Summer 1989

Detection of Outliers and Influential Observations in Regression Models

Anwar M. Hossain
Old Dominion University

DETECTION OF OUTLIERS AND INFLUENTIAL OBSERVATIONS IN REGRESSION MODELS

By

Anwar M. Hossain
M. Sc. 1976, Jahangirnagar University, Bangladesh
B. Sc. (Honors), 1975, Jahangirnagar University, Bangladesh

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfilment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY
Computional and Applied Mathematics

OLD DOMINION UNIVERSITY
August, 1989

Approved by:

_____
Dayanand N. Naik (Director)

_____

_____

ABSTRACT

DETECTION OF OUTLIERS AND INFLUENTIAL OBSERVATIONS

IN REGRESSION MODELS

ANWAR M. HOSSAIN

Old Dominion University, 1989

Director: Dr. Dayanand N. Naik

Observations arising from a linear regression model, lead one to believe that a particular observation or a set of observations are aberrant from the rest of the data. These may arise in several ways: for example, from incorrect or faulty measurements or by gross errors in either response or explanatory variables. Sometimes the model may inadequately describe the systematic structure of the data, or the data may be better analyzed in another scale. When diagnostics indicate the presence of anomalous data, then either these data are indeed unusual and hence helpful, or contaminated and, therefore, in need of modifications or deletions.

Therefore, it is desirable to develop a technique which can identify unusual observations, and determine how they influence the response variate. A large number of statistics are used, in the literature, to detect outliers and influential observations in the linear regression models. Two kinds of comparison studies to determine an optimal statistic are done in this dis-

sertation: (i) using several data sets studied by different authors, and (ii) a detailed simulation study. Various choices of the design matrix of the regression model are considered to study the performance of these statistics in the case of multicollinearity and other situations. Calibration points using the exact distributions and the Bonferroni's inequality are given for each statistic. The results show that, in general, a set of two or three statistics is needed to detect outliers, and a different set of statistics to detect influential observations.

Various measures have been proposed which emphasize different aspects of influence upon the linear regression model. Many of the existing measures for detecting influential observations in linear regression models have natural extensions to the multivariate regression. The measures of influence are generalized to the multivariate regression model and multivariate analysis of variance models. Several data sets are considered to illustrate the methods. The regression models with autocorrelated errors are also studied to develop diagnostic statistics based on intervention analysis.

ACKNOWLEDGEMENTS

I wish to express my sincere thanks to Dr. D. N. Naik for all the advice, encouragement, guidance, and support he has given me during the long hours I spent working on this dissertation. My special heartfelt thanks go to Professor Ram Dahiya for his invaluable help. Grateful appreciation is extended to my dissertation committee: Dr. Michael Doviak and Dr. Edward Markowski for their useful suggestions.

A special word of thanks are reserved for my parents and my elder brother, Dr. H. A. Howlader. Without their continual support and love, this study would not have been possible. I also thank the authorities at the University of Dhaka, Bangladesh for allowing me study leave to pursue higher studies at the Old Dominion University. I am thankful to Barbara Jeffrey, who assisted me in putting the pieces of this dissertation together.

And lastly, but mostly, thanks are due to my wife, Nilufar, daughters, Nusrat and Suhaila for providing the family support, and waiting patiently, knowing that sooner or later I would finish my work.

to

my parents

TABLE OF CONTENTS

PAGE

Chapter

LIST OF TABLES

PAGE

TABLE

LIST OF FIGURES

PAGE

FIGURE

1. INTRODUCTION

Researchers and data analysts are often faced with the problem of finding an observation or a set of observations in their sample which stand apart from the rest. It seems likely that such spurious observations should come from one or a combination of the following sources.

(i) The observation is in error. For example, an investigator might have recorded the response of a variate incorrectly.

(ii) The model is incorrectly specified.

(iii) The observation is inconsistent with the inherent variability of the system being investigated.

It must be emphasized that the first objective of a data analyst is to detect those aberrant observations, which are called outliers. This type of observations may or may not have an effect on the parameter estimation or prediction. The study of outliers has interested practicing statisticians and other scientists for a great number of years. Benjamin Pierce, in the Astronomical Journal (1852), produced an outlier criterion and applied it to "fifteen observations of the vertical semi-diameters of Venus made by Lieutenant Herndon in 1846". This data set has since become a classic in the literature and has been used by a number of authors to compare various outlier detection criteria. The authors of this period, including Chauvenet

(1863), Stone (1867), and Edgeworth (1887), in developing their criteria, assumed knowledge of the population mean and population standard deviation.

Thompson (1935) was the first author to drop both assumptions about population mean and standard deviation, and his paper was the basis upon which most modern day outlier theory grew. Anscombe (1960) and Daniel (1960) were among the first authors to propose the use of standardized residual for detecting a single outlier in linear regression models. Since then, many authors, Srikantan (1961), Ellenberg (1973), Teitjen, Moore and Beckman (1973), and Cook and Weisberg (1982) have considered the problem of detection of outliers in linear regression models.

In recent years, considerable interest has centered on a particular class of diagnostic methods that are intended to aid in assessing the role that individual observations play in determining a fitted model. Key features of a fitted model can be dominated by a single observation. It seems that spurious observations may not always be outliers. It is therefore important for an analyst to be able to identify such observations, and assess their effect on various aspects of the analysis. Such observations are called influential observations. A definition, which seems most appropriate, is given by Belseley, Kuh, and Welsch (1980): "An influential observation is the one which, either

individually or together with several other observations, has a considerably larger impact on the calculated values of various estimates... than is the case for most of the other observations".

Wood (1983) referred to such influential cases as "golden points" and related an actual application in which such cases actually led to an improvement of the physical process from which the data arose. This discovery eventually translated into millions of dollars in additional profit for the company.

An observation, however, may not have the same impact on all regression outputs. An observation may have influence on estimates of the regression coefficients, the estimated variance of these estimates and/or the fitted values. The primary goal of the researcher should determine which influence to consider. Once the influential observations are identified, he must make use of all available additional information about those data points in the context of actual application and exercise his best statistical and common sense judgements in deciding the appropriate action to take. An even more immediate task, after the detection of potential or actual influence, is the attempt to understand or explain the source of the influence.

The detection of influential observations from the linear regression models, known as regression diagnostics, received the attention of several authors

after the paper by Cook (1977). There have been many books and research papers on this topic since then. Some of these are by Belsley, Kuh, and Welsch (1980), Cook and Weisberg (1982) and Atkinson (1985) with their references. Pendleton (1985) has looked at the application of many of these methods to analysis of variance problems. For some work on detection of outliers and influential observations in the case of multivariate regression, see the papers of Naik (1986) and Hossain and Naik (1989). The number of techniques available for regression diagnostics is indeed very large. The analyst must try to understand the basis of these methods, and choose the set that seems most appropriate. Chatterjee and Hadi (1986) reviewed most of the available statistics for detection of influential observations. Hoaglin and Kempthorne (1986), pointed out that the calibration points considered in that paper, and in many earlier papers, are based on the rule of thumb. Recently, Balasooriya and Tse (1986) and Balasooriya, Tse and Liew (1987), have compared performance of several statistics, but many important and sensitive statistics were omitted in their study. With the analyses of several data sets and simulation study, it is noticed that, important influential observations can be overlooked if appropriate calibration points are not exercised.

In this dissertation, (i) a comparison of various statistics by using the calibration points which are obtained, from the exact distributions and Bon-

ferroni's inequality is done. (ii) Diagnostic statistics are developed for the multivariate regression model and MANOVA model. (iii) With use of the missing value techniques, some of the methods are extended to the regression model with correlated errors. For the regression model with correlated errors, an estimator of the parameter vector is obtained, using the whole data set and without the *ith* observation. The distance between the two estimators with and without the *ith* observation in some appropriate norm will give the measure of influence. Some diagnostic statistics are then found, using the intervention analysis, first introduced by Box and Tiao (1975).

The scheme of presentation of this dissertation is as follows:

(1) In Chapter 2 briefly reviewed the literature, introduce the notations, and summarize some available results.

(2) Chapter 3 deals with comparison studies in order to determine an optimal set of statistics for detection of outliers and influential observations. Two kinds of comparison studies are done: (a) Using several data sets studied by different authors, (b) A detailed simulation study using an IMSL subroutine to generate a linear regression model.

(3) In Chapter 4 diagnostic statistics are developed for multivariate regression models. Examples are presented to illustrate the methods. Further, it is shown that some of the statistics can easily be used for identifying the

outliers and influential observations from MANOVA models.

(4) Chapter 5 deals with the regression model with autocorrelated errors. The finding is that the likelihood displacements, along with some of the statistics, can be used for detecting the influential observations.

2. REVIEW

A large number of statistical measures have been proposed in the literature for detecting the outliers and influential observations in the linear regression model. These measures along with an examination of their interrelationships are presented in the following.

In Section 2.1, are presented the linear regression model and notations. In Sections 2.2 to 2.8, are described various statistics that have been used in the literature. In Section 2.9 is suggested a simple modification of the available statistics which performs slightly better in many cases for identifying the influential observations.

2.1 Notations

Consider the linear regression model

$$Y = X\beta + \epsilon, \tag{2.1}$$

where Y is an nx1 vector of response or dependent variable; X is an nxm matrix of predictors, including one constant predictor; β is a mx1 vector of unknown parameters to be estimated; and ϵ is an nx1 vector of random disturbances each with zero mean and unknown variance σ^2. Let y_i, x_i', $i = 1, 2, \ldots, n$, denote the ith element of Y and the ith row of X respectively, and $X_j, j = 1, 2, \ldots, m$, denote the jth column of X. By the ith observation

is meant the row of $(x_i' : y_i)$, that is, the ith row of the augmented matrix $X^* = (X : Y)$. The notation "(i)" or "$[j]$" written as a subscript to a quantity is used to indicate the ith observation or the jth variable deleted respectively. Thus, for example, $X_{(i)}$ is the matrix X with the ith row deleted, and $X_{[j]}$ is the matrix X with the jth column deleted, and $\hat{\beta}_{(i)}$ is the vector of estimated parameters when the ith observation is omitted.

2.2 Standard Estimation Results in Least Squares

The least squares theory gives $\hat{\beta} = (X'X)^{-1}X'Y$, $Cov(\hat{\beta}) = \sigma^2(X'X)^{-1}$. And $\hat{\beta} \sim N_m(\beta, \sigma^2(X'X)^{-1})$, $\hat{Y} = X\hat{\beta} = PY$, where $P = X(X'X)^{-1}X'$, $p_i = x_i'(X'X)^{-1}x_i$, $\hat{Y} \sim N_n(X\beta, \sigma^2 P)$ and $N_m(\mu, \Sigma)$ denotes a m - dimensional multivariate normal distribution with mean vector μ and covariance matrix Σ. Also, $e = Y - \hat{Y} = (I - P)Y$, $var(e) = \sigma^2(I - P)$, $e \sim N_n(0, \sigma^2(I - P))$, $SSE = e'e$, and $\hat{\sigma}^2 = e'e/(n - m)$, the residual mean square estimate of σ^2. Let $\hat{\beta}_{(i)}$ and $\hat{\sigma}^2_{(i)}$ be the estimates of β and σ^2 when the ith observation is deleted. It is assumed throughout that $(X'X)^{-1}$ exists.

2.3 Diagnostic Plots

Several diagnostic plots have been proposed in the literature. For example, see Belsley, Kuh and Welsch (1980), Cook and Weisberg (1982) and Atkinson (1985) and references therein. In Section 3.2 are to be considered a few plots through scaled values of different statistics against each observation.

Scaling is done so that the calibration point is unity.

2.4 Measures Based on Residuals

Many commonly used tests for detecting outliers are based on the standardized residuals

$$t_i = \frac{e_i}{\hat{\sigma}\sqrt{(1-p_i)}}, i = 1, 2, \ldots, n, \tag{2.2}$$

where $e_i = y_i - x_i\hat{\beta}$. If there are no outliers in the data it can be easily shown that $t_i^2/(n-m)$ has beta distribution with parameters 1/2 and (n-m-1)/2. Anscombe (1960) and Daniel (1960) were among the first authors to propose the use of t_i for detecting a single outlier in the linear regression models. It can be shown that, under normality, $\hat{\sigma}^2_{(i)}$ and e_i are independent (see Cook and Weisberg (1982) , pp.20 - 21), and the studentized residuals are defined as

$$t_i^* = \frac{e_i}{\hat{\sigma}_{(i)}\sqrt{(1-p_i)}}, i = 1, 2, \ldots, n, \tag{2.3}$$

where $\hat{\sigma}^2_{(i)} = \frac{Y'_{(i)}(I-P_{(i)})Y_{(i)}}{(n-m-1)}$. The distribution of t_i^* is student's t with (n-m-1) degrees of freedom. Computationally easier form for $\hat{\sigma}^2_{(i)}$ can be given as

$$\hat{\sigma}^2_{(i)} = \hat{\sigma}^2 \frac{n-m-t_i^2}{n-m-1}, i = 1, 2, \ldots, n. \tag{2.4}$$

2.5 Measures Based on Influence Curve

An important class of measures of the influence of the *ith* observation on the regression results is based on the idea of the influence curve or influence function introduced by Hampel (1968, 1974). The influence function is defined as

$$IF_i(x_i; y_i; F; T) = \lim_{\epsilon \to 0} \frac{T[(1-\epsilon)F + \epsilon \delta_{(x_i,y_i)}] - T(F)}{\epsilon}, \qquad (2.5)$$

where $T(.)$ is a vector-valued functional defined on the set of all cumulative distribution functions and $\delta_{(x_i,y_i)} = 1$ at (x_i, y_i) and zero elsewhere. The function IF_i measures the influence on T of adding one observation (x_i, y_i) to a large sample. Let $\hat{F}$ be the empirical distribution function based on the full sample and $\hat{F}_{(i)}$ be the empirical distribution function when the *ith* observation is deleted. The empirical influence curve (EIC) for $\hat{\beta}$ is found by substituting $\hat{F}_{(i)}$ for F and $\hat{\beta}_{(i)}$ for $T(\hat{F}_{(i)})$ in (2.5) and obtaining

$$EIC_i = (n-1)(X'_{(i)}X_{(i)})^{-1}x'_i(y_i - x_i\hat{\beta}_{(i)}), i = 1, 2, \ldots, n, \qquad (2.6)$$

where $\hat{\beta}_{(i)} = (X'_{(i)}X_{(i)})^{-1}X'_{(i)}Y_{(i)}$. The quantity $(y_i - x'_i\hat{\beta}_{(i)})$ is called the predicted residual, because the *ith* observation does not enter in the calculation of $\hat{\beta}_{(i)}$. Miller (1974) showed that

$$\hat{\beta} - \hat{\beta}_{(i)} = (X'X)^{-1}x'_i \frac{e_i}{1-p_i}, i = 1, 2, \ldots, n. \qquad (2.7)$$

Using the relation (2.7) the predicted residual can be written as

$$y_i - x_i'\hat{\beta}_{(i)} = e_i/1 - p_i$$

and substituting this in (2.6) one can obtain

$$EIC_i = (n-1)(X'X)^{-1}x_i' \frac{e_i}{(1-p_i)^2}, i = 1, 2, \ldots, n. \tag{2.8}$$

The sample influence curve (SIC) for $\hat{\beta}$ is found by omitting the limit in (2.5) and taking $F = \hat{F}$, $T(\hat{F}) = \hat{\beta}$, $\epsilon = -\frac{1}{(n-1)}$. This gives

$$SIC_i = (n-1)(X'X)^{-1}x_i'(y_i - x_i\hat{\beta}_{(i)})$$

$$= (n-1)(X'X)^{-1}x_i' \frac{e_i}{1-p_i}, i = 1, 2, \ldots, n. \tag{2.9}$$

The sensitivity curve (SC) for $\hat{\beta}$ is obtained by setting $F = \hat{F}_{(i)}$, $T(\hat{F}_{(i)}) = \hat{\beta}_{(i)}$ and $\epsilon = \frac{1}{n}$. This gives

$$SC_i = n(X'X)^{-1}x_i' \frac{e_i}{1-p_i}, i = 1, 2, \ldots, n. \tag{2.10}$$

Since IF_i is a vector, it must be normalized so that the observations can be ordered in a meaningful way. The class of norms which are location and scale invariant is given by

$$D_i(M, C) = \frac{(IF_{(i)})'M(IF_i)}{C} = \frac{(\hat{\beta}_{(i)} - \hat{\beta})'M(\hat{\beta}_{(i)} - \hat{\beta})}{C}, \tag{2.11}$$

for any appropriate choice of M and C. A large value of $D_i(M,C)$ indicates that the ith observation has strong influence on the estimated coefficients relative to M and C.

Cook's distance: Cook (1977) suggested the measure

$$C_i = \frac{(\hat{\beta} - \hat{\beta}_{(i)})'(X'X)(\hat{\beta} - \hat{\beta}_{(i)})}{m\hat{\sigma}^2}, i = 1, 2, \ldots, n, \qquad (2.12)$$

to assess the influence of the ith observation. At first sight, it might seem that computing $C_i, i = 1, 2, \ldots, n$, requires (n+1) regressions, one regression using the full data and n regressions using the reduced data . However, substituting (2.7) in (2.12) yields

$$C_i = \frac{x_i'(X'X)^{-1}(X'X)(X'X)^{-1}x_i}{m(1-p_i)} \quad \frac{e_i^2}{\hat{\sigma}^2(1-p_i)}$$

$$= \frac{p_i}{1-p_i} \quad \frac{t_i^2}{m}, i = 1, 2, \ldots, n. \qquad (2.12a)$$

This measure, called Cook's distance, can be thought of as the scaled distance between $\hat{\beta}$ and $\hat{\beta}_{(i)}$. Cook (1979) suggested that each C_i be compared to the percentiles of the central F-distribution with m and (n-m) d.f. It is clear from equation (2.12a) that C_i can be expressed as $C_i = \alpha_i t_i^2$, where, under the assumption that X is nonstochastic, the $\alpha_i, i = 1, 2, \ldots, n$, are known but unequal constants. Hence C_i is a monotonic function of t_i^2.

Welsch-Kuh's distance: The influence of the ith observation on the predicted value $\hat{y}_i$ can be measured by the change in the prediction at x_i when

the ith observation is deleted, that is,

$$\frac{|\hat{y}_i - \hat{y}_{i(i)}|}{\sigma\sqrt{p_i}} = \frac{|x_i'(\hat{\beta} - \hat{\beta}_{(i)}|}{\sigma\sqrt{p_i}}, \; i = 1, 2, \ldots, n. \tag{2.13}$$

Welsch and Kuh (1977), Welsch and Peters (1978) and Belsley et al.(1980) suggested using $\hat{\sigma}_{(i)}$ as an estimate of σ in (2.13) and named the expression in (2.13) as $DFFITS_i$. For simplicity, we will refer to (2.13) by WK_i. Thus

$$WK_i = \frac{|x_i(\hat{\beta} - \hat{\beta}_{(i)})|}{\hat{\sigma}_{(i)}\sqrt{p_i}} = \frac{|e_i \quad x_i'(X'X)^{-1}x_i|}{(1-p_i) \quad \hat{\sigma}_{(i)}\sqrt{p_i}}$$

$$= |t_i^*| \quad \sqrt{\frac{p_i}{1-p_i}}, \; i = 1, 2, \ldots, n. \tag{2.14}$$

Belsley, Kuh and Welsch (1980) suggested using $2\sqrt{\frac{m}{n}}$ as a calibration point for WK_i.

Welsch's distance: Using the empirical influence curve based on (n-1) observations, which is defined in (2.8) as an approximation to the influence curve for $\hat{\beta}$ and setting $M = X_{(i)}'X_{(i)}$ and $C = (n-1)\hat{\sigma}_{(i)}^2$, (2.11) becomes

$$W_i^2 = \frac{(n-1)\, p_i\, e_i^2}{\hat{\sigma}_{(i)}^2(1-p_i)^3} = (n-1)t_i^{*2}\frac{p_i}{(1-p_i)^2}, \; i = 1, 2, \ldots, n. \tag{2.15}$$

Comparing (2.14) and (2.15) yields

$$W_i = WK_i\sqrt{\frac{n-1}{1-p_i}}, \; i = 1, 2, \ldots, n. \tag{2.16}$$

Welsch (1982) suggested using W_i as a diagnostic tool for identifying the influential observations. The fact that WK_i is easier to interpret has led

some people to prefer WK_i over W_i. It is clear from (2.16) that W_i gives more emphasis to high leverage points. The equation (2.16) suggests that the calibration points for W_i can be obtained by multiplying the calibration points for WK_i by $[n(n-1)/(n-m)]^{1/2}$.

Modified Cook's distance: Atkinson (1981) has suggested using a modified version of C_i for the detection of influential observations. The proposed measure is

$$C_i^* = \sqrt{D_i((X'X), \frac{m(n-1)^2}{n-m}\hat{\sigma}^2_{(i)})}$$

$$= |t_i^*| \sqrt{\frac{p_i}{1-p_i}} \frac{n-m}{m} = WK_i \sqrt{\frac{n-m}{m}}, i = 1, 2, \ldots, n, \qquad (2.17)$$

which, apart from a constant factor, is the same as WK_i. C_i^* was originally proposed by Welsch and Kuh (1977) and subsequently by Welsch and Peters (1978) and Belsley et al. (1980). The suggested calibration point for C_i^* is $2\sqrt{\frac{n-m}{m}}$.

2.6 Measures Based on Volume of Confidence Ellipsoids

Under the normality, the $100(1-\alpha)$ percent joint confidence region for β is obtained from

$$\frac{(\beta - \hat{\beta})'(X'X)(\beta - \hat{\beta})}{m\hat{\sigma}^2} \leq F(\alpha; m, n-m), \qquad (2.18)$$

where $F(\alpha; m; n-m)$ is the upper α percentile point of the central F-distribution with m and (n-m) degrees of freedom. The diagnostic measures

based on the influence curve can be interpreted as measures which are based on the change in the center of the confidence ellipsoid given by (2.18) when the *ith* observation is deleted.

Andrews-Pregibon statistic: Andrews and Pregibon (1978) suggested the ratio which measures the influence of the *ith* observation on the volume of the confidence ellipsoid and the ratio is

$$\frac{SSE_{(i)}\, det(X'_{(i)}X_{(i)})}{SSE\, det(X'X)}, i = 1, 2, \ldots, n, \tag{2.19}$$

where

$$SSE_{(i)} = Y'_{(i)}(I - P_{(i)})Y_{(i)} = SSE - \frac{e_i^2}{1 - p_i}, i = 1, 2, \ldots, n. \tag{2.20}$$

Lemma 2.6.1: Let B and C be mxp matrices. If A is a mxm nonsingular matrix, then $det(A - BC') = det(A)det(I - C'A^{-1}B)$.

Substituting $A = X'X$ and B=C=x_i, in Lemma (2.6.1) one can obtain

$$det(X'_{(i)}X_{(i)}) = det(X'X) - x_i x'_i)$$

$$= det(X'X)(1 - x'_i(X'X)^{-1}x_i) = det(X'X)(1 - p_i). \tag{2.21}$$

Now AP_i can be written as $AP_i = 1 - (1 - p_i)\frac{SSE_{(i)}}{SSE}$, where $\frac{SSE_{(i)}}{SSE} \sim \beta(\frac{(n-m-1)}{2}, \frac{m}{2}), i = 1, 2, \ldots, n.$

Covariance Ratio: One can assess the influnce of the *ith* observation by comparing the estimated variance of $\hat{\beta}$ and the estimated variance of $\hat{\beta}_{(i)}$,

that is, by comparing $\hat{\sigma}^2 (X'X)^{-1}$ and $\hat{\sigma}^2_{(i)}(X'_{(i)}X_{(i)})^{-1}$. Belsley et al. (1980) suggested using the ratio

$$Covr_i = \frac{det[\hat{\sigma}^2_{(i)}(X'_{(i)}X_{(i)})^{-1}]}{det[\hat{\sigma}^2(X'X)^{-1}]} = \left(\frac{\hat{\sigma}^2_{(i)}}{\hat{\sigma}^2}\right)^m \frac{det(X'X)}{det(X'_{(i)}X_{(i)})}, i = 1, 2, \ldots, n, \tag{2.22}$$

for that purpose. After substituting (2.4) and (2.21) in (2.22) one can obtain

$$Covr_i = \frac{1}{1-p_i}\left(\frac{n-m-t_i^2}{n-m-1}\right)^m, i = 1, 2, \ldots, n. \tag{2.23}$$

A rough calibration point is $|Covr_i - 1| > \frac{3m}{n}$. Belsley, Kuh and Welsch (1980) called (2.23) COVRATIO.

Cook-Weisberg statistic: Under the assumption of normality, the $100(1-\alpha)$ percent joint confidence ellipsoid for β, when the ith observation is deleted is

$$\frac{(\hat{\beta} - \hat{\beta}_{(i)})'(X'_{(i)}X_{(i)})(\hat{\beta} - \hat{\beta}_{(i)})}{m\hat{\sigma}^2_{(i)}} \leq F(\alpha; m, n-m-1). \tag{2.24}$$

Cook and Weisberg (1980) proposed the logarithm of the ratio of the volume of the region in (2.24) to that in (2.18) as a measure of the influence of the ith observation on the volume of confidence ellipsoid for β which can be written as

$$CW_i = log\left(\left(\frac{det(X'_{(i)}X_{(i)})}{det(X'X)}\right)^{1/2}\left(\frac{\hat{\sigma}^2}{\hat{\sigma}^2_{(i)}}\right)^m \times \right.$$

$$\left.\left(\frac{F(\alpha; m, n-m)}{F(\alpha; m, n-m-1)}\right)^{m/2}\right), i = 1, 2, \ldots, n. \tag{2.25}$$

Substituting (2.4) and (2.21) in (2.25) one can obtain

$$CW_i = \frac{1}{2}log(1-p_i) + \frac{m}{2}log\left(\frac{n-m-1}{n-m-t_i^2}\right) + \frac{m}{2}\times$$

$$log\left(\frac{F(\alpha;m,n-m)}{F(\alpha;,m,n-m-1)}\right), i = 1,2,\ldots,n, \qquad (2.26)$$

where $F(\alpha;.,.)$ is the upper α percentile point of F- distribution with appropriate degrees of freedom. Cook and Weisberg (1980) say the following about CW_i; "if this quantity is large and positive, then deletion of the *ith* observation will result in a substantial decrease in volume...[and if it is] large and negative, it will result in a substantial increase in volume". From (2.23), it is seen that CW_i is related to $Covr_i$ by

$$CW_i = -\frac{1}{2}log(Covr_i)+\frac{m}{2}log\left(\frac{F(\alpha;m,n-m)}{F(\alpha;m,n-m-1)}\right), i = 1,2,\ldots,n. \quad (2.27)$$

The second term on the right hand side of (2.27) does not depend on i; thus, CW_i and $Covr_i$ are equivalent.

2.7 Measures based on the Likelihood Function

Let θ be a (mx1) parameter vector partitioned as $\theta' = (\theta_1', \theta_2')$ and the maximum likelihood estimator (mle) of θ based on n observations be $\hat{\theta}' = (\hat{\theta}_1', \hat{\theta}_2')$. Further, let $\hat{\theta}'_{(i)} = (\hat{\theta}'_{1(i)}, \hat{\theta}'_{2(i)})$ be the mle of θ obtained by using all but the *ith* observation. Let $l(\hat{\theta})$ and $l(\hat{\theta}_{(i)})$ be the log-likelihood functions when θ is repalced by $\hat{\theta}$ and $\hat{\theta}_{(i)}$ respectively. Cook and Weisberg

(1982) defined the likelihood displacement for θ by

$$LD_i(\theta) = 2[l(\hat{\theta}) - l(\hat{\theta}_{(i)})],\ i = 1, 2, \ldots, n. \tag{2.28}$$

Usually $LD_i(\theta)$ is compared with $\chi^2(\alpha; q)$ where $\chi^2(\alpha; q)$ is the α percentile point of the χ^2 distribution with q degrees of freedom and q is the dimension of θ. Suppose one is interested only in estimating θ_1, then the likelihood displacement for θ is defined as

$$LD_i(\theta_1|\theta_2) = 2[l(\hat{\theta}) - l(\hat{\theta}_{1(i)}), \theta_2(\hat{\theta}_{1(i)})],\ i = 1, 2, \ldots, n, \tag{2.29}$$

where $\theta_2(\hat{\theta}_{1(i)})$ is the mle of θ_2 when θ_1 is repalced by $\hat{\theta}_{1(i)}$. It is fairly straight forward to apply the general results (2.28) and (2.29) to the linear regression model (2.1). Let $l(\beta, \sigma^2)$ be the log-likelihood function based on all n observations. Let $\hat{\beta}$ and $\hat{\sigma}^2$ be the mle's of β and σ^2 respectively. Let the mle's of β and σ^2 be $\hat{\beta}_{(i)}$ and $\hat{\sigma}^2_{(i)}$ respectively when the *ith* observation is deleted. Then the likelihood displacement (2.28) for (β, σ^2) is given by

$$LD_i(\beta, \sigma^2) = 2[l(\hat{\beta}, \hat{\sigma}^2) - l(\hat{\beta}_{(i)}, \hat{\sigma}^2_{(i)})],\ i = 1, 2, \ldots, n,$$

where

$$l(\hat{\beta}, \hat{\sigma}^2) = -\frac{n}{2} log(\hat{\sigma}^2) - \frac{n}{2} - \frac{n}{2} log(2\pi) \tag{2.30}$$

and

$$l(\hat{\beta}_{(i)}, \hat{\sigma}^2_{(i)}) = -\frac{n}{2} log(\hat{\sigma}^2_{(i)}) - \frac{n}{2} log(2\pi) - \frac{1}{\hat{\sigma}^2_{(i)}} (Y - X\hat{\beta}_{(i)})'(Y - X\hat{\beta}_{(i)}),$$

$$= -\frac{n}{2}log(\hat{\sigma}^2_{(i)}) - \frac{n}{2}log(2\pi) - \frac{e_i^2}{\hat{\sigma}^2_{(i)}\,(1-p_i)^2} - (n-1),\ i = 1,2,\ldots,n. \quad (2.31)$$

Substituting (2.30) and (2.31) in (2.28), one can write

$$LD_i(\beta,\sigma^2) = nlog\left(\frac{\hat{\sigma}^2_{(i)}}{\hat{\sigma}^2}\right) + \frac{e_i^2}{\hat{\sigma}^2_{(i)}(1-p_i)^2} - 1,\ i = 1,2,\ldots,n. \quad (2.32)$$

Note that $LD_i(\beta,\sigma^2)$ is useful if we are interested in estimating both β and σ^2. If the only interest is in estimating β or σ^2, then the likelihood displacement can be obtained by using equation (2.29). The likelihood displacement for estimating only β is

$$LD_i(\beta|\sigma^2) = 2[l(\hat{\beta},\hat{\sigma}^2) - l(\hat{\beta}_{(i)},\sigma^2(\hat{\beta}_{(i)}))] \quad (2.33)$$

and that for estimating σ^2 is

$$LD_i(\sigma^2|\beta) = 2[l(\hat{\beta},\hat{\sigma}^2) - (l(\hat{\sigma}^2_{(i)},\beta(\hat{\sigma}^2_{(i)})] \quad (2.34)$$

Note that

$$l(\hat{\beta}_{(i)},\sigma^2(\hat{\beta}_{(i)})) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\sigma^2(\hat{\beta}_{(i)})) - \frac{n}{2}, \quad (2.35)$$

and hence by substituting (2.30) and (2.35) in (2.33) one can obtain

$$LD_i(\beta|\sigma^2) = nlog(\frac{\sigma^2(\hat{\beta}_{(i)})}{\hat{\sigma}^2}) = nlog(1 + \frac{m}{n-m}C_i),\ i = 1,2,\ldots,n, \quad (2.36)$$

where C_i is given by (2.12). Thus $LD_i(\beta|\sigma^2)$ is compared to the percentage points of $\chi^2(\alpha;m)$. Noting that

$$l(\hat{\sigma}^2_{(i)},\beta(\hat{\sigma}^2_{(i)})) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\hat{\sigma}^2_{(i)}) - n\frac{\hat{\sigma}^2}{2\hat{\sigma}^2_{(i)}}, \quad (2.37)$$

and substituting (2.30) and (2.37) in (2.34) yields

$$LD_i(\sigma^2|\beta) = nlog(\frac{\hat{\sigma}^2_{(i)}}{\hat{\sigma}^2}) + n\frac{\hat{\sigma}^2}{\hat{\sigma}^2_{(i)}} - n, \; i = 1, 2, \ldots, n. \tag{2.38}$$

It should be noted that the likelihood displacements are based on the probability model used, whereas the other measures of influence are numerical. An advantage of the likelihood displacement is that it can be extended to models other than the normal linear model.

2.8 Influnce of an Observation on a Single Coefficient

The influnce of the ith observation on the jth coefficient of β can be measured, as suggested in Cook and Weisberg (1980), by

$$D_{ij} = \frac{e_i^2}{\hat{\sigma}^2(1-p_i)} \frac{p_i - p_i[j]}{1-p_i} = t_i^2 \frac{p_i - p_i[j]}{1-p_i}, \; i = 1, 2, \ldots, n; \; j = 1, 2, \ldots, m, \tag{2.39}$$

where $p_i[j]$ is the ith diagonal element of $P_{[j]} = X_{[j]}(X'_{[j]}X_{[j]})^{-1}X'_{[j]}$, and $X_{[j]}$ is the matrix with the jth column deleted. The equation (2.39) can be simplified to

$$D_{ij} = \frac{t_i^2}{1-p_i} \frac{w_{ij}^2}{W_j'W_j}, \; i = 1, 2, \ldots, n; \; j = 1, 2, \ldots, m, \tag{2.40}$$

where w_{ij} is the ith element of $W_j = (I - P_{[j]})X'_j$, X_j is the jth column of X. Instead of D_{ij} one can use D^*_{ij} as suggested in Belsley et al. (1980), where

$$D^*_{ij} = \frac{t_i^*}{\sqrt{1-p_i}} \frac{w_{ij}}{\sqrt{W_j'W_j}}, \; i = 1, 2, \ldots, n; \; j = 1, 2, \ldots, m. \tag{2.41}$$

The suggested calibration point for D_{ij}^* is $\frac{2}{\sqrt{n}}$.

2.9 Modified Measures

(a) Many studies on outliers and influential observations in regression models are based on either t_i or t_i^* and p_i. The statistic t_i^* detects the significant differences between predicted and observed values of the response for the *ith* observation. But this t_i^* does not provide sufficient information about changes in the parameter estimates due to the deletion of the *ith* observation. Some of the most influential data points can have relatively small t_i^* and hence the calibration point corresponding to t_i^* can not detect these extreme points. One way of overcoming this situation is to scale t_i^* by $(1-p_i)$, which measure can be called as modified t_i^* (Mt_i^*). Thus modified t_i^* is

$$Mt_i^* = \frac{|t_i^*|}{1-p_i}, \; i = 1, 2, \ldots, n. \tag{2.42}$$

A large value of modified t_i^* indicates that the *ith* observation is outlier, influential or both. Hence, $max\,(Mt_i^*)$ may be used as a statistic for identifying an extreme observation in the data.

(b) A measure of goodness of fit of the regression is the multiple correlation coefficient, R^2, which is estimated as a correlation coefficient between the observed and predicted Y's. Usually outlying or influential observations have drastic adverse effect on R^2. The values of R^2 and $R_{(i)}^2$ differ considerably, especially when the *ith* observation is an influential one. Therefore,

the influence of the *ith* observation can be measured by

$$R^2_{(i)} = \frac{[Y'_{(i)}P_{(i)}Y_{(i)} - (n-1)\bar{Y}^2_{(i)}]}{[Y'_{(i)}Y_{(i)} - (n-1)\bar{Y}^2_{(i)}]}, \; i = 1, 2, \ldots, n, \tag{2.43}$$

where $\bar{Y}_{(i)}$ is the sample mean dropping the *ith* observation,

$$P_{(i)} = X_{(i)}(X'_{(i)}X_{(i)})^{-1}X'_{(i)},$$

and for any matrix A, $A_{(i)}$ is A after dropping the *ith* row. It may be noted that $R^2_{(i)}$ can be easily computed using the PROC MATRIX of SAS program. A very small value of $R^2_{(i)}$, (that is much smaller than R^2) means the *ith* observation is influential.

3. EMPERICAL STUDY IN LINEAR REGRESSION

In this Chapter two types of comparison studies have been made for determining an optimal statistic. This included: (i) a detailed simulation study, and (ii) using several data sets studied by different authors. Different choices of the design matrix of regression model are considered to study the performance of these statistics. The designs A and B are chosen to study the performance of the various statistics for detecting the outliers in the case of multicollinearity. Designs C and D are chosen to study the performance of the statistics for identifying the influential observations.

Hoaglin and Kempthorne (1986), mentioned in their comment on Chatterjee and Hadi (1986), that the calibration points considered in that paper are the rule of thumb. They also pointed out that one should use calibration points computed using the exact distributions. For example, Cook(1979) suggested that C_i, the Cook's distance of the ith observation, be compared with the quantiles of the central F distribution with m and (n-m) degrees of freedom, where m is the number of parameters and n is the number of observations in the regression model. In many cases this calibration point cannot identify the influential observation and hence it reduces the capability of C_i in detecting an influential observation. Another widely used regression diagnostic statistic is WK_i, which is defined by Belsley, Kuh and

Welsch (1980) and a calibration point suggested by them is $2\sqrt{\frac{m}{n}}$. With this calibration point, important observations can be overlooked, and some influential observations may not be detected. Therefore, the statistics require clear criteria and guidelines for identifying outliers and influential observations.

A limited study to compare certain statistics is done by Balasooriya and Tse (1986) and Balasooriya, Tse and Liew (1987). Many important and useful statistics were omitted in their comparison study. This motivated the simulation study for determining an optimal statistic for identifying outliers and influential observations taking most of the statistics available into consideration. In this study a large number of statistics are considered and their calibration points are recomputed. They are further compared by using their power to detect outliers and/or influential observations. Apart from certain constants, all the statistics are functions of t_i (standardized residual) or t_i^* (studentized residual) and p_i (the ith diagonal element of the prediction matrix). A reasonable rule of thumb for large p_i, as suggested by Hoaglin and Welsch (1978) and born out by our experience, is $\frac{2m}{n}$. The distribution of t_i^* is student's t with (n-m-1) degrees of freedom and $\frac{t_i^2}{n-m}$ follows a beta distribution with parameters $\frac{1}{2}$ and $\frac{(n-m-1)}{2}$. Using an appropriate cutoff value for p_i and the above exact distributions, the calibration points for var-

ious statistics can be obtained. In this study, various statistics are compared by studying their capability to detect outliers and/or influential observations using the calibration points which are obtained by using the exact distributions and Bonferoni's inequality. In many cases it was found that the performances of these statistics, using these almost exact calibration points, increased considerably.

Computation of calibration points using the exact distribution and Bonferoni's inequality is summarized below. Let T_i, $i = 1, 2, \ldots, n$ be statistics, such that the marginal distribution of each T_i is identical. If a decision criterion is based on $Max_i(T_i)$, for example a decision is made if $Max_i(T_i) > T_\alpha$, then the cutoff value T_α for $Max_i(T_i)$ can be obtained as follows: If T_α is such that

$$P[Max_i(T_i) > T_\alpha] \leq \alpha$$

then

$$P(T_i > T_\alpha \text{for some } i) \leq \alpha$$

that is

$$P[\bigcup_i (T_i > T_\alpha)] \leq \alpha.$$

But

$$P[\bigcup_i (T_i > T_\alpha)] \leq \sum_i P(T_i > T_\alpha).$$

Therefore, if T_α is such that

$$P[T_i > T_\alpha] \leq \frac{\alpha}{n},$$

then

$$P[Max_i(T_i) > T_\alpha] \leq \alpha,\ i = 1, 2, \ldots, n.$$

Such a point T_α we call as calibration point. Knowing the exact distribution of T_i it is, in practice, possible to find T_α for some choice of α. Although, α does not have the same meaning as in the testing of hypotheses problems, in this study α is chosen to be 0.05.

To illustrate the above method, consider the Cook's statistic C_i with $n = 19$, $m = 4$, and $\alpha = 0.05$. The criterion suggested by Cook is to identify the kth observation as influential if $C_k = Max_i(C_i)$. To find a calibration point for C_k the above method can be used. It is known that $C_i \sim \frac{n-m}{m} \frac{p_i}{1-p_i} \beta(\frac{1}{2}, \frac{n-m-1}{2})$, $i = 1, 2, \ldots, n$. Problem is to find C_α such that

$$P[Max_i(C_i) > C_\alpha] \leq \alpha$$

that is, to find C_α such that

$$P[C_i > C_\alpha] \leq \frac{\alpha}{n}.$$

But such a C_α using the cutoff value of beta distribution is

$$C_\alpha = \frac{n-m}{m} \frac{p_i}{1-p_i} (0.2323).$$

As it is mentioned before taking a common value of $\frac{2m}{n}$ for p_i, C_α can be taken to be 0.6335.

The statistics used for simulation study are summarized below. The Bonferoni's inequality and the exact distributions were used to compute the calibration points. In section 3.1, a simulation study is presented together with some general conclusions. In section 3.2 are described the data sets and the performances of the various statistics; also given are several plots of the statistics based on a unit calibration point.

Table 3.1 Various Statistics and Calibration Points

Statistics and Notations	Calibration Points Suggest by Different Authors	Calibration Points in this Study: Based on
$t_i = \frac{e_i}{\hat{\sigma}\sqrt{(1-p_i)}}$	$Z_{\frac{\alpha}{2}}$	$\beta(\frac{1}{2}, \frac{n-m-1}{2})$
$t_i^* = \frac{e_i}{\hat{\sigma_{(i)}}\sqrt{(1-p_i)}}$	$t_{\frac{\alpha}{2}}(n-m-1)$	$t(n-m-1)$
$AP_i = p_i + \frac{e_i^2}{e'e}$	$\frac{2(m+1)}{n}$	$\beta(\frac{1}{2}, \frac{n-m-1}{2})$
$C_i = \frac{p_i}{(1-p_i)} \frac{t_i^2}{m}$	$F_\alpha(m, n-m)$	$\beta(\frac{1}{2}, \frac{n-m-1}{2})$
$WK_i = t_i^* \sqrt{\frac{(n-1)}{(1-p_i)}}$	$2\sqrt{\frac{m}{n}}$	$t(n-m-1)$
$W_i = WK_i \sqrt{\frac{(n-1)}{(1-p_i)}}$	$3\sqrt{m}$	$t(n-m-1)$
$C_i^* = WK_i \sqrt{\frac{n-m}{m}}$	$2\sqrt{\frac{n-m}{n}}$	$t(n-m-1)$
$Mt_i^* = \frac{t_i^*}{(1-p_i)}$	$t_{\frac{\alpha}{2}}(n-m-1)$	

Table 3.1 Various Statistics and Calibration Points (contd.)

Statistics and Notations	Calib. Points Suggest by Diff. Authors	Calib. Points in this Study: Based on
$LD_i(\beta\|\sigma^2) = n\,log(\frac{m}{n-m}C_i + 1)$	$\chi^2_\alpha(m-1)$	$\beta(\frac{1}{2}, \frac{n-m-1}{2})$
$LD_i(\beta, \sigma^2) = nlog(\frac{n}{n-1})$ $+nlog(1-b_i) + \frac{b_i}{1-b_i}\frac{n-1}{1-p_i} - 1,$ $where\ b_i = \frac{t_i^2}{n-m}$	$\chi^2_\alpha(m)$	$\beta(\frac{1}{2}, \frac{n-m-1}{2})$
$LD_i(\sigma^2\|\beta) = nlog(\frac{n}{n-1})$ $+nlog(1-b_i) + \frac{nb_i-1}{1-b_i}$ $R_i^2 = \frac{[Y'_{(i)}P_{(i)}Y_{(i)} - (n-1)\bar{Y}^2_{(i)}]}{[Y'_{(i)}Y_{(i)} - (n-1)\bar{Y}^2_{(i)}]}$	$\chi^2_\alpha(1)$	$\beta(\frac{1}{2}, \frac{n-m-1}{2})$

Here $Z_{\frac{\alpha}{2}}$, $t_{\frac{\alpha}{2}}$, F_α, and χ^2_α are standard cutoff values of $N(0,1)$, t, F, and χ^2 distributions with appropriate degrees of freedom.

3.1 Performances of the Various Statistics for Simulated Data

In this section are to be compared the performances of the statistics under the following designs. The designs A and B are chosen so as to study the performances of the various statistics in detecting the outliers when the design matrix is multicollinear. In designs C and D influential observations

are introduced by considering the values of R^2, the multiple correlation coefficient. The different design matrices of the regression model considered in the study are as follows.

Design A : $X_{nxm} = P_{nxn}\ \Lambda_{nxm}\ Q_{mxm}$, where P and Q are orthogonal matrices, $\Lambda = \begin{bmatrix} \gamma_{mxm} \\ 0_{(n-m)xm} \end{bmatrix}$, and $\gamma = Diag(\gamma_1, \ldots, \gamma_m)$ is a diagonal matrix such that $\gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_m > 0$. If $\gamma_1 \div \gamma_m$ is large then there exists multicollinearity. Two sets of this ratio are chosen.

Design B : $X = (X_{ij})$;

$$X_{ij} = (1 - \alpha^2) Z_{ij} + \alpha\, Z_{i,m+1},\ i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, m,$$

where Z_{ij}, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m+1$, are all independent N(0,1) variables. Here $corr(X_{ij}, X_{ij'}) = \alpha^2$, $j \neq j'$. Large values of α indicates higher degree of multicollinearity. These different values are considered for α : 0, .1, .2, .3, .4, .5, .6, .7, .8, .9.

Design C : Consider the simple regression model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i,\ i = 1, 2, \ldots, n.$$

If y and x are standardized then $R^2 = (\sum x_i y_i)^2$, that is $\mid R \mid = \mid x_1 y_1 + x_2 y_2 + \ldots, + x_n y_n \mid$. To find $x_1, x_2, \ldots, x_n$ such that $M_{nxn}\, X_{nx1} = R_{(i)nx1}$, solve $M_{nxn}\, X_{nx1} = R_{(i)nx1}$ for a given M and $R_{(i)}$. The $R_{(i)}$ represents the multiple correlation coefficient.

Design D : Suppose it is known beforehand that a particular data set has influential observations. By using this design matrix the model (2.1) is simulated. Then the various statistics are computed and their performances noted.

The IMSL subroutine RNNOR is used to generate the standared normal variables ϵ. For $\sigma = 1$, and $\beta_i = i$, $i = 1, 2, \ldots, m$, and for different designs described above the model $Y = X\beta + \epsilon$ is simulated.

Mean Shift Outliers : A mean shift model is

$$Y = X\beta + d\sigma\delta_i + \epsilon \tag{3.1}$$

where δ_i is a nx1 column vector with one at the *ith* row and zeros elsewhere. For the designs A and B , 1000 random samples are generated with n=10, 15, 20, and 50. For each sample, one observation is selected at random and an outlier is created by adding $d\sigma$ to mean shift model (3.1). Values of d are chosen to be 0 (no outlier), 1, and 3.5. The proportions of the number of times out of 1000, different statistics detect outliers are tabulated in the Tables 3.2 - 3.9. To economize here on space, the results are presented for some values of sample sizes only.

For design A, values for n=10, 20, and 50 and $(\gamma_1, \gamma_2) = (1, 2)$ and (0.005, 2) are presented in Table 3.2. First, it is noted that the multi-collinearity introduced in the manner of design A does not reduce the power

to detect outliers, of any statistics. Next, kurtosis has about 40 - 50 percent capability of detecting the outliers for different sample sizes, which is higher or almost equal to what other statistics could do. For small sample (n=10) only t_i, t_i^* and $LD_i(\sigma^2|\beta)$ have about 30 percent power to detect outliers, whereas for moderate to high sample size (n=20, 50) along with the above, Mt_i^* and $LD_i(\beta, \sigma^2)$ have about 40 - 50 percent power of detecting the outliers.

For design B, values for $\alpha = 0, .9(.1)$ and n=10 are presented in Table 3.3, and that for n=20 and n=50 are presented in Table 3.4 and Table 3.5 respectively. One may note that different choices of α describe different levels of multicollinearity. However, examination of the Tables 3.3 - 3.5 reveals that the multicollinearity does not affect the power of these statistics in detecting outliers. Kurtosis in the present case has varying power 10 - 30 percent for small sample (n=10) and about 20 - 50 percent for n=20 and 50. All the statistics mentioned in the previous paragraph have about 20 - 30 percent power for small sample and about 40 - 50 percent for large sample. The conclusion, is that for detection of mean shift outliers, two or three of the above statistics along with kurtosis may be used.

Scale Shift Outliers : Consider the scale shift outlier model

$$Y = X\beta + \eta^{(d)} \tag{3.2}$$

where $\eta_k^{(d)} = \epsilon_k$ for all $k \neq i$ and $\eta_i^{(d)} = \epsilon_i e^{d\sigma}$. Values of d here are chosen to be 0, 1 and 3.5, with d=3.5 giving a very high value for the variance. Various results are given in the Tables 3.6 - 3.9.

For designs A and B it is observed that multicollinearity does not affect the power of these statistics. For a moderate increase in the variance and for different sample sizes the statistics; t_i, t_i^*, Mt_i^*, $LD_i(\sigma^2|\beta)$, $LD_i(\beta,\sigma^2)$, and kurtosis have about 50 - 60 percent power to detect outliers, in both designs A and B. For large variance these statsitics perform better in all the above cases. Thus, there is not much difference from the conclusions which are made for the mean shift model.

The results for designs C and D are presented in Table 3.10 and 3.11 respectively. Along with the influential observations, which are introduced through the selection of design itself, are introduced outliers of mean shift type (model (3.1)) for d=0, 1, and 3.5. Here d=0 means only influential observations in the data. The columns in Table 3.10 and 3.11 corresponding to d=0 for designs C and D show that, performance of W_i is better than the other statistics. However, C_i and $LD_i(\beta|\sigma^2)$ are also good competitors for W_i. Further, as expected, it is observed that the statistics; t_i, t_i^*, and $LD_i(\sigma^2|\beta)$ may be used to detect outliers (see columns corresponding to d=3.5) but not influential observations.

Table 3.2: Proportions of Detecting Outliers for Design A
(mean shift model for n=10, 20, and 50)

Statistics	n=10		n=20		n=50	
	γ=1,2	γ=.005,2	γ=1,2	γ=.005,2	γ=1,2	γ=.005,2
	.697(.269)	.697(.267)	.855(.395)	.855(.396)	.916(.466)	.918(.466)
t_i	.083(.046)	.082(.045)	.118(.047)	.120(.048)	.149(.061)	.148(.060)
	(.054)	(.053)	(.047)	(.046)	(.052)	(.048)
	.762(.279)	.760(.277)	.861(.401)	.861(.400)	.991(.467)	.991(.466)
t_i^*	.105(.050)	.105(.047)	.123(.049)	.125(.051)	.373(.062)	.375(.060)
	(.056)	(.054)	(.047)	(.049)	(.052)	(.050)
	0(.004)	0(.004)	.669(.074)	.640(.066)	.871(.029)	.870(.023)
C_i	0(.026)	0(.026)	.036(.099)	.028(.097)	.102(.047)	.104(.048)
	(.044)	(.044)	(.100)	(.093)	(.048)	(.050)
	.400(.009)	.398(.008)	.746(.063)	.727(.051)	.980(.047)	.980(.040)
WK_i	.025(.018)	.025(.047)	.056(.055)	.045(.052)	.308(.031)	.309(.032)
	(.022)	(.022)	(.060)	(.055)	(.030)	(.021)
	.400(.009)	.400(.008)	.745(.064)	.727(.051)	.980(.051)	.980(.051)
C_i^*	.026(.018)	.025(.017)	.056(.055)	.045(.052)	.308(.031)	.309(.032)
	(.022)	(.022)	(.060)	(.055)	(.032)	(.031)
	.317(.006)	.315(.005)	.721(.101)	.698(.085)	.980(.066)	.980(.057)
W_i	.017(.031)	.017(.031)	.047(.107)	.040(.101)	.307(.052)	.308(.036)
	(.046)	(.046)	(.080)	(.099)	(.052)	(.047)
	.617(.142)	.615(.142)	.836(.350)	.834(.348)	.991(.510)	.991(.509)
Mt_i^*	.058(.033)	.058(.033)	.109(.040)	.108(.039)	.370(.054)	.372(.056)
	(.047)	(.046)	(.046)	(.043)	(.042)	(.041)
	.264(0)	.264(0)	.668(.064)	.665(.060)	.824(.249)	.821(.244)
AP_i	.011(0)	.011(0)	.035(.006)	.079(.007)	.080(.014)	.082(.013)
	(0)	(0)	(.007)	(.006)	(.016)	(.013)
	0(.004)	0(.004)	.668(.064)	.665(.060)	.824(.249)	.821(.244)
$LD_i(\beta \mid \sigma^2)$	0(.026)	0(.026)	.091(.098)	.079(.097)	.015(.047)	.082(.013)
	(.044)	(.044)	(.100)	(.093)	(.048)	(.045)
	.569(.175)	.568(.175)	.819(.365)	.818(.362)	.925(.444)	.925(.436)
$LD_i(\beta, \sigma^2)$	.050(.053)	.050(.052)	.092(.085)	.091(.081)	.156(.085)	.158(.086)
	(.061)	(.061)	(.082)	(.083)	(.080)	(.068)
	.696(.269)	.695(.267)	.859(.395)	.860(.396)	.972(.464)	.972(.463)
$LD_i(\sigma^2 \mid \beta)$	.082(.046)	.081(.045)	.123(.047)	.123(.048)	.264(.060)	.265(.061)
	(.054)	(.053)	(.047)	(.046)	(.052)	(.049)
	(.440)	(.440)	(.417)	(.429)	(.470)	(.472)
Kurtosis	(.057)	(.057)	(.063)	(.060)	(.065)	(.065)
	(.040)	(.040)	(.052)	(.048)	(.064)	(.063)

The first two rows are the proportions of detecting the correct outlier. The figures are in percentages where first row corresponds to d=3.5 and second row to d=1. For d=3.5 and 1, the figures given in the parentheses are the proportions of detecting the outliers. For d=0, the figures in the third row, are the estimated level of significance.

Table 3.3: Proportions of Detecting Outliers for Design B
(mean shift model for n=10)

Statistics	α=0	α=.1	α=.2	α=.3	α=.4
	.616(.214)	.679(.257)	.545(.179)	.634(.248)	.591(.214)
t_i	.061(.041)	.081(.041)	.079(.038)	.069(.043)	.084(.039)
	(.040)	(.050)	(.056)	(.044)	(.054)
	.692(.221)	.761(.272)	.638(.191)	.706(.257)	.666(.224)
t_i^*	.088(.042)	.107(.045)	.104(.044)	.084(.044)	.101(.040)
	(.042)	(.057)	(.061)	(.050)	(.058)
	.686(.012)	.200(.065)	.812(.022)	.722(.017)	.722(.044)
C_i	.146(.037)	.098(.020)	.213(.039)	.102(.003)	.056(.109)
	(.031)	(.060)	(.016)	(.013)	(.020)
	.692(.060)	.566(.045)	.750(.114)	.714(.094)	.694(.069)
WK_i	.121(.017)	.100(.017)	.116(.020)	.114(.008)	.083(.035)
	(.019)	(.033)	(.027)	(.033)	(.031)
	.692(.060)	.567(.046)	.750(.114)	.714(.096)	.693(.069)
C_i^*	.121(.017)	.100(.018)	.163(.020)	.114(.008)	.084(.035)
	(.019)	(.033)	(.027)	(.033)	(.031)
	.692(.066)	.508(.068)	.783(.144)	.717(.108)	.705(.095)
W_i	.132(.037)	.100(.026)	.192(.038)	.112(.011)	.011(.084)
	(.035)	(.059)	(.048)	(.066)	(.039)
	.692(.216)	.662(.184)	.714(.252)	.709(.227)	.680(.255)
Mt_i^*	.108(.052)	.101(.041)	.143(.065)	.117(.026)	.091(.072)
	(.039)	(.054)	(.052)	(.049)	(.045)
	.294(0)	.289(0)	.305(0)	.327(0)	.304(0)
AP_i	.023(0)	.022(0)	.022(0)	.021(0)	.022(0)
	(0)	(0)	(0)	(0)	(0)
	.700(.010)	.234(.065)	.819(.022)	.732(.022)	.734(.043)
$LD_i(\beta\|\sigma^2)$	.150(.036)	.100(.020)	.218(.038)	.107(.003)	.061(.108)
	(.031)	(.058)	(.016)	(.013)	(.062)
	.615(.215)	.604(.219)	.639(.222)	.637(.256)	.609(.233)
$LD_i(\beta,\sigma^2)$	.071(.047)	.079(.051)	.107(.061)	.090(.040)	.047(.063)
	(.051)	(.060)	(.063)	(.061)	(.064)
	.616(.212)	.677(.257)	.545(.179)	.634(.248)	.591(.214)
$LD_i(\sigma^2\|\beta)$	.060(.041)	.081(.041)	.097(.038)	.097(.038)	.083(.039)
	(.039)	(.052)	(.056)	(.044)	(.054)
	(.120)	(.308)	(.135)	(.150)	(.162)
Kurtosis	(.042)	(.036)	(.047)	(.036)	(.036)
	(.044)	(.015)	(.048)	(.074)	(.047)

The first two rows are the proportions of detecting the correct outlier. The figures are in percentages where first row corresponds to d=3.5 and second row to d=1. For d=3.5 and 1, the figures given in the parentheses are the proportions of detecting the outliers. For d=0, the figures in the third row, are the estimated level of significance.

Table 3.3: (continued)

Statistics	α=.5	α=.6	α=.7	α=.8	α=.9
	.649(.257)	.631(.239)	.633(.229)	.684(.251)	.603(.225)
t_i	.085(.059)	.075(.050)	.093(.049)	.079(.055)	.085(.057)
	(.051)	(.042)	(.043)	(.055)	(.039)
	.721(.265)	.715(.253)	.708(.243)	.750(.257)	.681(.240)
t_i^*	.125(.062)	.103(.051)	.124(.052)	.106(.057)	.110(.059)
	(.053)	(.047)	(.052)	(.059)	(.043)
	.071(.103)	.331(.080)	.606(.119)	.143(.007)	.580(.020)
C_i	0(.074)	.015(.040)	0(.011)	.034(.021)	.031(.080)
	(.008)	(.080)	(.073)	(.086)	(.031)
	.509(.042)	.573(.043)	.665(.058)	.555(.024)	.635(.053)
WK_i	.024(.022)	.058(.028)	.041(.011)	.116(.030)	.073(.027)
	(.022)	(.044)	(.019)	(.043)	(.015)
	.510(.043)	.575(.043)	.665(.058)	.557(.024)	.638(.053)
C_i^*	.024(.022)	.058(.028)	.042(.011)	.116(.030)	.073(.027)
	(.022)	(.044)	(.019)	(.043)	(.015)
	.460(.072)	.528(.081)	.654(.116)	.494(.002)	.619(.058)
W_i	.018(.049)	.052(.042)	.031(.018)	.119(.044)	.068(.059)
	(.042)	(.067)	(.053)	(.067)	(.030)
	.622(.198)	.633(.228)	.679(.256)	.661(.177)	.650(.210)
Mt_i^*	.067(.029)	.073(.036)	.077(.030)	.110(.042)	.090(.055)
	(.049)	(.060)	(.049)	(.053)	(.044)
	.277(0)	.290(0)	.300(0)	.285(0)	.302(0)
AP_i	.014(0)	.014(0)	.015(0)	.020(0)	.019(0)
	(0)	(0)	(0)	(0)	(0)
	.099(.100)	.353(.080)	.633(.118)	.170(.007)	.597(.020)
$LD_i(\beta\|\sigma^2)$	0(.074)	.021(.038)	0(.01)	.131(.034)	.040(.079)
	(.008)	(.080)	(.073)	(.084)	(.031)
	.565(.227)	.578(.211)	.603(.221)	.596(.201)	.582(.212)
$LD_i(\beta,\sigma^2)$	.052(.061)	.058(.061)	.062(.056)	.083(.063)	.069(.073)
	(.055)	(.066)	(.056)	(.063)	(.055)
	.649(.257)	.631(.239)	.633(.229)	.683(.251)	.603(.225)
$LD_i(\sigma^2\|\beta)$	.084(.059)	.074(.050)	.093(.049)	.079(.055)	.085(.057)
	(.051)	(.042)	(.043)	(.055)	(.039)
	(.335)	(.277)	(.193)	(.350)	(.172)
Kurtosis	(.080)	(.049)	(.050)	(.053)	(.059)
	(.048)	(.071)	(.045)	(.058)	(.043)

The first two rows are the proportions of detecting the correct outlier. The figures are in percentages where first row corresponds to d=3.5 and second row to d=1. For d=3.5 and 1, the figures given in the parentheses are the proportions of detecting the outliers. For d=0, the figures in the third row, are the estimated level of significance.

Table 3.4: Proportions of Detecting Outliers for Design B
(mean shift model for n=20)

Statistics	α=0	α=.1	α=.2	α=.3	α=.4
	.790(.352)	.848(.387)	.815.(.390)	.848(.422)	.818(.361)
t_i	.114(.049)	.117(.056)	.118(.053)	.124(.052)	.130(.060)
	(.050)	(.052)	(.045)	(.050)	(.056)
	.804(.360)	.856(.394)	.824(.397)	.858(.430)	.831(.366)
t_i^*	.120(.052)	.125(.053)	.126(.056)	.134(.052)	.140(.065)
	(.051)	(.053)	(.048)	(.051)	(.058)
	.932(.300)	.893(.182)	.929(.202)	.718(.053)	.925(.179)
C_i	.192(.023)	.031(.007)	.186(.050)	.219(.047)	.040(.026)
	(.059)	(.069)	(.029)	(.015)	(.048)
	.919(.033)	.886(.219)	.920(.279)	.757(.063)	.917(.264)
WK_i	.176(.015)	.046(.008)	.179(.049)	.201(.027)	.065(.020)
	(.024)	(.037)	(.020)	(.021)	(.035)
	.919(.332)	.886(.219)	.920(.297)	.757(.064)	.917(.265)
C_i^*	.176(.015)	.046(.008)	.179(.048)	.200(.027)	.065(.020)
	(.035)	(.037)	(.020)	(.021)	(.035)
	.932(.386)	.889(.296)	.929(.324)	.741(.083)	.927(.297)
W_i	.188(.029)	.041(.012)	.182(.070)	.218(.052)	.057(.030)
	(.060)	(.065)	(.045)	(.035)	(.054)
	.867(.481)	.865(.413)	.865(.489)	.833(.372)	.872(.470)
Mt_i^*	.187(.047)	.110(.038)	.157(.065)	.155(.061)	.155(.053)
	(.048)	(.043)	(.045)	(.042)	(.045)
	.815(.108)	.760(.109)	.815(.132)	.683(.079)	.815(.113)
AP_i	.082(.003)	.041(.002)	.079(.002)	.115(.003)	.046(0)
	(.007)	(.005)	.(002)	(.003)	(.005)
	.957(.300)	.930(.182)	.953(.202)	.837(.202)	.952(.178)
$LD_i(\beta\|\sigma^2)$	.270(.023	.071(.007)	.157(.065)	.291(.047)	.097(.026)
	(.058)	(.068)	(.060)	(.030)	(.048)
	.878(.486)	.872(.437)	.889(.476)	.807(.384)	.890(.456)
$LD_i(\beta,\sigma^2)$	.152(.069)	.092(.055)	.162(.080)	.167(.075)	.106(.053)
	(.062)	(.064)	(.050)	(.060)	(.062)
	.802(.352)	.854(.385)	.821(.390)	.858(.422)	.829(.361)
$LD_i(\sigma^2\|\beta)$	.127(.052)	.137(.055)	.128(.053)	.131(.049)	.138(.049)
	(.050)	(.051)	(.048)	(.050)	(.056)
	(.213)	(.328	(.283)	(.328)	(.261)
Kurtosis	(.041)	(.052)	(.058)	(.044)	(.049)
	(.045)	(.056)	(.049)	(.047)	(.052)

The first two rows are the proportions of detecting the correct outlier. The figures are in percentages where first row corresponds to d=3.5 and second row to d=1. For d=3.5 and 1, the figures given in the parentheses are the proportions of detecting the outliers. For d=0, the figures in the third row, are the estimated level of significance.

Table 3.4: (continued)

Statistics	α=.5	α=.6	α=.7	α=.8	α=.9
	.794(.330)	.844(.393)	.887(.480)	.863(.419)	.833(.360)
t_i	.131(.049)	.134(.051)	.136(.054)	.132(.053)	.133(.050)
	(.057)	(.044)	(.054)	(.051)	(.046)
	.804(.337)	.854(.402)	.892(.485)	.868(.426)	.845(.371)
t_i^*	.139(.064)	.142(.053)	.141(.022)	.141(.056)	.139(.050)
	(.059)	(.047)	(.055)	(.055)	(.050)
	.935(.144)	.702(.035)	.882(.069)	.747(.033)	.935(.092)
C_i	.281(.016)	.198(.032)	082(.028)	.015(.021)	.031(.018)
	(.057)	(.053)	(.037)	(.060)	(.057)
	.918(.234)	.761(.042)	.862(.049)	.787(.067)	.927(.222)
WK_i	.253(.023)	.189(.023)	.092(.017)	.047(.013)	.022(.020)
	(.035)	(.032)	(.025)	(.024)	(.049)
	.918(.235)	.759(.044)	.862(.049)	.787(.068)	.926(.223)
C_i^*	.253(.023)	.189(.023)	.192(.018)	.047(.013)	.022(.020)
	(.035)	(.032)	(.025)	(.024)	(.049)
	.935(.258)	.843(.108)	.896(.085)	.774(.067)	.934(.223)
W_i	.281(.026)	.196(.038)	.086(.032)	.060(.029)	.041(.027)
	(.060)	(.060)	(.049)	(.060)	(.080)
	.847(.445)	.848(.398)	.854(.390)	.854(.400)	.886(.559)
Mt_i^*	.151(.050)	.196(.038)	.154(.058)	.163(.047)	.176(.054)
	(.050)	(.053)	(.043)	(.045)	(.046)
	.788(.106)	.733(.074)	.678(.077)	.697(.082)	.804(.105)
AP_i	.132(.007)	.092(.004)	.061(.002)	.047(.004)	.045(.002)
	(.006)	(.004)	(.003)	(.003)	(.009)
	.956(.144)	.903(.039)	.936(.069)	.865(.033)	.961(.092)
$LD_i(\beta\|\sigma^2)$	.268(.144)	.903(.039)	.936(.069)	.865(.033)	.961(.092)
	(.057)	(.035)	(.037)	(.065)	(.070)
	.872(.435)	.844(.387)	.816(.360)	.835(.405)	.903(.438)
$LD_i(\beta,\sigma^2)$	.195(.057)	.164(.057)	.140(.052)	.135(.058)	.125(.057)
	(.065)	(.056)	(.062)	(.057)	(.072)
	.801(.329)	.484(.392)	.891(.480)	.868(.418)	.841(.360)
$LD_i(\sigma^2\|\beta)$	.139(.048)	.141(.051)	.140(.058)	.158(.053)	.137(.048)
	(.057)	(.045)	(.056)	(.051)	(.046)
	(.214)	(.380)	(.482)	(.428)	.(291)
Kurtosis	(.045)	(.043)	(.053)	(.052)	(.062)
	(.064)	(.050)	(.049)	(.059)	(.055)

The first two rows are the proportions of detecting the correct outlier. The figures are in percentages where first row corresponds to d=3.5 and second row to d=1. For d=3.5 and 1, the figures given in the parentheses are the proportions of detecting the outliers. For d=0, the figures in the third row, are the estimated level of significance.

Table 3.5: Proportions of Detecting Outliers for Design B
(mean shift model for n=50)

Statistics	α=0	α=.1	α=.2	α=.3	α=.4
	.918(.483)	.907(.418)	.923.(502)	.934(.490)	.915(.465)
t_i	.150(.044)	.168(.045)	.176(.050)	.185(.054)	.178(.046)
	(.050)	(.050)	(.048)	(.047)	(.054)
	.980(.484)	.975(.485)	.981(.502)	.989(.490)	.982(.466)
t_i^*	.362(.046)	.412(.045)	.398(.050)	.418(.054)	.408(.048)
	(.050)	(.050)	(.048)	(.047)	(.053)
	.819(.047)	.893(.060)	.708(.019)	.852(.011)	.911(.013)
C_i	.013(.022)	.016(.019)	.030(.004)	.036(.043)	.021(.270)
	(.028)	(.036)	(.042)	(.016)	(.018)
	.961(.037)	.974(.080)	.939(.019)	.979(.021)	.978(.016)
WK_i	.164(.018)	.186(.019)	.119(.005)	.233(.031)	.196(.020)
	(.024)	(.022)	(.028)	(.012)	(.026)
	.961(.038)	971(.084)	.938(.018)	.975(.024)	.976(.068)
C_i^*	.164(.018)	.186(.019)	.117(.005)	.231(.030)	.194(.020)
	(.026)	(.025)	(.031)	(.014)	(.023)
	.961(.068)	.974(.106)	.938(.029)	.978(.026)	.978(.070)
W_i	.155(.022)	.181(.019)	.206(.014)	.229(.039)	.186(.024)
	(.034)	(.043)	(.046)	(.042)	(.046)
	.976(.446)	.975(.470)	.980(.456)	.988(.469)	.982(.451)
Mt_i^*	.374(.052)	.398(.046)	.381(.047)	.407(.049)	.393(.063)
	(.045)	(.048)	(.049)	(.046)	(.050)
	.804(.253)	.837(.287)	.803(.217)	.830(.236)	.840(.254)
AP_i	.034(.008)	.050(.008)	.038(.005)	.063(.022)	.005(.008)
	(.023)	(.020)	(.180)	(.016)	(.027)
	.670(.409)	.635(.060)	.551(.021)	.520(.012)	.648(.013)
$LD_i(\beta\|\sigma^2)$	.030(.018)	.040(.017)	.010(.002)	.120(.036)	.020(.019)
	(.028)	(.039)	(.042)	(.016)	(.026)
	.917(.443)	.915(.481)	.895(.419)	.932(.443)	.927(.452)
$LD_i(\beta,\sigma^2)$	.091(.020)	.125(.021)	.116(.016)	.147(.025)	.131(.020)
	(.073)	(.061)	(.069)	(.061)	(.057)
	.961(.480)	.944(.478)	.964(.501)	.972(.489)	.952(.461)
$LD_i(\sigma^2\|\beta)$	.255(.044)	.290(.044)	.287(.049)	.297(.054)	.294(.054)
	(.050)	(.045)	(.042)	(.046)	(.056)
	(.484)	(.455	(.513)	(.497)	(.455)
Kurtosis	(.042)	(.048)	(.054)	(.053)	(.051)
	(.057)	(.046)	(.051)	(.045)	(.059)

The first two rows are the proportions of detecting the correct outlier. The figures are in percentages where first row corresponds to d=3.5 and second row to d=1. For d=3.5 and 1, the figures given in the parentheses are the proportions of detecting the outliers. For d=0, the figures in the third row, are the estimated level of significance.

Table 3.5: (continued)

Statistics	α=.5	α=.6	α=.7	α=.8	α=.9
	.928(.514)	.902(.454)	.926(.4681)	.914(.497)	.923(.476)
t_i	.150(.047)	.157(.047)	.168(.042)	.170(.054)	.163(.056)
	(.045)	(.046)	(.057)	(.051)	(.045)
	.989(.514)	.976(.455)	.984(.470)	.080(.497)	.985(.478)
t_i^*	.372(.047)	.370(.047)	.168(.042)	.394(.054)	.409(.058)
	(.046)	(.045)	(.058)	(.051)	(.045)
	.898(.026)	.944(.135)	.810(.003)	.633(.008)	.822(.017)
C_i	.304(.017)	.089(.026)	286(.040)	.071(.025)	.007(.028)
	(.026)	(.022)	(.018)	(.033)	(.019)
	.886(.016)	.987(.203)	.961(.010)	.924(.012)	.973(.027)
WK_i	.403(.016)	.430(.017)	.496(.029)	.267(.022)	.138(.021)
	(.035)	(.028)	(.044)	(.026)	(.017)
	.876(.017)	.984(.210)	.961(.010)	.921(.012)	.971(.030)
C_i^*	.402(.016)	.319(.017)	.493(.027)	.265(.022)	.137(.021)
	(.040)	(.029)	(.046)	(.028)	(.020)
	.876(.027)	.988(.229)	.961(.011)	.922(.020)	.971(.037)
W_i	.417(.016)	.318(.022)	.502(.033)	.259(.024)	.120(.026)
	(.052)	(.046)	(.052)	(.039)	(.045)
	.984(.480)	.979(.473)	.982(.442)	.979(.441)	.985(.551)
Mt_i^*	.396(.049)	.365(.050)	.401(.061)	.381(.054)	.384(.065)
	(.048)	(.046)	(.052)	(.038)	(.044)
	.773(.230)	.881(.301)	.812(.213)	.792(.206)	.790(.262)
AP_i	.197(.008)	.062(.015)	.182(.010)	.071(.011)	.050(.014)
	(.018)	(.020)	(.037)	(.037)	(.022)
	.645(.026)	.795(.141)	.507(.003)	.589(.008)	.030(.040)
$LD_i(\beta\|\sigma^2)$	.115(.014)	.050(.020)	.108(.027)	.024(.022)	.030(.040)
	(.023)	(.025)	(.031)	(.033)	(.059)
	.883(.443)	.946(.498)	.916(.405)	.887(.395)	.920(.438)
$LD_i(\beta,\sigma^2)$	.276(.017)	.164(.022)	.271(.022)	.159(.020)	.112(.027)
	(.060)	(.057)	(.062)	(.066)	(.065)
	.967(.510)	.954(.447)	.963(.466)	.965(.494)	.968(.472)
$LD_i(\sigma^2\|\beta)$	.261(.047)	.276(.047)	.271(.049)	.273(.053)	.286(.053)
	(.045)	(.046)	(.055)	(.050)	(.050)
	(.529)	(.423)	(.469)	(.498)	(.454)
Kurtosis	(.052)	(.060)	(.050)	(.059)	(.053)
	(.046)	(.045)	(.055)	(.055)	(.054)

The first two rows are the proportions of detecting the correct outlier. The figures are in percentages where first row corresponds to d=3.5 and second row to d=1. For d=3.5 and 1, the figures given in the parentheses are the proportions of detecting the outliers. For d=0, the figures in the third row, are the estimated level of significance.

Table 3.6: Proportions of Detecting Outliers for Design A
(scale shift model for n=10, 20, 50)

Statistics	n=10		n=20		n=50	
	γ=1,2	γ=.005, 2	γ=1,2	γ=.005,2	γ=1,2	γ=.005,2
	.720(.555)	.720(.552)	.765(.634)	.768(.635)	.768(.646)	.767(.646)
t_i	.319(.149)	.319(.149)	.420(.191)	.425(.194)	.456(.226)	.451(.225)
	(.054)	(.053)	(.047)	(.045)	(.052)	(.049)
	.740(.565)	.740(.565)	.769(.635)	.770(.640)	.846(.646)	.846(.646)
t_i^*	.365(.161)	.364(.156)	.428(.194)	.433(.196)	.625(.227)	.624(.225)
	(.056)	(.054)	(.047)	(.046)	(.052)	(.049)
	0(.012)	0(.012)	.701(.048)	.696(.041)	.743(.332)	.742(.318)
C_i	0(.024)	0(.023)	.295(.088)	.276(.081)	.398(.039)	.394(.034)
	(.044)	(.044)	(.100)	(.094)	(.048)	(.045)
	.604(.215)	.603(.215)	.727(.418)	.725(.406)	.824(.454)	.823(.444)
WK_i	.190(.012)	.189(.012)	.328(.073)	.320(.068)	.590(.063)	.587(.052)
	(.022)	(.022)	(.060)	(.056)	(.030)	(.021)
	.604(.215)	.604(.215)	.727(.418)	.725(.406)	.824(.456)	.822(.451)
C_i^*	.190(.012)	.190(.012)	.328(.073)	.319(.068)	.588(.066)	.587(.052)
	(.022)	(.022)	(.060)	(.055)	(.032)	(.024)
	.518(.184)	.581(.184)	.721(.419)	.711(.405)	.824(.460)	.821(.451)
W_i	.158(.027)	.158(.027)	.317(.115)	.307(.103)	.587(.079)	.585(.056)
	(.046)	(.046)	(.048)	(.056)	(.052)	(.047)
	.688(.548)	.688(.547)	.763(.597)	.764(.596)	.844(.620)	.843(.621)
Mt_i^*	.281(.157)	.281(.147)	.409(.163)	.409(.187)	.624(.188)	.622(.196)
	(.047)	(.046)	(.046)	(.044)	(.042)	(.041)
	.548(0)	.548(0)	.699(.459)	.699(.462)	.725(.50)	.725(.567)
AP_i	.132(0)	.132(0)	.295(.054)	.292(.052)	.353(.129)	.353(.127)
	(0)	(0)	(.007)	(.006)	(.016)	(.013)
	0(.012)	(.012)	.753(.048)	.747(.040)	.642(.335)	.638(.321)
$LD_i(\beta \mid \sigma^2)$	0(.023)	0(.023)	.382(.088)	.292(.052)	.205(.039)	.197(.034)
	(.044)	(.041)	(.043)	(.099)	(.048)	(.045)
	.677(.516)	.666(.515)	.755(.622)	.757(.621)	.770(.635)	.770(.630)
$LD_i(\beta, \sigma^2)$	.256(.119)	.256(.116)	.384(.208)	.385(.203)	.467(.232)	.465(.223)
	(.061)	(.060)	(.082)	(.068)	(.080)	(.068)
	.720(.555)	.720(.552)	.768(.634)	.769(.635)	.811(.646)	.810(.645)
$LD_i(\sigma^2 \mid \beta)$	.319(.149)	.319(.148)	.428(.191)	.431(.194)	.549(.224)	.547(.223)
	(.054)	(.053)	(.047)	(.045)	(.052)	(.049)
	(.630)	(.630)	(.634)	(.646)	(.640)	(.640)
Kurtosis	(.215)	(.215)	(.213)	(.217)	(.215)	(.215)
	(.040)	(.040)	(.052)	(.051)	(.064)	(.064)

The first two rows are the proportions of detecting the correct outlier. The figures are in percentages where first row corresponds to d=3.5 and second row to d=1. For d=3.5 and 1, the figures given the parentheses are the proportions of detecting the outliers. For d=0, figures in the third row, are the estimated level of significance.

Table 3.7: Proportions of Detecting Outliers for Design B
(scale shift model for n=10)

Statistics	α=0	α=.1	α=.2	α=.3	α=.4
	.691(.543)	.656(.503)	.677(.5139)	.647(.508)	.682(.538)
t_i	.317(.162)	.315(.138)	.277(.144)	.331(.150)	.304(.149)
	(.040)	(.050)	(.056)	(.044)	(.054)
	.725(.546)	.676(.509)	.705(.523)	.679(.516)	.704(.545)
t_i^*	.366(.268)	.360(.152)	.319(.148)	.371(.160)	.348(.153)
	(.042)	(.057)	(.061)	(.050)	(.056)
	0(.073)	.832(.088)	.658(.023)	.656(.007)	.487(.030)
C_i	0(.010)	.352(.304)	.380(.060)	.325(.050)	.139(.045)
	(.031)	(.060)	(.016)	(.013)	(.020)
	.504(.162)	.761(.435)	.687(.361)	.670(.352)	.637(.303)
WK_i	.147(.016)	.357(.120)	.352(.076)	.237(.063)	.265(.039)
	(.019)	(.033)	(.027)	(.033)	(.031)
	.504(.163)	761(.435)	.687(.435)	.671(.352)	.639(.303)
C_i^*	.148(.016)	.357(.120)	.353(.076)	.237(.063)	.266(.030)
	(.019)	(.033)	(.027)	(.033)	(.031)
	.465(.156)	.778(.489)	.678(.356)	.668(.347)	.623(.285)
W_i	.109(.022)	.357(.279)	.365(.115)	.296(.101)	.243(.058)
	(.035)	(.059)	(.048)	(.066)	(.039)
	.639(.430)	.741(.499)	.696(.513)	.674(.498)	.676(.525)
Mt_i^*	.267(.147)	.357(.136)	.339(.140)	.227(.139)	.302(.152)
	(.039)	(.054)	(.052)	(.049)	(.045)
	.517(0)	.583(0)	.552(0)	.536(0)	.545(0)
AP_i	.125(0)	.146(0)	.152(0)	.109(0)	.139(0)
	(0)	(0)	(0)	(0)	(0)
	0(.073)	.791(.087)	.666(.023)	.664(.007)	.498(.028)
$LD_i(\beta\|\sigma^2)$	0(.008)	.357(.301)	.389(.060)	.309(.055)	.150(.050)
	(.031)	(.059)	(.016)	(.013)	(.02)
	.623(.500)	.708(.545)	.672(.503)	.642(.503)	.656(.527)
$LD_i(\beta,\sigma^2)$	.248(.136)	.310(.196)	.299(.164)	.261(.145)	.275(.143)
	(.051)	(.060)	(.063)	(.061)	(.064)
	.691(.543)	.671(.485)	.677(.513)	.647(.508)	.682(.538)
$LD_i(\sigma^2\|\beta)$	.316(.162)	.315(.138)	.275(.144)	.331(.150)	.304(.149)
	(.039)	(.050)	(.056)	(.044)	(.054)
	(.464)	(.492)	(.488)	(.451)	(.565)
Kurtosis	(.252)	(.111)	(.124)	(.166)	(.177)
	(.044)	(.075)	(.048)	(.074)	(.047)

The first two rows are the proportions of detecting the correct outlier. The figures are in percentages where first row corresponds to d=3.5 and second row to d=1. For d=3.5 and 1, the figures given in the parentheses are the proportions of detecting the outliers. For d=0, the figures in the third row, are the estimated level of significance.

Table 3.7: (continued)

Statistics	α=.5	α=.6	α=.7	α=.8	α=.9
	.677(.534)	.680(.518)	.8698(.531)	.708(.558)	.683(.552)
t_i	.310(.154)	.314(.143)	.326(.154)	.345(.147)	.309(.147)
	(.051)	(.042)	(.043)	(.055)	(.039)
	.705(.539)	.707(.525)	.723(.543)	.732(.566)	.711(.558)
t_i^*	.360(.149)	.361(.149)	.365(.161)	.381(.159)	.345(.155)
	(.053)	(.047)	(.047)	(.052)	(.059)
	.608(.004)	.714(.009)	.638(.078)	0(.008)	(.158)
C_i	.287(.003)	.042(.039)	0(.116)	.113(.037)	.225(.002)
	(.008)	(.050)	(.073)	(.086)	(.031)
	.664(.350)	.711(.378)	.692(.347)	.536(.154)	.603(.255)
WK_i	.322(.041)	.233(.030)	.171(.028)	.279(.042)	.294(.034)
	(.022)	(.044)	(.019)	(.043)	(.015)
	.664(.352)	.711(.379)	.692(.347)	.536(.154)	.603(.256)
C_i^*	.322(.041)	.234(.030)	.174(.028)	.282(.042)	.295(.034)
	(.022)	(.044)	(.019)	(.043)	(.015)
	.654(.341)	.713(.389)	.684(.360)	.490(.116)	.581(.302)
W_i	.315(.041)	.209(.048)	.136(.085)	.255(.051)	.280(.033)
	(.042)	(.067)	(.053)	(.067)	(.030)
	.685(.523)	.708(.495)	.705(.509)	.673(.550)	.661(.515)
Mt_i^*	.340(.146)	.299(.123)	.268(.140)	.341(.150)	.322(.132)
	(.049)	(.060)	(.049)	(.053)	(.044)
	.550(0)	.563(0)	.553(0)	.542(0)	.539(0)
AP_i	.156(0)	.130(0)	.124(0)	.150(0)	.150(0)
	(0)	(0)	(0)	(0)	(0)
	.615(.003)	.718(.009)	.650(.077)	0(.007)	0(.157)
$LD_i(\beta\|\sigma^2)$	.298(.003)	.058(.038)	0(.116)	.128(.037)	.234(.001)
	(.008)	(.080)	(.073)	(.084)	(.031)
	.657(.526)	.681(.523)	.684(.525)	.663(.514)	.650(.520)
$LD_i(\beta,\sigma^2)$	.300(.146)	.263(.126)	.259(.133)	.298(.136)	.281(.136)
	(.055)	(.066)	(.056)	(.063)	(.055)
	.677(.534)	.680(.518)	.698(.531)	.707(.588)	.683(.552)
$LD_i(\sigma^2\|\beta)$	.310(.154)	.313(.143)	.326(.154)	.345(.147)	.308(.147)
	(.051)	(.042)	(.043)	(.055)	(.039)
	(.533)	(.423)	(.518)	(.663)	.(603)
Kurtosis	(.139)	(.192)	(.231)	(.181)	(.144)
	(.048)	(.071)	(.045)	(.058)	(.043)

The first two rows are the proportions of detecting the correct outlier. The figures are in percentages where first row corresponds to d=3.5 and second row to d=1. For d=3.5 and 1, the figures given in the parentheses are the proportions of detecting the outliers. For d=0, the figures in the third row, are the estimated level of significance.

Table 3.8: Proportions of Detecting Outliers for Design B
(scale shift model for n=20)

Statistics	α=0	α=.1	α=.2	α=.3	α=.4
	.763(.604)	.752(.609)	.745(.599)	.756(.626)	.748(.624)
t_i	.416(.200)	.440(.208)	.427(.223)	.409(.214)	.397(.203)
	(.050)	(.052)	(.045)	(.050)	(.056)
	.768(.607)	.755(.615)	.752(.599)	.759(.627)	.766(.611)
t_i^*	.425(.205)	.446(.211)	.434(.224)	.418(.219)	.407(.207)
	(.051)	(.053)	(.048)	(.051)	(.058)
	.791(.514)	.754(.477)	.755(.492)	.636(.301)	.759(.452)
C_i	.310(.026)	.279(.038)	.58(.026)	.091(.052)	.254(.023)
	(.059)	(.069)	(.029)	(.015)	(.048)
	.791(.514)	.754(.477)	.755(.492)	.636(.301)	.759(.453)
WK_i	.332(.044)	.327(.051)	.186(.023)	.212(.028)	.296(.043)
	(.034)	(.037)	(.020)	(.021)	(.035)
	.791(.514)	754(.477)	.754(.492)	.636(.802)	.759(.453)
C_i^*	.332(.044)	.327(.051)	.186(.023)	.212(.028)	.296(.040)
	(.035)	(.037)	(.020)	(.021)	(.035)
	.794(.518)	.753(.471)	.755(.514)	.616(.34)	.758(.452)
W_i	.324(.054)	.314(.059)	.161(.034)	.196(.057)	.282(.039)
	(.060)	(.065)	(.045)	(.035)	(.054)
	.775(.602)	.755(.593)	.753(.585)	.756(.599)	.763(.590)
Mt_i^*	.409(.211)	.420(.205)	.382(.203)	.377(.208)	.375(.200)
	(.048)	(.043)	(.045)	(.042)	(.045)
	.726(.472)	.705(.456)	.699(.439)	.683(.465)	.720(.446)
AP_i	.282(.057)	.301(.068)	.267(.058)	.276(.050)	.271(.062)
	(.07)	(.005)	(.002)	(.003)	(.005)
	.826(.398)	.796(.456)	.788(.232)	.621(.081)	.803(.060)
$LD_i(\beta\|\sigma^2)$	.411(.026)	.376(.038)	.141(.026)	.198(.052)	.342(.023)
	(.058)	(.068)	(.060)	(.030)	(.048)
	.783(.617)	.753(.613)	.752(.618)	.753(.613)	.762(.607)
$LD_i(\beta,\sigma^2)$	.388(.198)	.397(.196)	.347(.180)	.346(.182)	.354(.191)
	(.062)	(.064)	(.050)	(.060)	(.062)
	.767(.603)	.753(.609)	.750(.599)	.783(.626)	.764(.606)
$LD_i(\sigma^2\|\beta)$	.425(.200)	.446(.207)	.434(.223)	.424(.214)	.407(.203)
	(.050)	(.051)	(.048)	(.050)	(.056)
	(.585)	(.603)	(.581)	(.655)	(.602)
Kurtosis	(.207)	(.215)	(.248)	(.250)	(.204)
	(.045)	(.056)	(.048)	(.047)	(.052)

The first two rows are the proportions of detecting the correct outlier. The figures are in percentages where first row corresponds to d=3.5 and second row to d=1. For d=3.5 and 1, the figures given in the parentheses are the proportions of detecting the outliers. For d=0, the figures in the third row, are the estimated level of significance.

Table 3.8: (continued)

Statistics	α=.5	α=.6	α=.7	α=.8	α=.9
	.748(.624)	.756(.600)	.758(.621)	.722(.630)	.759(.601)
t_i	.437(.202)	.380(.193)	.422(.214)	.415(.215)	.404(.187)
	(.057)	(.044)	(.054)	(.051)	(.046)
	.757(.625)	.760(.602)	.763(.623)	.777(.631)	.757(.604)
t_i^*	.447(.206)	.392(.197)	.428(.218)	.426(.221)	.413(.193)
	(.059)	(.047)	(.055)	(.055)	(.050)
	.606(.086)	.580(.092)	.787(.347)	.716(.150)	.655(.090)
C_i	.348(.004)	.288(.009)	.086(.028)	.191(.008)	.541(.100)
	(.057)	(.053)	(.037)	(.060)	(.057)
	.640(.342)	.569(.325)	.784(.523)	.750(.346)	.682(.341)
WK_i	.374(.050)	.318(.034)	.203(.025)	.265(.018)	.522(.134)
	(.035)	(.032)	(.025)	(.024)	(.049)
	.640(.342)	658(.325)	.784(.533)	.749(.346)	.682(.341)
C_i^*	.374(.050)	.318(.034)	.703(.026)	.265(.018)	.520(.134)
	(.035)	(.032)	(.025)	(.024)	(.049)
	.632(.323)	.641(.304)	.786(.532)	.781(.432)	.675(.328)
W_i	.364(.046)	.307(.034)	.179(.034)	.244(.021)	.541(.149)
	(.060)	(.060)	(.049)	(.060)	(.080)
	.731(.613)	.745(.589)	.768(.603)	.752(.618)	.733(.519)
Mt_i^*	.425(.198)	.373(.188)	.396(.196)	.389(.201)	.457(.198)
	(.050)	(.053)	(.043)	(.045)	(.046)
	.655(.453)	.679(.438)	.736(.489)	.685(.453)	.673(.428)
AP_i	.317(.063)	.207(.051)	.261(.059)	.275(.055)	.409(.073)
	(.006)	(.004)	(.003)	(.003)	(.003)
	.653(.086)	.675.(.050)	.809(.347)	.798(.135)	.700(.100)
$LD_i(\beta\|\sigma^2)$	.438(.004)	.369(.009)	.179(.028)	.277(.008)	.595(.100)
	(.057)	(.035)	(.037)	(.065)	(.070)
	.713(.610)	.730(.570)	.769(.631)	.731(.591)	.719(.578)
$LD_i(\beta,\sigma^2)$	.412(.182)	.355(.176)	.366(.189)	.361(.185)	.481(.231)
	(.065)	(.056)	(.062)	(.057)	(.072)
	.756(.624)	.759(.600)	.763(.621)	.774(.630)	.755(.602)
$LD_i(\sigma^2\|\beta)$	.445(.200)	.389(.193)	.426(.212)	.423(.215)	.413(.197)
	(.057)	(.045)	(.056)	(.051)	(.046)
	(.629)	(.628)	(.609)	(.640)	.(618)
Kurtosis	(.207)	(.191)	(.245)	(.232)	(.159)
	(.064)	(.050)	(.049)	(.059)	(.055)

The first two rows are the proportions of detecting the correct outlier. The figures are in percentages where first row corresponds to d=3.5 and second row to d=1. For d=3.5 and 1, the figures given in the parentheses are the proportions of detecting the outliers. For d=0, the figures in the third row, are the estimated level of significance.

Table 3.9: Proportions for Detecting Outliers for Design B
(Scale shift model for n=50)

Statistics	α=0	α=.1	α=.2	α=.3	α=.4
	.791(.635)	.762(.631)	.787(.642)	.790(.641)	.775(.634)
t_i	.451(.207)	.464(.247)	.475(.236)	.459(.246)	.473(.238)
	(.050)	(.050)	(.048)	(.047)	(.054)
	.864(.635)	.840(.631)	.858(.642)	.854(.642)	.844(.634)
t_i^*	.625(.207)	.621(.247)	.651(.236)	.619(.246)	.635(.238)
	(.050)	(.050)	(.048)	(.047)	(.043)
	.771(.401)	.698(.187)	.762(.362)	.599(.110)	.680(.080)
C_i	.406(.059)	.241(.037)	.323(.053)	.404(.030)	.264(.043)
	(.028)	(.036)	(.042)	(.016)	(.018)
	.860(.480)	.796(.381)	.851(.468)	.761(.239)	.795(.376)
WK_i	.587(.074)	.460(.032)	.545(.047)	.581(.064)	.487(.041)
	(.024)	(.022)	(.028)	(.012)	(.026)
	.854(.482)	.794(.382)	.850(.472)	.755(.243)	.793(.377)
C_i^*	.582(.074)	.450(.035)	.539(.050)	.575(.067)	.482(.044)
	(.026)	(.025)	(.031)	(.014)	(.023)
	.858(.493)	.796(.384)	.851(.473)	.755(.398)	.794(.388)
W_i	.586(.098)	.451(.047)	.541(.075)	.581(.070)	.483(.058)
	(.034)	(.043)	(.046)	(.042)	(.046)
	.863(.641)	.835(.651)	.857(.638)	.851(.659)	.837(.613)
Mt_i^*	.652(.188)	.613(.241)	.643(.211)	.615(.218)	.624(.230)
	(.045)	(.048)	(.049)	(.046)	(.050)
	.747(.576)	.702(.550)	.737(.577)	.711(.565)	.716(.560)
AP_i	.366(.137)	.337(.148)	.338(.158)	.371(.153)	.339(.165)
	(.023)	(.020)	(.018)	(.016)	(.027)
	.661(.404)	.572(.188)	.652(.366)	.574(.112)	.530(.089)
$LD_i(\beta\|\sigma^2)$	.208(.059)	.160(.037)	.156(.054)	.222(.031)	.188(.043)
	(.028)	(.039)	(.042)	(.016)	(.026)
	.799(.640)	.750(.656)	.794(.644)	.761(.655)	.763(.652)
$LD_i(\beta,\sigma^2)$	.457(.221)	.425(.233)	.454(.243)	.470(.246)	.438(.230)
	(.073)	(.061)	(.069)	(.061)	(.057)
	.826(.635)	.801(.631)	.830(.642)	.825(.641)	.811(.633)
$LD_i(\sigma^2\|\beta)$	.542(.206)	.548(.245)	.570(.235)	.549(.240)	.555(.237)
	(.050)	(.045)	(.042)	(.046)	(.056)
	(.631)	(.659)	.646)	(.652)	.(625)
Kurtosis	(.211)	(.260)	(.244)	(.248)	.(251)
	(.057)	(.046)	(.051)	(.045)	(.059)

The first two rows are the proportions of detecting the correct outlier. The figures are in percentages where first row corresponds to d=3.5 and second row to d=1. For d=3.5 and 1, the figures given in the parentheses are the proportions of detecting the outliers. For d=0, the figures in the third row, are the estimated level of significance.

Table 3.9 (continued)

Statistics	α=.5	α=.6	α=.7	α=.8	α=.9
	.773(.647)	.771(.618)	.775(.636)	.792(.660)	.780(.640)
t_i	.469(.223)	.463(.211)	.437(.233)	.446(.217)	.430(.206)
	(.045)	(.046)	(.057)	(.051)	(.045)
	.851(.647)	.843(.618)	.866(.636)	.871(.661)	.854(.640)
t_i^*	.646(.225)	.644(.211)	.615(.233)	.642(.217)	.593(.207)
	(.046)	(.045)	(.058)	(.051)	(.045)
	.711(.158)	.809(.262)	.683(.085)	.725(.190)	.695(.095)
C_i	.426(.111)	.167(.032)	.434(.042)	.453(.050)	.553(.149)
	(.033)	(.018)	(.026)	(.033)	(.019)
	.804(.383)	.871(.369)	.798(.360)	.829(.416)	.809(.357)
WK_i	.614(.104)	.424(.026)	.610(.076)	.643(.076)	.685(.149)
	(.035)	(.028)	(.044)	(.026)	(.017)
	.801(.388)	.867(.371)	.789(.368)	.822(.416)	.804(.360)
C_i^*	.616(.107)	.445(.028)	.604(.080)	.637(.080)	.682(.153)
	(.040)	(.029)	(.046)	(.028)	(.020)
	.801(.398)	.872(.480)	.790(.369)	.829(.428)	.804(.365)
W_i	.610(.161)	.415(.037)	.610(.085)	.643(.085)	.696(.169)
	(.052)	(.046)	(.050)	(.039)	(.045)
	.846(.637)	.850(.617)	.863(.611)	.868(.645)	.846(.630)
Mt_i^*	.643(.212)	.634(.198)	.615(.230)	.643(.213)	.630(.205)
	(.048)	(.046)	(.052)	(.038)	(.044)
	.725(.582)	.767(.584)	.705(.550)	.736(.604)	.715(.564)
AP_i	.306(.128)	.317(.111)	.368(.141)	.378(.144)	.476(.150)
	(.018)	(.020)	(.037)	(.022)	(.024)
	.574(.174)	.590(.363)	.539(.087)	.589(.193)	.553(.097)
$LD_i(\beta\|\sigma^2)$	.217(.111)	.126(.032)	.268(.042)	.273(.050)	.390(.150)
	(.023)	(.025)	(.031)	(.033)	(.059)
	.767(.659)	.806(.645)	.762(.616)	.785(.654)	.769(.657)
$LD_i(\beta,\sigma^2)$	.474(.259)	.415(.208)	.471(.239)	.482(.228)	.534(.252)
	(.060)	(.057)	(.062)	(.066)	(.065)
	.817(.645)	.809(.616)	.822(.636)	.842(.658)	.825(.638)
$LD_i(\sigma^2\|\beta)$	.576(.219)	.562(.210)	.574(.232)	.548(.217)	.521(.200)
	(.045)	(.046)	(.055)	(.050)	(.050)
	(.638)	(.605)	(.632)	(.660)	(.633)
Kurtosis	(.217)	(.225)	(.227)	(.213)	(.183)
	(.046)	(.045)	(.055)	(.055)	(.054)

The first two rows are the proportions of detecting the correct outlier. The figures are in percentages where first row corresponds to d=3.5 and second row to d=1. For d=3.5 and 1, the figures given in the parentheses are the proportions of detecting the outliers. For d=0, the figures in thei third row, are the estimated level of significance.

Table 3.10: Proportions of Detecting Outliers and Influential Observations for Design C

Statistics	d=0	d=1	d=3.5
t_i	.048(.030)	.108(.035)	.737(.260)
t_i^*	.059(.044)	.135(.052)	.785(.345)
C_i	.025(.722)	.059(.703)	.589(.795)
WK_i	.036(.516)	.095(.540)	.699(.663)
C_i^*	.036(.516)	.089(.511)	.689(.661)
W_i	.033(.831)	.088(.796)	.676(.884)
Mt_i^*	.050(.622)	.121(.631)	.751(.754)
$LD_i(\beta\|\sigma^2)$	.025(.722)	.016(.703)	.593(.794)
$LD_i(\beta,\sigma^2)$	.040(.248)	.097(.271)	.712(.463)
$LD_i(\sigma^2\|\beta)$	.048(.031)	.106(.036)	.734(.264)
Kurtosis	(.081)	(.067)	(.390)

The figures are in percentages. For d=1 and d=3.5, the figures are the proportions of detecting the correct outliers. The figures given in the parentheses for d=1 and d=3.5, are the proportions of detecting the outliers. For d=0, the figures in the parentheses are the proportions of identifying the influential observations.

Table 3.11: Proportions of Detecting Outliers and Influential Observations for Design D

Statistics	d=0	d=1	d=3.5
t_i	.041(.042)	.134(.045)	.833(.471)
t_i^*	.042(.042)	.144(.045)	.888(.472)
C_i	.009(.516)	.026(.514)	.619(.421)
WK_i	.013(.413)	.047(.417)	.720(.438)
C_i^*	.013(.417)	.048(.418)	.717(.435)
W_i	.011(.648)	.043(.616)	.705(.658)
Mt_i^*	.038(.216)	.129(.312)	.874(.590)
$LD_i(\beta\|\sigma^2)$	.009(.516)	.025(.511)	.614(.419)
$LD_i(\beta,\sigma^2)$	.030(.243)	.092(.230)	.855(.551)
$LD_i(\sigma^2\|\beta)$	.041(.042)	.141(.045)	.883(.471)
Kurtosis	(.055)	(.051)	(.502)

The figures are in percentages. For d=1 and d=3.5, the figures are the proportions of detecting the correct outliers. The figures given in the parentheses for d=1 and d=3.5, are the proportions of detecting the outliers. For d=0, the figures in the parentheses are the proportions of identifying the influential observations.

3.2 Performances of the Various Statistics for Real Data Sets

Below are nine selected data sets previously used by many researchers for comparing the performances of various statistics to detect outliers and influential observations.

Table 3.12 Several Data Sets

Data Sets	Studied by Different Authors
Mickey, Dunn and Clark (1967) ($n = 21$, $m = 2$)	Andrews and Pregibon (1978) Draper and John (1981)
Snedecor and Cochran (1967) ($n = 21$, $m = 3$)	Cook and Weisberg (1982)
Weisberg (1980) ($n = 19$, $m = 4$)	Cook and Weisberg (1982)
Forbes (1857) ($n = 17$, $m = 2$)	Weisberg (1980)
Chatterjee and Price (1977) ($n = 30$, $m = 2$)	Chatterjee and Price (1977)
Brownlee (1965) ($n = 21$, $m = 4$)	Cook (1979) Daniel and Wood (1971)
Moore (1975) ($n = 20$, $m = 6$)	Chatterjee and Hadi (1986)
Draper and Stoneman (1966) ($n = 10$, $m = 3$)	Tukey(1977a, b)
Aitchinson and Dunsmore (1975) ($n = 16$, $m = 2$)	Geisser (1987)

The computed statistics are summarized in Table 3.13. In Table 3.13, are presented the values of the statistics corresponding to those observations with extreme values for at least one of the statistics.

The results from Table 3.13 show that only the observation number 19, 17, and 12, belonging to the data sets 1, 2, and 4 respectively, have t_i, t_i^*, and $LD_i(\sigma^2|\beta)$ exceeding the corresponding calibration points. It is noted further that deletions of these observations do not change the values of R^2 and the estimates of the regression parameters in any significant way. Hence, we declare these observations to be outliers. Although, several other statistics identify the observations 17 and 12 belonging to the data sets 2 and 4 respectively as outliers, they fail to detect observation 19 of the data set 1 as an outlier. Hence, it is suggested that the statistics t_i, t_i^*, and $LD_i(\sigma^2|\beta)$ be used for detecting outliers.

Next it is observed that the statistics C_i, WK_i, C_i^*, W_i, Mt_i^*, AP_i, $LD_i(\sigma^2|\beta)$, and $LD_i(\beta,\sigma^2)$, have exceeded the corresponding calibration points for the observations 18, 3, 21, (1,17), and 16 respectively belonging to the data sets 1, 3, 6, 7, and 9. Further calculations have shown that these observations have strong influence on the parameter estimates and on the values of R^2. Hence, a marked claim is indicated that these observations are influential ones. The first observation belonging to the data set

Table 3.13: Computed Statistics for Detecting Outliers and Influential Observations for Real Data sets

Data Sets	Obs.	t_i	t_i^*	C_i	WK_i	C_i^*	W_i	Mt_i^*
Mickey, Dunn & Clark	18	.85	.84	.68*	1.15*	3.56*	8.75*	2.43
(n=21, m=2)	19	2.82*	3.6*	.22	.85	2.63	3.92	3.8
Snedecor & Cochran	6	.79	.78	.17	.72	1.61	4.05	1.45
(n=18, m=3)	17	3.17*	5.36*	.83*	2.67	5.98*	12.33*	6.69*
Weisberg	3	.81	.79	.92*	1.9*	3.68*	20.9*	5.34*
(n=19, m=4)	19	1.92	2.13	.20	.99	1.93	4.56	2.6
Forbes	12	3.7*	12.4*	.46	3.24*	8.87*	13.4*	13.25*
(n=17, m=2)	17	.25	.25	.009	.13	.36	.60	.32
Chatterjee & Price	27	1.80	1.88	.19	.65	2.45	3.75	2.1
(n=30, m=2)	29	2.2	2.38	.26	.78	2.94	4.45	2.64
Brownlee	17	.60	.59	.06	.50	1.03	2.9	1.01
(n=21, m=4)	21	2.62	3.30	.69*	2.10*	4.32*	11.10*	4.65*
Moore	1	2.63	3.58	.59*	2.55*	3.9*	13.67*	5.4*
(n=20, m=6)	17	.97	.97	1.77*	3.26*	4.9*	49.7*	11.9*
Draper & Stoneman	1	2.1	3.25	1.07*	2.75	4.21	10.84*	5.6
(n=10, m=3)	4	.96	.96	.48	1.18	1.81	5.66	2.4
Aitchinson & Dunsmore	10	2.07	2.40	.41	1.05	2.78	4.44	2.86
(n=16, m=2)	16	1.84	2.04	.81*	1.45*	3.72*	6.62*	3.06

Notes: (1) The numbers are the extreme values of the statistics for at least one of the statistics corresponding to these observations.

(2) Asterisk denots that the observations of exceeding the calibration point at 5% level of significance.

Table 3.13: (continued)

Data Sets	AP_i	$LD_i(\beta \mid \sigma^2)$	$LD_i(\beta, \sigma^2)$	$LD_i(\sigma^2 \mid \beta)$	$R^2_{(i)}$	R^2
Mickey, Dunn & Clark	.66*	1.44*	1.48	.01	.112	.410
(n=21, m=)2	.45	.48	3.86	3.05*	.572	
Snedecor & Cochran	.48	.10	.11	.001	.503	.482
(n=18, m=3)	.73*	2.78*	23.5*	14.8*	.525	
Weisberg	.85*	4.2*	4.66	.008	.021	.364
(n=19, m=4)	.38	.98	1.81	.53	.457	
Forbes	.92*	1.03	145.6*	133.6*	.9996	.995
(n=17, m=2)	.22	.02	.05	.03	.994	
Chatterjee & Price	.21	.42	.58	.11	.345	.396
(n=30, m=2)	.25	.56	1.05	.38	.355	
Brownlee	.42	.32	.32	.02	.911	.914
(n=21, m=4)	.58*	3.16*	8.34*	2.8	.949	
Moore	.66*	4.50*	14.6*	5.05	.831	.810
(n=20, m=6)	.92*	11.3*	15.5*	.003	.824	
Draper & Stoneman	.78	3.77*	17.17	5.77	.933	.900
(n=10, m=3)	.66	1.85	2.12	.005	.941	
Aitchinson & Dunsmore	.42	.91	2.08	.81	.9622	.9554
(n=16, m=2)	.49*	1.75	2.67	.38	.9586	

Notes: (1) The numbers are the extreme values of the statistics for at least one of the statistics corresponding to these observations.

(2) Asterisk denotes that the observations of exceeding the calibration point at 5% level of significance.

FIGURE 3.1
SENSITIVITY OF DIFFERENT STATISTICS
BASED ON UNIT CALIBRATION POINTS
MICKEY, DUNN AND CLARK DATA

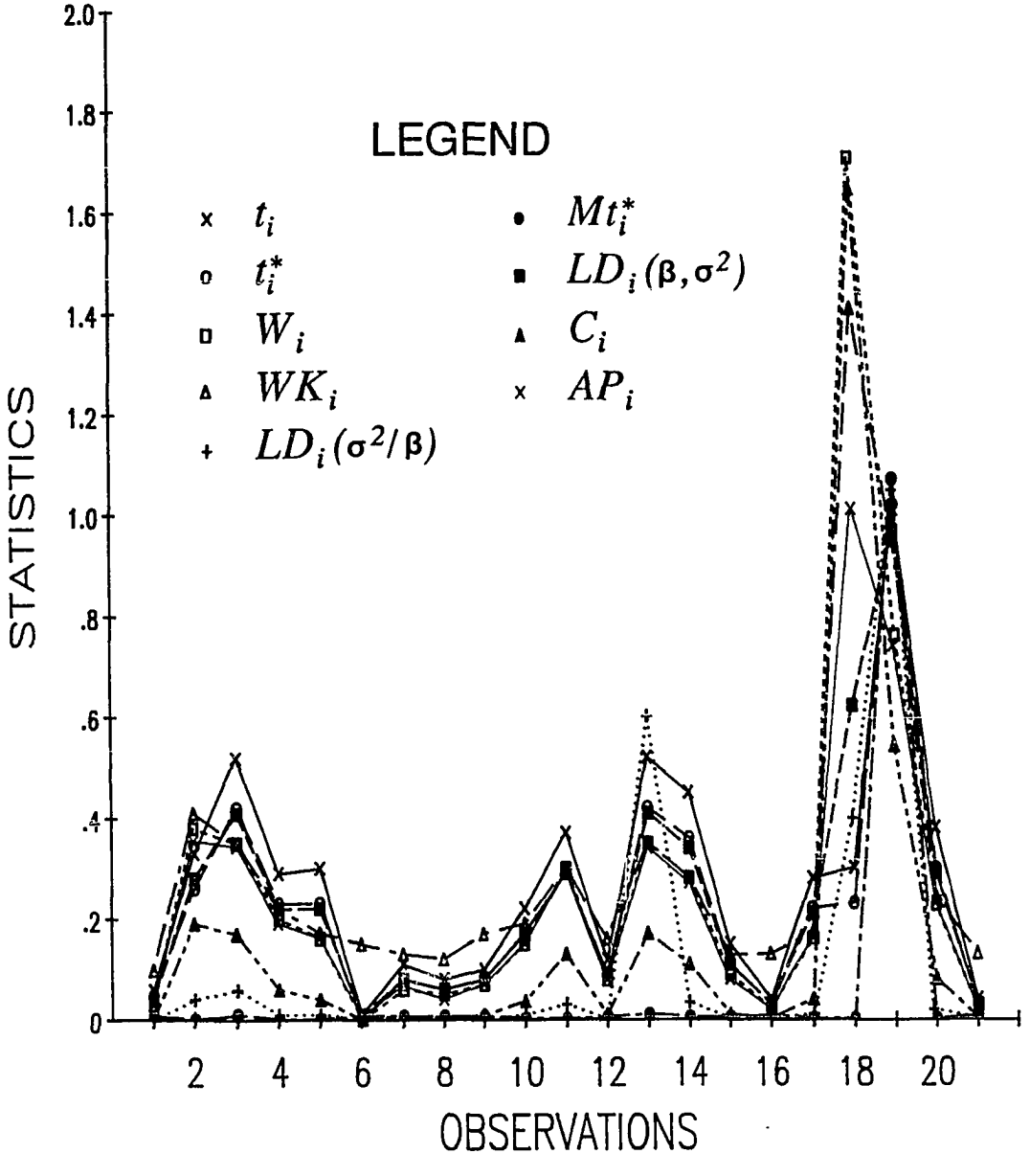


FIGURE 3.2
SENSITIVITY OF DIFFERENT STATISTICS
BASED ON UNIT CALIBRATION POINTS
WEISBERG DATA

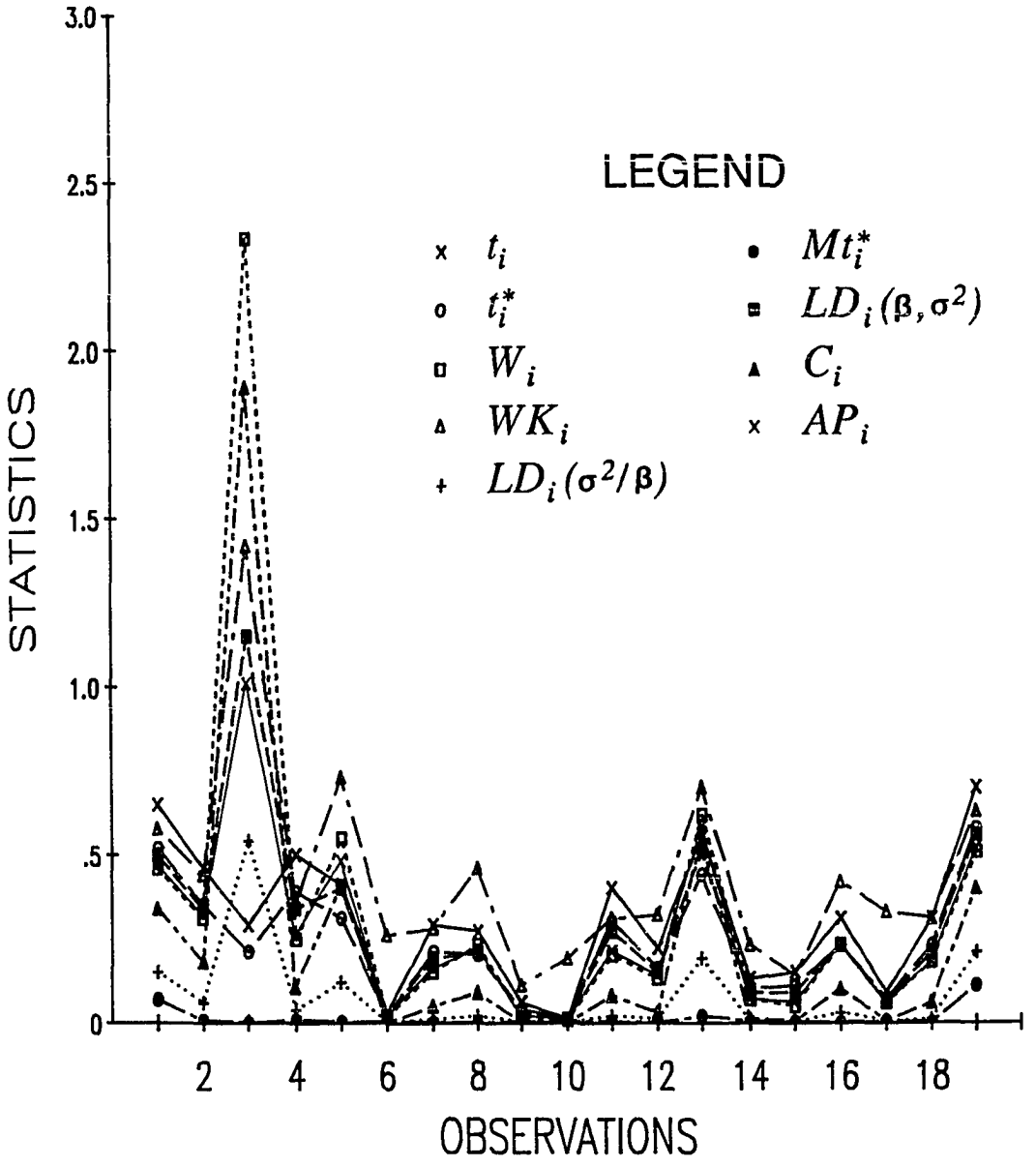


FIGURE 3.3
SENSITIVITY OF DIFFERENT STATISTICS
BASED ON UNIT CALIBRATION POINTS
MOORE DATA

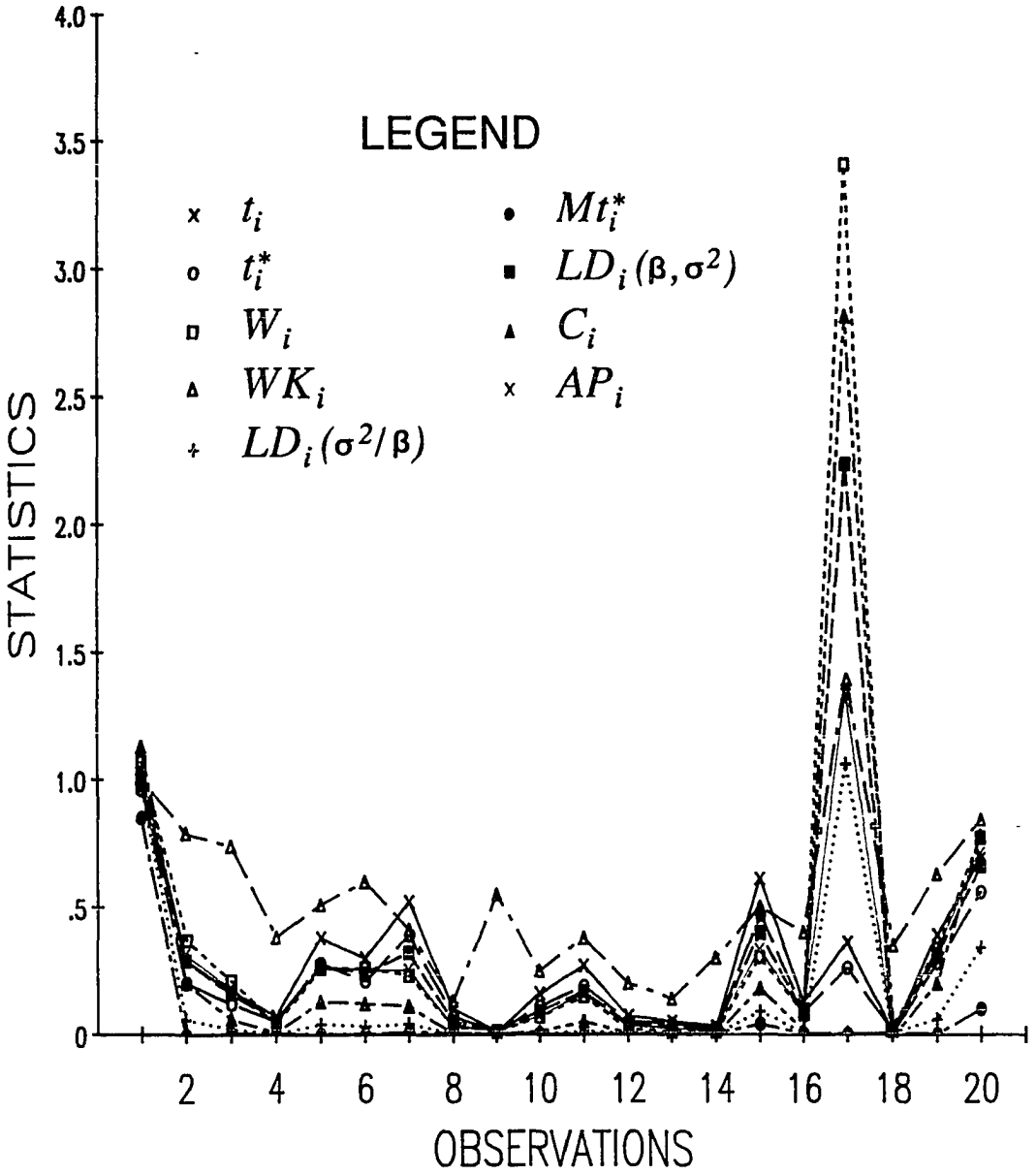
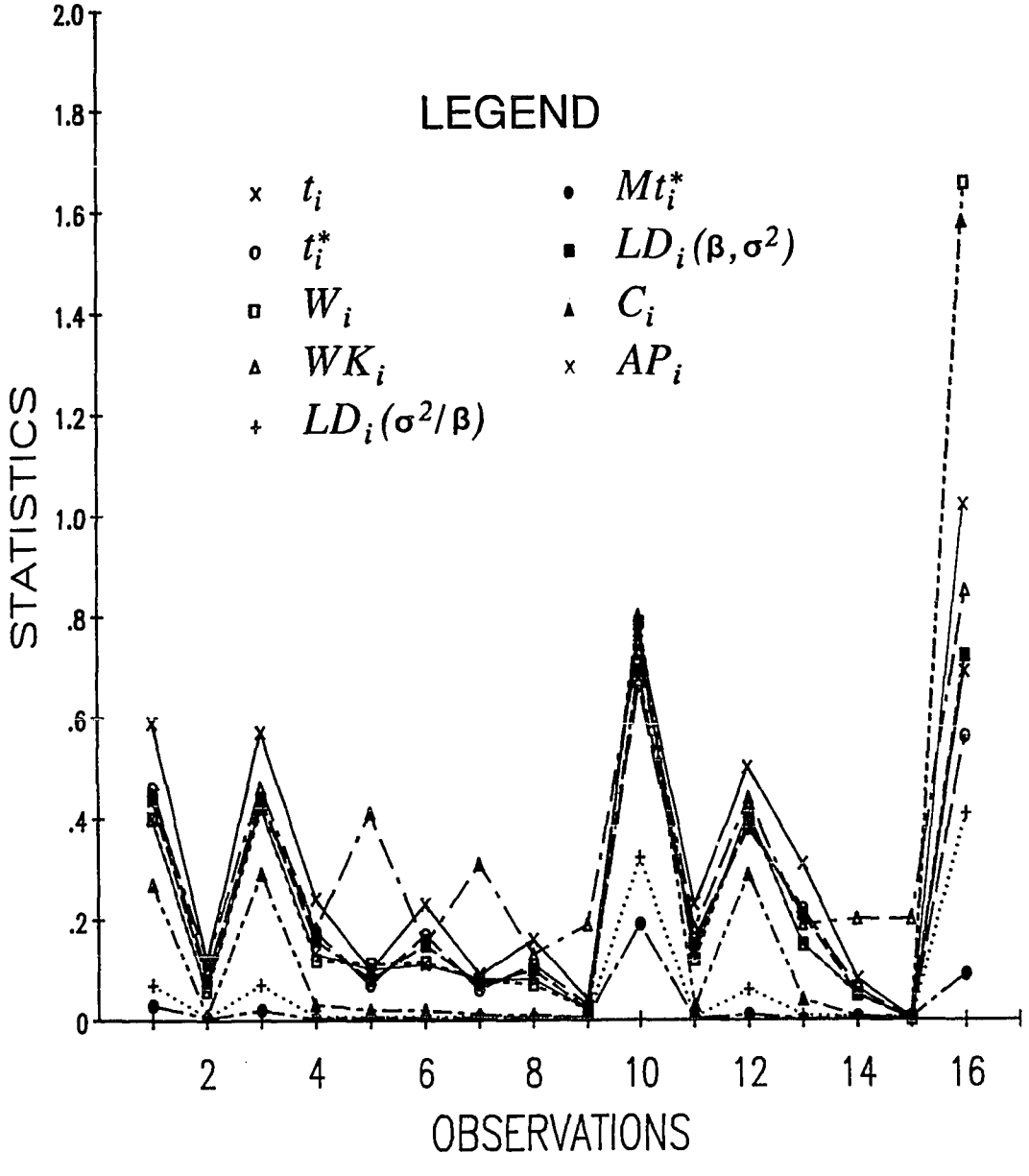


FIGURE 3.4
SENSITIVITY OF DIFFERENT STATISTICS
BASED ON UNIT CALIBRATION POINTS
AITCHINSON AND DUNSMORE DATA



8 is identified as influential by C_i, W_i, and $LD_i(\beta|\sigma^2)$. The data set 5 does not seem to have any aberrant observations. One sees that most of these statistics with the present calibration points are able to detect influential observations.

However, it is observed through a few plots that some of these statistics are more sensitive than the others. The four plots corresponding to the data sets 1, 3, 7, and 9 are examined by scaled values of different statistics drawn against each observation. Scaling is done so that the calibration point is unity. By examining the plots I - IV it is noted that W_i is more sensitive to influential observations than the rest of the statistics. The simulation study of section 3.1 also exhibits a similar pattern. Hence, it can be claimed that the procedure based on W_i is the best for detecting the influential observations. Further examination of these plots and simulation results yields that C_i, the Cook's distance with the new calibration point performs almost as good as W_i. Hence one may prefer to use C_i because of its computational simplicity.

3.3 Influential Observations in Analysis of Variance

This section is concerned with the methods for identifying the influential observations in analysis of variance models (ANOVA). Very little attention has been given for ANOVA models for identifying the influential observations. Gentleman and Wilk (1975, 1980) and Pendleton (1985) have considered the

problem of detection of outliers and influential observations respectively from the two-way classification ANOVA models. Pendleton (1985) has mentioned that diagnostic statistics, which depend upon the "delete one observation" principle such as Cook's distance, may no longer apply. To examine these problems let us consider a balanced ANOVA model. The diagonals (p_i) of the prediction matrix are all equal to the constant $\frac{1}{m}$. Since p_i's are constants, in the balanced situation, most of the regression diagnostic statistics of Table 3.1 are reduced to simple functions of the residuals. Therefore, the statistics t_i and t_i^* contain all the information regarding an observation's influence on the parameter estimates. However, in unbalanced ANOVA, the diagonals (p_i) of the prediction matrix are not the same and it becomes essential to examine the diagnostic statistics of Table 3.1 to identify the influential observations. In the model

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij},\ j = 1, 2, ..., n_i;\ i = 1, 2, ...k, \tag{3.3}$$

where

n_i is the number of observations in the ith treatment, Y_{ij} is the response or dependent variable, μ, the overall mean, α_i, the ith treatment effect and ϵ_{ij}, the random error. The following examples will illustrate the use of these diagnostic statistics for identifying the influential observations. This set of data are collected by Federer (1955) and is taken from Searle (1971, P. 165).

Federer (1955) reported an analysis of ruber producing plants called guayule, for which the piant weights were available for 54 plants of three different kinds; 27 of the normal, 15 off-types, and 12 aberrants. Only 6 plants are used for the purpose of illustration: $n_1 = 3$, normals; $n_2 = 2$, off-types; and $n_3 = 1$, aberrant. The diagonals of prediction matrix is $p_i = 1/n_i$, and the p_i for the last observation is 1.0. In this example, the residual of the last observation is zero, whereas C_i is infinity. Since the p_i of a single response is 1.0, most of the diagnostic statistics of Table 3.1 can no longer be applied. However, the statistics $R^2_{(i)}$, and likelihood displacements can be used for identifying the influential observations and the statistics are given in Table 3.14 below.

Table 3.14 Diagnostic Statistics for Federer's Data

observation	t_i^*	$R^2_{(i)}$	$LD_i(\beta,\sigma^2)$	$LD_i(\beta\|\sigma^2)$	$LD_i(\sigma^2\|\beta)$
1	0.2093	0.9786	0.1282	0.1272	0.0734
2	0.1.5191	0.9892	4.1442	2.5739	1.2596
3	−2.5981	0.9953	16.5510	3.4300	8.01134
4	−0.5080	0.9926	0.6561	1.2351	0.0109
5	0.5080	0.9835	0.6560	1.2351	0.0109
6	0.0000	0.8033	4681.24	45.2132	0.0939

Inspection of the Table 3.14 shows that the observation 6 is most influential. Also the values of R^2 and $R^2_{(6)}$ are 0.9803 and 0.8033 respectively. The goodness of fit of the regression hardly changes when the other observations are omitted, but changes substantially when the observation 6 is omitted. The observation 6 is the only point that is individually influential.

Another set of data was analyzed for balanced ANOVA. (See Pendleton (1985)). The model for this data is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}, \tag{3.4}$$

where $i = 1,2,\ldots,4$ treatments, $j = 1,2$ directions, $k = 1,2,\ldots,16$ subjects. The experiments consists of 8 treatment-direction combinations. Sixteen subjects were used for each combination. Using SAS and manipulating

with PROC MATRIX of SAS, different statistics are obtained which are given for a subset of the data in Table 3.15. From Table 3.15 is shown that the values of t_i and t_i^* corresponding to the observations 43 and 47 exceed their calibration points. Other statistics such as WK_i, $LD_i(\beta|\sigma^2)$, $LD_i(\beta,\sigma^2)$, $LD_i(\sigma^2|\beta)$ and $R^2_{(i)}$ also exhibit that the observations 43 and 47 are influential.

Sometimes ANOVA models may include restrictions on the elements of the parameter vector. Such kind of restrictions are quite different from the usual constraints frequently put for obtaining a solution to the normal equations. These restrictions are considered as part of the model, and these models are called restricted models. In general, the restrictions of the form $H'\beta = \delta$ are considered as a part of the model, where H' has full row rank. The restricted model is then $Y = X\beta + \epsilon$ with restriction $H'\beta = \delta$. Most of the computer programs do not provide all the diagnostic statistics listed in Table 3.1. Regression procedures of SAS using the RESTRICT statement can provide a few of these statistics. However, one can easily compute other statistics. For example, likelihood displacements, and $R^2_{(i)}$ by manipulating with the PROC MATRIX of SAS can be computed for identifying the influential observations from the restricted ANOVA models.

Table 3.15: Influence Measures for a Subset of the Pendleton's Data

Obs.	t_i	t_i^*	WK_i	$LD_i(\sigma^2\|\beta)$	$LD_i(\beta,\sigma^2)$	$LD_i(\beta\|\sigma^2)$
24	-0.0950	-0.0946	-0.0244	0.0038	0.0044	0.0006
25	-0.8673	-0.8664	-0.2237	0.0001	0.0535	0.0534
26	0.11880	0.1183	0.0305	0.0038	0.0048	0.0010
27	-0.9267	-0.9261	-0.2391	0.0001	0.0610	0.0610
28	-0.0950	-0.0946	-0.0244	0.0038	0.0044	0.0006
29	-1.0574	-1.0579	-0.2731	0.0002	0.0797	0.0794
30	-0.6415	-0.6399	-0.1652	0.0012	0.0303	0.0292
31	-0.9267	-0.9261	-0.2391	0.0001	0.0610	0.0610
32	-0.2851	-0.2840	-0.0733	0.0032	0.0090	0.0057
33	0.4633	0.4618	0.1192	0.0023	0.0175	0.0152
34	0.1188	0.1180	0.0305	0.0038	0.0048	0.0010
35	-0.926	-0.9261	-0.2391	0.0001	0.0610	0.0610
36	1.0455	1.0452	0.2700	0.0001	0.0779	0.0777
37	-1.2474	-1.2503	-0.3228	0.0017	0.1129	0.1106
38	-1.0217	-1.0210	-0.2638	0.0001	0.0743	0.0742
39	-1.1168	-1.1179	-0.2886	0.0004	0.0893	0.0886
40	-0.0950	-0.0946	-0.0244	0.0038	0.0044	0.0006
41	0.2732	0.2722	0.0702	0.0033	0.0086	0.0053
42	0.8791	0.8783	0.2267	0.0001	0.0550	0.0549
43	5.5365	6.3895	1.6497	5.8193	8.7240	2.1614
44	0.2851	1.2840	0.0733	0.0032	0.0090	0.0057
45	1.0334	1.0339	0.2669	0.0002	0.0761	0.0759
46	0.6890	0.6875	0.1775	0.0009	0.0345	0.0337
47	3.8256	4.0656	1.0497	0.9962	2.1723	1.0365
48	0.0950	0.0946	0.0244	0.0038	0.0044	0.0006
49	0.2732	0.2722	0.0702	0.0033	0.0086	0.0053
50	0.1188	0.1183	0.0305	0.0038	0.0048	0.0010
51	-0.5465	-0.5449	-0.1406	0.0018	0.0229	0.0212
52	-0.4752	-0.4736	-0.1223	0.0022	0.0182	0.0160
53	0.4633	0.4618	0.1192	0.0023	0.0175	0.0152

In conclusions: (i) A set of statistics C_i, W_i, and $LD_i(\beta|\sigma^2)$ which performs better are obtained for identifying the influential observations. (ii) The conclusion based on simulation study, real data analysis and the plots of different statistics based on unit calibration points, the conclusion is that the statistic W_i is the best among all the statistics. (iii) However, C_i, the Cook's distance with the modified calibration point performs almost as good as W_i. Hence one may prefer to use C_i because of its computational simplicity. (iv) This study also suggests that a set of statistics t_i, t_i^*, and $LD_i(\sigma^2|\beta)$ can be used for detection of outliers. Further, it is noted that the statistic t_i^* with the modified calibration point performs better than the other statistics. (v) Finally, the statistics WK_i, $LD_i(\beta, \sigma^2)$, and $LD_i(\sigma^2|\beta)$ may be used for detection of influential observations for the ANOVA models.

4. MULTIVARIATE REGRESSION MODEL

In recent years, there has been given much attention to the detection of outliers and influential observations within the framework of the usual linear regression model. Various measures have been proposed which emphasize different aspects of influence upon the regression. (For example, see Chapter 2). In this chapter, generalization is made of some of the univariate measures of influence to the multivariate regression model. In this study several data sets are considered to illustrate the methods.

4.1 The Model and Notations

Consider the multivariate regression model

$$Y_{nxk} = X_{nxm}B_{mxk} + E_{nxk},\ rank(X) = m. \tag{4.1}$$

Assume rows of E to be independent, normally distributed each with kx1 mean vector zero, and kxk covariance matrix Σ, that is, $vec(E) \sim N(0, \Sigma \otimes I_n)$. We write (4.1) in the form

$$(Y_1 : \ldots : Y_k) = (X\beta_1 : \ldots : X\beta_k) + (E_1 : \ldots : E_k). \tag{4.2}$$

The BLUE of β_i is $\hat{\beta}_i = (X'X)^{-1}X'Y_i$, $i = 1, 2, \ldots, k$. The residual vectors, $\hat{E}_i = Y_i - X\hat{\beta}_i, i = 1, 2, \ldots, k$, are correlated, that is, if $\hat{E} = (\hat{E}_1 : \ldots : \hat{E}_k)$ and $P = X(X'X)^{-1}X'$, then $vec(\hat{E}) \sim N(0, \Sigma \otimes (I - P))$. Note that, if $\hat{B} =$

$(\hat{\beta}_1 : \ldots : \hat{\beta}_k)$ then $\hat{E} = Y - X\hat{B}$ and $\hat{B} = (X'X)^{-1}X'Y$. An estimator of Σ is $\hat{\Sigma} = S/n-m$, where $S = \hat{E}'\hat{E}$. Let $\hat{\Sigma}_{(i)} = \hat{E}'_{(i)}\hat{E}_{(i)}/(n-m-1)$ where $\hat{E}_{(i)} = Y_{(i)} - X_{(i)}\hat{\beta}_{(i)}$, $\hat{B}_{(i)} = (\hat{\beta}_{1(i)} : \ldots : \hat{\beta}_{k(i)})$ and $\hat{\beta}_{r(i)} = (X'_{(i)}X_{(i)})^{-1}X'_{(i)}Y_{r(i)}$, $r = 1, 2, \ldots, k$, $i = 1, 2, \ldots, n$, is the BLUE of β_r calculated by deleting the *ith* observation.

4.2 Measures based on Residual

Recently, Naik (1986) has proposed methods for detecting outliers, from the models of the form (4.1), on using the multivariate kurtosis of transformed residuals. Here, the methods are proposed for detection of influential observations, which are extensions of the methods of chapter 2 for univariate models. Define

$$\tau_i^2 = \frac{1}{1-p_i} e_i'\hat{\Sigma}^{-1}e_i,\ i = 1, 2, \ldots, n, \tag{4.3}$$

and

$$T_i^2 = \frac{1}{1-p_i} e_i'\hat{\Sigma}_{(i)}^{-1}e_i,\ i = 1, 2, \ldots, n, \tag{4.4}$$

where $e'_1, e'_2, \ldots, e'_n$ are the rows of $\hat{E}$, each of dimension 1xk. It can be shown that T_i^2 has Hotelling's T^2 - distribution and $\frac{\tau_i^2}{(n-m-k)}$ has a beta distribution with parmeters $\frac{k}{2}$ and $\frac{(n-m-k)}{2}$. (See Mardia, Kent and Bibby (1979)).

A modified form of T_i^2 which is similar to Mt_i^* (cf: (2.42)) can be defined

as

$$MT_i^* = \frac{T_i^2}{(1-p_i)}, i = 1, 2, \ldots n. \tag{4.5}$$

This measure has some appealing performance for detecting the influenctial observations and outliers.

4.3 Measures Based on Influence Curve

Influence of (y'_{1xk}, x'_{1xm}) on least square estimate of B in the model (4.1) is $IF = (X'X)^{-1}x(y' - x'B)$ (cf: Cook and Weisberg (1982), P.107). It is evident that IF is an mxk matrix. One can measure the influence of (y', x') on B by defining

$$D(M, C) = tr(IF'MIFC^{-1}) \tag{4.6}$$

for any appropriate choice of mxm symmetric matrix M and kxk nonsingular matrix C. The sample influence curve of the ith observation (y'_i, x'_i) can be defined as

$$SIF_i = (n-1)(X'X)^{-1}x_i(y'_i - x'_i\hat{B}_{(i)})$$

$$= (n-1)(X'X)^{-1}x_i e'_i/1 - p_i. \tag{4.7}$$

If $D_i(M, C)$ is $D(M, C)$ as defined in (4.5) when IF is replaced by SIF_i then for $M = X'X$ and $C = (n-1)^2 m\hat{\Sigma}$ one obtain

$$D_i(X'X, (n-1)^2 m\hat{\Sigma}) = C_i, \; i = 1, 2, \ldots, n,$$

where

$$C_i = p_i e_i' \hat{\Sigma}^{-1} e_i / m(1-p_i)^2$$

$$= p_i r_i^2 / m(1-p_i),\ i = 1, 2, \ldots, n. \tag{4.8}$$

It can be observed that C_i measures the influence of the ith observation on $\hat{B}$ and is similar to the Cook's distance for univariate regression. A calibration point for C_i can be obtained using beta distribution.

Again if IF is replaced by SIF_i and $M = X'X$, $C = (n-1)^2 \hat{\Sigma}_{(i)}$ one can get

$$D_i(X'X, (n-1)^2 \hat{\Sigma}_{(i)}) = WK_i, i = 1, 2, \ldots, n,$$

where

$$WK_i = \frac{p_i}{(1-p_i)^2} e_i' \hat{\Sigma}_{(i)} e_i = \frac{p_i}{(1-p_i)} T_i^2,\ i = 1, 2, \ldots, n. \tag{4.9}$$

This measures the influence of the ith observation $(y_i' : x_i')$ on the ith predicted value $\hat{y}_i'$, where $\hat{y}_i'$ is the ith row of $\hat{Y} = X\hat{B}$, is indicated by large values of WK_i. A calibration point for this statistic can be taken to be $\frac{mk(n-m-1)}{(n-m-k-1)} F(\alpha; k, n-m-k-1)$.

Similarly if $M = X'X$ and $C = \frac{(n-1)^2 m}{(n-m)} \hat{\Sigma}_{(i)}$ it gives

$$D_i(X'X, \frac{(n-1)^2 m}{(n-m)} \hat{\Sigma}_{(i)}) = C_i^*,\ i = 1, 2, \ldots, n,$$

where

$$C_i^* = \frac{p_i(n-m)}{m(1-p_i)^2} e_i' \hat{\Sigma}_{(i)}^{-1} e_i$$

$$C_i^* = \frac{p_i(n-m)}{m(1-p_i)} T_i^2 = \frac{n-m}{m} WK_i,\ i = 1, 2, \ldots, n. \qquad (4.10)$$

This statistic is similar to the modified Cook's distance and the calibration point for (4.10) can be obtained by multiplying calibration points for WK_i by $\frac{n-m}{m}$.

Finally if $M = (X_{(i)}' X_{(i)}$ and $C = (n-1)\hat{\Sigma}_{(i)}$ then

$$D_i(X_{(i)}' X_{(i)}, (n-1)\hat{\Sigma}_{(i)}) = W_i, i = 1, 2, \ldots, n,$$

where

$$W_i = \frac{(n-1)}{(1-pi)^4} e_i' \hat{\Sigma}_{(i)}^{-1} e_i x_i' (X'X)^{-1} (X'X - x_i x_i')(X'X)^{-1} x_i,$$

A little simplication gives,

$$W_i = \frac{(n-1)p_i}{(1-p_i)^2} T_i^2 = \frac{n-1}{1-p_i} WK_i,\ i = 1, 2, \ldots, n. \qquad (4.11)$$

This statistic measures the influence on both $\hat{B}$ and $\hat{\Sigma}$ and is similar to the Welsch distance (1982). The equation (4.11) suggests that the calibration points for W_i can be obtained by multiplying the calibration points for WK_i by $\frac{n(n-1)}{n-m}$.

4.4 Measures Based on Volume of Confidence Ellipsoids

Andrews and Pregibon (1978) proposed measure (2.19), generalization is accomplished by replacing Y_{nx1} with the multivariate observation matrix Y_{nxk}. Consider the measure, given in (2.19) and define $X^* = [X : Y]$. Now

$$X^{*'}X^* = \begin{pmatrix} X'X & X'Y \\ Y'X & Y'Y \end{pmatrix} \tag{4.12}$$

and

$$det(X^{*'}X^*) = det(X'X)det(Y'Y - Y'X(X'X)^{-1}X'Y),$$

that is

$$det(X^{*'}X^*) = det(X'X)det(\hat{E}'\hat{E}). \tag{4.13}$$

Similarly

$$det(X^{*'}_{(i)}X^*_{(i)}) = det(X'_{(i)}X_{(i)})det(\hat{E}'_{(i)}\hat{E}_{(i)}). \tag{4.14}$$

Substituting (4.13) and (4.14) into (2.19) gives

$$AP_i = \frac{det(X'_{(i)}X_{(i)})det(\hat{E}'_{(i)}\hat{E}_{(i)})}{det(X'X)det(\hat{E}'\hat{E})}, \; i = 1, 2...., n. \tag{4.15}$$

Note that

$$det(X'_{(i)}X_{(i)}) = det(X'X)(1 - p_i). \tag{4.16}$$

Therefore, applying (4.16) to the (4.15), one obtains

$$AP_i = (1 - p_i)\frac{det(\hat{E}'_{(i)}\hat{E}_{(i)})}{det(\hat{E}'\hat{E})}, \; i = 1, 2, \ldots, n. \tag{4.17}$$

Covariance Ratio type statistic: Influence of the *ith* observation on the Covariance of $\hat{B}$ can be measured by

$$\frac{det(Cov(vec(\hat{B}_{(i)})))}{det(Cov(vec(\hat{B})))} = \frac{det[\hat{\Sigma}_{(i)} \otimes (X'_{(i)}X_{(i)})^{-1}]}{det[\hat{\Sigma} \otimes (X'X)^{-1}]}.$$

A simplification of which gives

$$\frac{[det((X'_{(i)}X_{(i)})^{-1}]^k}{[det(X'X)^{-1}]^k} \frac{[det\hat{\Sigma}_{(i)}]^m}{[det\hat{\Sigma}]^m}$$

$$= \left[\frac{1}{1-p_i}\right]^k \left[\frac{det\hat{\Sigma}_{(i)}}{det\hat{\Sigma}}\right]^m, i = 1, 2, \ldots, n. \tag{4.18}$$

The low and high value of (4.18) are considered significant. A lower calibration point can be obtained using the fact that (cf: Rao (1973))

$$\frac{det(\hat{\Sigma}_{(i)})}{det(\hat{\Sigma})} = (1 + \frac{T_i^2}{n-m-1})^{-1}. \tag{4.19}$$

Cook-Weisberg type statistic: The multivariate version of the statistic defined in (2.25) is

$$CW_i = log(\left(\frac{det(X'_{(i)}X_{(i)})}{det(X'X)}\right)^{1/2} \left(\frac{det(\hat{\Sigma})}{det(\hat{\Sigma}_{(i)})}\right) \times$$

$$\left(\frac{F(\alpha; m, n-m)}{F(\alpha; m, n-m-1)}\right)^{m/2}), \tag{4.20}$$

where $F(\alpha; .,.)$ is the upper α percentile point of F distribution with appropriate degrees of freedom.

$F_{(i)}$ statistic: Here an influence measure based on the Wilks Λ statistic is introduced. Let

$$\lambda_{(i)} = \frac{det(\hat{E}'_{(i)}\hat{E}_{(i)})}{det(\hat{E}'_{(i)}\hat{E}_{(i)} + Y'_{(i)}P_{(i)}Y_{(i)})}, i = 1, 2, \ldots, n.$$

Influence of the *ith* observation on the test statistic F (which is used to test the hypothesis that the regression parameters are zero) is obtained by computing

$$F_{(i)} = \frac{(1 - \lambda_{(i)})/(m-1)}{\lambda_{(i)}/(n-m-1)}, i = 1, 2, \ldots, n. \tag{4.21}$$

4.5 Influence on rows of B

Influence of the *ith* observation on the *jth* row of B can be measured by the statistic

$$\frac{T_i^2}{1-p_i}\frac{w_{ij}^2}{W_j'W_j}, i = 1, 2, ..., n; j = 1, 2, ..., m, \tag{4.22}$$

which is similar to (2.41) .

A calibration point can be suggested using the approximate value of 1/n for $\frac{w_{ij}^2}{W_j'W_j}$. A diagnostic strategy for the multivariate case which is similar to the univariate case (cf: Hoaglin and Kempthorne (1986)) begins with the following: (i) Plot the data to look at the scatter plots of Y_i against X_j for $i = 1, 2, \ldots, k$, $j = 1, 2, \ldots, m$. (ii) Inspect the diagonal elements p_i of prediction matrix for high leverage. As suggested by Huber (1981), consider

the observations with p_i larger than 0.5 as influential. (iii) Examine the partial leverage: the term $\delta_{ij}^2 = w_{ij}^2/W_j'W_j$ which appears in (4.22) represents the contribution of the *jth* variable to the leverage of the *ith* observation. (iv) Study the influence of the individual observations through (4.8) - (4.11). (v) Study the influence of individual observations on covariance matrix of $\hat{B}$ through $COVR_i$ (4.18). (vi) Study the influence of the individual observations on the estimated coefficients through D_{ij}^* (4.22).

4.6 Measures based on likelihood function

For the model (4.1) the likelihood displacement for B given Σ (cf: equation 2.29) is given by $LD_i(B|\Sigma) = 2[l(\hat{B},\hat{\Sigma}) - l(\hat{B}_{(i)}, \Sigma(\hat{B}_{(i)}))]$, where $\Sigma(\hat{B}_{(i)})$ is the estimate of Σ when B is estimated by $\hat{B}_{(i)}$. Simplifications yield

$$LD_i(B|\Sigma) = n\,log\left(1 + \frac{1}{(n-m)}\frac{p_i}{(1-p_i)^2}\,e_i'\hat{\Sigma}^{-1}e_i\right)$$

$$= n\,log(1 + \frac{m}{n-m}\,C_i),\ i = 1,2,\ldots,n, \qquad (4.23)$$

where C_i is the statistic defined in (4.8). Thus the likelihood displacement leads to C_i. Similarly the statistic WK_i can be obtained by taking the difference between two likelihood displacements. The joint likelihood displacement for (B,Σ) is

$$LD_i(B,\Sigma) = 2[l(\hat{B},\hat{\Sigma}) - l(\hat{B}_{(i)},\hat{\Sigma}_{(i)})]$$

$$= n\,log\left[\frac{det(\hat{\Sigma}_{(i)})}{det(\hat{\Sigma})}\right] + \frac{T_i^2}{1-\bar{p}_i} - k,\ i = 1,2,\ldots,n. \qquad (4.24)$$

If Σ is the only parameter of interest, then the displacement is

$$LD_i(\Sigma|B) = 2[l(\hat{B},\hat{\Sigma}) - l(B(\hat{\Sigma}_{(i)}),\hat{\Sigma}_{(i)})],$$

where $B(\hat{\Sigma}_{(i)})$ is the estimate of B when Σ is estimated by $\hat{\Sigma}_{(i)}$. However, $\hat{B}(\hat{\Sigma}_{(i)}) = \hat{B}$. Thus,

$$LD_i(\Sigma|B) = n\,log\left[\frac{det(\hat{\Sigma}_{(i)})}{det(\hat{\Sigma})}\right] + n\,[tr(\hat{\Sigma}_{(i)}^{-1}\hat{\Sigma}) - k],\ i = 1,2,\ldots,n. \quad (4.25)$$

Here tr(A) means the trace of the matrix A. The difference between (4.24) and (4.25) gives

$$LD_i(B,\Sigma) - LD_i(\Sigma|B) = \frac{n-1}{n-m-1} WK_i,\ i = 1,2,\ldots,n, \quad (4.26)$$

where WK_i is the statistic defined in (4.9).

4.7 Examples

In the following two multivariate data sets are considered to illustrate the methods. All the calculations that follow are done using SAS programs and manipulating with the PROC MATRIX of SAS.

1. Anderson's data: The following data are taken from Anderson (1984, p.369). The dependent variables are weight of grain (Y_1), and weight of straw (Y_2). The independent variable is the amount of fertilizer X_1.

Weight of grain (Y_1) : 40 17 9 15 6 12 5 9

Weight of straw (Y_2) : 53 19 10 29 13 27 19 30

Amount of fertilizer (X_1) : 24 11 5 12 7 14 11 18

Model: $Y_{8x2} = X_{8x2}\, B_{2x2} + E_{8x2}$, rows of E are independently distributed as $N_2(0, \Sigma)$. The first column of X above is a vector of ones and the second column is X_1. For the dependent variables Y_1 and Y_2 the fit is significant with the F- values of 7.878 and 71.026 respectively. Regression estimates along with the standard errors, individual t-values, and the corresponding p-values are given below.

variable	$\hat{\beta}_1$	$\hat{\beta}_2$	$S.E.(\hat{\beta}_1)$	$S.E.(\hat{\beta}_2)$	t_1	$p-val$	t_2	$p-val$
constant	−3.752	−2.296	6.967	3.543	−.539	.609	−.648	.541
X_1	1.402	2.141	0.500	0.254	2.807	.031	8.428	.0002

Estimate of Σ is $\hat{\Sigma}$=S/ 6, where

$$S = \begin{bmatrix} 382.554 & 143.023 \\ 143.023 & 98.928 \end{bmatrix}.$$

Different statistics for testing $\Gamma = 0$, where Γ is the matrix B after omitting the first row which corresponds to the constant term, are given below. For the definitions and uses of these statistics one can refer to Anderson (1984), Rao (1973) or Timm (1975).

statistics	*values*	*F−value*	*p−value*
Wilks lambda :	0.059	40.01	.0008
Pillai's trace :	0.942	40.01	.0008
Hotelling-Lawley trace :	16.004	40.01	.0008
Roy's largest root :	16.004	40.01	.0008

All the criteria consistently reject the hypothesis. Next the several influence measures described in section 3-6 are examined. Different influence measures are given in Table 4.1.

Table 4.1: Influence Measures for Anderson's Data

Observation	p_i	C_i	WK_i	$COVR_i$	MT_i^*	$LD_i(B,\Sigma)$
1	0.620	0.923	7.825	3.720	31.198	34.172
2	0.137	0.406	0.818	0.376	6.950	3.563
3	0.360	0.548	0.450	2.009	0.540	0.730
4	0.127	0.402	0.958	0.284	9.051	5.231
5	0.254	0.470	0.006	1.788	0.023	0.145
6	0.131	0.404	0.048	1.189	0.363	0.163
7	0.137	0.406	0.123	1.045	0.914	0.240
8	0.233	0.457	1.700	0.402	9.842	6.571

Inspection of the Table 4.1 shows that based on all the different methods, the first observation is influential. In the following it can be seen how various statistics change by dropping the first observation. The Regression estimates along with the standard errors, individual t-values, and the corresponding p-values are given below.

variable	$\hat{\beta}_1$	$\hat{\beta}_2$	$S.E.(\hat{\beta}_1)$	$S.E.(\hat{\beta}_2)$	t_1	$p-val$	t_2	$p-val$
constant	7.858	2.204	5.361	3.845	1.466	.203	0.573	.592
X_1	0.231	1.686	0.453	0.325	0.509	.632	5.190	.0035

The deletion of observation one increases the significance of constant

but decreases the significance of the regression coefficients. Further, it may be noted that there is a drastic change in the estimate of Σ. The estimate now is $\hat{\Sigma}$= S/5, where

$$S = \begin{bmatrix} 113.816 & 38.865 \\ 38.865 & 58.558 \end{bmatrix}.$$

Various statistics for testing $\Gamma = 0$, are given below.

statistics	*values*	*F−value*	*p−value*
Wilks lambda :	0.1355	12.764	.0184
Pillai's trace :	0.8645	12.764	.0184
Hotelling-Lawley trace :	6.3819	12.764	.0184
Roy's largest root :	6.3819	12.764	.0184

The results show that at 5 percent level of significance, the null hypothesis is rejected based on all the criteria. It may be noted that the p - values of the tests increased considerably.

2. Rohwer's data: The data given in Table 4.2 were collected by Dr. W. D. Rohwer of University of California at Berkley and reproduced here from Timm (1975). Thirty two students from an upper-class, white, residential school, were selected at random to determine how well data from a set of paired -associated (PA), learning-proficiency tests may be used to predict childrens performances on the Peabody picture vocabulary test (Y_1). The

Table 4.2: Rohwer's Data

y_1	y_2	y_3	x_1	x_2	x_3	x_4	x_5
68	15	24	0	10	8	21	22
82	11	8	7	3	21	28	21
82	13	88	7	9	17	31	30
91	18	82	6	11	16	27	25
82	13	90	20	7	21	28	16
100	15	77	4	11	18	32	29
100	13	58	6	7	17	26	23
96	12	14	5	2	11	22	23
63	10	1	3	5	14	24	20
91	18	98	16	12	16	27	30
87	10	8	5	3	17	25	24
105	21	88	2	11	10	26	22
87	14	4	1	4	14	25	19
76	16	14	11	5	18	27	22
66	14	38	0	0	3	16	11
74	15	4	5	8	11	12	15
68	13	64	1	6	10	28	23
98	16	88	1	9	12	30	18
63	15	14	0	13	13	19	16
94	16	99	4	6	14	27	19
82	18	50	4	5	16	21	24
89	15	36	1	6	15	23	28
80	19	88	5	8	14	25	24
61	11	14	4	5	11	16	22
102	20	24	5	7	17	26	15
71	12	24	0	4	8	16	14
102	16	24	4	17	21	27	31
96	13	50	5	8	20	28	26
55	16	8	4	7	19	20	13
96	18	98	4	7	10	23	19
74	15	98	2	6	14	25	17
78	19	50	5	10	18	27	26

other dependent variables are student achievement test (Y_2), and the ravin progressive matrices test (Y_3). The independent variables are the sum of the number of items correct out of 20 (on two exposures) to five types of PA tasks. The basic tasks are named (X_1), still (X_2), named still (X_3), named action (X_4), and sentence still (X_5). The model: $Y_{32x6}B_{6x3} + E_{32x3}$, rows of $E \sim N_3(0, \Sigma)$. The regression summary of Rohwer's data are presented in Table 4.3a and 4.3b.

Table 4.3a: Regression Summary

variable	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$S.E.(\hat{\beta}_1)$	$S.E.(\hat{\beta}_2)$	$S.E.(\hat{\beta}_3)$
constant	39.697	13.243	−28.467	12.268	2.614	25.719
x_1	0.067	0.059	3.257	0.618	0.131	1.295
x_2	0.369	0.492	2.996	0.715	0.152	1.500
x_3	−0.374	−0.164	−5.859	0.736	0.157	1.544
x_4	1.523	0.118	5.666	0.638	0.136	1.338
x_5	0.7410	−0.121	−0.622	0.544	0.115	1.140

Table 4.3b: Regression Summary (continued)

variable	t_1	$p-val$	t_2	$p-val$	t_3	$p-val$
constant	3.236	0.0033	5.066	0.0001	1.107	0.2785
x_1	0.109	0.9142	0.451	0.6560	2.514	0.0185
x_2	0.517	0.6095	3.230	0.0034	1.998	0.0563
x_3	−.508	0.6157	−1.044	0.3059	−3.792	0.0008
x_4	2.385	0.0246	0.875	0.3898	4.234	0.0003
x_5	0.754	0.4577	−1.045	0.3056	−0.546	0.5898

Estimates of β_1, β_2, and β_3 along with standard errors (S.E.) and values of t - statistic, are given in Tables 4.3a and 4.3b. For dependent variables Y_1 and Y_3 the fit is significant with F-values of 2.846 and 6.539; where as an F-value of 2.32 for Y_2 shows an insignificant fit. Observing the p-values we see that the variables X_2 is significant for fitting Y_2 on X. An estimate of Σ is $\hat{\Sigma} = S/26$, where

$$S = \begin{pmatrix} 3898.990 & 281.386 & 1279.600 \\ 281.386 & 177.031 & 623.773 \\ 1279.600 & 623.773 & 17134.100 \end{pmatrix}.$$

Different statistics for testing $\Gamma = 0$, are given in Table 4.4.

Table 4.4

Stastistics	*value*	*F−value*	*p−value*	*calibration point*
Wilks lambda:	0.243	2.97	0.0012	0.388
Pillai's trace:	1.039	2.75	0.0019	0.818
Hotelling-Lawley trace:	2.062	3.12	0.0007	1.229
Roy's largest root:	1.465	7.62	0.0002	0.477

The p-values are based on the F-distribution. All the criteria, consistently reject the hypothesis. Now an examination of the several influence measures described in section 3-6 is made. Different influence measures are given in Table 4.5 and D_{ij}^{*} are given in Table 4.6. Inspection of Table 4.5 shows that the observation 5 is most influential using all the methods. Further, WK_i and $COVR_i$ detect some other observations as influential also. But there is no other value which is detected as influential by all the criteria. Further, inspection of D_{ij}^{*} in Table 4.6 shows that observation 5 is influential jointly with the variable 1. If variable 1 is removed from the analysis, observation 5 is not influential.

Rohwer's second set of data for the low-socioeconomic status area were also analyzed and no influential observations were found. However, D_{ij}^{*} revealed some observations to be influential in one dimension.

Table 4.5 : Computed Influence Measures for Rohwer's Data

Obs.	$LD_i(\Sigma\|B)$	$LD_i(B,\Sigma)$	$LD_i(B\|\Sigma)$	T_i^2	WK_i	C_i	$Covr_i$
1	0.0272	1.1201	0.8069	3.6493	0.7316	0.0612	1.9343
2	0.0667	0.2955	0.2629	0.7605	0.2125	0.0652	1.2377
3	0.0031	0.7221	0.5426	2.8881	0.4769	0.0594	1.5045
4	0.0905	0.0815	0.0476	0.4804	0.0379	0.0550	0.6950
5	0.0701	7.4234	5.7112	4.3593	5.7367	0.1181	16.085
6	0.0916	0.1409	0.1075	0.4697	0.0857	0.0603	0.9126
7	0.0209	0.4020	0.1862	3.5061	0.1664	0.0534	1.2468
8	0.0993	1.6228	1.0723	4.7223	1.0129	0.0619	2.4971
9	0.1457	0.7994	0.2969	5.2071	0.2816	0.0537	1.7990
10	0.0932	0.4966	0.4647	0.4521	0.3723	0.0930	3.3330
11	0.0155	0.3970	0.3355	1.6508	0.2809	0.0596	1.1608
12	0.0323	1.1829	0.8474	3.7541	0.7716	0.0614	2.0022
13	0.0006	0.4188	0.3135	2.3244	0.2690	0.0569	1.1689
14	0.7447	2.9201	1.1906	8.8648	1.2837	0.0583	4.5757
15	0.1208	0.1510	0.1119	0.1772	0.0882	0.0764	1.7313
16	0.0236	0.9375	0.8620	1.4544	0.7223	0.0763	2.3232
17	0.1001	1.5976	1.0495	4.7321	0.9913	0.0616	2.4712
18	0.0527	0.4534	0.4160	0.9489	0.3395	0.0692	1.5453
19	0.0001	1.4729	1.2541	2.5937	1.1029	0.0726	2.5839
20	0.0019	0.4196	0.2745	2.7996	0.2395	0.0553	1.1938
21	0.3595	2.1688	1.1006	6.8286	1.1138	0.0593	3.3077
22	0.0489	0.3323	0.2958	1.0046	0.2415	0.0632	1.1932
23	0.0746	0.5746	0.2233	4.4197	0.2060	0.0533	1.5030
24	0.0127	0.6098	0.5341	1.7298	0.4499	0.0642	1.4755
25	1.5215	5.0784	1.8651	11.871	2.2129	0.0605	8.4837
26	0.0245	0.3635	0.3131	1.4352	0.2599	0.0602	1.1369
27	0.0396	3.0433	2.4079	3.8898	2.2578	0.0806	4.6405
28	0.0147	0.3107	0.2517	1.6718	0.2106	0.0574	1.0391
29	0.1017	2.9474	2.1599	4.7502	2.0774	0.0733	4.1627
30	0.0070	0.5113	0.3309	3.0810	0.2919	0.0558	1.3007
31	0.4661	1.8492	0.7126	7.4627	0.7310	0.0559	3.1320
32	0.1052	0.8469	0.4038	4.7892	0.3783	0.0550	1.7748

Table 4.6: Values of D_{ij}^{*} for Rohwer's Data

Obs.	Const.	x_1	x_2	x_3	x_4	x_5
1	0.0922	0.0023	0.1381	0.2132	0.0022	0.018
2	0.0583	0.0072	0.0665	0.0618	0.0022	0.001
3	0.3753	0.0081	0.0074	0.0217	0.0574	0.116
4	0.0035	0.0014	0.0122	0.0011	0.0013	0.000
5	0.0666	3.4972	0.0234	0.0013	0.0461	1.189
6	0.4413	0.0075	0.0023	0.0000	0.0173	0.003
7	0.0043	-0.0013	0.0143	0.0213	-0.0011	0.001
8	0.0511	0.0532	0.4181	0.0932	0.0194	0.274
9	0.0073	0.0323	0.0651	0.0037	0.0022	-0.001
10	0.0011	2.5746	0.0152	0.0417	0.0025	0.025
11	0.0023	0.0081	0.1593	0.0318	0.0061	0.047
12	0.0021	0.0031	0.2169	0.2941	0.1351	0.0161
13	0.0013	0.0834	0.0658	0.0164	0.0254	0.0092
14	-0.0011	0.3710	0.1843	-0.0021	0.0132	-0.0042
15	1.1295	0.0013	0.0061	0.0171	0.0000	0.0011
16	1.5783	0.0312	0.0513	0.0046	1.3934	0.0024
17	0.0411	0.0181	0.0276	0.3051	0.3686	0.0071
18	0.0114	0.0130	0.0243	0.0314	0.1725	0.0342
19	0.3082	0.1541	1.1323	0.0671	0.1373	0.3864
20	0.0024	0.0033	0.0046	0.0076	0.1123	0.0555
21	0.0423	0.0372	0.2233	0.2172	0.4165	0.3121
22	0.0024	0.0395	0.0375	0.0171	0.0357	0.1171
23	0.0026	0.0091	0.0032	0.0254	0.0026	0.0373
24	0.4415	0.0316	0.0583	-0.0013	0.7324	0.3721
25	0.0152	0.0421	0.0414	0.2142	0.2263	1.2555
26	0.6424	0.0110	0.0012	0.0041	0.0333	0.0063
27	0.1272	0.1131	0.5111	0.2015	0.1282	0.0722
28	0.0434	0.0344	0.0091	0.0854	0.0011	0.0112
29	0.1962	0.1181	0.0262	0.7856	0.1543	0.4524
30	0.0482	0.0331	0.0133	0.1513	0.0212	0.0145
31	0.0084	0.1211	0.0000	0.0167	0.1272	0.2636
32	0.0592	0.0235	0.0184	0.0533	0.0011	0.0258

4.8 Detection of Influential Observations in MANOVA

In this section, it will be shown that some of the influence measures developed for the multivariate regression model can be used for MANOVA model for detecting the influential observations. Several examples are presented to illustrate the diagnostic statistics.

Examples 4.9

Two examples will be presented to reflect the use of these diagnostics in identifying the influential observations.

Dental Data: A certain measurement in a dental study was made on each of 11 girls and 16 boys at ages 8, 10, 12 and 14. The data given in Table 4.7 is taken from Potthoff and Roy (1964). The model:

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij},\ j = 1, 2, \ldots, n_{ij},\ j = 1, 2, \tag{4.27}$$

where $\epsilon_{ij} \sim N_k(0, \Sigma)$, with $k = 4,\ m = 2,\ n_1 = 11,\ n_2 = 16$. An estimate of $\Sigma = S/25$, where

$$S = \begin{pmatrix} 135.386 & 67.920 & 97.756 & 67.756 \\ 67.920 & 104.619 & 73.179 & 82.929 \\ 97.756 & 73.179 & 161.393 & 103.268 \\ 67.756 & 82.929 & 103.268 & 124.643 \end{pmatrix}.$$

The different statistics to test

$$\mu_j = \mu + \alpha_j = 0 \quad for\ j = 1, 2,$$

are given below.

Table 4.7: Dental Data

Obs.	y_1	y_2	y_3	y_4
1	21.0	20.0	21.5	23.0
2	21.0	21.5	24.0	25.5
3	20.5	24.0	24.5	26.0
4	23.5	24.5	25.0	26.5
5	21.5	23.0	22.5	23.5
6	20.0	21.0	21.0	22.5
7	21.5	22.5	23.0	25.0
8	23.0	23.0	23.5	24.0
9	20.0	21.0	22.0	21.5
10	16.5	19.0	19.0	19.5
11	24.5	25.0	28.0	28.0
12	26.0	25.0	29.0	31.0
13	21.5	22.5	23.0	26.5
14	23.0	22.5	24.0	27.5
15	25.5	27.5	26.5	27.0
16	20.0	23.5	22.5	26.0
17	24.5	25.5	27.0	28.5
18	22.0	22.0	24.5	26.5
19	24.0	21.5	24.5	25.5
20	23.0	20.5	31.0	26.0
21	27.5	28.0	31.0	31.5
22	23.0	23.0	23.5	25.0
23	21.5	23.5	24.0	28.0
24	17.0	24.5	26.0	29.5
25	22.5	25.5	25.5	26.0
26	23.0	24.5	26.0	30.0
27	22.0	21.5	23.5	25.0

Statistics	*value*	*p−values(based on F)*
Wilks lambda:	0.602	0.02
Pillai's trace:	0.398	0.02
Hotelling-Lawley trace:	0.660	0.02
Roy's Largest root:	0.660	0.02

These values suggest rejection of the hypothesis. Influence measures using different diagnostic statistics are given in Table 4.8. Examination of Table 4.8 shows that the statistics $LD_i(\Sigma|B)$, $LD_i(B,\Sigma)$, WK_i, $COVR_i$ and T_i^2 corresponding to the observations 20 and 24 exceed their respective calibration points. Further, examination of Table 4.8 reveals that in the MANOVA situation, WK_i, $COVR_i$ and $LD_i(\Sigma|B)$ can be used for detection of influential observations.

Example 2: This set of data consists of 120 Medicare and non - Medicare benificiaries admitted during the years 1982 - 85, with the primary medical diagnosis of congestive heart failure. There are four dependent variables, namely: Y_1=total length of stay; Y_2= severity of illness; Y_3= readmission; and Y_4= referral to home health agencies. The model for a vector response consisting of k=4 components is

Table 4.8: Computed Influence Measures for Dental Data

Obs.	p_i	C_i	WK_i	$COVR_i$	$LD_i(\Sigma \mid B)$	$LD_i(B,\Sigma)$	$LD_i(B \mid \Sigma)$
1	0.091	0.063	0.247	1.277	0.032	0.395	0.263
2	0.091	0.063	0.198	1.229	0.061	0.308	0.214
3	0.091	0.063	0.237	1.268	0.037	0.373	0.251
4	0.091	0.063	0.176	1.207	0.077	0.278	0.192
5	0.091	0.063	0.083	1.117	0.174	0.154	0.093
6	0.091	0.063	0.075	1.111	0.183	0.146	0.085
7	0.091	0.063	0.051	1.089	0.214	0.118	0.059
8	0.091	0.063	0.105	1.140	0.145	0.178	0.119
9	0.091	0.063	0.211	1.243	0.052	0.332	0.226
10	0.091	0.063	0.770	1.851	0.268	1.854	0.668
11	0.091	0.063	0.492	1.535	0.025	0.960	0.472
12	0.063	0.059	0.324	1.351	0.023	0.761	0.312
13	0.063	0.059	0.114	1.063	0.083	0.209	0.124
14	0.063	0.059	0.199	1.175	0.011	0.383	0.206
15	0.063	0.059	0.870	2.262	1.335	3.678	0.649
16	0.063	0.059	0.198	1.179	0.011	0.382	0.206
17	0.063	0.059	0.055	0.989	0.172	0.122	0.063
18	0.063	0.059	0.065	1.001	0.155	0.131	0.073
19	0.063	0.059	0.269	1.272	0.002	0.577	0.267
20	0.063	0.059	6.499	25.276	53.992	72.928	1.418
21	0.063	0.059	0.494	1.611	0.234	1.473	0.435
22	0.063	0.059	0.153	1.115	0.040	0.272	0.163
23	0.063	0.059	0.124	1.077	0.069	0.226	0.135
24	0.063	0.059	2.679	6.928	13.494	21.149	1.119
25	0.063	0.059	0.389	1.449	0.079	1.011	0.362
26	0.063	0.059	0.206	1.185	0.008	0.399	0.212
27	0.063	0.059	0.125	1.0719	0.074	0.214	0.131

$$Y_{ijr} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijr}, \quad i = 1,2,3,4; \quad j = 1,2; \; r = 1,2,\ldots,60. \tag{4.28}$$

The vectors are all of order 4x1 and ϵ_{ijr} is assumed to be an $N_4(0,\Sigma)$ random vector. Thus the responses consists of 4 measurements replicated 60 times at each of the possible combinations of levels of factors 1 (years 1982 - 85) and 2 (Medicare and non - Medicare). The factor 1 has 4 levels and factor 2 has 2 levels. A test for the hypothesis

$$H: \; \gamma_{11} = \gamma_{12} = \ldots = \gamma_{42} = 0$$

is conducted and the results using different criteria are given in Table 4.9.

Table 4.9

Statistics	*value*	*F−value*	*p−value*
Wilks Lambda:	0.9394	1.21	0.2742
Pillai's trace:	0.0613	1.20	0.2755
Hotelling-Lawley trace:	0.0636	1.21	0.2728
Roy's Largest root:	0.0478	2.76	0.2756

It is observed based on all the criteria that there is no interaction effect. Now to test

$$H: \; \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$$

the criteria are summarized in Table 4.9a.

Table 4.9a

statsitcs	*value*	*F−value*	*p−value*
Wilks lambda:	0.9204	1.61	0.0852
Pillai's trace:	0.0812	1.61	0.0846
Hotelling-Lawley trace:	0.0845	1.60	0.0859
Roy's Largest root:	0.0532	3.07	0.0860

The results from Table 4.9a show that at 10 percent level of significance, we do reject the null hypothesis. In a similar manner the test criteria for testing the hypothesis $H: \beta_1 = \beta_2 = 0$, for factor 2 are given in Table 4.9b.

Table 4.9b

Statistics	*value*	*F−value*	*p−value*
Wilks Lambda:	0.8909	7.00	0.0001
Pillai's trace:	0.1090	7.00	0.0001
Hotelling-Lawley trace:	0.1223	7.00	0.0001
Roy's largest root:	0.1223	7.00	0.0001

The conclusion follows that there is a significant difference between Medicare and non - Medicare subjects. Examination of the different diagnostic statistics for identifying the influential observations is now conducted.

The influence measures for this analysis are shown for a subset of the data in the Tables 4.10 and 4.11 for Medicare and non - Medicare respectively. For economy of space, only T_i^2, WK_i, C_i, $LD_i(B,\Sigma)$, $LD_i(B|\Sigma)$, and $LD_i(\Sigma|B)$ are presented in the Tables. The measures T_i^2 and WK_i pinpoint that the observations 1, 5, 11, 22, 28, 30, 43, 62, 63, 90, and 100 are different from the rest of the observations. Observation number 22 is declared to be most influential by $LD_i(B,\Sigma)$, $LD_i(\Sigma|B)$, and WK_i. From Table 4.11 it is clear that the observations 14, 26, 90, and 116 are different from all the others and the observation number 26 is influential. From this experience with several data sets , examining WK_i, $LD_i(B,\Sigma)$, and $LD_i(\Sigma|B)$ seems sufficient for detecting the influential observations in MANOVA situation.

In summary: (i) A generalization of the available statistics are carried out for the multivariate regression model. Based on several data analyses it is concluded that the statistics similar to Cook's and Welsch's statistics can be used for identifying the influential observations. (ii) The statistics WK_i, $Covr_i$, and $LD_i(\Sigma|B)$ are useful measures for detection of influential observations in MANOVA models.

Table 4.10: Influence measures for Medical Data (Medicare)

Obs.	$LD_i(\Sigma\|B)$	$LD_i(B,\Sigma)$	$LD_i(B\|\Sigma)$	T_i^2	WK_i	C_i
10	0.0329	0.0537	0.0409	1.1492	0.0396	0.0345
11	0.0018	0.1516	0.1137	3.2526	0.1121	0.0345
12	0.0054	0.2655	0.1739	5.0512	0.1741	0.0345
13	0.0156	0.0888	0.0708	2.0032	0.0690	0.0345
14	0.1886	0.7956	0.3519	10.707	0.3692	0.0345
15	0.0004	0.1699	0.1246	3.5731	0.1232	0.0345
16	0.0168	0.0857	0.0684	1.9343	0.0667	0.0345
17	0.0063	0.1218	0.0946	2.6943	0.0928	0.0345
18	0.0030	0.1416	0.1075	3.0704	0.1058	0.0345
19	0.0127	0.3099	0.1938	5.6566	0.1950	0.0345
20	0.0105	0.2980	0.1886	5.4983	0.1895	0.0345
21	0.0226	0.0722	0.0574	1.6204	0.0558	0.0345
22	0.0245	0.3633	0.2158	6.3350	0.2184	0.0345
23	0.3317	1.0781	0.4188	12.9782	0.4475	0.0345
24	0.0232	0.3580	0.2137	6.2695	0.2161	0.0345
25	0.0246	0.0682	0.0540	1.5225	0.0525	0.0345
26	3.6751	5.8226	0.9787	35.8214	1.2352	0.0345
27	0.0130	0.0963	0.0765	2.1688	0.0747	0.0345
28	0.0001	0.1834	0.1323	3.8010	0.1310	0.0345
29	0.0009	0.1627	0.1204	3.4492	0.1189	0.0345
30	0.0004	0.2109	0.1471	4.2422	0.1462	0.0345
80	0.0468	0.0353	0.0222	0.6231	0.0214	0.0345
81	0.0325	0.0543	0.0414	1.1650	0.0401	0.0345
82	0.0057	0.1251	0.0968	2.7569	0.0950	0.0345
83	0.0061	0.1232	0.0954	2.7175	0.0937	0.0345
84	0.0040	0.2546	0.1688	4.8964	0.1688	0.0345
85	0.0190	0.0802	0.0640	1.8078	0.0623	0.0345
86	0.0025	0.1456	0.1100	3.1433	0.1083	0.0345
87	0.0005	0.2126	0.1480	4.2694	0.1472	0.0345
88	0.0742	0.5243	0.2734	8.1483	0.2809	0.0345
89	0.0333	0.0530	0.0403	1.1324	0.0390	0.0345
90	0.7537	1.7924	0.5518	17.7451	0.6119	0.0345
91	0.0123	0.0986	0.0782	2.2180	0.0764	0.0345
92	0.0169	0.0853	0.0681	1.9261	0.0664	0.0345
107	0.0034	0.1386	0.1056	3.0157	0.1039	0.0345
108	0.0040	0.1347	0.1031	2.9420	0.1014	0.0345
109	0.0030	0.2452	0.1643	4.7595	0.1641	0.0345
110	0.0147	0.3202	0.1981	5.7904	0.1996	0.0345
111	0.0018	0.1520	0.1140	3.2602	0.1124	0.0345
112	0.0001	0.1830	0.1320	3.7942	0.1308	0.0345
113	0.0229	0.0715	0.0568	1.6015	0.0552	0.0345
114	0.0032	0.1403	0.1067	3.0469	0.1050	0.0345
115	0.0077	0.2810	0.1810	5.2675	0.1816	0.0345
116	0.8597	1.9584	0.5779	18.7226	0.6456	0.0345
117	0.0200	0.3447	0.2083	6.1039	0.2108	0.0345
118	0.0015	0.1551	0.1159	3.3158	0.1143	0.0345
119	0.0002	0.1767	0.1285	3.6895	0.1272	0.0345
120	0.0001	0.2030	0.1429	4.1188	0.1420	0.0345

Table 4.11:Influence Measures for medical Data (non - Medicare)

Obs.	$LD_i(\Sigma\|B)$	$LD_i(B,\Sigma)$	$LD_i(B\|\Sigma)$	T_i^2	WK_i	C_i
1	0.2297	0.8807	0.3733	11.4248	0.3939	0.0305
2	0.0003	0.2069	0.1450	4.1795	0.1441	0.0305
3	0.1368	0.6811	0.3209	9.6847	0.3339	0.0305
4	0.0401	0.0435	0.0309	0.8667	0.0298	0.0305
5	0.7517	1.7890	0.5515	17.7260	0.6112	0.0305
6	0.0002	0.1786	0.1296	3.7204	0.1281	0.0305
7	0.0018	0.2328	0.1584	4.5778	0.1573	0.0305
8	0.0017	0.1523	0.1146	3.2647	0.1128	0.0305
9	0.0034	0.2492	0.1662	4.8185	0.1666	0.0305
10	0.3228	1.0616	0.4153	12.8523	0.4431	0.0305
11	0.0100	0.1064	0.0839	2.3828	0.0821	0.0305
12	0.0523	0.4602	0.2518	7.4613	0.2577	0.0305
13	0.0086	0.1119	0.0878	2.4945	0.0860	0.0305
14	0.0213	0.0750	0.0597	1.6863	0.0581	0.0305
15	0.0176	0.0835	0.0666	1.8843	0.0649	0.0305
16	0.0008	0.1635	0.1207	3.4628	0.1194	0.0305
17	0.0044	0.1323	0.1017	2.8961	0.0998	0.0305
18	0.0236	0.0709	0.0557	1.5695	0.0541	0.0305
19	0.0010	0.1687	0.1192	3.4246	0.1185	0.0305
20	0.0195	0.0795	0.0637	1.7806	0.0614	0.0305
21	0.0176	0.0836	0.0666	1.8848	0.0649	0.0305
22	5.6092	8.2626	1.1446	44.0699	1.5196	0.0305
23	0.0102	0.1064	0.0839	2.3831	0.0821	0.0305
24	0.0009	0.2209	0.1524	4.3962	0.1515	0.0305
25	0.0135	0.0948	0.0754	2.1356	0.0736	0.0305
26	0.0001	0.1869	0.1345	3.8521	0.1330	0.0305
27	0.0019	0.2345	0.1590	4.6064	0.1586	0.0305
28	0.3633	1.1319	0.4313	13.4103	0.4624	0.0305
29	0.0317	0.0557	0.0427	1.2018	0.0414	0.0305
30	0.0179	0.0828	0.0661	1.8673	0.0644	0.0305
31	0.3676	1.1406	0.4331	13.4681	0.4644	0.0305
32	0.0388	0.0452	0.0325	0.9131	0.0314	0.0305

Table 4.11 (continued)

Obs.	$LD_i(\Sigma\|B)$	$LD_i(B,\Sigma)$	$LD_i(B\|\Sigma)$	T_i^2	WK_i	C_i
33	0.0248	0.0677	0.0536	1.5103	0.0521	0.0305
34	0.0310	0.0568	0.0438	1.2319	0.0424	0.0305
35	0.1125	0.6233	0.3042	9.1383	0.3151	0.0305
36	0.0467	0.0354	0.0224	0.6261	0.0215	0.0305
37	0.0174	0.3328	0.2034	5.9531	0.2052	0.0305
38	0.0223	0.3544	0.2122	6.2249	0.2146	0.0305
39	0.0345	0.0513	0.0386	1.0842	0.0373	0.0305
40	0.0116	0.1008	0.0799	2.2654	0.0781	0.0305
41	0.0434	0.0393	0.0265	0.7437	0.0256	0.0305
42	0.0176	0.3338	0.2039	5.9663	0.2057	0.0305
43	0.0329	0.0537	0.0409	1.1499	0.0396	0.0305
44	0.5753	1.5035	0.5027	15.942	0.5497	0.0305
45	0.0428	0.0405	0.0273	0.7663	0.0264	0.0305
46	0.0162	0.0877	0.0695	1.9673	0.0678	0.0305
47	0.0539	0.0271	0.0137	0.3832	0.0132	0.0305
48	0.0477	0.4455	0.2467	7.2994	0.2517	0.0305
58	0.0293	0.0597	0.0464	1.3053	0.0450	0.0305
59	0.0026	0.2414	0.1625	4.7043	0.1622	0.0305
60	0.0027	0.2428	0.1631	4.7244	0.1629	0.0305
61	0.0092	0.2903	0.1852	5.3941	0.1860	0.0305
62	0.3750	1.1574	0.4358	13.5663	0.4677	0.0305
63	0.2362	0.8940	0.3765	11.5332	0.3979	0.0305
64	0.0045	0.1318	0.1012	2.8861	0.0995	0.0305
65	0.0365	0.0483	0.0357	1.0027	0.0345	0.0305
66	0.0772	0.5325	0.2761	8.2335	0.2836	0.0305
67	0.0024	0.1467	0.1107	3.1631	0.1095	0.0305
68	0.0362	0.4072	0.2327	6.8599	0.2361	0.0305
69	0.0402	0.0434	0.0308	0.8628	0.02975	0.0305
70	0.0002	0.1944	0.1383	3.9807	0.1376	0.0305

Table 4.11 (continued)

Obs.	$LD_i(\Sigma\|B)$	$LD_i(B,\Sigma)$	$LD_i(B\|\Sigma)$	T_i^2	WK_i	C_i
88	0.0442	0.0383	0.0254	0.7148	0.0246	0.0305
89	0.0387	0.0453	0.0327	0.9189	0.0316	0.0305
90	0.0317	0.0556	0.0427	1.2009	0.0414	0.0305
91	0.2802	0.9806	0.3970	12.2283	0.4216	0.0305
92	0.0859	0.5557	0.2835	8.4716	0.2921	0.0305
93	0.1201	0.6418	0.3096	9.3158	0.3212	0.0305
94	0.0294	0.0594	0.0462	1.3001	0.0448	0.0305
95	0.0001	0.1928	0.1375	3.9552	0.1363	0.0305
96	0.0439	0.0387	0.0259	0.7266	0.0250	0.0305
97	0.0377	0.0466	0.0340	0.9555	0.0329	0.0305
98	0.0403	0.0431	0.0305	0.8556	0.0295	0.0305
99	0.0244	0.3632	0.2157	6.3334	0.2183	0.0305
100	0.5110	1.3955	0.4829	15.2298	0.5251	0.0305
101	0.0042	0.1334	0.1023	2.9180	0.1006	0.0305
102	0.0001	0.1951	0.1387	3.9918	0.1376	0.0305

5. REGRESSION MODEL WITH AUTOCORRELATED ERRORS

The study of outliers and influential observations have been done mostly for uncorrelated data such as the linear regresssion model. The main idea behind this study is to compute the measures as differences between the quantities obtained with and without the *ith* observation. The limitation of this approach is that it can not be easily generalized to the time series data, in which the deletion of one observation changes the error structure. However, it is a well known fact that in the regression situation the deletion procedure is equivalent to treating the observation as a missing value. There exists an extensive literature on estimation in linear regression models with first - order correlated errors when some observations are missing. Wansbeck and Kapteyn (1985) showed that the most efficient estimator is the maximum likelihood estimator (mle) for a linear regression model with correlated errors when the observations are missing. In this chapter this missing value technique is applied to build some influence measures. The model and the estimators considered are described in the next section.

5.1 Model and Estimators

Consider the model

$$Y = X\beta + U, \tag{5.1}$$

where U is an nx1 vector, X is an nxm matrix, β is a mx1 vector of parameters

to be estimated, and Y is an nx1 vector of dependent variables. It is assumed that E(U)=0 and $E(UU') = \sigma^2 V$, where

$$V = \begin{pmatrix} 1 & \rho & \dots & \rho^{n-1} \\ \rho & 1 & \dots & \rho^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \dots & 1 \end{pmatrix}. \tag{5.2}$$

It can be easily shown that $|V| = (1-\rho^2)^{n-1}$ and

$$V^{-1} = \frac{1}{(1-\rho^2)} \begin{pmatrix} 1 & -\rho & 0 & \dots & 0 & 0 \\ -\rho & 1+\rho^2-\rho & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{pmatrix}$$

$$= \frac{1}{(1-\rho^2)}(I + \rho^2 C_1 - \rho C_2), \tag{5.2a}$$

where

$$C_1 = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & 1 & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{pmatrix} \tag{5.2b}$$

and

$$C_2 = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 1 & 0 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \ddots & 1 \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 \end{pmatrix} \tag{5.2c}$$

The model (5.1) includes a scale parameter σ^2, regression parameter β and error structural parameter ρ. Let $\theta' = (\beta', \sigma^2, \rho)$ and $\hat{\theta}'$ be the mle of θ' using the full data and $\hat{\theta}'_{(i)}$ be the mle assuming that the *ith* observation is missing and is computed, using the intervention analysis developed by Box

and Tiao (1975). The likeiihood function for the model (5.1) can be written as

$$L(\beta,\sigma^2,\rho)$$

$$= \left(\frac{1}{2\pi}\right)^{n/2}(\sigma^2)^{-n/2}(1-\rho^2)^{-(n-1)/2}e^{-1/2[(Y-X\beta)'(\sigma^2V)^{-1}(Y-X\beta)]} \quad (5.3)$$

and the log likelihood as

$$l(\beta,\sigma^2,\rho) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\sigma^2) - \frac{(n-1)}{2}log(1-\rho^2)$$

$$-\frac{1}{2\sigma^2(1-\rho^2)}(Y-X\beta)'(I+\rho^2C_1-\rho C_2)(Y-X\beta). \quad (5.4)$$

The equation (5.4) can be rewritten in the form

$$l(\beta,\sigma^2,\rho)$$

$$= -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\sigma^2) - \frac{(n-1)}{2}log(1-\rho^2) - \frac{A_1}{2\sigma^2(1-\rho^2)} - \frac{\rho^2A_2}{2\sigma^2(1-\rho^2)}$$

$$+\frac{\rho A_3}{2\sigma^2(1-\rho^2)}, \quad (5.5)$$

where

$$A_1 = (Y-X\beta)'(Y-X\beta), \quad (5.5a)$$

$$A_2 = (Y-X\beta)'C_1(Y-X\beta), \quad (5.5b))$$

$$A_3 = (Y-X\beta)'C_2(Y-X\beta). \quad (5.5c)$$

The likelihood function (5.5) yields the following non - linear maximum likelihood (ML) equations which can be solved for the unknown β, ρ, and σ^2. The equations are:

$$\hat{\beta} = [(X'X)^{-1} + \hat{\rho}^2 X'C_1X - \hat{\rho}X'C_2X]^{-1}[X'Y + \hat{\rho}^2 X'C_1Y - \hat{\rho}X'C_2Y], \quad (5.6a)$$

$$\hat{\sigma}^2 = \frac{1}{n(1-\hat{\rho}^2)}[A_1 + \hat{\rho}^2 A_2 - \hat{\rho}A_3], \quad (5.6b)$$

$$-\hat{\rho}^3(n-1)\hat{\sigma}^2 + \hat{\rho}^2 A_3 + \hat{\rho}(2A_1 - 2A_2 + 2(n-1)\hat{\sigma}^2) + A_3 = 0. \quad (5.6c)$$

The matrices C_1 and C_2 are as defined in (5.2b) and (5.2c) respectively. To compute the estimators, when the *ith* observation is missing, we write the model (5.1) as

$$Y = X\beta + \omega\delta_{(i)} + U, \; i = 1, 2, \ldots, n, \quad (5.7)$$

where $\delta_{(i)}$ has one in the *ith* position and zero elsewhere. The effect of an intervention at the *ith* observation can be estimated by the parameter ω. The log likelihood for (5.7) is

$$l(\beta, \rho, \sigma^2, \omega) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\sigma^2) - \frac{(n-1)}{2}log(1-\rho^2)$$

$$-\frac{1}{2\sigma^2(1-\rho^2)}[(Y - X\beta - \omega\delta_{(i)})'(I + \rho^2 C_1 - \rho C_2)(Y - X\beta - \omega\delta_{(i)})]. \quad (5.8)$$

Denoting the mles of β, σ^2, ρ, and ω by $\hat{\beta}_{(i)}$, $\hat{\sigma}^2_{(i)}$, $\hat{\rho}_{(i)}$, and $\hat{\omega}_{(i)}$ when the *ith* observation is missing, the ML equations using (5.8) can be written as

follows:

$$\hat{\beta}_{(i)} = (X'X + \hat{\rho}^2_{(i)}X'C_1X - \hat{\rho}_{(i)}X'C_2X)^{-1}$$

$$[X'Y + \hat{\rho}^2_{(i)}X'C_1Y - \hat{\rho}_{(i)}X'C_2Y - \hat{\omega}_{(i)}(X'\delta_{(i)} + \hat{\rho}^2_{(i)}X'C_1\delta_{(i)} - \hat{\rho}_{(i)}X'C_2\delta_{(i)}).$$

(5.9a)

$$\hat{\sigma}^2_{(i)} = \frac{1}{n(1-\hat{\rho}^2_{(i)})}(D_1 + \hat{\rho}^2_{(i)}D_2 - \hat{\rho}_{(i)}D_3). \quad (5.9b)$$

$$\hat{\omega}_{(i)} = [\delta'_{(i)}\delta_{(i)} + \hat{\rho}^2_{(i)}\delta'_{(i)}C_1\delta_{(i)} - \hat{\rho}_{(i)}\delta'_{(i)}C_2\delta_{(i)}]^{-1}$$

$$[\delta'_{(i)}(Y - X\hat{\beta}_{(i)}) + \hat{\rho}^2_{(i)}\delta'_{(i)}C_1(Y - X\hat{\beta}_{(i)}) - \hat{\rho}_{(i)}\delta'_{(i)}C_2(Y - X\hat{\beta}_{(i)})]. \quad (5.9c)$$

$$-\hat{\rho}^3_{(i)}(n-1)\hat{\sigma}^2_{(i)} + \hat{\rho}^2_{(i)}D_3 + \hat{\rho}_{(i)}(2D_1 - 2D_2 + 2(n-1)\hat{\sigma}^2_{(i)}) + D_3 = 0 \quad (5.9d)$$

where

$$D_1 = (Y - X\hat{\beta}_{(i)} - \hat{\omega}_{(i)}\delta_{(i)})'(Y - X\hat{\beta}_{(i)} - \hat{\omega}_{(i)}\delta_{(i)}). \quad (5.9e)$$

$$D_2 = (Y - X\hat{\beta}_{(i)} - \hat{\omega}_{(i)}\delta_{(i)})'C_1(Y - X\hat{\beta}_{(i)} - \hat{\omega}_{(i)}\delta_{(i)}). \quad (5.9f)$$

$$D_3 = (Y - X\hat{\beta}_{(i)} - \hat{\omega}_{(i)}\delta_{(i)})C_2(Y - X\hat{\beta}_{(i)} - \hat{\omega}_{(i)}\delta_{(i)}). \quad (5.9g)$$

To obtain the mles from the equations (5.6a) - (5.6c), the following iterative steps are required: (1) Set $\rho = 0$ and obtain estimator of β, denote the resulting estimator by b(1). (2) Set $\beta = b(1)$ and solve for ρ and denote the estimator of that by r(1). (3) Set $\rho = r(1)$ and obtain estimator for β and denote the estimator by b(2). Repeat steps 2 and 3. Iteration is continued until the changes in the estimators of β and ρ are sufficiently small or until

the largest percentage change in any estimate is sufficiently small. (4) Now the stable estimates of β and ρ are the mles, that is, $\hat{\beta}$ and $\hat{\rho}$. (5) Using the above mles for β and ρ, mle of σ^2 is obtained. In the same one can solved the equations (5.9a) - (5.9d), to obtain $\hat{\beta}_{(i)}$, $\hat{\rho}_{(i)}$,$\hat{\omega}_{(i)}$ and $\hat{\sigma}^2_{(i)}$.

5.2 Influence Measures

Let $\hat{\beta}$ and $\hat{\beta}_{(i)}$ be as in the previous section. The statistic suggested by Cook (1977) (cf: equation 2.12) can be written as

$$C_i = \frac{(\hat{\beta} - \hat{\beta}_{(i)})'(X'X)(\hat{\beta} - \hat{\beta}_{(i)})}{m\hat{\sigma}^2}, i = 1, 2, \ldots, n. \tag{5.10}$$

Similarly, WK_i, W_i, and C_i^* can be defined as:

$$WK_i \text{ type statistic} = \frac{|x_i(\hat{\beta} - \hat{\beta}_{(i)})|}{\hat{\sigma}_{(i)}\sqrt{p_i}}, i = 1, 2, \ldots, n; \tag{5.11}$$

$$C_i^* \text{ type statistic} = WK_i\sqrt{\frac{n-m}{m}}, i = 1, 2, \ldots, n; \tag{5.12}$$

$$W_i \text{ type statistic} = WK_i\sqrt{\frac{n-1}{1-p_i}}, i = 1, 2, \ldots, n. \tag{5.13}$$

Further, $Covr_i$ type statistic can be suggested as

$$Covr_i = \left(\frac{\hat{\sigma}^2_{(i)}}{\hat{\sigma}^2}\right)^m \frac{det(X'X)}{det(X'_{(i)}X_{(i)})}, i = 1, 2, \ldots, n. \tag{5.14}$$

5.3 Measures Based on the Likelihood Function

In section (2.7) the influence of a single observation on the likelihood function has been described. The distances (2.32), (2.36), and (2.38) can be

obtained for the model (5.1) in a similar manner. For, the mles of β, σ^2, and ρ based on the full data are substituted for β, σ^2, and ρ in (5.4). Which yields:

$$l(\hat{\beta},\hat{\sigma}^2,\hat{\rho}) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\hat{\sigma}^2) - \frac{(n-1)}{2}log(1-\hat{\rho}^2) - \frac{n}{2}. \qquad (5.15)$$

When the *ith* observation is deleted, the mles of β, σ^2, and ρ are given by (5.9a) - (5.9d) and the log-iikelihood function in that case is:

$$l(\hat{\beta}_{(i)},\hat{\sigma}^2_{(i)},\hat{\rho}_{(i)})$$

$$= -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\hat{\sigma}^2_{(i)}) - \frac{(n-1)}{2}log(1-\hat{\rho}^2_{(i)}) - \frac{(n-1)}{2}. \qquad (5.16)$$

Substituting (5.15) and (5.16) in (2.28), yields

$$LD_i(\beta,\sigma^2,\rho) = nlog\left(\frac{\hat{\sigma}^2_{(i)}}{\hat{\sigma}^2}\right) + (n-1)log\left(\frac{1-\hat{\rho}^2_{(i)}}{1-\hat{\rho}^2}\right) - 1. \qquad (5.17)$$

If the interest is only in estimating β, then the likelihood displacement is $LD_i(\beta|\sigma^2,\rho)$, which is anlogus to (2.33). Therefore,

$$\max_{\sigma^2,\rho} l(\beta,\sigma^2,\rho)$$

$$= l(\beta,\sigma^2(\beta),\rho^2(\beta)) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\sigma^2(\beta)) - \frac{(n-1)}{2}log(1-\rho^2(\beta))$$

$$-\frac{1}{2\sigma^2(\beta)}\frac{1}{\rho^2(\beta)}[(Y-X\beta)'(I+\rho^2(\beta)C_1-\rho(\beta)C_2)(Y-X\beta)]. \qquad (5.18)$$

The equation (5.18) is maximized over the parameter space for σ^2 and ρ keeping β fixed, which yields:

$$l(\beta, \sigma^2(\beta), \rho(\beta)) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\sigma^2(\beta)) - \frac{(n-1)}{2}log(1-\rho^2(\beta)) - \frac{n}{2}. \tag{5.19}$$

Now setting $\beta = \hat{\beta}_{(i)}$ in (5.19) the following is obtained:

$$l(\hat{\beta}_{(i)}, \sigma^{\bar{2}}(\hat{\beta}_{(i)}), \rho(\hat{\beta}_{(i)}))$$

$$= -\frac{n}{2}log(2\pi) - \frac{n}{2}log\sigma^2(\hat{\beta}_{(i)}) - \frac{(n-1)}{2}log(1-\rho^2(\hat{\beta}_{(i)})) - \frac{n}{2}, \tag{5.20}$$

where

$$\sigma^2(\hat{\beta}_{(i)})$$

$$= \frac{1}{n(1-\rho^2(\hat{\beta}_{(i)}))}[(Y - X\hat{\beta}_{(i)})'(I + \rho^2(\hat{\beta}_{(i)})C_1 - \rho(\hat{\beta}_{(i)})C_2)(Y - X\hat{\beta}_{(i)})], \tag{5.20a}$$

and

$$\rho^3(\hat{\beta}_{(i)})(n-1)\,\sigma^2(\hat{\beta}_{(i)}) + \rho^2(\hat{\beta}_{(i)})E_3$$

$$+2\rho(\hat{\beta}_{(i)})(E_1 - E_2 + (n-1)\sigma^2(\hat{\beta}_{(i)})) + E_3 = 0, \tag{5.20b}$$

with

$$E_1 = (Y - X\hat{\beta}_{(i)})'(Y - X\hat{\beta}_{(i)}) \tag{5.20c}$$

$$E_2 = (Y - X\hat{\beta}_{(i)})'C_1(Y - X\hat{\beta}_{(i)}) \tag{5.20d}$$

$$E_3 = (Y - X\hat{\beta}_{(i)})'C_2(Y - X\hat{\beta}_{(i)}). \tag{5.20e}$$

Sustituting (5.15) and (5.20) in (2.29) yields:

$$LD_i(\beta|\sigma^2,\rho) = nlog\left(\frac{\sigma^2(\hat{\beta}_{(i)})}{\hat{\sigma}^2}\right) + (n-1)log\left(\frac{1-\rho^2(\hat{\beta}_{(i)})}{1-\hat{\rho}^2}\right). \tag{5.21}$$

Similarly, $LD_i(\sigma^2|\beta,\rho)$ is obtained by maximizing the log likelihood for β and ρ keeping σ^2 fixed. Let $\max_{\beta,\rho} l(\sigma^2,\beta,\rho) = l(\sigma^2,\beta(\sigma^2),\rho(\sigma^2))$. Then,

$$l(\sigma^2,\beta(\sigma^2),\rho(\sigma^2)) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\sigma^2) - \frac{(n-1)}{2}log(1-\rho^2(\sigma^2))$$

$$-\frac{1}{2\sigma^2(1-\rho^2(\sigma^2))}[(Y - X\beta(\sigma^2))'(I + \rho^2(\sigma^2)C_1 - \rho(\sigma^2)C_2)(Y - X\beta(\sigma^2))]. \tag{5.22}$$

Setting $\sigma^2 = \hat{\sigma}^2_{(i)}$ in (5.22) yields:

$$l(\hat{\sigma}^2_{(i)},\beta(\hat{\sigma}^2_{(i)}),\rho(\hat{\sigma}^2_{(i)})) = -\frac{n}{2}log(\hat{\sigma}^2_{(i)}) - \frac{(n-1)}{2}log(1-\rho^2(\hat{\sigma}_{(i)}))$$

$$-\frac{1}{2\hat{\sigma}^2_{(i)}}(1-\rho^2(\hat{\sigma}^2_{(i)}))[(Y - X\beta(\hat{\sigma}^2_{(i)}))'W(Y - X\beta(\hat{\sigma}^2_{(i)})), \tag{5.23}$$

where

$$W = (I + \rho^2(\hat{\sigma}^2_{(i)})C_1 - \rho(\hat{\sigma}^2_{(i)})C_2).$$

Note that the cubic equation for $\rho(\hat{\sigma}^2_{(i)})$ is

$$-\rho^3(\hat{\sigma}^2_{(i)})(n-1)\hat{\sigma}^2_{(i)} + \rho^2(\hat{\sigma}^2_{(i)})A_3 + 2\rho(\hat{\sigma}^2_{(i)})(A_1 - A_2 + (n-1)\hat{\sigma}^2_{(i)}) + A_3 = 0, \tag{5.23a}$$

where A_1, A_2 and A_3 are as defined previously, but estimate of ρ is replaced by $\rho(\hat{\sigma}^2_{(i)})$. Then the right hand side of (5.23) becomes:

$$-\frac{n}{2}log(\hat{\sigma}^2_{(i)}) - \frac{(n-1)}{2}log(1-\rho^2(\hat{\sigma}^2_{(i)})) - \frac{[A_1+\rho^2(\hat{\sigma}^2_{(i)})A_2 - \rho(\hat{\sigma}^2_{(i)})A_3]}{2\hat{\sigma}^2_{(i)}(1-\rho^2(\hat{\sigma}^2_{(i)}))}. \tag{5.24}$$

Substituting (5.15) and (5.24) in (2.29) one finds

$$LD_i(\sigma^2|\beta,\rho) = n\left(\frac{\hat{\sigma}^2_{(i)}}{\hat{\sigma}^2}\right) + (n-1)\left(\frac{1-\rho^2(\hat{\sigma}^2_{(i)})}{1-\hat{\rho}^2}\right)$$

$$+\frac{(A_1+\rho^2(\hat{\sigma}^2_{(i)})A_2 - \rho(\hat{\sigma}^2_{(i)})A_3)}{\hat{\sigma}^2_{(i)}(1-\rho^2(\hat{\sigma}^2_{(i)}))} - n. \tag{5.25}$$

The likelihood displacement for ρ given β, σ^2 is obtained using the relation,

$$LD_i(\rho|\beta,\sigma^2) = 2[l(\hat{\beta},\hat{\rho},\hat{\sigma}^2) - l(\beta(\hat{\rho}_{(i)}),\sigma^2(\hat{\rho}_{(i)}),\hat{\rho}_{(i)})]. \tag{5.26}$$

Note that,

$$l(\hat{\rho}_{(i)},\sigma^2(\hat{\rho}_{(i)}),\beta(\hat{\rho}_{(i)})) = -\frac{n}{2}log\,\sigma^2(\hat{\rho}_{(i)}) - \frac{(n-1)}{2}log(1-\hat{\rho}^2_{(i)}) - \frac{n}{2}. \tag{5.27}$$

Substituting (5.15) and (5.27) in (5.26) yields

$$LD_i(\rho|\beta,\sigma^2) = nlog\left(\frac{\sigma^2(\hat{\rho}_{(i)})}{\hat{\sigma}^2}\right) + (n-1)log\left(\frac{1-\hat{\rho}^2_{(i)})}{1-\hat{\rho}^2}\right), \tag{5.28}$$

where

$$\sigma^2(\hat{\rho}_{(i)}) = \frac{1}{n(1-\hat{\rho}^2_{(i)})}(A_1 + \hat{\rho}^2_{(i)}A_2 - \hat{\rho}_{(i)}A_2) \tag{5.28a}$$

and

$$\hat{\beta} = (X'X + \hat{\rho}^2_{(i)}X'C_iX - \hat{\rho}_{(i)}X'C_2X)^{-1}(X'Y + \hat{\rho}^2_{(i)}X'C_1Y - \hat{\rho}_{(i)}X'C_2Y). \tag{5.28b}$$

Again, if there is interest only in σ^2 and ρ, then $LD_i(\sigma^2, \rho|\beta)$ is

$$LD_i(\sigma^2, \rho|\beta) = 2[l(\hat{\beta}, \hat{\sigma}^2, \hat{\rho}) - l(\beta(\hat{\rho}_{(i)}, \hat{\sigma}^2_{(i)}), \hat{\sigma}^2_{(i)}, \hat{\rho}_{(i)})]. \tag{5.29}$$

That is,

$$\max_{\beta} l(\beta, \sigma^2, \rho) = -\frac{n}{2}log(\sigma^2) - \frac{(n-1)}{2}log(1-\rho^2)$$

$$-\frac{1}{2\sigma^2(1-\rho^2)}[(Y - X\beta(\rho, \sigma^2))'(I + \rho^2C_1 - \rho C_2)(Y - X\beta(\rho, \sigma^2))]. \tag{5.30}$$

The equation (5.30) reduces to

$$-\frac{n}{2}log(\sigma^2) - \frac{(n-1)}{2}log(1-\rho^2) - \frac{n}{2}. \tag{5.31}$$

Next by setting $\sigma^2 = \hat{\sigma}^2_{(i)}$ and $\rho = \hat{\rho}_{(i)}$ in (5.31) one can obtain

$$l(\hat{\sigma}^2_{(i)}, \hat{\rho}_{(i)}, \hat{\beta}(\hat{\rho}_{(i)}, \hat{\sigma}^2_{(i)})) = \frac{n}{2}log(\hat{\sigma}^2_{(i)}) - \frac{(n-1)}{2}log(1-\hat{\rho}^2_{(i)}) - \frac{n}{2}. \tag{5.32}$$

Substituting (5.15) and (5.32) in (5.29)

$$LD_i(\sigma^2, \rho|\beta) = n\left(\frac{\hat{\sigma}^2_{(i)}}{\hat{\sigma}^2}\right) + n\,log\left(\frac{1-\hat{\rho}^2_{(i)})}{(1-\hat{\rho}^2}\right) \tag{5.33}$$

is obtained.

For algebric simplicity, if the estimate of ρ employing the full data is used, then

$LD_i(\beta|\sigma^2,\rho)$, $LD_i(\sigma^2|\beta,\rho)$ and $LD_i(\sigma^2,\rho|\beta)$ reduce to the following:

$$LD_i(\beta|\sigma^2,\rho) = n\,log\left(\frac{\sigma^2(\hat{\beta}_{(i)})}{\hat{\sigma}^2}\right), i = 1,2,\ldots,n, \tag{5.34}$$

where

$$\sigma^2(\hat{\beta}_{(i)}) = \frac{1}{n(1-\hat{\rho}^2)}[(Y - X\hat{\beta}_{(i)})'(I + \hat{\rho}^2 C_1 - \hat{\rho}C_2)(Y - X\hat{\beta}_{(i)})], \tag{5.34a}$$

and $\hat{\rho}$ is same as in (5.6c);

$$LD_i(\sigma^2|\beta,\rho) = nlog\left(\frac{\hat{\sigma}^2_{(i)}}{\hat{\sigma}^2}\right) + \frac{[A_1 + \hat{\rho}^2 A_2 - \hat{\rho}A_3]}{\hat{\sigma}^2_{(i)}(1-\hat{\rho}^2_{(i)})} - n,\ i = 1,2,\ldots,n; \tag{5.35}$$

and

$$LD_i(\sigma^2,\rho|\beta) = n\,log\left(\frac{\hat{\sigma}^2_{(i)}}{\hat{\sigma}^2}\right), i = 1,2,\ldots,n. \tag{5. 36}$$

As noted earlier, the likelihood displacement $LD(.,.)$ can be interpreted in terms of the asymptotic confidence region. (See Cox and Hinkley, 1974, Chapter 9). That is

$$LD(\theta) = 2[l(\hat{\theta}) - l(\theta)] \leq \chi^2(\alpha; q),$$

where $\chi^2(\alpha; q)$ is the upper α percentile point of the chi-squared distribution with q degrees of freedom and q is the dimension of θ. Therefore, $LD_i(.,.)$ can be compared with the $\chi^2(\alpha; q)$.

Some diagnostic statistics are obtained based on the likelihood function using the missing value technique for detection of influential observations for the regression model with correlated errors.

REFERENCES

Aitchinson, T. and Dunsmore, I. R. (1975). Statistical Prediction Analysis. Cambridge University Press.

Anderson, T. W. (1984). An Introduction to Multivariate Statistical Analysis. John Wiley, New York.

Andrews, D. F. and Pregibon, D. (1978). Finding outliers that matter. Journal of the Royal Statistical Society, Series B, 40, 87 - 93.

Anscombe, F. J. (1960). Rejection of outliers. Technometrics 2, 123 - - 147.

Atkinson, A. C. (1981). Two graphical displays for outlying and influential observations in regression. Biometrika, 68, 13 - 20.

_____ (1985). Plots, Transformation, and Regression. University Press, Oxford.

Balasooriya, U., Tse, Y. K. and Liew, Y. S. (1987). An empirical comparison of some statistics for identifying outliers and influential observations in linear regression models. Journal of Applied Statistics, 14, No. 2, 177 - 184.

Balasooriya, U. and Tse, Y. K. (1986). Outlier detection in linear models: A comparative study in simple linear regression. Communiactions in Statistics: Theory and Methods, 15(12), 3589 - 3597.

Belsley, D. A., Kuh, E. and Welsch, R. E. (1980). Regression Diagnostics: Identifying influential data and sources of collinearity. John Wiley, New York.

Box, G. E. P. and Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. Journal of the Amer. Statist. Assoc. 70, 70 - 79.

Brownlee, K. A. (1965). Statistical Theory and Methodology in Science and Engineering. Second Edn. John Wiley, New York.

Chatterjee, S. and Price, B. (1977). Regression analysis by example. John Wiley, New York.

Chatterjee, S. and Hadi, A. S. (1986). Influential observations, high leverage points, and outlier in linear regression. Statistical Science, 1, 379 - 416.

Chauvenet, W. (1863). Methods of least squares. Appendix to Manual of Spherical and Practical Astrnomy, Vol 2, Lippincott, Philadelphia, 469 - 566.

Cook, R. D. (1977). Detection of influential observations in linear regression. Technometrics, 19, 15 - 18.

_____ (1979). Influential observations in linear regression. Journal of the Amer. Statist. Assoc., 74, 169 - 174.

Cook, R. D. and Weisberg, S. (1980). Characterization of an empirical influence function for detecting influential cases in regression. Technometrics, 22, 495 - 508.

_____ (1982). Residuals and Influence in Regression. Chapman and Hall, London.

Cox, D. R. and Hinkley, D. V. (1974). Theoretical Statistics. Chapman and Hall, London.

Daniel, C. (1960). Locating outliers in factorial experiments . Technometrics, 2, 149 - 156.

Daniel, C. and Wood, F. S. (1971). Fitting Equations to Data: Computer Analysis of Multifactor Data, John Wiley, New York.

Draper, N. R. and Stoneman, D. M. (1966). Testing for the inclusion of variables in linear regression by a randomization technique. Technometrics, 8, 695 - 699.

Draper, N. R. and John, J. A. (1981). Influential observations and outliers in regression. Technometrics, 23, 21 - 26.

Edgeworth, F. Y. (1887). On discordant observations. Philosophical Magagine, 23, 364 - 375.

Ellenberg, J. H. (1973). The joint distribution of the standardized least squares residuals from a general linear regression. Journal of the Amer. Statist. Assoc., 68, 941 - 943.

Federer, W. T. (1955). Experimental Design, Macmillan, New York.

Forbes, J. D. (1857). Further experiments and remarks on the measurements of heights by the boiling point of water. Transaction of the Royal Society of Edinburgh, 21, 135 - 143.

Geisser, S. (1987). Influential observations, diagnostics and discovery test. Journal of Applied Statistics, 14, No.2, 133 - 142.

Gentlemen, J. F. and Wilk, M. B. (1975). Detecting outliers in two- way Tables. Technometrics, 17, 1 - 14.

Gentlemen, J. F. (1980). Finding the K-most likely outliers in two- way Tables. Technometrics, 22, 591 - 600.

Hampel, F. R. (1968). Contributions to the Theory of Robust Estimation. Ph.D, thesis, University of California, Berkley.

_____ (1974). The influence curve and its role in robust estimation. Journal of the Amer. Statist. Assoc., 62, 1179 - 1186.

Hoaglin, D. C. and Kempthorne, P. J. (1986). Comment on Chatterjee and Hadi's paper. Statistical Science, 1, 408 - 412.

Hoaglin, D. C. and Welsch, R. E. (1978). The hat matrix in regression and ANOVA. The American Statistician, 32, 117 - 122.

Hossain, A. and Naik, D. N. (1989). Detection of influential observations in multivariate regression. Journal of Applied Statistics, Vol 16, No.1, 25 - 37.

Huber P. J. (1981). Robust Statistics. John Wiley, New York.

Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979). Multivariate Analysis. Academic Press, London.

Mickey, M. R., Dunn, O. J. and Clark, V. (1967). Note on the use of stepwise regression in detecting outliers. Computers and Biomedical Research, 1, 105 - 111.

Miller, R. G. (1974). An unbalanced Jackknife. The Annals of Statistics, 2, 880 - 891.

Moore, J. (1975). Total biochemical oxygen demand of diary manures. Ph.D. thesis, University of Minnesota, Department of Agricultural Engineering.

Naik, D. N. (1986). Detection of outliers in the multivariate linear regression model. Tech, Report No. 86 - 11, University of Pittsburgh, USA.

Pendleton, O. J. (1985). Influential observations in the analysis of variance. Communication Statist. Theory Method, 14(3), 551 - 565.

Potthoff, R. F. and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems, Biometrika, 51, 313 - 326.

Rao, C. R. (1973). Linear Statistical Inference and Its Applications. John Wiley, New York.

Searle, S. R. (1971). Linear models. John Wiley, New York.

Snedecor, G. W. and Cochran, W. G. (1967). Statistical Methods. 6th Edn. Ames, Iowa, Iowa State University.

Srikantan, K. S. (1961). Testing for the single outlier in a regression model. Sankhya, Series A, 23, 251 - 260.

Stone, E. J. (1867). On the regression of discordant observations. Monthly Notices Royal Astronomical Society, 28,165 - 168.

Tietjen, G. L., Moore, R. L. and Beckmen, R. J. (1973). Testing for a single outlier in simple linear regression, Technometrics, 15, 717 - 721.

Thompson, W. R. (1935). On a criterion for the rejection of observations and the distribution and the sample mean in samples of n from a normal universe. Biometrika, 32, 301 - 310.

Timm, N. H. (1975). Multivariate Analysis with Applications in Education and Psychology, Brooks/Cole, California.

Tukey, J. W. (1972a). Data Analysis, Computation and Mathematics. Quarterly Journal of Applied Mathematics, 30, 51 - 65.

_____ (1972b). Some graphic and semigraphic displays. In statistical papers in Honor of George W. Snedecor. ed, T. A. Bancroft, Ames, Iowa. Iowa State University Press.

Wansbeck, T. and Kapteyn, A. (1985). Estimation in a linear model with serially correlated errors when observations are missing. International Economic Review, Vol. 26, No. 2, 469 - 490.

Weisberg, S. (1980). Applied linear regression. John Wiley, New York.

Welsch, R. E. and Kuh, E. (1977). Linear regression diagnostics. Technical report 923 - 77, Sloan School of Management, Massachusetts Institute of Technology.

Welsch, R. E. and Peters, S. C. (1978). Finding influential subsets of data in regression models. Proceedings of the eleventh interface symposium on computer science and statistics, Raleigh, Institute of Statistics, North Carolina State University.

Welsch, R. E. (1982). Influence function and regression diagnostics. In modern data analysis. R. L. Launer and A. F. Siegel, Eds. Academic, New York.

Wood, F. S. (1983). Measurements of observations far-out in influence and/or factor space. Unpublished manuscript (presented at the Econometrics and Statistics Colloquium, University of Chicago, Graduate School of Business.

Autobiographical Statement

Place and Date of Birth	:	Barisal (Bangladesh) July 1st 1953
Educational	:	⋆ M. Sc. (Statistics) 1976 Jahangirnagar University, Bangladesh
	:	⋆ B. Sc. (Honors) Statistics, 1975 Jahangirnagar University, Bangladesh
Employment	:	⋆ Lecturer (January 1977 - July 1980) University of Dhaka, Bangladesh
	:	⋆ Assistant Professor (July 1980 - on leave) University of Dhaka, Bangladesh
	:	⋆ Part - time Faculty (July 1978 - September 1982) ICMA, Dhaka, Bangladesh
	:	⋆ Teaching Assistant (August 1983 - August 1989) Dept. of Mathematics and Statistics Old Dominion University, Virginia
Academic	:	⋆ Merit scholarship (1970 - 1975) Jessore Board, Bangladesh
	:	⋆ Merit scholarship (1975 - 1976) Jahangirnagar University, Bangladesh