irregularity. But it does not appear abnormal for a sample of eleven observations from a normal distribution. For judging plots like this, one can make use of a table of random normal deviates. An extensive table is published by the Rand Corporation but shorter tables are given in some statistical textbooks. A number of observations of selected size (here eleven) can be taken from the table at random and

plotted in the way illustrated above. Then the two figures are compared. One can superimpose to see any irregularity in the residuals.

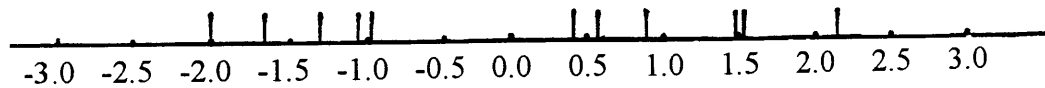


Figure 2.1: Overall Plot Of residuals.

An alternative procedure is to construct a normal plot of the residuals on standard probability paper. The points should fall, approximately, on a straight line. Once again, however, standards of judgment are required for assessing the plot.

Thus there is no particular advantage to be gained by using these alternatives, although they may be preferred by some practitioners.

The problem of assessing the assumption of normality is discussed throughly by Johnson and Wichern (1992, p-152).

Plot Against $\hat{Y}_i$

In section 2.3 we saw that $\hat{e}$ and Y are correlated. We proved that $r_{\hat{e}, Y}$, the correlation between $\hat{e}$ and Y , is given by

$$r_{\hat{e}, Y} = (1 - R^2)^{1/2}.$$

This will be zero only when the model is a perfect fit. Otherwise, the residuals plotted against Y_i , $i = 1, 2, \dots, n$ will show a slope of $(1 - R^2)$. However, it has also been observed in section 2.3 that $\hat{e}$ and $\hat{Y}$ are not correlated. Thus if the residuals plotted

against $\hat{Y}_i$ will show no slope. The slope will be zero. In view of this, we plot the residuals against the $\hat{Y}_i$ and not against the Y_i .

We plot the residuals obtained in table.1. against the corresponding values of $\hat{Y}_i$. We obtain the diagram shown in Figure 2.2. The "horizontal band" indicates no abnormality, and our least squares analysis would not appear to be invalidated.

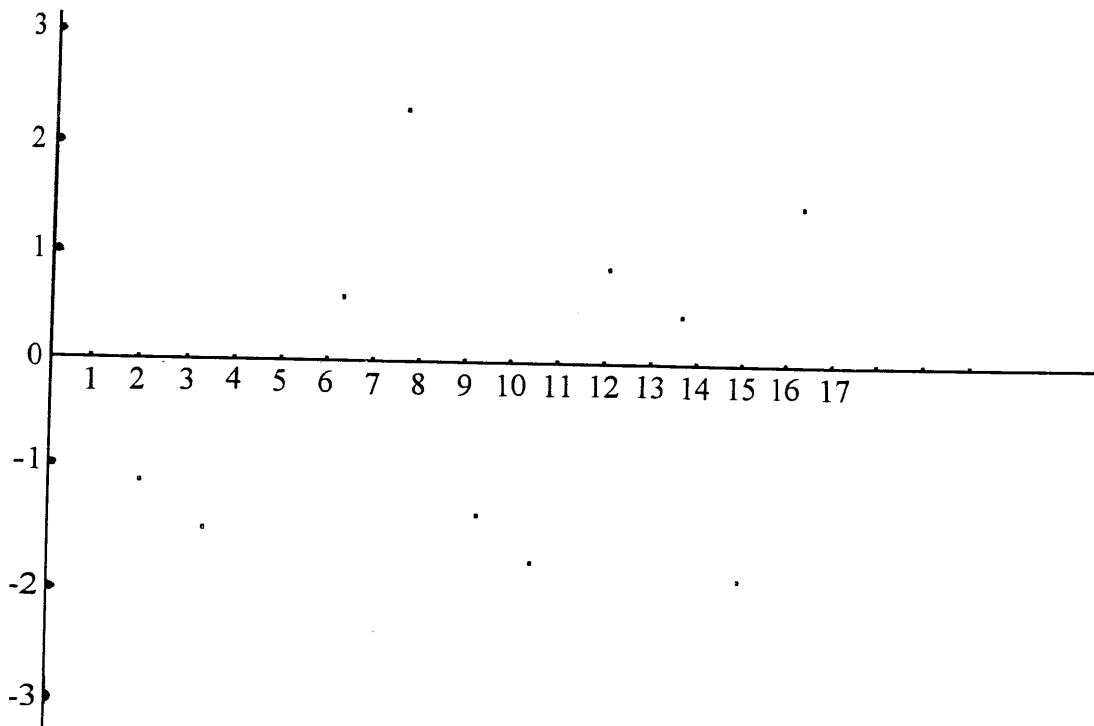


Figure 2.2: Residuals Plotted against Fitted Values

Abnormality would be indicated by plots of the form shown as (1), (2), and (3) in figure 2.3. Here the plots would indicate:

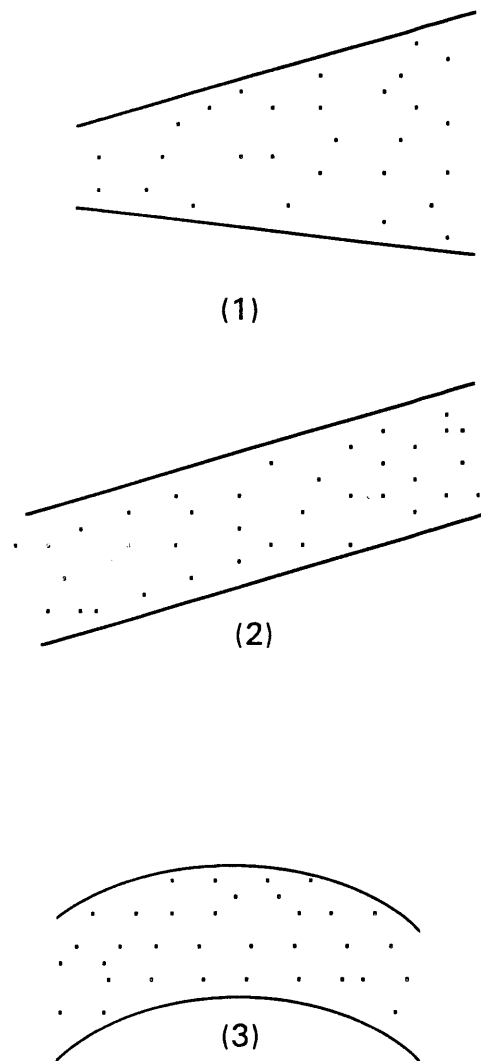


Figure 2.3: Examples of
Characteristic shown by
unsatisfactory residuals behavior.

(1) Variance is not constant, as assumed; analysis calls for weighted least squares or an appropriate transformation of the observed values is needed.

(2) Departure from the fitted equation is systematic. Negative residual correspond to high $\hat{Y}$, positive residual correspond to low $\hat{Y}$. The effect can also be caused by wrongly omitting a β_0 term from the model.

(3) Underlying model is inadequate. It needs extra terms (e.g. square or cross-product terms).

Plot Against The Predictor Values

The form of these plots is the same as that against the $\hat{Y}_i$, except that we use (instead of the values of the corresponding $\hat{Y}_i$) the values of the corresponding X_j , namely x_{1j} , x_{2j} , ..., x_{nj} . Once again an overall impression of a horizontal band of residuals is regarded as satisfactory.

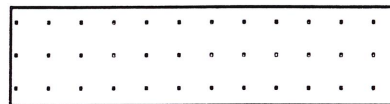


Figure 2.4: A satisfactory residual plot

The anomalies illustrated in Figure 2.3 indicate here

- (1) Variance is not constant, need for weighted least squares or a preliminary transformation on the Y's.
- (2) Error in calculations.
- (3) Need for extra terms, for example, a quadratic term, in X_j in the model or a transformation on the Y's.

In small regression problems which involve two or three X's only, it is possible to draw a diagram of the two or three dimensional space in which the data points occur. Where this is possible, it often provides a good visual grasp of the situation. When more than three X-variables occur it is possible to make such diagrams for subsets of the variables and this is sometimes appropriate.

Maximum Normed Residual

Shiffler (1988) showed that the Z score associated with the largest value of X in a data set of size n , is bounded by $\frac{(n-1)}{\sqrt{n}}$. Gray and Woodall (1994) used this result to obtain the maximum size of standardized residuals in regression. Here we briefly reproduce their results.

Let $Y_1, Y_2, \dots, Y_{n-1}$ be a random sample of size $(n-1)$ from a population with unknown mean and variance, and without loss of generality, we assume that Y_i are already ordered and $\left[\sum_{i=1}^{n-1} Y_i / (n-1) \right] = 0$. Suppose another observation Y_n is added to the sample such that $Y_n > Y_{n-1}$. The mean of the n values is

$$\bar{Y} = Y_n / n$$

and the total sum of squares,

$$SS_Y = \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

Let

$$\begin{aligned} MSS_Y &= \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{(n-1)} \\ &= \frac{1}{n-1} \left[\sum_{i=1}^n Y_i^2 - n\bar{Y}^2 \right] \\ &= \frac{1}{n-1} \left[\sum_{i=1}^{n-1} Y_i^2 + Y_n^2 - \frac{Y_n^2}{n} \right] \\ &= \frac{1}{n-1} \left[\sum_{i=1}^{n-1} Y_i^2 + \{(n-1)/n\} Y_n^2 \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n-1} \sum_{i=1}^{n-1} Y_i^2 + \frac{Y_n^2}{n} \\
&\geq \frac{Y_n^2}{n}.
\end{aligned}$$

Equality holds iff $\sum_{i=1}^{n-1} Y_i^2 = 0$

The largest positive Z score for the n values

$$\begin{aligned}
Z_{(n)} &= \frac{(Y_n - \bar{Y})}{\sqrt{MSS_Y}} \\
&= \frac{[Y_n - Y_n / n]}{\sqrt{MSS_Y}} \\
&= \frac{\left(\frac{n-1}{n}\right) Y_n}{\sqrt{MSS_Y}}
\end{aligned}$$

is maximized when MSS_Y is minimized.

Therefore, the largest possible value of Z score is

$$\begin{aligned}
Z_{(n)} &= \frac{\left(\frac{n-1}{n}\right) Y_n}{\sqrt{Y_n^2 / n}} \\
&= \frac{(n-1)}{\sqrt{n}}
\end{aligned}$$

Applying this result Gray and Woodall (1994) showed that

$$|\hat{e}_i / S| \leq \sqrt{\frac{(n-p)(n-1)}{n}}, \quad i = 1, 2, 3, \dots, n.$$

They also discussed the effects of these bounds on outlier identification in regression analysis.

CHAPTER 3

INFLUENTIAL OBSERVATION

3.1 Introduction

Sometimes, while most of the observations fit the model and meet Gauss-Markov conditions at least some of the observations do not. This occurs when there is something wrong with the observations or if the model is faulty. As far as the observations are concerned, there could have been a mistake in inputting or recording data. A few observations may reflect conditions different from those under which other observations were obtained. Observations that do not fit the model may also indicate deficiencies in the model. There could be an independent variable which should have been included in the model but was not. Such data points may also show us that the algebraic form of the model is not correct, e.g., we might need to make some transformations to the data or consider forms which require the use of non-linear regression. Observations that do not belong to the model are called outliers. Outliers frequently have an inordinate influence on least squares estimates. Observations that do not fit the model do not necessarily have large residuals, nor are they necessarily influential. There is little hope of finding such nonconforming points if they do not stand out either as outliers or as influential points. In this chapter we shall confine our attention to influential observations. In chapter 6, we shall address the problems associated with outliers.

3.2 Definition Of Influential Observations

An influential observation is one which either individually or together with several other observations has a demonstrably larger impact on the estimates of parameters of a model than in the case for most of the other observations. One obvious means for examining such an impact is to delete each row one at a time and note the resultant effect on the estimates. Rows whose deletion produces relatively large changes in the estimates are deemed influential. Moreover, residuals and the leverages are also used to assess the influence of observations. We shall examine leverages and residuals of observations in the next section. In section 3.4 we shall study the impacts of deletion of rows on the estimates.

3.3 The Leverage

The leverage of the i^{th} observation is i^{th} diagonal element of the prediction matrix $H = X(X'X)^{-1}X'$. The i^{th} diagonal element of the matrix H is given by

$$h_{ii} = x_i'(X'X)^{-1}x_i,$$

where x_i' is the i^{th} row of the design matrix X . The key feature of h_{ii} is that it describes how far away the i^{th} data point is from the centroid of all data points in the space of independent variables.

In the case of a lever, greater influence can be generated at a point far removed from the fulcrum, than at a point closer to it. This is because for a given change at a point remote from the fulcrum the corresponding changes in close-in points are relatively small. The situation is similar for the leverage h_{ii} with fulcrum at the centroid when

$\beta_0 \neq 0$. When $\beta_0 = 0$ the fulcrum is obviously the origin and leverage measures the distance from the origin.

A large value of h_{ii} indicates undue influence of the i^{th} observation. Belsey, Kuh and Welsch (1980, p-17) offered $2(k+1)/n$, where $k+1$ is the number of unknown regression coefficients in the model and n is the number of observations, as a possible cut-off points for h_{ii} . A slight shortcoming of the leverage as a diagnostic tool is that it treats all independent variables the same regardless of how each one affects the dependent variable. For example, a variable that we might ultimately discard from the model because of its inefficacy would affect h_{ii} as much, or as little, as if it were the most important predictor.

Let Z be the centered version of X , i.e. $Z' = (x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_n - \bar{x})$, where $\bar{x}_i (i = 1, 2, \dots, n)$ is the i^{th} row of X and $\bar{x}$ is the vector of means of variables.

Then the diagonal elements of $\tilde{H} = Z(Z'Z)^{-1} Z'$ are

$$h_{ii} = (x_i - \bar{x})'(Z'Z)^{-1}(x_i - \bar{x}), \quad i=1, 2, \dots, n.$$

Thus, h_{ii} also describes distance between x_i and $\bar{x}$.

For the purpose of detecting observations that do not belong to the model, more valuable than residuals are the studentized residuals, often called RSTUDENT and defined by

$$t_i = \frac{\hat{e}_i}{\left[S(i)\sqrt{1-h_{ii}} \right]}, \quad i = 1, 2, \dots, n,$$

where $\hat{e}_i$ is the residual associated with the i^{th} observation and $S^2(i)$ is the mean sum of squares of residuals when i^{th} row is deleted. The statistic t_i is a standardized

measure of the distance between the i^{th} observation and the model estimated on the basis of the remaining observations. It can serve as a test statistic to decide if the i^{th} point belongs to the model. Cook and Weisberg (1982, p-21) showed that t_i has a t -distribution with $(n-k-2)$ degrees of freedom when the errors are Gaussian.

3.4 Measures Of Influence

Here we describe diagnostic measures used to identify influential observations. Each observation, of course, is closely associated with a single row of the data matrix X and the corresponding element of Y . Since the estimated coefficients are often of primary interest to users of regression models, we look first at the change in the estimated regression coefficients that would occur if the i^{th} row were deleted. Denoting the coefficients estimated with the i^{th} row deleted by $b(i)$, this change is easily computed from the formula

$$\text{DFBETA}_i = b - b(i) = \frac{(X'X)^{-1} x_i \hat{e}_i}{1 - h_{ii}}.$$

The proof of this equation is given below:

We denote as $Y(i)$ and $X(i)$ the results of deleting the i^{th} row from Y and X .

Therefore,

$$X'(i) X(i) = X'X - x_i x_i',$$

and

$$\begin{aligned} [X'(i) X(i)]^{-1} &= [X'X - x_i x_i']^{-1} \\ &= (X'X)^{-1} + \frac{(X'X)^{-1} x_i x_i' (X'X)^{-1}}{1 - h_{ii}}. \end{aligned}$$

$$X'(i)Y(i) = X'Y - x_i y_i .$$

Now, by definition

$$\begin{aligned} b(i) &= [X'(i)X(i)]^{-1} X'(i)Y(i) \\ &= \left[(X'X)^{-1} + \frac{(X'X)^{-1} x_i x_i' (X'X)^{-1}}{1 - h_{ii}} \right] [X'Y - x_i y_i] \\ &= b + \frac{(X'X)^{-1} x_i x_i' b}{1 - h_{ii}} - (X'X)^{-1} x_i y_i \\ &\quad - \frac{(X'X)^{-1} x_i x_i' (X'X)^{-1} x_i y_i}{1 - h_{ii}} \\ &= b - \frac{(X'X)^{-1} x_i}{1 - h_{ii}} [-x_i' b + (1 - h_{ii}) y_i + h_{ii} y_i] \\ &= b - \frac{(X'X)^{-1} x_i}{1 - h_{ii}} [y - x_i' b] \\ &= b - \frac{(X'X)^{-1} x_i \hat{e}_i}{1 - h_{ii}} . \end{aligned}$$

Therefore,

$$b - b(i) = \frac{(X'X)^{-1} x_i \hat{e}_i}{1 - h_{ii}} , \quad i = 1, 2, 3, \dots, n.$$

Whether the change in b_j , the j^{th} component of b , that results from the deletion of the i^{th} row is large or small is often most usefully assessed relative to the variance of b_j , that is $\sigma^2 (X'X)^{-1}_{jj}$. If we let

$$C = (X'X)^{-1} X' ,$$

then

then

$$b_j - b_j(i) = \frac{c_{ji} \hat{e}_i}{1 - h_{ii}} .$$

Since

$$\sum_{k=1}^n (X'X)^{-1} x_k x_k' (X'X)^{-1} = (X'X)^{-1} .$$

It follows that

$$\text{Var}(b_j) = \sigma^2 \sum_{k=1}^n c_{jk}^2 .$$

Hence a scaled measure of change can be defined as

$$\begin{aligned} \text{DFBETAS}_{ij} &= \frac{b_j - b_j(i)}{S(i) \sqrt{(X'X)^{-1}_{jj}}} \\ &= \frac{c_{ji}}{\sqrt{\sum_{k=1}^n c_{jk}^2}} \cdot \frac{\hat{e}_i}{S(i) (1 - h_{ii})} . \end{aligned}$$

Large values of $|\text{DFBETAS}_{ij}|$ indicate observations that are influential in the determination of the j^{th} coefficient, b_j .

Another way to summarize coefficient changes and, at the same time, to gain insight into forecasting effects when an observation is deleted is by change in fit, defined by

$$\begin{aligned} \text{DFFIT} &= \hat{y}_i - \hat{y}_i(i) \\ &= x_i' [b - b(i)] \\ &= h_{ii} \hat{e}_i / (1 - h_{ii}) \end{aligned}$$

This diagnostic measure has the advantage that it does not depend on the particular coordinate system used to form the regression model. For scaling purposes, it is natural to divide by $\sigma\sqrt{h_{ii}}$, the standard deviation of the fit, $\hat{y}_i = x_i'b$ giving

$$DFFITS_i = \left[\frac{h_{ii}}{1 - h_{ii}} \right]^{1/2} \frac{\hat{e}_i}{S(i)\sqrt{(1 - h_{ii})}} ,$$

where σ has been estimated by $S(i)$.

3.5 Other Measures Of Influence

The measures given in the last section are by no means the only ones available for detecting influential observations. Some measures take into account the effect on the estimated covariance matrix of b . One such measure is the ratio of the determinants of the estimated covariance matrix of $b(i)$ and b . This is called the covariance ratio and is given by

$$\text{Covariance ratio} = \frac{\det [S^2(i) (X'(i) X(i))^{-1}]}{\det [S^2 (X'X)^{-1}]} ,$$

where $X(i)$, $b(i)$, and $S^2(i)$ are defined as before.

A value of this ratio close to 1 would indicate lack of influence of the i^{th} data point.

We have,

$$X'(i)X(i) = X'X - x_i x_i', \text{ where } x_i' \text{ is the } i^{\text{th}} \text{ row of } X,$$

and

$$\begin{aligned} \det [X'(i) X(i)] &= \det (X'X - x_i x_i') \\ &= \det (X'X) \cdot \det [I - (X'X)^{-1} x_i x_i'] \end{aligned}$$

$$\begin{aligned}
&= \det(X'X) \cdot \det[1 - x_i'(X'X)^{-1}x_i] \\
&= (1 - h_{ii}) \det(X'X).
\end{aligned}$$

Therefore

$$\text{Covariance ratio} = [S^2(i) / S^2]^{k+1} [1 - h_{ii}]^{-1}.$$

A very frequently used measure of influence can be defined as the distance between the vectors b and $b(i)$ standardized by the estimated covariance matrix $S^2(X'X)^{-1}$ of b . This distance, known as Cook's (1977) distance, is given by

$$D(b, b(i)) = \left[(b - b(i))' X'X(b - b(i)) / [(k+1)S^2] \right].$$

Since

$$\begin{aligned}
b - b(i) &= \frac{(X'X)^{-1}x_i \hat{e}_i}{(1 - h_{ii})}, \text{ hence} \\
D(b, b(i)) &= \frac{\hat{e}_i' x_i' (X'X)^{-1} X'X(X'X)^{-1} x_i \hat{e}_i}{(1 - h_{ii})^2 [(k+1)S^2]} \\
&= \frac{h_{ii} \hat{e}_i^2}{(1 - h_{ii})^2 [(k+1)S^2]}.
\end{aligned}$$

Ghosh (1983) used the Cook's distance to detect a set of $t(\geq 1)$ influential observations. Let X_1 be the design matrix and Y_1 be the vector of observations obtained by deleting any set of t rows from X and Y respectively. Further, let X_2 and Y_2 be the matrices formed by the rows deleted from X and Y , respectively.

For this situation the Cook's distance is given by

$$D(b, b^*) = \frac{(b - b^*)' X'X(b - b^*)}{(k+1)S^2},$$

where

$$b = (X'X)^{-1} X'Y \quad \text{and} \quad b^* = (X_1'X_1)^{-1} X_1'Y_1.$$

It can be checked that

$$b^* - b = (X'X)^{-1} X_2' B (\hat{Y}_2 - Y_2),$$

where

$$\hat{Y}_2 = X_2 b \quad \text{and} \quad B = [I - X_2 (X'X)^{-1} X_2']^{-1}.$$

Therefore,

$$\begin{aligned} D(b, b^*) &= \frac{(\hat{Y}_2 - Y_2)' B X_2 (X'X)^{-1} X'X (X'X)^{-1} X_2' B (\hat{Y}_2 - Y_2)}{(k+1) S^2} \\ &= \frac{(\hat{Y}_2 - Y_2)' B X_2 (X'X)^{-1} X_2' B (\hat{Y}_2 - Y_2)}{(k+1) S^2} \\ &= \frac{(\hat{Y}_2 - Y_2)' B (I - B^{-1}) B (\hat{Y}_2 - Y_2)}{(k+1) S^2} \\ &= \frac{(\hat{Y}_2 - Y_2)' (B^2 - B) (\hat{Y}_2 - Y_2)}{(k+1) S^2}. \end{aligned}$$

Ghosh (1983) also proposed some other measures to detect a set of t influential observations. One of his measures is given by

$$I(X_2, L'\beta) = \frac{\text{Var}(L' b^*) - \text{Var}(L' b)}{\text{Var}(L' b)},$$

where $b^* = (X_1'X_1)^{-1} X_1'Y_1$ and L is a given $(k+1)$ vector. It can be shown that

$$I(X_2, L'\beta) = \frac{L' (X'X)^{-1} X_2' B X_2 (X'X)^{-1} L}{L' (X'X)^{-1} L} .$$

The larger the value of $I(X_2, L'\beta)$ indicates Y_2 is influential.

Some other measures of influence of a set of t observations in Y_2 are

$$I_1(Y_2) = \det [\text{Var}(b^*) / \sigma^2] = |(X_1'X_1)^{-1}|,$$

$$I_2(Y_2) = \text{Trace of } [\text{Var}(b^*) / \sigma^2]$$

$$= \text{Tr}(X_1'X_1)^{-1},$$

and

$$I_3(Y_2) = [\text{Trace Variance of } X_2 b^*] / \sigma^2$$

$$= \text{Tr}[X_2 (X_1'X_1)^{-1} X_2'] .$$

Again small values of $I_1(Y_2)$, $I_2(Y_2)$, and $I_3(Y_2)$ indicate Y_2 is less influential vector of observations.

Other measures of influence may be found in Belsley, Kuh and Welsch (1980).

Results 1:

The comparison of influence of two sets of t observations with respect to the measure $I_1(Y_2)$ is equivalent to the comparison with respect to $|I_t - M|^{-1}$, where I_t is the $t \times t$ identity matrix and $M = X_2 (X'X)^{-1} X_2'$.

Proof :

We have,

$$X'X = X_1'X_1 + X_2'X_2.$$

Hence,

$$\begin{aligned} X_1'X_1 &= X'X - X_2'X_2 \\ &= (X'X) \left[I_{k+1} - (X'X)^{-1} X_2'X_2 \right]. \end{aligned}$$

Therefore,

$$\begin{aligned} |X_1'X_1| &= |X'X| \left| I_{k+1} - (X'X)^{-1} X_2'X_2 \right| \\ &= |X'X| \left| I_t - X_2(X'X)^{-1} X_2' \right| \\ &= |X'X| |I_t - M|. \end{aligned}$$

Note that for a given design $|X'X|$ is constant. Therefore, comparison of $|(X_1'X_1)^{-1}|$ is equivalent to comparison of $|I_t - M|^{-1}$. This completes the proof.

Result 2:

The comparison of influence of two sets of t observations with respect to the measure $I_2(Y_2)$ is equivalent to the comparison with respect to $\text{Tr}(I_t - M)^{-1}Q$, where $Q = X_2(X'X)^{-2}X_2'$, and $M = X_2(X'X)^{-1}X_2'$.

Proof :

We observe that

$$(X_1'X_1)^{-1} = (X'X)^{-1} + (X'X)^{-1} X_2'(I_t - M)^{-1} X_2(X'X)^{-1}.$$

Thus,

$$\text{Tr}(\mathbf{X}'_1 \mathbf{X}_1)^{-1} = \text{Tr}(\mathbf{X}' \mathbf{X})^{-1} + \text{Tr}(\mathbf{I}_t - \mathbf{M})^{-1} \mathbf{Q} .$$

The rest is obvious.

Result 3:

The comparison of influence of two sets of t observations with respect to the measure

$I_3(\mathbf{Y}_2)$ is equivalent to the comparison with respect to $\text{Tr}(\mathbf{I} - \mathbf{M})^{-1}$.

Proof :

We notice that

$$I_3(\mathbf{Y}_2) = \text{Tr} \mathbf{X}_2 (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_2 .$$

Now,

$$\begin{aligned} \mathbf{X}_2 (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_2 &= \mathbf{X}_2 \left[(\mathbf{X}' \mathbf{X})^{-1} + (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}'_2 (\mathbf{I}_t - \mathbf{M})^{-1} \mathbf{X}_2 (\mathbf{X}' \mathbf{X})^{-1} \right] \mathbf{X}'_2 \\ &= \mathbf{M} + \mathbf{M} (\mathbf{I}_t - \mathbf{M})^{-1} \mathbf{M} \\ &= \mathbf{M} + [\mathbf{I}_t - (\mathbf{I}_t - \mathbf{M})] (\mathbf{I}_t - \mathbf{M})^{-1} \mathbf{M} \\ &= \mathbf{M} + (\mathbf{I}_t - \mathbf{M})^{-1} \mathbf{M} - \mathbf{M} \\ &= (\mathbf{I}_t - \mathbf{M})^{-1} \mathbf{M} \\ &= (\mathbf{I}_t - \mathbf{M})^{-1} [\mathbf{I}_t - (\mathbf{I}_t - \mathbf{M})] \\ &= (\mathbf{I}_t - \mathbf{M})^{-1} - \mathbf{I}_t \end{aligned}$$

The rest is obvious.

CHAPTER IV

ERRORS OF MEASUREMENT IN REGRESSION

4.1 Introduction

Cochran (1968) presented a review of the literature on errors of measurement. His attention was centered on the type of mathematical models used to represent errors of measurement, on the extent to which standard techniques of analysis become erroneous and misleading if certain types of errors are present. He also described various techniques that are available for the numerical study of errors of measurement. Chatterjee and Hadi (1988) studied the effects of errors of measurement on the regression coefficients.

They considered the following approaches:

- (i) Asymptotic Approach.
- (ii) Perturbation Approach.
- (iii) Simulation Approach.

In the asymptotic approach the effects the errors on the regression coefficients are examined for large samples. This is essentially a large sample theory and is favoured by the econometricians. the asymptotic approach may provide good approximations for the bias in some cases in which the sizes are not very large. However, Chatterjee and Hadi (1988, p. 246) pointed out that from a practical point of view the results using the asymptotic approach are of limited use, they do not shed much light on the problem at hand, a particular data set which is often far from being large.

The perturbation approach, which is favoured by numerical analysts, is more attractive. In this approach the effects on the regression coefficients of perturbing the

data by small amounts are studied. These perturbations correspond to errors of measurement. Chatterjee and Hadi (1988) showed that although the effects cannot be measured exactly, it is often possible to get an upper bound for the relative errors in the estimated regression coefficients.

In the simulation approach rather than deriving these bounds asymptotically, they simulated the error generating process and empirically studied the behavior of the calculated regression coefficients. Here we shall not describe these approaches. These approaches were described extensively by Chatterjee and Hadi (1988).

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.

In this chapter 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. We obtain upper bounds for the relative bias and mean squared error (MSE) of the least squares estimates of estimable linear function of parameters of a linear model when both the independent variables and the dependent variable are subject to measurement errors. An approach to reduce the impacts of the errors of measurement on the estimates of the parameters is also considered.

4.2 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.

We consider the general linear regression model:

$$Y = X\beta + e, \quad E(e) = 0, \quad E(ee') = \sigma^2 I_n, \quad (4.2.1)$$

where Y and e are $n \times 1$ vectors of observations and random errors respectively, β is a $p \times 1$ vector of unknown regression coefficients, X is an $n \times p$ design matrix with $\text{rank}(X) = p < n$, σ^2 is an unknown positive constant, 0 denotes the null vector of appropriate dimension and I_n stands for the $n \times n$ identity matrix. Suppose that instead of observing X in (4.2.1) we observe

$$X_* = X + U,$$

where X_* , X and U are fixed matrices. We also assume that $\text{Ran}(X_*) = \text{Rank}(X) = p$. Then an arbitrary linear combination of the regression coefficients, $L'\beta$, where L is a given $p \times 1$ vector, is estimated by

$$L'b = L'(X_*'X_*)^{-1} X_*'Y.$$

Clearly ,

$$\begin{aligned} E[L'(b - \beta)] &= L'(X_*'X_*)^{-1} X_*'X\beta - L'\beta \\ &= L'(X_*'X_*)^{-1} X_*'(X_* - U)\beta - L'\beta \\ &= -L'(X_*'X_*)^{-1} X_*'U\beta. \end{aligned} \quad (4.2.2)$$

The variance of $L'b$ is given by

$$\text{Var}(L'b) = \sigma^2 L'(X_*'X_*)^{-1} L. \quad (4.2.3)$$

The mean square error (MSE) of $L'b$ is

$$\begin{aligned} \text{MSE}(L'b) &= \text{Var}(L'b) + [\text{bias}(L'b)]^2 \\ &= \sigma^2 L'(X_*'X_*)^{-1} L + L'(X_*'X_*)^{-1} X_*'U\beta\beta'U'X_*(X_*'X_*)^{-1} L. \end{aligned} \quad (4.2.4)$$

We define relative squared bias of $L'b$ by

$$\text{Relative squared bias} = \frac{[\text{bias}(L'b)]^2}{\text{Var}(L'b)}.$$

Therefore,

$$\begin{aligned} \text{Relative squared bias} &= \frac{L'(X_*'X_*)^{-1} X_*'U\beta\beta'U'X_*(X_*'X_*)^{-1} L}{\sigma^2 L'(X_*'X_*)^{-1} L} \\ &= \frac{L'(X_*'X_*)^{-1} X_*'U\beta\beta'U'X_*(X_*'X_*)^{-1} L}{\sigma^2 L'(X_*'X_*)^{-1} X_*'X_*(X_*'X_*)^{-1} L}. \end{aligned} \quad (4.2.5)$$

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

$$\sup_L [L'GL / (L'L)] = \text{Largest eigenvalue of } G,$$

and

$$\inf_L [L'GL / (L'L)] = \text{Smallest eigenvalue of } G.$$

Therefore

$$\begin{aligned} \sup_L \frac{L' (X_*' X_*)^{-1} X_*' U \beta \beta' U' X_* (X_*' X_*)^{-1} L}{L' (X_*' X_*)^{-1} X_*' X_* (X_*' X_*)^{-1} L} \\ = \text{Largest eigenvalue of } (U \beta \beta' U'), \end{aligned} \quad (4.2.6)$$

and

$$\begin{aligned} \inf_L \frac{L' (X_*' X_*)^{-1} X_*' U \beta \beta' U' X_* (X_*' X_*)^{-1} L}{L' (X_*' X_*)^{-1} X_*' X_* (X_*' X_*)^{-1} L} \\ = \text{Smallest eigenvalue of } (U \beta \beta' U'). \end{aligned} \quad (4.2.7)$$

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

$$\text{tr } (U \beta \beta' U') = \text{Sum of the eigenvalues of } (U \beta \beta' U'),$$

and

$$\text{tr } (U \beta \beta' U') = \text{tr } (\beta' U' U \beta) = \beta' U' U \beta,$$

where $\text{tr } (G)$ denotes the trace of a square matrix G . Hence,

$$\text{Largest eigenvalue of } (U\beta\beta'U') = \beta'U'U\beta, \quad (4.2.8)$$

and

$$\text{Smallest eigenvalue of } (U\beta\beta'U') = 0. \quad (4.2.9)$$

Relations (4.2.5), (4.2.6), (4.2.7), (4.2.8), and (4.2.9) produce the following result:

$$0 \leq \frac{[\text{bias}(L'b)]^2}{\text{Var}(L'b)} \leq (\beta'U'U\beta / \sigma^2). \quad (4.2.10)$$

Clearly, the relative bias in point estimation of $L'\beta$ caused by errors, U , in the independent variables will be negligible when $(\beta'U'U\beta / \sigma^2)$ is very small.

We now consider the mean square error of $L'b$. It follows from (4.2.4) that,

$$\text{MSE}(L'b) = \sigma^2 L'(X_*'X_*)^{-1} L \left[1 + \frac{[\text{bias}(L'b)]^2}{\text{Var}(L'b)} \right]. \quad (4.2.11)$$

Using (4.2.10) in (4.2.11) we get

$$\sigma^2 L'(X_*'X_*)^{-1} L \leq \text{MSE}(L'b) \leq \sigma^2 L'(X_*'X_*)^{-1} L \left[1 + \beta'U'U\beta / \sigma^2 \right]. \quad (4.2.12)$$

4.3 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

$$\theta = W\beta, \quad (4.3.1)$$

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

The least squares estimator of θ is

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

Clearly,

$$\begin{aligned}
 E(\hat{\theta} - \theta) &= W(X'_*X_*)^{-1} X'_*X\beta - W\beta \\
 &= W(X'_*X_*)^{-1} X'_*(X_* - U)\beta - W\beta \\
 &= -W(X'_*X_*)^{-1} X'_*U\beta,
 \end{aligned}$$

and

$$[E(\hat{\theta} - \theta)][E(\hat{\theta} - \theta)]' = WRU\beta\beta'U'R'W', \quad (4.3.2)$$

where

$$R = (X'_*X_*)^{-1} X'_*.$$

The dispersion matrix of the least squares estimator, $\hat{\theta}$, of θ is

$$\begin{aligned}
 D(\hat{\theta}) &= E\left[(\hat{\theta} - E(\hat{\theta}))(\hat{\theta} - E(\hat{\theta}))'\right] \\
 &= \sigma^2 W(X'_*X_*)^{-1} W'.
 \end{aligned} \quad (4.3.3)$$

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

$$M(\hat{\theta}) = D(\hat{\theta}) + [E(\hat{\theta}) - \theta][E(\hat{\theta}) - \theta]'. \quad (4.3.4)$$

From (4.3.2) and (4.3.3), we obtain

$$M(\hat{\theta}) = \sigma^2 W(X'_*X_*)^{-1} W' + WRU\beta\beta'U'R'W', \quad (4.3.4)$$

where R is defined in (4.3.2).

Theorem 4.1:

For any full rank matrices X_* and W ,

$$\left| M(\hat{\theta}) / \sigma^2 \right| \leq \left| W(X_*'X_*)^{-1} W' \right| \left(1 + \frac{\beta' U' U \beta}{\sigma^2} \right), \quad (4.3.5)$$

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

Proof:

We rewrite (4.3.4) as

$$M(\hat{\theta}) = \sigma^2 Q_* Q_*' + Q_* U \beta \beta' U' Q_*',$$

where

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

Hence,

$$\begin{aligned} \left| M(\hat{\theta}) / \sigma^2 \right| &= \left| Q_* Q_*' + Q_* U \beta \beta' U' Q_*' / \sigma^2 \right| \\ &= |Q_* Q_*'| \left[1 + \beta' U' Q_*' (Q_* Q_*')^{-1} Q_* U \beta / \sigma^2 \right], \end{aligned}$$

see, Rao (1973, p.32).

Therefore,

$$\left| M(\hat{\theta}) / \sigma^2 \right| \leq |Q_* Q_*'| \left[1 + (\beta' U' U \beta / \sigma^2) \text{Chmax}(F_*) \right],$$

where

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

and $\text{Chmax}(F_*)$ stands for the largest eigenvalue of F_* . Since F_* is an idempotent

matrix, therefore,

$$\text{Chmax}(F_*) \leq 1.$$

Hence,

$$\begin{aligned} |M(\hat{\theta}) / \sigma^2| &\leq |Q_* Q_*'| [1 + \beta' U' U \beta / \sigma^2] \\ &= |W(X_*' X_*)^{-1} W'| [1 + \beta' U' U \beta / \sigma^2]. \end{aligned}$$

This completes the proof.

4.4 Error In The Dependent Variable of a General Linear Model

Suppose that some of the observations of the dependent variable of the model (4.2.1) are subject to shift in means and variances, and the observations of the independent variables are free from measurement error. So, we assume that instead of Y in the model (4.2.1), we observe

$$Y_* = Y + Z$$

where Z is an $n \times 1$ vector of random variables with $E(Z) = \upsilon$,

$$E[(Z - \upsilon)(Z - \upsilon)'] = \sigma_1^2 I_n,$$

and

$$E[(Z - \upsilon)e'] = E[e(Z - \upsilon)'] = 0.$$

The $n \times 1$ vector e is given in (4.2.1). We further assume that one or more of the elements of υ are nonzero and σ_1^2 is a positive unknown constant.

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

$$L'b = L'(X'X)^{-1} X'Y_*.$$

Clearly,

$$\begin{aligned}\text{bias}(L'b) &= E(L'b) - L'\beta \\ &= L'(X'X)^{-1}X'u.\end{aligned}\quad (4.4.1)$$

The variance of $L'b$ is given by

$$\begin{aligned}\text{Var}(L'b) &= L' \text{Var}(b) L \\ &= L'(X'X)^{-1}X'[\sigma^2 I_n + \sigma_1^2 I_n]X(X'X)^{-1}L \\ &= (\sigma^2 + \sigma_1^2)L'(X'X)^{-1}L.\end{aligned}\quad (4.4.2)$$

The mean square error of $L'b$ is

$$\begin{aligned}\text{MSE}(L'b) &= \text{Var}(L'b) + [\text{bias}(L'b)]^2 \\ &= (\sigma^2 + \sigma_1^2)L'(X'X)^{-1}L + L'(X'X)^{-1}X'u u'X(X'X)^{-1}L\end{aligned}\quad (4.4.3)$$

The relative squared bias of $L'b$ is

$$\begin{aligned}\text{Relative squared bias} &= \frac{[\text{bias}(L'b)]^2}{\text{Var}(L'b)} \\ &= \frac{L'(X'X)^{-1}X'u u'X(X'X)^{-1}L}{(\sigma^2 + \sigma_1^2)L'(X'X)^{-1}L} \\ &= \frac{L'(X'X)^{-1}X'u u'X(X'X)^{-1}L}{(\sigma^2 + \sigma_1^2)L'(X'X)^{-1}X'X(X'X)^{-1}L}.\end{aligned}$$

Hence,

$$0 \leq \frac{[\text{bias}(L'b)]^2}{\text{Var}(L'b)} \leq \left[\frac{(u'u)}{(\sigma^2 + \sigma_1^2)} \right]. \quad (4.4.4)$$

Therefore, the relative squared bias in point estimation of $L'b$ caused by error, Z , in

the dependent variable will be negligible when $\left[\frac{(v'v)}{(\sigma^2 + \sigma_1^2)} \right]$ is very small.

We now consider the mean square error of $L'b$. It follows from (4.4.3) that

$$\text{MSE}(L'b) = (\sigma^2 + \sigma_1^2) L'(X'X)^{-1} L \left[1 + \frac{[\text{bias}(L'b)]^2}{\text{Var}(L'b)} \right]. \quad (4.4.5)$$

Using (4.4.4) in (4.4.5) we obtain,

$$L'(X'X)^{-1} L \leq \frac{\text{MSE}(L'b)}{\sigma^2 + \sigma_1^2} \leq L'(X'X)^{-1} L \left[1 + \frac{v'v}{\sigma^2 + \sigma_1^2} \right]. \quad (4.4.6)$$

4.5 Simultaneous Estimation of Linear Functions of Parameters when Dependent Variable is Subject to Measurement Errors

Here we generalize the results obtained in section (4.4). We consider the problem of estimation of $\theta = W\beta$, where W is a given $r \times p$ matrix with rank $r \leq p$, when instead of Y in the model (4.2.1) we observe,

$$Y_* = Y + Z.$$

as in section (4.4) we assume that

$$E(Z) = v, \quad D(Z) = \sigma_1^2 I_n, \text{ and}$$

$$E[(Z - v)e'] = E[e(Z - v)'] = 0.$$

The least squares estimator of θ based on Y_* is

$$\hat{\theta}_* = W(X'X)^{-1} X'Y_*.$$

Clearly

$$\begin{aligned}
 \text{bias}(\hat{\theta}_*) &= E(\hat{\theta}_*) - \theta \\
 &= W(X'X)^{-1} X'(X\beta + v) - W\beta \\
 &= W(X'X)^{-1} X'v.
 \end{aligned}$$

The dispersion matrix of $\hat{\theta}_*$ is given by

$$\begin{aligned}
 D(\hat{\theta}_*) &= E \left[\left(\hat{\theta}_* - E(\hat{\theta}_*) \right) \left(\hat{\theta}_* - E(\hat{\theta}_*) \right)' \right] \\
 &= (\sigma^2 + \sigma_v^2) W(X'X)^{-1} W'.
 \end{aligned}$$

The mean squared error matrix of $\hat{\theta}_*$ is

$$\begin{aligned}
 M(\hat{\theta}_*) &= D(\hat{\theta}_*) + \left[E(\hat{\theta}_*) - \theta \right] \left[E(\hat{\theta}_*) - \theta \right]' \\
 &= (\sigma^2 + \sigma_v^2) W(X'X)^{-1} W' + W(X'X)^{-1} X'v v' X (X'X)^{-1} W'. \quad (4.5.1)
 \end{aligned}$$

Theorem 4.2:

For any full rank matrices X and W ,

$$\left| M(\hat{\theta}_*) / (\sigma_v^2 + \sigma^2) \right| \leq \left| W(X'X)^{-1} W' \right| \left[1 + v'v / (\sigma_v^2 + \sigma^2) \right]. \quad (4.5.2)$$

Proof:

We rewrite (4.5.1) as

$$M(\hat{\theta}_*) = (\sigma^2 + \sigma_v^2)(QQ') + Qvv'Q',$$

where $Q = W(X'X)^{-1} X'$.

Next,

$$\begin{aligned}
\left| M(\hat{\theta}_*) / (\sigma_1^2 + \sigma^2) \right| &= \left| QQ' + Qvv'Q' / (\sigma_1^2 + \sigma^2) \right| \\
&= |QQ'| \left[1 + v'Q'(QQ')^{-1}Qv / (\sigma_1^2 + \sigma^2) \right] \\
&\leq |QQ'| \left[1 + \frac{v'v}{\sigma_1^2 + \sigma^2} \text{Ch max}(F) \right],
\end{aligned}$$

where $F = Q'(QQ')^{-1}Q$ is an idempotent matrix. Hence, its largest eigenvalue, $\text{Chmax}(F) = 1$.

Therefore,

$$\begin{aligned}
\left| M(\hat{\theta}_*) / (\sigma_1^2 + \sigma^2) \right| &\leq |QQ'| \left[1 + v'v / (\sigma_1^2 + \sigma^2) \right] \\
&= |W(X'X)^{-1}W'| \left[1 + v'v / (\sigma_1^2 + \sigma^2) \right].
\end{aligned}$$

This completes the proof.

4.6 Remarks

Errors of measurement affect data analysis. It is a field of growing importance. The recent rapid increase in data gathering projects in the social and medical sciences is producing large bodies of data containing variables that cannot be measured without error. Moreover, sometimes we deliberately measure the wrong quantity using a substitute measurement that is cheaper or more convenient, as for example, in the standard manometer used for measuring blood pressure by a sleeve wrapped round the arm instead of a direct measurement of intra-arterial blood pressure. Sometimes the quantity being measured may be fixed at a point in time, but vary with time. When the quantity being measured varies with time, we may in effect be using the wrong measurement if we treat measurements made some months earlier as they applied to the situation today. It is known that the yield of wheat increase by 4-5% owing to moisture when the rainy season begins.

It is important to know the amount of harm done by errors of measurement in producing imprecise estimates of the parameters under consideration. The results obtained in (4.2.10), (4.2.12), (4.3.5), (4.4.4), (4.4.6), and (4.5.2) may serve this purpose. From (4.2.2) and (4.4.1) we see that with errors of measurement present in both independent and dependent variables we do not get unbiased estimates.

We observe from (4.4.1) and (4.4.2) that measurement errors in Y can decrease the precision of the estimates significantly, because the errors in the measurement of Y contribute to the bias as well as to the variance.

If we set $W = I_p$, the $p \times p$ identity matrix, in (4.3.5) and (4.5.2) we can study the effects of collinearity and the errors of measurement simultaneously. Clearly, collinear relationship among the X variables will cause the MSE to increase. Collinearity contributes greatly to inaccuracies of the estimates. It should be clear now that the effects of measurement errors become more acute in a collinear situation than in a non-collinear situation. An analyst should particularly worry about the effects of measurement errors on the magnitude of the MSE of the estimates of the regression coefficients in situations where collinearity is a problem.

For some time there has been much interest in investigating the robustness of standard estimation techniques against non-normality and the use of an incorrect model.

Obviously, research should also extend to robustness against the presence of errors of measurement.

CHAPTER V

ERRORS OF MEASUREMENT AND PRIOR INFORMATION IN LINEAR MODELS

5.1 Introduction

Theil and Goldberger (1961) and Theil (1963) found the mixed regression estimation technique by mixing both the sample and the prior information. In this chapter the use of the mixed regression approach when data are subject to measurement errors is discussed. The effects of the prior information on the least squares estimator of the vector of unknown regression coefficient are examined when there are errors of measurement either in the independent variables or in the dependent variable. It is found that the mixed regression approach, in some sense, reduces the effects of errors of measurement on the estimator of the vector of unknown regression coefficients.

5.2 Prior Information and Errors in the Independent Variables

We consider the general linear model:

$$Y = X\beta + e, E(e) = 0, E(ee') = \sigma^2 I_n, \quad (5.2.1)$$

where Y and e are $n \times 1$ vectors of observations and random errors respectively, β is $p \times 1$ vector of unknown regression coefficients, X is an $n \times p$ design matrix with $\text{rank}(X) = p < n$, σ^2 is an unknown positive constant, 0 denotes the null vector of appropriate dimension and I_n stands for the $n \times n$ identity matrix. Suppose that instead of observing X in (5.2.1), we observe

$$X_* = X + U,$$

where X_* , X and U are fixed matrices. We also assume that $\text{Rank}(X_*) = p$. Then the least squares estimator of β is

$$b = (X_*'X_*)^{-1} X_*'Y. \quad (5.2.2)$$

The risk of the estimator, b , is defined as

$$R(b, A) = E \left[(b - \beta)' A (b - \beta) \right], \quad (5.2.3)$$

where A is a positive semi-definite matrix of order $p \times p$.

Applying the trace operator, we have

$$R(b, A) = \text{tr} \left\{ A E \left[(b - \beta)(b - \beta)' \right] \right\} = \text{tr} \{ A \text{MSE}(b) \}. \quad (5.2.4)$$

As

$$\begin{aligned} E(b) - \beta &= \text{bias}(b) \\ &= (X_*'X_*)^{-1} X_*'X\beta - \beta \\ &= (X_*'X_*)^{-1} X_*'(X_* - U)\beta - \beta \\ &= -(X_*'X_*)^{-1} X_*'U\beta, \end{aligned} \quad (5.2.5)$$

and

$$\text{Var}(b) = \sigma^2 (X_*'X_*)^{-1}, \quad (5.2.6)$$

we have

$$\text{MSE}(b) = \sigma^2 (X_*'X_*)^{-1} + (X_*'X_*)^{-1} X_*'U\beta\beta'U'X_*(X_*'X_*)^{-1}, \quad (5.2.7)$$

and

$$R(b, A) = R(b) = \text{tr} \left\{ A \left[\sigma^2 V + V X_*'U\beta\beta'U'X_*V \right] \right\}, \quad (5.2.8)$$

where

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

We now assume that we have auxiliary information on the vector of regression coefficients. If the auxiliary information is such that it can be written as linear equalities of the form

$$r = R\beta + \phi \quad (5.2.9)$$

where r is a $m \times 1$ stochastic vector which is assumed to be known, R is $m \times p$ fixed matrix with rank $m < p$, and it is assumed that the disturbance term has

$$\phi \sim N\left(0, \frac{\sigma^2}{k} I_m\right). \quad (5.2.10)$$

The constant k measures the precision of the prior information. The auxiliary information is assumed to be independent of the sample information (5.2.1), that is

$$E(\phi e') = E(e\phi') = 0. \quad (5.2.11)$$

Combining (5.2.1) with (5.2.9) giving the mixed model

$$\begin{pmatrix} Y \\ r \end{pmatrix} = \begin{pmatrix} X \\ R \end{pmatrix} \beta + \begin{pmatrix} e \\ \phi \end{pmatrix}. \quad (5.2.12)$$

Let

$$\tilde{Y} = \begin{pmatrix} Y \\ r \end{pmatrix}, \quad \tilde{X} = \begin{pmatrix} X \\ R \end{pmatrix} \text{ and } \tilde{e} = \begin{pmatrix} e \\ \phi \end{pmatrix}.$$

Then the model (5.2.12) can be rewritten as

$$\tilde{Y} = \tilde{X}\beta + \tilde{e} \quad (\text{where rank } \tilde{X} = p). \quad (5.2.13)$$

Using assumption (5.2.11), we have

$$E(\tilde{e}) = 0, \quad (5.2.14)$$

and

$$E(\tilde{\mathbf{e}}\tilde{\mathbf{e}}') = \sigma^2 \mathbf{W}, \quad (5.2.15)$$

where

$$\mathbf{W} = \begin{pmatrix} \mathbf{I}_n & 0 \\ 0 & \frac{\mathbf{I}_m}{k} \end{pmatrix}.$$

Since $\mathbf{W}$ is a positive definite matrix, then $\mathbf{W}$ can be written as

$$\mathbf{W} = \mathbf{G}'\mathbf{G}, \quad (5.2.16)$$

where $\mathbf{G}$ is nonsingular matrix. Premultiplying (5.2.13) by $\mathbf{G}'^{-1}$, we have

$$\mathbf{G}'^{-1}\tilde{\mathbf{Y}} = \mathbf{G}'^{-1}\tilde{\mathbf{X}}\beta + \mathbf{G}'^{-1}\tilde{\mathbf{e}}. \quad (5.2.17)$$

This can be written as

$$\mathbf{Y}_1 = \mathbf{X}_1\beta + \mathbf{e}_1, \quad (5.2.18)$$

where

$$\mathbf{Y}_1 = \mathbf{G}'^{-1}\tilde{\mathbf{Y}}, \quad \mathbf{X}_1 = \mathbf{G}'^{-1}\tilde{\mathbf{X}}, \quad \text{and} \quad \mathbf{e}_1 = \mathbf{G}'^{-1}\tilde{\mathbf{e}}. \quad (5.2.19)$$

Obviously,

$$E(\mathbf{e}_1) = \mathbf{G}'^{-1}E(\tilde{\mathbf{e}}) = \mathbf{0}, \quad (5.2.20)$$

and

$$\begin{aligned} E(\mathbf{e}_1\mathbf{e}_1') &= E(\mathbf{G}'^{-1}\tilde{\mathbf{e}}\tilde{\mathbf{e}}'\mathbf{G}^{-1}) \\ &= \mathbf{G}'^{-1}E(\tilde{\mathbf{e}}\tilde{\mathbf{e}}')\mathbf{G}^{-1} \\ &= \mathbf{G}'^{-1}\sigma^2\mathbf{W}\mathbf{G}^{-1} \\ &= \sigma^2\mathbf{G}'^{-1}\mathbf{G}'\mathbf{G}\mathbf{G}^{-1} \\ &= \sigma^2\mathbf{I}. \end{aligned} \quad (5.2.21)$$

Thus the least squares estimator, $\hat{\beta}_k$, of β in the model (5.2.18) is given by

$$\hat{\beta}_k = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y}_1 \quad (5.2.22)$$

Using (5.2.19), we have

$$\begin{aligned}
 \hat{\beta}_k &= (\tilde{X}'G^{-1}G'^{-1}\tilde{X})^{-1}\tilde{X}'G^{-1}G'^{-1}\tilde{Y} \\
 &= [\tilde{X}'(G'G)^{-1}\tilde{X}]^{-1}\tilde{X}'(G'G)^{-1}\tilde{Y} \\
 &= (\tilde{X}'W^{-1}\tilde{X})^{-1}\tilde{X}'W^{-1}\tilde{Y}.
 \end{aligned} \tag{5.2.23}$$

Since

$$W^{-1} = \begin{bmatrix} I_n & 0 \\ 0 & kI_m \end{bmatrix}, \tag{5.2.24}$$

$$\begin{aligned}
 \tilde{X}'W^{-1}\tilde{X} &= (X' \ R') \begin{bmatrix} I_n & 0 \\ 0 & kI_m \end{bmatrix} \begin{bmatrix} X \\ R \end{bmatrix} \\
 &= (X' \ kR') \begin{bmatrix} X \\ R \end{bmatrix} \\
 &= X'X + kR'R,
 \end{aligned} \tag{5.2.25}$$

and

$$\begin{aligned}
 \tilde{X}'W^{-1}\tilde{Y} &= (X' \ R') \begin{bmatrix} I_n & 0 \\ 0 & kI_m \end{bmatrix} \begin{bmatrix} Y \\ r \end{bmatrix} \\
 &= X'Y + kR'r.
 \end{aligned} \tag{5.2.26}$$

Using (5.2.25) and (5.2.26) in (5.2.23), we get

$$\hat{\beta}_k = (X'X + kR'R)^{-1}(X'Y + kR'r). \tag{5.2.27}$$

Therefore, the mixed estimator, b_k of β (depending on our assumption that we observed X_* instead of observing X) is given by

$$b_k = (X_*'X_* + kR'R)^{-1}(X_*'Y + kR'r). \tag{5.2.28}$$

Clearly,

$$\begin{aligned}
 E(b_k) - \beta &= \text{bias}(b_k) \\
 &= (X_*'X_* + kR'R)^{-1} (X_*'X\beta + kR'R\beta) - \beta \\
 &= (X_*'X_* + kR'R)^{-1} (X_*'X_*\beta + kR'R\beta - X_*'U\beta) - \beta \\
 &= \beta - (X_*'X_* + kR'R)^{-1} X_*'U\beta - \beta \\
 &= -(X_*'X_* + kR'R)^{-1} X_*'U\beta, \tag{5.2.29}
 \end{aligned}$$

and

$$\begin{aligned}
 \text{Var}(b_k) &= (X_*'X_* + kR'R)^{-1} (\sigma^2 X_*'X_* + \sigma^2 kR'R) (X_*'X_* + kR'R)^{-1} \\
 &= \sigma^2 (X_*'X_* + kR'R)^{-1}. \tag{5.2.30}
 \end{aligned}$$

Hence,

$$\begin{aligned}
 \text{MSE}(b_k) &= E \left[(b_k - \beta)(b_k - \beta)' \right] \\
 &= \sigma^2 (X_*'X_* + kR'R)^{-1} + (X_*'X_* + kR'R)^{-1} X_*'U\beta\beta'U'X_* (X_*'X_* + kR'R)^{-1}. \tag{5.2.31}
 \end{aligned}$$

Let

$$(X_*'X_*)^{-1} = V \text{ and } (k^{-1}I + RVR')^{-1} = S_k. \tag{5.2.32}$$

Applying the matrix inversion formula to the matrix $(X_*'X_* + kR'R)^{-1}$, we have

$$\begin{aligned}
 (X_*'X_* + kR'R)^{-1} &= (V^{-1} + kR'R)^{-1} \\
 &= V - VR'(k^{-1}I + RVR')^{-1}RV \\
 &= V - VR'S_kRV. \tag{5.2.33}
 \end{aligned}$$

This gives (5.2.31) as

$$\text{MSE}(b_k) = \sigma^2 V - \sigma^2 VR'S_k RV + (V - VR'S_k RV)X'_*U\beta\beta'U'X_*(V - VR'S_k RV). \quad (5.2.34)$$

The risk of the estimator, b_k , is given by

$$\begin{aligned} R(b_k) &= \text{tr}\{\text{AMSE}(b_k)\} \\ &= \text{tr}\left\{A\left[\sigma^2 V - \sigma^2 VR'S_k RV + (V - VR'S_k RV)X'_*U\beta\beta'U'X_*(V - VR'S_k RV)\right]\right\} \end{aligned} \quad (5.2.35)$$

where A is a given positive semi-definite matrix of order $p \times p$.

From (5.2.8) and (5.2.35), we have,

$$\begin{aligned} R(b) - R(b_k) &= \text{tr}\left\{A\left[\sigma^2 V + VX'_*U\beta\beta'U'X_*V\right]\right\} \\ &\quad - \text{tr}\left\{A\left[\sigma^2 V - \sigma^2 VR'S_k RV + (V - VR'S_k RV)X'_*U\beta\beta'U'X_*(V - VR'S_k RV)\right]\right\} \\ R(b) - R(b_k) &= \text{tr}\left\{A\left[\sigma^2 VR'S_k RV - VR'S_k RVX'_*U\beta\beta'U'X_*VR'S_k RV\right]\right\} \\ &\quad + \text{tr}\left\{A\left[VX'_*U\beta\beta'U'X_*VR'S_k RV + VR'S_k RVX'_*U\beta\beta'U'X_*V\right]\right\} \\ &\geq \text{tr}\left\{A\left[\sigma^2 VR'S_k RV - VR'S_k RVX'_*U\beta\beta'U'X_*VR'S_k RV\right]\right\}. \end{aligned}$$

Therefore,

$$R(b) - R(b_k) \geq \text{tr}\left\{A\left[VR'S_k(\sigma^2 S_k^{-1} - RVX'_*U\beta\beta'U'X_*VR')S_k RV\right]\right\}. \quad (5.2.36)$$

Since A is a positive semi-definite matrix, then (5.2.36) is positive semi-definite iff

$$\sigma^2 S_k^{-1} - RVX'_*U\beta\beta'U'X_*VR' \geq 0.$$

Since

$$\sigma^2 S_k^{-1} - RVX_*' U \beta \beta' U' X_* VR' = \sigma^2 S_k^{-1/2} \left[I - \sigma^{-2} S_k^{1/2} RVX_*' U \beta \beta' U' X_* VR' S_k^{1/2} \right] S_k^{-1/2},$$

then

$$\sigma^2 S_k^{-1} - RVX_*' U \beta \beta' U' X_* VR' \geq 0 \text{ iff}$$

$$I - \sigma^{-2} S_k^{1/2} RVX_*' U \beta \beta' U' X_* VR' S_k^{1/2} \geq 0.$$

Now this matrix has eigenvalue 1 with multiplicity (m-1), and eigenvalue

$1 - \sigma^{-2} \beta' U' X_* VR' S_k RVX_*' U \beta$ with multiplicity one. Therefore,

$$1 - \sigma^{-2} S_k^{1/2} RVX_*' U \beta \beta' U' X_* VR' S_k^{1/2} \geq 0 \text{ iff}$$

$$\sigma^{-2} \beta' U' X_* VR' S_k RVX_*' U \beta \leq 1. \quad (5.2.37)$$

Let

$$\sigma^{-2} \beta' U' X_* VR' S_k RVX_*' U \beta = f(k). \quad (5.2.38)$$

Differentiating (5.2.38) with respect to k, we obtain

$$\frac{\partial f(k)}{\partial k} = \sigma^{-2} k^{-2} \beta' U' X_* VR' S_k S_k RVX_*' U \beta > 0.$$

(see, Toutenburg p. 201).

Thus (5.2.38) is an increasing function of the constant k.

Since $\lim_{k \rightarrow 0^+} f(k) = 0$, then there exists a value k^* such that $f(k^*) = 1$. So we have the

$$k \rightarrow 0^+$$

following theorem.

Theorem 5.1:

If the independent variables are subject to errors of measurement, then for all k lies in

$$[0, k^*], \quad R(b) - R(b_k) \geq 0.$$

5.3 Prior Information and Errors in Dependent Variable

Suppose that some of the observations of the dependent variable of the model (5.2.1) are subject to shift in means and variances. So instead of Y in the model (5.2.1), we observe

$$Y_* = Y + Z, \quad (5.3.1)$$

where Z is an $n \times 1$ vector of random variables with

$$E(Z) = v, \quad E[(Z - v)(Z - v)'] = \sigma_1^2 I_n, \quad (5.3.2)$$

and

$$E[(Z - v)e'] = E[e(Z - v)'] = 0. \quad (5.3.3)$$

We also assume that one or more of the elements of v are nonzero and σ_1^2 is an unknown positive constant.

Then the least squares estimator, b^* , of the vector of parameters, β , is given by

$$b^* = (X'X)^{-1} X'Y_*. \quad (5.3.4)$$

Clearly,

$$\begin{aligned} E(b^*) - \beta &= \text{bias}(b^*) \\ &= (X'X)^{-1} X'(X\beta + v) - \beta \\ &= (X'X)^{-1} X'v, \end{aligned} \quad (5.3.5)$$

and

$$\begin{aligned}
 \text{Var}(b^*) &= E\left\{\left[b^* - E(b^*)\right]\left[b - E(b^*)\right]'\right\} \\
 &= (X'X)^{-1}X'(\sigma^2 + \sigma_1^2)X(X'X)^{-1} \\
 &= (\sigma^2 + \sigma_1^2)(X'X)^{-1}.
 \end{aligned} \tag{5.3.6}$$

Hence

$$\begin{aligned}
 \text{MSE}(b^*) &= E\left[\left(b^* - \beta\right)\left(b^* - \beta\right)'\right] \\
 &= (\sigma^2 + \sigma_1^2)(X'X)^{-1} + (X'X)^{-1}X'\mathbf{v}\mathbf{v}'X(X'X)^{-1},
 \end{aligned} \tag{5.3.7}$$

and

$$\begin{aligned}
 R(b^*) &= \text{tr}\{A \text{MSE}(b^*)\} \\
 &= \text{tr}\left\{A\left[(\sigma^2 + \sigma_1^2)(X'X)^{-1} + (X'X)^{-1}X'\mathbf{v}\mathbf{v}'X(X'X)^{-1}\right]\right\} \\
 &= \text{tr}\left\{A\left[(\sigma^2 + \sigma_1^2)T + TX'\mathbf{v}\mathbf{v}'XT\right]\right\},
 \end{aligned} \tag{5.3.8}$$

where A is positive semi-definite matrix, and

$$T = (X'X)^{-1}. \tag{5.3.9}$$

Suppose that in addition to the classical model (5.2.1) we have the auxiliary information of the type (5.2.9).

Then the mixed estimator, b_k^* , of the vector of parameters, β , (based on Y_*) is given by

$$b_k^* = (X'X + kR'R)^{-1}(X'Y_* + kR'r). \tag{5.3.10}$$

We observe that

$$\begin{aligned}
 E(\mathbf{b}_k^*) - \beta &= \text{bias}(\mathbf{b}_k^*) \\
 &= (\mathbf{X}'\mathbf{X} + k\mathbf{R}'\mathbf{R})^{-1} (\mathbf{X}'\mathbf{X}\beta + \mathbf{X}'\mathbf{v} + k\mathbf{R}'\mathbf{R}\beta) - \beta \\
 &= (\mathbf{X}'\mathbf{X} + k\mathbf{R}'\mathbf{R})^{-1} \mathbf{X}'\mathbf{v}, \tag{5.3.11}
 \end{aligned}$$

and

$$\begin{aligned}
 V(\mathbf{b}_k^*) &= (\mathbf{X}'\mathbf{X} + k\mathbf{R}'\mathbf{R})^{-1} \left[\mathbf{X}'(\sigma^2 + \sigma_1^2)\mathbf{X} + k\sigma^2\mathbf{R}'\mathbf{R} \right] (\mathbf{X}'\mathbf{X} + k\mathbf{R}'\mathbf{R})^{-1} \\
 &= (\mathbf{X}'\mathbf{X} + k\mathbf{R}'\mathbf{R})^{-1} \left[\sigma^2(\mathbf{X}'\mathbf{X} + k\mathbf{R}'\mathbf{R}) + \sigma_1^2\mathbf{X}'\mathbf{X} \right] (\mathbf{X}'\mathbf{X} + k\mathbf{R}'\mathbf{R})^{-1} \\
 &= (\sigma^2 + \sigma_1^2)(\mathbf{X}'\mathbf{X} + k\mathbf{R}'\mathbf{R})^{-1} - k\sigma_1^2(\mathbf{X}'\mathbf{X} + k\mathbf{R}'\mathbf{R})^{-1} \mathbf{R}'\mathbf{R}(\mathbf{X}'\mathbf{X} + k\mathbf{R}'\mathbf{R})^{-1}. \tag{5.3.12}
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 \text{MSE}(\mathbf{b}_k^*) &= E \left[(\mathbf{b}_k^* - \beta)(\mathbf{b}_k^* - \beta)' \right] \\
 &= (\sigma^2 + \sigma_1^2)(\mathbf{X}'\mathbf{X} + k\mathbf{R}'\mathbf{R})^{-1} - k\sigma_1^2(\mathbf{X}'\mathbf{X} + k\mathbf{R}'\mathbf{R})^{-1} \mathbf{R}'\mathbf{R}(\mathbf{X}'\mathbf{X} + k\mathbf{R}'\mathbf{R})^{-1} \\
 &\quad + (\mathbf{X}'\mathbf{X} + k\mathbf{R}'\mathbf{R})^{-1} \mathbf{X}'\mathbf{v}\mathbf{v}'\mathbf{X}(\mathbf{X}'\mathbf{X} + k\mathbf{R}'\mathbf{R})^{-1}, \tag{5.3.13}
 \end{aligned}$$

and

$$\mathbf{R}(\mathbf{b}_k^*) = \text{tr} \left\{ \mathbf{A} \left[(\sigma^2 + \sigma_1^2)\mathbf{Z}_k - k\sigma_1^2\mathbf{Z}_k\mathbf{R}'\mathbf{R}\mathbf{Z}_k + \mathbf{Z}_k\mathbf{X}'\mathbf{v}\mathbf{v}'\mathbf{X}\mathbf{Z}_k \right] \right\}, \tag{5.3.14}$$

where

$$\mathbf{Z}_k = (\mathbf{X}'\mathbf{X} + k\mathbf{R}'\mathbf{R})^{-1}. \tag{5.3.15}$$

Let

$$(\mathbf{k}^{-1}\mathbf{I} + \mathbf{R}\mathbf{T}\mathbf{R}')^{-1} = \mathbf{F}_k \tag{5.3.16}$$

and applying matrix inversion formula to Z_k , we have

$$Z_k = T - TR'F_kRT, \quad (5.3.17)$$

where T is given by (5.3.9).

Then (5.3.14) can be rewritten as

$$\begin{aligned} R(b_k^*) &= \text{tr} \left\{ A \left[(\sigma^2 + \sigma_1^2) T + TX' \mathbb{U} \mathbb{U}' XT \right] \right\} \\ &\quad + \text{tr} \left\{ A \left[-\sigma^2 TR'F_kRT - \sigma_1^2 TR'F_kRT \right. \right. \\ &\quad \left. \left. - k^{-1} \sigma_1^2 TR'F_kF_kRT - TX' \mathbb{U} \mathbb{U}' XTR'F_kRT \right. \right. \\ &\quad \left. \left. - TR'F_kRTX' \mathbb{U} \mathbb{U}' XT + TR'F_kRTX' \mathbb{U} \mathbb{U}' XTR'F_kRT \right] \right\}. \end{aligned} \quad (5.3.18)$$

Using (5.3.8) and (5.3.18), we have,

$$\begin{aligned} R(b^*) - R(b_k^*) &= \text{tr} \left\{ A \left(\sigma^2 + \sigma_1^2 \right) TR'F_kRT - TR'F_kRTX' \mathbb{U} \mathbb{U}' XTR'F_kRT \right\} \\ &\quad + \text{tr} \left\{ A \left[k^{-1} \sigma_1^2 TR'F_kF_kRT + TX' \mathbb{U} \mathbb{U}' XTR'F_kRT \right. \right. \\ &\quad \left. \left. + TR'F_kRTX' \mathbb{U} \mathbb{U}' XT \right] \right\} \\ &\geq \text{tr} \left\{ A \left[(\sigma^2 + \sigma_1^2) TR'F_kRT - TR'F_kRTX' \mathbb{U} \mathbb{U}' XTR'F_kRT \right] \right\}. \end{aligned} \quad (5.3.19)$$

Therefore,

$$R(b^*) - R(b_k^*) \geq \text{tr} \left\{ A \left[TR'F_k \left((\sigma^2 + \sigma_1^2) F_k^{-1} - RTX' \mathbb{U} \mathbb{U}' XTR' \right) F_kRT \right] \right\}. \quad (5.3.20)$$

As in section (5.2), it can be shown that

$$R(b^*) - R(b_k^*) \geq 0 \text{ iff } (\sigma^2 + \sigma_1^2)^{-1} \mathbb{U}' XTR'F_kRTX' \mathbb{U} \leq 1.$$

Let

$$(\sigma^2 + \sigma_1^2)^{-1} \mathbb{U}' XTR'F_kRTX' \mathbb{U} = h(k). \quad (5.3.21)$$

Differentiating (5.3.21) with respect to k , we obtain

$$\frac{\partial h(k)}{\partial k} = (\sigma^2 + \sigma_1^2)^{-1} k^{-2} \mathbf{v}' \mathbf{X} \mathbf{T} \mathbf{R}' \mathbf{F}_k \mathbf{F}_k \mathbf{R} \mathbf{T} \mathbf{X}' \mathbf{v} > 0.$$

Thus (5.3.21) is an increasing function of the constant k . Since $\lim_{k \rightarrow 0^+} h(k) = 0$, then there exists a value k_1 , such that $h(k_1) = 1$. So we have the following theorem.

Theorem 5.2:

If some of the observations of the dependent variable are subject to shift in means and variances, then for all k lies in $[0, k_1]$, $R(b^*) - R(b_k^*) \geq 0$.

CHAPTER VI

IDENTIFICATION OF LOCATION SHIFT OUTLIERS

6.1 Introduction

Data collected for analysis often contain discordant observations. In regression analysis such observations can occur either in the response variable, or in the explanatory variables or in both. Whatever be the source of these observations, it is important to assess them.

There are two major approaches by which discordant observations in regression can be handled. One is by robust regression analysis and the other is by dual applications of informal diagnostics and formal testing procedures. Robust regression approaches are employed to obtain robust estimators for the vector of unknown regression coefficients. These estimators are relatively unaffected by discordant observations. The residuals obtained from a robust fit should help to flag cases off the regression line. Huber (1981), Rousseeuw and Leroy (1987), Yohai and Zamar (1988) and Simpson, Ruppert and Carroll (1992) discussed many efficient highly robust approaches. However, most of the robust approaches are computationally intensive. Also the distribution of residuals from these fits is unknown, particularly for small samples. So, criteria for identification of an observation as outlier can be constructed using Monte Carlo technique only.

The second approach has also received considerable attention of authors. Belsley, Kuh and Welsch (1980), Cook (1977), Cook and Weisberg (1982) and Chatterjee and Hadi (1986, 1988) described most of the informal diagnostics. These diagnostics are used to select candidate outliers.

Then each candidate outlier is tested for outlyingness. Details on the construction of the various test statistics and outlier identification criteria can be found in Marasinghe (1985) and Kianifard and Swallow (1989, 1990).

Often the available outlier identification methods do not succeed in detecting multiple outliers because they are affected by masking and swamping. Hadi and Simonoff (1993) proposed two test procedures for the detection of multiple outliers that appear to be less sensitive to masking and swamping effects. Their procedures also attempt to separate the data into a set of clean observations and a set of potential outliers. The potential outlier are then tested for outlyingness using appropriate scaled version of the prediction error.

Although an observation may be an outlier, it may contain information that is useful to the analysis. The available approaches for detecting outlier cannot assess the information contained in the outlying observations. Rather than identifying an observation as an outlier, it is important to examine whether the outlying observation is detrimental, in some sense, to the analysis or not. Ali (1994) assessed the effect of the outliers and the effect of deleting observation on the least squares estimate of the vector of unknown regression coefficients. He classified location shift outliers into

two categories - moderate location shift outliers and the extreme location shift outliers.

He developed a test for identifying extreme location shift outliers.

In this chapter the problem of identification and accommodation of location shift outliers is considered. Three simple measures are proposed to identify potential location shift outliers. To illustrate our approach a simple example is considered.

6.2 Location Shift Outliers

A general approach of outlier detection is to form a subset of the observations that is presumably free of outliers, and then test the outlyingness of the remaining observations. The subset of the observations that is free of outliers will be called the subset of genuine observations. The complementary subset will be called the subset of outlying observations. One way of finding the subset of outlying observations is to determine the subset of size k , the deletion of which produces the largest reduction in the residual sum of squares. Another way of finding the subset of genuine observations was proposed by Simonoff (1991).

We consider a location shift model:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ X_2 & I_k \end{bmatrix} \begin{bmatrix} \beta \\ v_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} . \quad (6.2.1)$$

Where Y_1 and Y_2 are $(n-k) \times 1$ and $k \times 1$ vectors of genuine and location shift observations respectively, β is a $p \times 1$ vector of unknown regression coefficients, v_2 is a $k \times 1$ vector of non-stochastic aberrations, $X = (X_1' \ X_2')'$ is an $n \times p$ design matrix, 0

stands for the null matrix, $e = (e_1' \ e_2')'$ is a vector of random errors with $E(e) = 0$ and $E(e \ e') = \sigma^2 I_n$, σ^2 is a positive constant.

We also assume that $\text{rank}(X) = \text{rank}(X_1) = p < (n - k)$. Under this model outliers may occur in Y_2 , X_2 or both, see Cook and Weisberg (1982, p. 20).

The weighted mean squared error of $\hat{\beta} = (X'X)^{-1} X'Y$, the least squares estimator of β based on $Y = (Y_1' \ Y_2')'$ is

$$\begin{aligned} D(\hat{\beta}) &= E\left[(\hat{\beta} - \beta)' W(\hat{\beta} - \beta)\right] \\ &= v_2' R W R' v_2 + \sigma^2 \text{tr}\left[W(X'X)^{-1}\right] \end{aligned} \quad (6.2.2)$$

where $R = X_2(X'X)^{-1}$, W is a given $p \times p$ non-negative definite matrix, and $\text{tr}(F)$ denotes the trace of a square matrix F . Similarly, the weighted mean squared error of $b = (X_1'X_1)^{-1} X_1'Y_1$, the best linear unbiased estimator of β based on the vector of genuine observations Y_1 is

$$\begin{aligned} D(b) &= E\left[(b - \beta)' W(b - \beta)\right] \\ &= \sigma^2 \text{tr}\left[W(X_1'X_1)^{-1}\right]. \end{aligned}$$

It can be checked from Rao (1973, P.33) that

$$(X_1'X_1)^{-1} = (X'X - X_2'X_2)^{-1} = (X'X)^{-1} + R'(I_k - M)^{-1}R, \quad (6.2.3)$$

where $M = X_2(X'X)^{-1}X_2'$ and R is defined in (6.2.2).

The weighted mean squared error of b can be expressed as

$$D(b) = \sigma^2 \text{tr}\left[W(X'X)^{-1}\right] + \sigma^2 \text{tr}\left[WR'(I_k - M)^{-1}R\right]. \quad (6.2.4)$$

We observe that

$$\begin{aligned} |X_1'X_1| &= |X'X - X_2'X_2| \\ &= |X'X||I_k - M|. \end{aligned}$$

Since X_1 is a matrix of full rank, $|X_1'X_1| > 0$. We also have $|X'X| > 0$.

Thus, $|I_k - M| > 0$.

Furthermore, $I_n - X(X'X)^{-1}X'$ is a positive semi-definite matrix. The matrix $(I_k - M)$ is a block diagonal matrix of $I_n - X(X'X)^{-1}X'$. Hence $(I_n - M)$ is positive definite.

Theorem 6.1

For any given non-negative definite matrix W , if $v_2'(I_k - M)v_2 / \sigma^2 \leq 1$,

then $[D(b) - D(\hat{\beta})] / \sigma^2 \geq 0$.

Proof:

We observe from (6.2.2) and (6.2.4) that

$$[D(b) - D(\hat{\beta})] / \sigma^2 = \text{tr} \left\{ (RWR') \left[(I_k - M)^{-1} - v_2 v_2' / \sigma^2 \right] \right\}.$$

Since $(I_k - M)^{-1}$ is a positive definite, there exists a $k \times k$ non-singular matrix Q such that $Q'Q = [I_k - M]^{-1}$.

Thus,

$$(I_k - M)^{-1} - v_2 v_2' / \sigma^2 = Q' [I_k - Q'^{-1} v_2 v_2' Q^{-1} / \sigma^2] Q.$$

The matrix $I_k - Q'^{-1}v_2v_2'Q^{-1} / \sigma^2$ has eigenvalue 1 with multiplicity $(k-1)$, and eigenvalue $1 - v_2'(I_k - M)v_2 / \sigma^2$ with multiplicity one. Therefore, if

$$1 - v_2'(I_k - M)v_2 / \sigma^2 \geq 0, \text{ then } \left[D(b) - D(\hat{\beta}) \right] / \sigma^2 \geq 0.$$

This completes the proof.

If $v_2'(I_k - M)v_2 / \sigma^2 \leq 1$, then outright rejection of Y_2 may result in poorer least squares estimation. On the other hand, if $v_2'(I_k - M)v_2 / \sigma^2 > 1$, then Y_2 is detrimental to the analysis.

6.3 Measures for Identification of Potential Location Shift Outliers

In this section we consider three measures to identify potential location shift outliers.

We call Y_2 the vector of potential location shift outliers, if

$$v_2'(I_k - M)v_2 > \sigma^2.$$

If a set of vector of observations is not potential outliers, then its accommodation in the analysis will increase the precision of the least squares estimates of the unknown regression coefficients. Accommodation of potential location shift outliers will result in poorer least squares estimation.

Theorem 6.2

If $v_2'(I_k - M)v_2 \leq \sigma^2$, then

$$E \left[(Y - Xb)' (I_n - H)(Y - Xb) \right] \leq \sigma^2(n - p + 1),$$

where

$$Y = (Y_1' \ Y_2')', \quad X' = (X_1' \ X_2'), \quad H = X(X'X)^{-1}X',$$

$$b = (X_1'X_1)^{-1}X_1'Y_1 \text{ and } M \text{ is given in (6.2.3).}$$

Proof:

Letting $v' = (0' \ v_2')$, and $e' = (e_1' \ e_2')$, we can rewrite (6.2.1) as

$$Y = X\beta + v + e.$$

Now,

$$\begin{aligned} Y - Xb &= (X\beta + v + e) - X(X_1'X_1)^{-1}X_1'Y_1 \\ &= (X\beta + v + e) - X(X_1'X_1)^{-1}X_1'(X_1\beta + e_1) \\ &= (v + e) - X(X_1'X_1)^{-1}X_1'e_1. \end{aligned}$$

Thus,

$$(I_n - H)(Y - Xb) = (I_n - H)(v + e),$$

and

$$(Y - Xb)'(I_n - H)(Y - Xb) = (v + e)'(I_n - H)(v + e).$$

Therefore,

$$\begin{aligned} E\left[(Y - Xb)'(I_n - H)(Y - Xb)\right] &= E\left[(v + e)'(I_n - H)(v + e)\right] \\ &= \text{tr}\left[(I_n - H)(vv' + \sigma^2 I_n)\right] \\ &= v'(I_n - H)v + (n - p)\sigma^2 \\ &= v_2'(I_k - M)v_2 + (n - p)\sigma^2. \end{aligned}$$

Hence if $\mathbf{v}_2'(\mathbf{I}_k - \mathbf{M})\mathbf{v}_2 \leq \sigma^2$, then

$$E\left[(\mathbf{Y} - \mathbf{Xb})'(\mathbf{I}_n - \mathbf{H})(\mathbf{Y} - \mathbf{Xb})\right] \leq (n - p + 1)\sigma^2.$$

This completes the proof.

Next,

$$E\left[(\mathbf{Y}_1 - \mathbf{X}_1\mathbf{b})'(\mathbf{Y}_1 - \mathbf{X}_1\mathbf{b})\right] = (n - p - k)\sigma^2$$

Hence if

$$(\mathbf{Y} - \mathbf{Xb})'(\mathbf{I}_n - \mathbf{H})(\mathbf{Y} - \mathbf{Xb}) \leq \left(\frac{n - p + 1}{n - p - k}\right)(\mathbf{Y}_1 - \mathbf{X}_1\mathbf{b})'(\mathbf{Y}_1 - \mathbf{X}_1\mathbf{b}), \quad (6.3.1)$$

then accommodation of $\mathbf{Y}_2$, the vector of location shift observations may be beneficial to the analysis. Otherwise, $\mathbf{Y}_2$ may be detrimental to the analysis.

Theorem 6.3

If $\mathbf{v}_2'(\mathbf{I}_k - \mathbf{M})\mathbf{v}_2 \leq \sigma^2$, then

$$E\left[(\mathbf{Y}_2 - \mathbf{X}_2\hat{\beta})'(\mathbf{I}_k - \mathbf{M})^{-1}(\mathbf{Y}_2 - \mathbf{X}_2\hat{\beta})\right] \leq (k + 1)\sigma^2,$$

where

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \mathbf{Y}' = (\mathbf{Y}_1' \mathbf{Y}_2'), \text{ and } \mathbf{M} = \mathbf{X}_2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_2'.$$

Proof:

It follows from (6.2.1) that

$$\mathbf{Y}_2 = \mathbf{X}_2\beta + \mathbf{v}_2 + \mathbf{e}_2.$$

Hence,

$$\begin{aligned} Y_2 - X_2 \hat{\beta} &= Y_2 - X_2 (X'X)^{-1} X'Y \\ &= (v_2 + e_2) - X_2 (X'X)^{-1} (X'v + X'e) \end{aligned}$$

Now

$$X'v = X'_2 v \quad , \quad X'e = X'_1 e_1 + X'_2 e_2,$$

and

$$\begin{aligned} Y_2 - X_2 \hat{\beta} &= (v_2 + e_2) - X_2 (X'X)^{-1} (X'_2 v + X'_1 e_1 + X'_2 e_2) \\ &= (I_k - M)(v_2 + e_2) - RX'_1 e_1, \end{aligned}$$

where

$$R = X_2 (X'X)^{-1}.$$

Therefore,

$$\begin{aligned} (Y_2 - X_2 \hat{\beta})' (I_k - M)^{-1} (Y_2 - X_2 \hat{\beta}) &= (v_2 + e_2)' (I_k - M)(v_2 + e_2) - e'_1 X_1 R' (v_2 + e_2) \\ &\quad - (v_2 + e_2)' RX'_1 e_1 + e'_1 X_1 R' (I_k - M)^{-1} RX'_1 e_1. \end{aligned}$$

Thus,

$$\begin{aligned} E \left[(Y_2 - X_2 \hat{\beta})' (I_k - M)^{-1} (Y_2 - X_2 \hat{\beta}) \right] &= \sigma^2 \text{tr}(I_k - M) + v'_2 (I_k - M) v_2 \\ &\quad + \sigma^2 \text{tr} \left[(I_k - M)^{-1} (I_k - M) M \right] \\ &= k\sigma^2 + v'_2 (I_k - M) v_2. \end{aligned}$$

If $v'_2 (I_k - M) v_2 \leq \sigma^2$, then

$$E \left[(Y_2 - X_2 \hat{\beta})' (I_k - M)^{-1} (Y_2 - X_2 \hat{\beta}) \right] \leq (k+1)\sigma^2.$$

This completes the proof.

Hence, if

$$\left(Y_2 - X_2 \hat{\beta} \right)' (I_k - M)^{-1} (Y_2 - X_2 \hat{\beta}) \leq \left(\frac{k+1}{n-p-k} \right) (Y_1 - X_1 b)' (Y_1 - X_1 b), \quad (6.3.2)$$

then accommodation of Y_2 may be beneficial to the analysis. Otherwise, Y_2 may be detrimental to the analysis.

Theorem 6.4

If $v_2'(I_k - M)v_2 \leq \sigma^2$, then

$$E \left[(Y_2 - X_2 b)' (I_k - M)(Y_2 - X_2 b) \right] \leq (k+1)\sigma^2.$$

Proof:

Clearly

$$Y_2 - X_2 b = (v_2 + e_2) - X_2 (X_1' X_1)^{-1} X_1' e_1,$$

and

$$\begin{aligned} (Y_2 - X_2 b)' (I_k - M)(Y_2 - X_2 b) &= (v_2 + e_2)' (I_k - M)(v_2 + e_2) \\ &\quad - e_1' X_1 (X_1' X_1)^{-1} X_2' (I_k - M)(v_2 + e_2) \\ &\quad - (v_2 + e_2)' (I_k - M) X_2 (X_1' X_1)^{-1} X_1' e_1 \\ &\quad + e_1' X_1 (X_1' X_1)^{-1} X_2' (I_k - M) X_2 (X_1' X_1)^{-1} X_1' e_1. \end{aligned}$$

Hence,

$$\begin{aligned} E \left[(Y_2 - X_2 b)' (I_k - M)(Y_2 - X_2 b) \right] &= v_2' (I_k - M) v_2 + \sigma^2 \text{tr}(I_k - M) \\ &\quad + \sigma^2 \text{tr} \left[(I_k - M) X_2 (X_1' X_1)^{-1} X_1' X_1 (X_1' X_1)^{-1} X_2' \right] \end{aligned}$$

$$\begin{aligned}
&= \mathbf{v}_2' (\mathbf{I}_k - \mathbf{M}) \mathbf{v}_2 + \sigma^2 \text{tr}(\mathbf{I}_k - \mathbf{M}) \\
&\quad + \sigma^2 \text{tr} \left[(\mathbf{I}_k - \mathbf{M}) \mathbf{X}_2 (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_2' \right]. \tag{6.3.3}
\end{aligned}$$

Now using (6.2.3) in (6.3.3), we obtain

$$\begin{aligned}
\mathbb{E} \left[(\mathbf{Y}_2 - \mathbf{X}_2 \mathbf{b})' (\mathbf{I}_k - \mathbf{M}) (\mathbf{Y}_2 - \mathbf{X}_2 \mathbf{b}) \right] &= \mathbf{v}_2' (\mathbf{I}_k - \mathbf{M}) \mathbf{v}_2 + \sigma^2 \text{tr}(\mathbf{I}_k - \mathbf{M}) + \sigma^2 \text{tr}[(\mathbf{I}_k - \mathbf{M}) \mathbf{M}] \\
&\quad + \sigma^2 \text{tr}[(\mathbf{I}_k - \mathbf{M}) \mathbf{M} (\mathbf{I}_k - \mathbf{M})^{-1} \mathbf{M}] \\
&= \mathbf{v}_2' (\mathbf{I}_k - \mathbf{M}) \mathbf{v}_2 + \sigma^2 \text{tr}(\mathbf{I}_k - \mathbf{M}) \\
&\quad + \sigma^2 \text{tr}[(\mathbf{I}_k - \mathbf{M}) \mathbf{M} (\mathbf{I}_k - \mathbf{M})^{-1}] \\
&= \mathbf{v}_2' (\mathbf{I}_k - \mathbf{M}) \mathbf{v}_2 + \sigma^2 \text{tr}(\mathbf{I}_k) \\
&= \mathbf{v}_2' (\mathbf{I}_k - \mathbf{M}) \mathbf{v}_2 + K \sigma^2.
\end{aligned}$$

Therefore if $\mathbf{v}_2' (\mathbf{I}_k - \mathbf{M}) \mathbf{v}_2 \leq \sigma^2$, then

$$\mathbb{E} \left[(\mathbf{Y}_2 - \mathbf{X}_2 \mathbf{b})' (\mathbf{I}_k - \mathbf{M}) (\mathbf{Y}_2 - \mathbf{X}_2 \mathbf{b}) \right] \leq (k+1) \sigma^2.$$

This completes the proof.

As before, if

$$(\mathbf{Y}_2 - \mathbf{X}_2 \mathbf{b})' (\mathbf{I}_k - \mathbf{M}) (\mathbf{Y}_2 - \mathbf{X}_2 \mathbf{b}) \leq \left(\frac{k+1}{n-p-k} \right) (\mathbf{Y}_1 - \mathbf{X}_1 \mathbf{b})' (\mathbf{Y}_1 - \mathbf{X}_1 \mathbf{b}),$$

then accommodation of $\mathbf{Y}_2$ may be beneficial to the analysis. Otherwise, $\mathbf{Y}_2$ may be detrimental to the analysis.

$$Y_1'X_1 = (69.7190 \quad 24.6642 \quad 18.8555 \quad -5.6128),$$

$$\hat{\beta}' = (5.0184 \quad 1.7899 \quad 1.3683 \quad -1.7119),$$

$$b' = (4.7911 \quad 1.7899 \quad 1.3683 \quad -1.2633),$$

$$I_k - M = 0.7278,$$

$$(I_k - M)^{-1} = 1.3740, \quad x_2'\hat{\beta} = 7.9286,$$

$$x_2'b = 6.9387,$$

$$\begin{aligned} (Y - X\hat{\beta})' &= (-0.2987 \quad 0.8565 \quad -0.1091 \quad 0.0961 \quad 0.3141 \quad -0.4477 \quad 1.5727 \quad -0.7231 \\ &\quad -0.8176 \quad -0.7176 \quad -0.8653 \quad -0.6225 \quad 2.6464 \quad 1.6478 \quad -1.3314 \quad -1.2014) \end{aligned}$$

$$\begin{aligned} (Y - Xb)' &= (-0.5200 \quad 1.5324 \quad -0.3304 \quad 0.7720 \quad 0.0928 \quad 0.2282 \quad 1.3514 \quad 0.0472 \\ &\quad -0.5903 \quad -0.4899 \quad -0.6380 \quad -0.3952 \quad 3.6363 \quad 1.1125 \quad -1.1041 \quad -0.9741), \end{aligned}$$

$$(\hat{\beta} - b)' = (0.2273 \quad 0 \quad 0 \quad -0.4486),$$

$$(Y - Xb)'(Y - Xb) = 22.9929,$$

$$(\hat{\beta} - b)'X'X(\hat{\beta} - b) = 3.5997,$$

$$\begin{aligned} (Y_1 - X_1b)'(Y_1 - X_1b) &= (Y - Xb)'(Y - Xb) - (Y_2 - X_2b)'(Y_2 - X_2b) \\ &= 9.7702 \end{aligned}$$

$$\begin{aligned} (Y - Xb)'(I_n - H)(Y - Xb) &= (Y - Xb)'(Y - Xb) - (\hat{\beta} - b)'X'X(\hat{\beta} - b) \\ &= 19.3932 \end{aligned}$$

$$(Y_2 - X_2\hat{\beta})'(I_k - M)^{-1}(Y_2 - X_2\hat{\beta}) = 9.6227, \text{ and}$$

$$(Y_2 - X_2b)'(I_k - M)(Y_2 - X_2b) = 9.6235.$$

Clearly,

$$(Y - Xb)' (I_n - H)(Y - Xb) > \frac{(n - p + 1)}{(n - p - k)} (Y_1 - X_1 b)' (Y_1 - X_1 b) ,$$

$$(Y_2 - X_2 \hat{\beta})' (I_k - M)^{-1} (Y_2 - X_2 \hat{\beta}) > \frac{k + 1}{n - p - k} (Y_1 - X_1 b)' (Y_1 - X_1 b) ,$$

and

$$(Y_2 - X_2 b)' (I_k - M)(Y_2 - X_2 b) > \frac{k + 1}{n - p - k} (Y_1 - X_1 b)' (Y_1 - X_1 b) .$$

Therefore, all the measures support the fact that a shift in mean of observation Y_{13} has occurred. These measures can be used to identify potential outliers. The main advantage of these measures is that they are computationally very simple.

REFERENCES

- Ali, M. A. (1994). Identifying the location shift outliers that matter. *Sankhya, Series A, Vol. 5b, Pt. 3*, 500-511.
- Andrews, D. F. & Pregibon, D. (1978). Finding outliers that matter. *J. Royal Statist. Soc. Series B*, 40, 85-93.
- Andrews, F. J. & Tukey, J. W. (1963). The examination and analysis of residuals. *Technometrics*, 5, 141-160.
- Anscombe, F. J. & Tukey, J. W. (1963). The examination and analysis of residuals. *Technometrics*, 5, 141-160.
- Beckman, R. & Trusell, H. (1974). The distribution of an arbitrary studentized residual and the effects of updating in multiple regression. *J. Amer. Statist. Assoc.* 69, 199-201.
- Behnken, D. W. & Draper, N. R. (1972). Residuals and their variance patterns. *Technometrics*, 14, 102-111.
- Belsley, D. A., Kuh, E. & Welsch, R. E. (1980). *Regression Diagnostics: Identifying influential data and sources of collinearity*. John Wiley & Sons, New York.
- Chatterjee, S. & Hadi, A. S. (1986). Influential observation, high leverage points, and outliers in linear regression. *Statistical Science* 1, No. 3, 379-416.
- Chatterjee, S. & Hadi, A. S. (1988). *Sensitivity analysis in linear regression*. Wiley, New York.
- Cochran, W. G. (1968). Errors of measurement in statistics. *Technometrics*, 10, 637-666.
- Cook, R.D. (1977). Detection of influential observations in linear regression. *Technometrics*, 19, 15-18.
- Cook, R.D. & Weisberg, S. (1980). Characterizations of an empirical influence function for detecting influential cases in regression. *Technometrics*, 22, 495-508.
- Cook, R.D. & Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman and Hall, N.Y.
- Ellenberg, J. H. (1973). The joint distribution of the standardized least squares residuals from a general linear regression. *J. Amer. Statist. Assoc.* 68, 941-943.
- Ellenberg, J. H. (1976). Testing for a single outlier from a general linear regression model. *Biometrics*, 32, 637-645.
- Ghosh, S. (1983). Influential observations in view of design and inference. *Commun. Statist. Theor. Meth.* 12 (14), 1675 - 1683.

- Graybill, F. A. (1976). *Theory and Application of the Linear Model*. Duxbury Press.
- Gray, J. B. & Woodall, W. H. (1994). The maximum size of standardized and internally studentized residuals in regression analysis. *The American Statistician*, 48, 111-113.
- Hadi, A. S. & Simonoff, J. S. (1993). Procedures for the identification of multiple outliers in linear models. *J. Amer. Statist. Assoc.* Vol. 88, 1264-1272.
- Hoaglin, D. C. & Welsch, R. (1978). The Hat Matrix in Regression and ANOVA. *American Statistician*, 32, 17-22.
- Huber, P. J. (1973). Robust Regression: Asymptotics, Conjectures, and Monte Carlo; *Annals of statistics*, 1, 799-821.
- Huber, P. (1981). *Robust Statistics*, John Wiley & Sons, New York.
- Johnson, R. A. & Wichern (1992). *Applied Multivariate Statistical Analysis*, (3rd Edition). Printice- Hall International, Inc.
- Kianifard, F. & Swallow, W.F. (1989). Using recursive residuals calculated on adaptively ordered observations, to identify outliers in linear regression. *Biometrics*, 45, 571-585.
- Kianifard, F. & Swallow, W.F. (1990). A Monte Carlo comparison of five procedures for identifying outliers in linear regression. *Commun, Statist. Theory and Methods*, 19, 1913-1938.
- Lave, L. B. & Seskin, E. P. (1977). Does air pollution shorten lives? *Statistics and public policy*, (W.B. Fairley & F. Mosteller, eds.) Reading: MA: Addison Wesley.
- Marasinghe, M. G. (1985). A multistage procedure for detecting several outliers in linear regression. *Technometrics*, 27, 395-399.
- Prescott, P. (1975). An approximate test for outliers in linear models. *Technometrics*, 17, 129-132.
- Rao, C. R. (1973). *Linear Statistical Inference and its Applications* (Second Edition): John Wiley & Sons. New York.
- Rousseeuw, P. J. (1985). "Multivariate estimation with high breakdown point" in *Mathematical Statistics and Applications* (Vol. B), eds. W. Grossmann, G. Pflug, I. Vineze and W. Wertz, Dordrecht. Reidel, pp. 283-297.
- Rousseeuw, P. J. & Leroy, A. M. (1987). *Robust regression and outlier detection*, New York; John Wiley.
- Searle, S. R. (1982). *Matrix Algebra Useful for Statistics*; John Wiley & Sons, New York.
- Shiffler, R. E. (1988). Maximum Z scores and outliers. *The American Statistician*, 42, 79-80.

- Simonoff, J. S. (1991). "General approaches to stepwise identification of unusual values in data analysis" in *Directions in Robust Statistics and Diagnostics. Part II*, eds. W. Stahel & S. Weisberg, New York: Springer, Verlag, pp. 223-242.
- Simpson, D. G. Ruppert, D. & Carroll, R. J. (1992). On one-step GM estimates and stability of inferences in linear regression. *J. Amer. Statist. Assoc.*, 86, 439-450.
- Srikantan, K. S. (1961). Testing for a single outlier in a regression model. *Sankhyā A*, 23, 251-260.
- Theil, H. & Goldberger, A. S. (1961). On pure and mixed estimation in economics. *Inter. Economic Review*, 2, 65-78.
- Theil, H. (1963). On the use of incomplete prior information in regression analysis. *J. Amer. Statist. Assoc.* 58, 401-414.
- Toutenburg, H. (1982). *Prior Information in Linear Models*. John Wiley and Sons, N.Y.
- Tukey, J. W. (1977). *Exploratory Data Analysis*: Reading Mass. Addison-Wesley, N.Y.
- Yohai, V. J. & Zamar, R. H. (1988). High breakdown point estimates of regression by means of the minimization of an efficient scale. *J. Amer. Statist. Assoc.*, 83, 406-413.