Identifying Outliers and Influential Observations in General Linear Regression Models

A thesis presented

by

Dominique Katshunga

to

The Department of Statistical Sciences

in fulfillment of the requirements

for the degree of

Masters in Science

in the subject of

Mathematical Statistics

University of Cape Town

August 2004

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

Acknowledgments

I wish to express my sincere gratitude to the following:

- First of all to my God and creator, without him my life is senseless. When things went wrong or when they went right, he did not change and he was always there for me. all the time my strength was drawn from him. thank you Father!
- to my supervisor, Professor Casper Troskie who always made time to answer my questions and supply me with good and new ideas. He also supplied me with the data used in this thesis.
- To Professor Tim Dunne for proofreading for spelling and grammatical errors, as English is not my first language.
- To Allan Clark who often helped me in developing R-codes used in this thesis.
- To Ian Durbach for his support and encouraging me to work hard.
- To my family for their continued support and love.
- To the National Research Foundation and UCT for their financial support during my studies.

Abstract

Identifying outliers and/or influential observations is a fundamental step in any statistical analysis, since their presence is likely to lead to erroneous results. Numerous measures have been proposed for detecting outliers and assessing the influence of observations on least squares regression results. Since outliers can arise in different ways, the above mentioned measures are based on motivational arguments and they are designed to measure the influence of observations on different aspects of various regression results. In what follows, we investigate how can one combine different test statistics based on residuals and diagnostic plots to identify outliers and influential observations (both in the single and multiple case) in general linear regression models.

Contents

Chapter 1

INTRODUCTION

The problem of identification of outliers and eventually influential observations is one of the most important aspects of regression theory. It is counted among the oldest but also most active problems in statistics. Several books have been written and major results about how to detect and treat outliers or influential data have been found as well. However, it appears that more research on the same topics are not useless, as long as the outliers appear in different contexts. A good illustration of this reality is the Masking and Swamping problems when one is dealing with a sample containing more than one outlier.

Many years ago the linear regression model became one of the statistically useful quantitative tools in many applied and physical sciences. As for the techniques used in the relevant linear regression, we can obviously realize the wide popularity of the ordinary least squares, this holds to the fact that this technique allows an easy statistical inference, a relatively low computational cost, a simple way of testing hypotheses , giving good results in interpretation and prediction. In practice, however, when it comes to assessing the stability or effectiveness of the model and possibly the influence of the data, the above characteristics of the ordinary least squares almost disappear. In fact, inferences based on the ordinary least squares regression are widely affected by influential data even if these are very few. Nevertheless, the reader should note that such data are not necessarily bad; they can provide some more important features for the analysis. In this regard, Neyman, J., and Scott. E.L.(1971) suggest that "apparent outliers" need sustained attention as well as careful study because they might be the most important observations. However, it doesn't matter whether the influential data are harmful or not, the investigator needs to identify them and hence improve the confidence about the model and its offshoot by drawing specific information contained in such data.

1.1 OUTLIERS

According to the encyclopedia of statistical sciences, the intuitive definition of an outlier is some observation whose discordancy from the majority of the sample is excessive in relation to the assumed distributional model for the model, thereby leading to the suspicion that it is not generated by this model. N.R. Drapper & Smith (1981) provide a similar and interesting definition of outliers: " the outlier is a peculiarity and indicates a data point which is not at all typical of the rest of the data". From the residual point of view, they define an outlier as one that is far greater than the rest in absolute value and perhaps lies three or four standard deviation or further from the mean of the residuals.

With regard to the above definitions, an outlier can be considered as an observation which deviates considerably from the others and seems to be generated by a mechanism different from that generating the rest of the observations.

1.2 Origins of Outliers

Two main reasons are generally retained as being the origin of outliers among the data. The first one is that the data come from some heavy tailed distribution known as "Outlier-Prone" (Green, 1976). These are statistical distributions with tails that go to zero slowly (e.g. t-distribution, Cauchy distribution, etc...). The second reason is that the data are from two different distributions. One of which, the basic distribution, with a given probability p generates good observations and the other one, the contaminating distribution, generating bad observations with probability $(1 - p)$.

When there is evidence of outliers in the sample, the knowledge of its origin might be of great interest for the statistician in handling and making good decisions about them. For example, if the experimentalist is convinced that the second alternative happens to be the origin of outliers in the sample, he can choose the rejection method as a palliative without risk of compromising the results . However, if the first reason is the cause of outliers in the sample, some caution should be taken because in this case the outlier is a valid observation.

1.3 Notation

This section introduces the symbols and notation adopted in this thesis; these are presented in the form of tables indicating in respective order the symbol, the dimension and the description. We use the expression "reduced model" to indicate that one or a group of regressors have been omitted in the regression model, whereas the expression "reduced data" refers to the regressor matrix W, with the i-th observation or group of m observations indexed by some set I deleted. Omission of the i-th case is indicated by the subscript (i), while omission of the i-th variable is indicated by the subscript $[i]$. Thus, W_{nxq} being the regressor matrix, one has

$$W_{(i)} = \begin{pmatrix} W_1^T \\ W_2^T \\ . \\ . \\ . \\ W_{i-1}^T \\ W_{i+1}^T \\ . \\ . \\ . \\ W_n^T \end{pmatrix}$$

and

$$W_{[i]}^T = (1, W_1, W_2, ..., W_{i-1}, W_{i+1}, ..., W_q)$$

Finally, the regression of W_j on the remaining regressors $W_{[j]}$ is referred to as the "secondary linear regression".

Table 1.1 **Matrices**

#	symbol	Dimension	Description
1	W	nxq	Matrix of regressors
2	$W_{(i)}$	(n-1)xq	Matrix of regressors based on the reduced data
3	$W_{(I)}$	(n-m)xq	Matrix of regressors based on the reduced data
4	$W_{[j]}$	nx(q-1)	Matrix of regressors based on the reduced model
5	$\widetilde{W}$	nx(q-1)	Matrix of centered regressors
6	$\widetilde{W}_{(i)}$	(n-1)x(q-1)	Matrix of centered regressors based on the reduced data
7	Z	nx(q+1)	Augmented matrix $[W\ X]$
8	V	nxn	Hat or projection matrix
9	V_1	nxn	Projection matrix of the column space of W_1
10	$(V_1)_I$	mxm	Submatrix of V_1 indexed by the set I with $\#(I) = m$
11	M	nxn	$M = I - V$, I is the identity matrix

Table 1.2 Vectors

#	Symbol	Dimension	Description
1	X	nx1	Vector of responses
2	$\hat{X}$	nx1	Vector of estimated responses
3	$\widetilde{X}$	nx1	Vector of centered responses
4	β	qx1	Vector of (unknown) regression parameters
5	$\hat{\beta}$	qx1	Vector least squares estimators
6	$\hat{\beta}_{(i)}$ & $\hat{\beta}_{(I)}$	qx1	Least squares estimators based on the reduced data
7	e	nx1	Vector of residuals
8	e_{W_j}	nx1	Vector of residuals from the secondary regression
9	d_i	nx1	unit vector with 1 in i-th position and zero elsewhere
10	$\hat{\alpha}_j$	(q-1)x1	least squares estimators from the reduced model
11	W_j	nx1	j-th regressor
12	$\hat{W}_j$	nx1	Vector of secondary fitted values
13	$\overline{W}$	qx1	Vector of column means of W
14	$\overline{W}_{(i)}$	qx1	Vector of column means based on the reduced data
15	w_j	qx1	j-th row of W
16	$\widetilde{\beta}$	qx1	Vector of ridge estimate
17	$\widetilde{e}$	nx1	Vector of ridge residuals
18	$\widetilde{\beta}_{(i)}$ & $\widetilde{\beta}_{(I)}$	qx1	Vectors of ridge estimates based on the reduced data
19	β_{pc}	qx1	Vector of the principal component estimator
20	e^{-}	nx1	Principal component residuals
21	β^{*}	qx1	Estimator under the stochastic prior information
22	r^{*}	nx1	Residuals from the stochastic prior information model

Table 1.3 Residuals and Sums of squares

#	Symbol	Dimension	Description
1	e_{i_X}	1	i-th residual
2	$e_{(i)X}$	1	i-th predicted residuals based on the reduced data
3	r_i	1	Standardized residual
4	t_i	1	Studentized residual
5	t_I^2	1	Generalized studentized residual
6	e_{iW_j}	1	i-th residual from the secondary regression
7	r_{iW_j}	1	standardized residual from the secondary regression
8	t_{iW_j}	1	Studentized residual from the secondary regression
9	$e_{(i)W_j}$	1	i-th predicted residual from the secondary regression
			and based on the reduced data
10	$RSS\ or\ RSS_X$	1	Residual sum of squares
11	$RSS_{X_{(i)}}$	1	Residual sum of squares based on the reduced data
12	RSS_{W_j}	1	Residual sum of squares from the secondary regression
13	$RSS_{W_{(i)j}}$	1	Residual sum of squares from the secondary regression
			and based on the reduced data

Table 1.4 Leverage and influence related measures

#	Symbol	Dimension	Description
1	$v_i = v_{ii}$	1	i-th diagonal element of V
2	v_{ii}^j	1	i-th diagonal element of the projection matrix $(V_{[j]})$ from
			the reduced model
3	$v_{ii} - v_{ii}^j$	1	measure of the i-th partial leverage
4	O_{i_X}	1	i-th orientator for X
5	O_{iW_j}	1	i-th orientator for W_j
6	S_{i_X}	1	i-th scale inflator for X
7	S_{iW_j}	1	i-th scale inflator for W_j
8	P_{iW_j}	1	Primary potential for W_j
9	ld_i	1	likelihood distance due to the deletion of the i-th observation
10	IF_I	qx1	Influence function with m observations indexed by I deleted
11	Δ_i	1	measure of influence of the i-th observation
12	Δ_I	1	measure of influence of a group of m observations indexed by I

Chapter 2

DETECTION OF A SINGLE OUTLIER

As an introduction to this chapter we briefly look at some important and often used central and non-central distributions when one is dealing with outliers in the linear model. Following this, we will present some tests and procedures for the identification of outliers.

2.1 Distributions

2.1.1 Univariate normal distribution

The random variable X follows a univariate normal distribution with mean μ *and variance* σ^2 if its probability density function is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \text{ for } -\infty < x < \infty$$

As shorthand for the statement "X follows a normal distribution with parameters μ *and* σ^2, it is convenient to use the following notation: $X \sim N(\mu, \sigma^2)$. The special case for which $\mu = 0$ and $\sigma^2 = 1$ is called the **standard normal** distribution.

2.1.2 Multivariate normal distribution

Consider an n-random vector $X' = [X_1, X_2, ..., X_n]$. The random variables in X' follow a multivariate normal distribution if their probability density function is given by

$$\mathrm{f}(x_1, x_2, ..., x_n) = \frac{\exp\left\{-\frac{1}{2}(X-\mu)' \sum^{-1}(X-\mu)\right\}}{(2\pi)^{n/2}\left|\sum\right|^{1/2}}$$

where $E(X) = \mu$ and $E\left(X-\mu\right)\left(X-\mu\right) = \sum$

2.1.3 The Chi-square distribution

A random variable U is said to have a Chi-square distribution with n degrees of freedom if its probability density function is

$$f(u) = \frac{u^{\frac{n}{2}-1}\exp\{-\frac{u}{2}\}}{2^{n/2}\Gamma\left(\frac{1}{2}n\right)} \; for\; u > 0$$

Note: If $X_i \sim N(0,1)$ then the random variable

$$U = \sum_{k=0}^{n} X_i^2$$

follows a Chi-square distribution with n degrees of freedom and we denote it as $U \sim \chi^2_{(n)}$.

2.1.4 The F-distribution

A random variable V has the F-distribution with parameters n_1 and n_2 if its probability density function is given by

$$f(v) = \frac{\Gamma(\frac{n_1+n_2}{2})n_1^{\frac{n_1}{2}}n_2^{\frac{n_2}{2}}v^{\frac{n_1}{2}-1}}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})(n_2+n_1 v)^{\frac{n_1+n_2}{2}}} \; for\; v \geq 0$$

If the random variables U_1 and U_2 follow the Chi-square distributions with n_1 and n_2 degrees of freedom respectively, then

$$V = \frac{U_1/n_1}{U_2/n_2}$$

follows an F-distribution with n_1 *and* n_2 degrees of freedom.

Notation: $V \sim F(\,n_1, n_2)$

2.1.5 The t-distribution

The random variable W follows the t-distribution with n degrees of freedom if its probability density function is given by the expression

$$f(w) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1+\frac{w^2}{n}) \; for \; -\infty < w < \infty$$

Notation: $W \sim t(n)$

Let consider the following random variables: $X \sim N(0,1)$ *and* $U \sim \chi^2(n)$. The random variable

$$W = \frac{X}{\sqrt{\frac{U}{n}}} \sim t(n)$$

2.1.6 The Beta-distribution

The random variable X has the beta distribution with parameters α *and* β if its probability density function is given by

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \; for \; 0 < x < 1$$

Notation: $X \sim Be(a, \beta)$

2.1.7 Non-central chi-square distribution

When $X \sim N(\mu, I)$ and $U = X'X$, the distribution of the random variable U is the non-central χ^2 with n degrees of freedom and non-centrality parameter γ given by

$$\gamma = \frac{\mu'\mu}{2}$$

The probability density function of this distribution is given by

$$f(u) = \sum_{i=0}^{\infty} \frac{e^{-\gamma}\gamma^i u^{\frac{n+2i-1}{2}} e^{-\frac{u}{2}}}{i! 2^{\frac{n}{2}+i}\Gamma(\frac{n}{2}+i)} \; for \; u > 0$$

Notation: $U \sim \chi^2(n, \gamma)$

2.1.8 Non-central F-distribution

Let U_1 *and* U_2 be independent random variables such that $U_1 \sim \chi^2(n, \gamma)$ *and* $U_2 \sim \chi^2(n)$. the random variable

$$V = \frac{U_1/n_1}{U_2/n_2}$$

has the non-central F-distribution with n_1 *and* n_2 degrees of freedom and non-centrality parameter γ, its probability density function is

$$f(v) = \sum_{i=0}^{n} \frac{e^{-\gamma}\gamma^i \Gamma(\frac{n_1+n_2+2i}{2}) n_1^{\frac{n_1}{2}+i} n_2^{\frac{n_2}{2}} v^{\frac{n_1}{2}+i-1}}{i!\Gamma(\frac{n_1}{2}+i)\Gamma(\frac{n_2}{2})(n_2+n_1 v)^{\frac{n_1+n_2}{2}+i}} \; for \; v > 0$$

Notation: $U \sim F(n_1, n_2, \gamma)$

2.1.9 Non-central t-distribution

X and U being independent random variables. if $X \sim N(\mu, 1)$ and $U \sim \chi^2_{(n)}$ then the random variable

$$T = \frac{X}{\sqrt{\frac{U}{n}}}$$

has the non-central t-distribution with n degrees of freedom and non-centrality parameter μ; its probability density function is

$$f(t) = \frac{n^{\frac{n}{2}} e^{-\frac{1}{2}\mu^2}}{\Gamma(\frac{n}{2})(n+t^2)^{\frac{n+1}{2}}} \sum_{i=0}^{\infty} \frac{\Gamma(\frac{n+i+1}{2})\mu^i 2^{i/2} t^i}{i!(n+t)^{i/2}} \quad for \quad -\infty < t < \infty$$

Notation: $T \sim t(n, \mu)$

The techniques used in this work for outliers detection will be essentially based on residuals .

2.2 Definition of residuals

Given X an nx1 vector of dependant variables, W an nxq matrix of independent variables, β a qx1 vector of unknown parameters and ε an nx1 vector of uncorrelated random variables with mean $\emptyset$ and covariance matrix $\Sigma = \sigma^2 I$, the standard form of the general linear model is:

$$X = W\beta + \varepsilon \tag{2.1}$$

Almost all the information about outliers is carried by the vector of residuals below:

$$e = X - W\hat{\beta}$$

Where $\hat{\beta} = (W^T W)^{-1} W^T X$ is the vector of the least squares estimated of the $\beta' s$.

Thus

$$e = X - W(W^T W)^{-1} W^T X$$

and

$$e = (I - W(W^T W)^{-1} W^T)X$$

setting

$$V = W(W^TW)^{-1}W^T$$

one can write:

$$e = (I - V)\varepsilon \tag{2.2}$$

which scalar form is

$$e_i = \varepsilon_i - \sum_{j=1}^{n} v_{ij}\varepsilon_j \; for \; i = 1, ..., n \tag{2.3}$$

V is known as the Hat matrix. We also define $M = I - V$.

From (2.3) it appears that the $e_{i's}$ will be closer to the $\varepsilon_{i's}$ if the $v_{ij's}$ are closer to zero. Hence , a good insight and understanding of the hat matrix V is essential for most regression diagnostics.

If we partition W as follows

$$W = (W_1, W_2)$$

where W_1 is an nxp matrix, one has

1°

$$W_1^* = W_1(W_1^TW_1)^{-1}W_1^T$$

the projection matrix for the column space of W_1.

2°

$$W_2^* = W_2 - W_1^*W_2 = (I - W_1^*)W_2$$

is the component of W_2 orthogonal to W_1.

Hence,

$$U = W_2^*(W_2^{*^T}W_2^*)^{-1}W_2^{*^T}$$

is the projection matrix of the column space of W orthogonal to the column space of W_1.

Since, V may be regarded as a sum of two components :

$$V = W_1^* + U \tag{2.4}$$

which is the orthogonal form of V.

2.2.1 Properties of the Hat matrix, V

Property1.

V is symmetric, that is $V^T = V$

Property2.

V is idempotent, that is $V^2 = V$

$$\begin{aligned} \sum_{j=1}^{n} v_{ij}^2 &= v_{ii}\ ,\ i = 1, 2, ..., n \\ trace(V) &= rank(V) = q \end{aligned}$$

Let denote the eigenvalues of V by $\lambda_1, \lambda_2, ..., \lambda_n$. Given that V is idempotent, these eigenvalues are either 0 or 1. Thus

$$\begin{aligned} trace(V) &= \sum_{i=1}^{n} \lambda_i \\ &= rank\,(V) \\ &= rank\,(W) \\ &= q \end{aligned}$$

Property3

V is invariant under a non-singular linear transformation; namely if V_W and V_{WF} are the projection matrices of W and $W * F$ respectively, with F non-singular, one has

$$\begin{aligned} V_{WF} &= WF(W^T F^T WF)^{-1} F^T W^T \\ &= WFF^{-1}(W^T W)^{-1} F^{-1} F^T W^T \\ &= W(W^T W)^{-1} W^T \\ &= V_W \end{aligned}$$

Property4. The eigenvalues of V are either 0 or 1.

Property5.

The Hat or projection matrix V can be written as

$$V = UU^T$$

with U an orthonormal matrix, that is $U^T U = I$.

Let consider the singular value decomposition of W, say

$$W = UDV^T$$

we have that

$$\begin{aligned} V &= W(W^TW)^{-1}W^T \\ &= UDV^T\left(VDU^TUDV^T\right)^{-1}VDU^T \\ &= UDV^T\left(VD^2V^T\right)^{-1}VDU^T \\ &= UU^T \end{aligned}$$

Property6.

$$0 \leq v_{ii} \leq 1$$

Proof: From property 2 we have

$$\begin{aligned} v_{ii} &= \sum_{j=1}^{n} v_{ij}^2 \\ &= v_{ii}^2 + \sum_{j\neq i}^{n} v_{ij}^2 \end{aligned}$$

which shows that $0 \leq v_{ii} \leq 1$

Property7.

In the hat matrix V, if $v_{ii} = 0$ then $v_{ij} = 0$ and if $v_{ii} = 1$ then $v_{ij} = 0\ \forall j \neq i$.

In 6 above we wrote

$$v_{ii} = v_{ii}^2 + \sum_{j\neq i}^{n} v_{ij}^2$$

substituting $v_{ii} = 0$ for v_{ii} we get $0 = 0 + \sum_{j\neq i}^{n} v_{ij}^2$ and it follows that $v_{ij} = 0$. The same, substituting $v_{ii} = 1$ for v_{ii} one obtains $1 = 1^2 + \sum_{j\neq i}^{n} v_{ij}^2$ which leads $v_{ij} = 0\ \forall j \neq i$.

Property8.

$$-\tfrac{1}{2} \leq v_{ij} \leq \tfrac{1}{2},\ \forall j \neq i.$$

expanding the expression in property 6, one can also write

$$v_{ii} = v_{ii}^2 + v_{ij}^2 + \sum_{k\neq i,j}^{n} v_{ik}^2$$

equivalently

$$v_{ij}^2 = v_{ii} - v_{ii}^2 - \sum_{k\neq i,j}^{n} v_{ik}^2$$

and

$$v_{ij}^2 \leq v_{ii} - v_{ii}^2$$

The maximum value of $v_{ii} - v_{ii}^2$ being $\frac{1}{4}$, the result follows.

Property9.

If V_k is a kxk submatrix of V obtained by the intersection of the k rows and k columns of V indexed by k, then the k eigenvalues of V_k and $(I - V_k)$ have values between 0 and 1.

Without loss of generality, suppose that V_k is the submatrix of V consisting of the first k rows and k columns of V. One can write V as

$$V = \begin{pmatrix} V_k & V_{k,n-k}^T \\ V_{k,n-k} & V_{n-k} \end{pmatrix}$$

V_k being symmetric, one has

$$V_k = NDN^T$$

with N an orthonormal matrix containing the eigenvectors of V_k, and D a diagonal matrix containing in the main diagonal the eigenvalues of V_k.

$$V_k = V_k V_k + V_{k,n-k}^T V_{k,n-k} \ \ (V \text{ is idempotent})$$

from the above

$$V_k \geq V_k V_k$$

or

$$NDN^T \geq ND^2N^T$$

which implies that

$$D - D^2 \geq 0$$

This last expression shows that the eigenvalues of V_k have values between 0 and 1. The following can be deduced:

$I - V_k$ is positive definite when the maximum eigenvalue of V_k is less than 1, and positive semidefinite otherwise.

Let assume that $\lambda_i \ (i = 1, 2, ..., k)$, the eigenvalues of V_k, satisfy $1 - \lambda_1 \geq 1 - \lambda_2 \geq 1 - \lambda_3 .... \geq 1 - \lambda_k$. The eigenvalues of $(I - V_k)$ satisfy $(1 - \lambda_1) \leq (1 - \lambda_2) \leq (1 - \lambda_3) .... \leq (1 - \lambda_k)$; letting $\lambda_1 = 1$ in this last expression gives $(1 - \lambda_1) = 0$ which shows that $(1 - \lambda_i) \geq 0 \ i = 1, 2, ..., k$

Property10

$$v_{ii} = 1 - \frac{\det\left(W_{(i)}^T W_{(i)}\right)}{\det\left(W^T W\right)}$$

Property11

If the regressor matrix has a column of ones then

- $v_{ii} \geq \frac{1}{n}$
- If an observation w_i occurs k times, then $v_{ii} \leq \frac{1}{k}$.

In (2.4) above if the regressor matrix is partitioned as follows

$$W = (\mathbf{1}\ W_2)$$

where $\mathbf{1}$ is a vector of ones, the projection matrix W_1^* is given by

$$\begin{aligned} W_1^* &= \mathbf{1}\left(\mathbf{1}^T\mathbf{1}\right)^{-1}\mathbf{1}^T \\ &= \left(\frac{1}{n}\right)\mathbf{1}\mathbf{1}^T \end{aligned}$$

Hence, following the same reasoning as previously, one can write

$$\begin{aligned} V &= W_1^* + U \\ &= W_1^* + (I_n - W_1^*)\,W_2\left(W_2^T\,(I_n - W_1^*)\,W_2\right)^{-1} W_2^T\,(I_n - W_1^*) \\ &= \left(\frac{1}{n}\right)\mathbf{1}\mathbf{1}^T + \widetilde{W}\left(\widetilde{W^T}\widetilde{W}\right)^{-1}\widetilde{W}^T \\ &= \left(\frac{1}{n}\right)\mathbf{1}\mathbf{1}^T + \widetilde{V} \end{aligned}$$

Since, the diagonal elements of V can be expressed as

$$v_{ii} = \frac{1}{n} + \widetilde{v}_{ii}$$

with $\widetilde{v}_{ii}$ nonnegative; the first result follows, namely $v_{ii} \geq \frac{1}{n}$.
Secondly, let define

$$S = \{j : w_j = w_i,\ j = 1, 2, 3, ..., k\}.$$

Knowing that

$$v_{ij} = w_i^T\left(W^T W\right)^{-1} w_j$$

one has $v_{ij} = v_{ii}$ for $j \in S$ and the expression from property 6 becomes

$$v_{ii} = \sum_j v_{ij}^2$$

$$
\begin{aligned}
&= \sum_{j \in S} v_{ij}^2 + \sum_{j \in S} v_{ij}^2 \\
&= k v_{ii}^2 + \sum_{j \notin S} v_{ij}^2
\end{aligned}
$$

From which $v_{ii} \geq k v_{ii}^2$ and finally $v_{ii} \leq \frac{1}{k}$

Property12

The diagonal elements of the hat matrix v_{ii} are non-decreasing in the number of independent variables q.

Let partition the regressor matrix as follows

$$
W^* = \left(W_{[j]} \; W_j \right)
$$

with respect to the earlier notation. Reasoning similar as above, the projection matrices are

$$
V_{[j]} = W_{[j]} \left(W_{[j]}^T W_{[j]} \right)^{-1} W_{[j]}^T
$$

and

$$
\begin{aligned}
V^* - V_{[j]} &= V - V_{[j]} \\
&= \left(I - V_{[j]} \right) W_j \left(W_j^T \left(I - V_{[j]} \right) W_j \right)^{-1} W_j^T \left(I - V_{[j]} \right) \\
&= \frac{e_{W_j} e_{W_j}^T}{RSS_{W_j}}
\end{aligned}
$$

Hence,

$$
\begin{aligned}
V &= V_{[j]} + \left(V - V_{[j]} \right) \\
&= V_{[j]} + \frac{e_{W_j} e_{W_j}^T}{RSS_{W_j}}
\end{aligned}
$$

In scalar form this can be written as

$$
v_{ii} = v_{ii}^j + \frac{e_{iW_j}^2}{RSS_{W_j}}
$$

where $\frac{e_{iW_j}^2}{RSS_{W_j}}$, the partial leverage of the i-th observation, is the contribution of the j-th regressor to the leverage of the i-th observation. Since $\frac{e_{iW_j}^2}{RSS_{W_j}} \geq 0$, one has $v_{ii} \geq v_{ii}^j$, which proves the result of the above property.

Property13

Deletion of the ith observation in the data matrix denoted by W leads the following results:

1.

$$v_{jk(i)} = v_{jk} + \frac{v_{ji}v_{ik}}{1 - v_{ii}} \quad j, k \neq i$$

2.

$$v_{kk(i)} = v_{kk} + \frac{v_{ik}^2}{1 - v_{ii}}$$

3

$$v_{ii(i)} = \frac{v_{ii}}{1 - v_{ii}}$$

The proof of these results is made possible by means of appendix A1,

$$\left(W_{(i)}^T W_{(i)}\right)^{-1} = \left(W^T W\right)^{-1} + \frac{\left(W^T W\right)^{-1} w_i w_i^T \left(W^T W\right)^{-1}}{1 - v_{ii}}$$

and

$$\begin{aligned} v_{jk(i)} &= w_j^T \left(W_{(i)}^T W_{(i)}\right)^{-1} w_k \\ &= w_j^T \left(\left(W^T W\right)^{-1} + \frac{\left(W^T W\right)^{-1} w_i w_i^T \left(W^T W\right)^{-1}}{1 - v_{ii}}\right) w_k \\ &= w_j^T \left(W^T W\right)^{-1} w_k + \frac{\left[w_j^T \left(W^T W\right)^{-1} w_i\right] \left[w_i^T \left(W^T W\right)^{-1} w_k\right]}{1 - v_{ii}} \\ &= v_{jk} + \frac{v_{ji}v_{ik}}{1 - v_{ii}} \end{aligned}$$

in the same way

$$v_{kk(i)} = v_{kk} + \frac{v_{ik}^2}{1 - v_{ii}}$$

and

$$\begin{aligned} v_{ii(i)} &= v_{ii} + \frac{v_{ii}^2}{1 - v_{ii}} \\ &= \frac{v_{ii}}{1 - v_{ii}} \end{aligned}$$

Property 14

Let V_k be the kxk submatrix of V indexed by the k cases to be deleted from W. On the condition that W is of full rank q, one has

$$V_k = I_k \Leftrightarrow rank\left(W_{(k)}\right) = q - k$$

Proof:

Let start by proving the property from the left to the right hand side

with regard to property 7, $v_{ii} = 1$ implies that $v_{ij} = 0 \; \forall j \neq i$. The projection matrix V can be expressed as

$$V = \begin{pmatrix} W_{(k)} \left(W^T W\right)^{-1} W_{(k)}^T & 0 \\ 0 & I_k \end{pmatrix}$$

because V is idempotent this last expression shows that $W_{(k)} \left(W^T W\right)^{-1} W_{(k)}^T$ is idempotent and

$$\begin{aligned} rank\left(W_{(k)} \left(W^T W\right)^{-1} W_{(k)}^T\right) &= rank\left(W_{(k)}\right) \\ &= q - k \end{aligned}$$

Now to prove the property in the other direction, one should remember that the expression $rank\left(W_{(k)}\right) = q - k$ simply means that they are k columns of $W_{(k)}$ that are linearly dependent; without loss of generality assume that these columns are the k last columns of $W_{(k)}$, hence the k rows to be left out are the last k rows of W. one can write

$$\begin{aligned} W &= \begin{pmatrix} C & 0 \\ 0 & I_k \end{pmatrix} F \\ &= MF \end{aligned}$$

where $F : qxq$ is non-singular, and $C : (n-k)\text{x}(q-k)$ has rank $q-k$. Following property 3 which state that the projection matrix V is invariant under a non-singular linear transformation, one can write

$$\begin{aligned} V_W &= V_{MF} \\ &= V_M \\ &= M\left(M^T M\right)^{-1} M^T \\ &= \begin{pmatrix} C\left(C^T C\right)^{-1} C^T & 0 \\ 0 & I_k \end{pmatrix} \\ &= \begin{pmatrix} I_{n-k} & 0 \\ 0 & I_k \end{pmatrix} \end{aligned}$$

2.3 Standardized residuals

With respect to the earlier notation, the standardized (internal studentized) residuals are defined as:

$$r_i = \frac{e_i}{m_i s} \quad i = 1, 2, ... n \tag{2.5}$$

where e_i is the i-th element of e, m_i the square root of the i-th diagonal element m_{ii} of M and $s^2 = \frac{e'e}{n-q}$ (the residual mean squared error) is an unbiased estimator of the variance σ^2.

Let's define

$$b_i = r_i = \frac{e_i}{m_i s} \tag{2.6}$$

Each b_i^2 is distributed as a beta random variable with parameters $\frac{1}{2}$ and $(n-q-1)/2$, say $Be[\frac{1}{2}, (n-q-1)/2]$; that is the distribution of the $b_i's$ is given by

$$f(b_i) = \frac{\Gamma(\frac{n-q}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{n-q-1}{2})}(1-b_i^2)^{\frac{n-q-2}{2}-\frac{1}{2}} \quad -1 < b_i < 1 \tag{2.7}$$

That is the standardized residuals r_i are related to the beta statistic, i.e. r_i^2 has a beta distribution.

Another version of the standardized residuals is the "external studentized residuals", Cook and Weisberg (1982), or "RSTUDENT", Belsley (1980), or jack-knife residuals, Atkinson (1981). They are defined as

$$t_i = \frac{e_i}{s_{(i)} m_i} \tag{2.8}$$

where $s_{(i)}^2$, the residual mean squared error with the i-th row deleted is the estimator of the variance σ^2. The numerator and the denominator of t_i are independent, so that t_i has the Student's t distribution under Gaussian assumption, that is $t_i \sim t(n-q-1)$.

Using the deletion formulae (Rao, 1973)

$$(n-q-1)s_{(i)}^2 = (n-q)s^2 - \frac{e_i^2}{m_{ii}} \tag{2.9}$$

leads to the following relationship between the two versions of standardized residuals,

$$t_i = r_i\sqrt{\frac{n-q-1}{n-q-r_i^2}} \tag{2.10}$$

Chen, J. (1978) considers

$$F_i = \frac{(n-q-1)r_i^2}{n-q-r_i^2} \tag{2.11}$$

from which

$$r_i^2 = \frac{(n-q)F_i}{(n-q-1+F_i)} \tag{2.12}$$

The standardized residuals r_i and t_i will be large when e_i is large or m_i is small; they are sensitive to outliers in the dependent space and/or outliers in the regressor space.

2.4 Test procedures

The test for a single outlier in the data set, based on the studentized residuals uses the maximum-absolute studentized residuals, say

$$R_n = Max\{|r_i|\}$$

From the above relationship between F_i and r_i in (2.11) and (2.12), one has

$$F_n = \frac{(n-q-1)R_n^2}{(n-q-R_n^2)} \tag{2.13}$$

or

$$R_n^2 = \frac{(n-q)F_n}{(n-q-1+F_n)} \tag{2.14}$$

In these last relations, F_n denotes the maximum of the different F_i $(i = 1, 2, ..., n)$.

We have the following two different cases: In the first, a specified observation is suspected as being an outlier and in the second, the observation which could possibly be an outlier is not known.

2.4.1 Test for a specified observation

From model (2.1) we have

$$E(X_i) = W_i\beta \quad i = 1, 2, ..., n$$

If observation X_k, k known, is suspected as being in error by an amount λ_k, then we can set up the following hypotheses:

$$H_o \quad : \quad \lambda_k = 0$$
$$H_1 \quad : \quad \lambda_k \neq 0 \qquad k \; is \; specified$$

These hypotheses derive from the following:

$$H_o \quad : \quad E(X_i) = \mu_i \qquad i = 1, 2, ..., n$$
$$H_1 \quad : \quad E(X_i) = \begin{cases} \mu_i & i \neq k \\ \mu_k + \lambda_k & i = k, \; k \; is \; specified \end{cases}$$

If X_k is an outlier, then the corresponding studentized residual r_k will be large in absolute value. Let $F(\alpha, 1, n-q-1)$ denote the upper $100\alpha\%$ point of the $F-$*variate* with 1 and $(n-q-1)$ degrees of freedom, then in regard to (2.12) the test for outlier will reject if $F_k > F(\alpha, 1, n-q-1)$

2.4.2 Test for an unspecified observation

Most of the time, the potential outlier is not known. We are interested in this case to test the hypotheses:

$$H_o \quad : \quad There \; is \; no \; outlier \; in \; the \; data \; set.$$
$$H_1 \quad : \quad There \; is \; one \; outlier \; in \; the \; data \; set.$$

If H_1 is true, then the absolute-maximum studentized R_n would be large. For fixed values of α, n *and* q, let $R_{(\alpha,1)}$ be the critical value from Lund (1975), the rejection region is given by:

$$R_n > R_{(\alpha,1)}$$

or equivalently (see 2.13 above)

$$F_n > F_{n_\alpha}$$

where

$$F_{n_\alpha} = \frac{(n-q-1)R^2_{(\alpha,1)}}{(n-q-R^2_{(\alpha,1)})}$$

2.5 The mean shift outlier model

The mean shift outlier model for a single row detection is given by

$$X = W\beta + d_i\phi + \varepsilon \tag{2.15}$$

With

$$E(\varepsilon) = \emptyset \; and \; Var(\varepsilon) = \sigma^2 I.$$

d_i is an n-vector with all components equal to zero except the i-th which is equal to 1.

As in (2.4), putting the added variable in the orthogonal form to the columns of W, one can write:

$$X = W\beta^* + (I - V)d_i\phi + \varepsilon \tag{2.16}$$

Thus, parameters in (2.16) can be fitted in two steps as shown below, given the orthogonality:

Step 1: $\hat{\beta}^* = (W^TW)^{-1}W^TX$, estimator from the regression of X on W, ignoring the added variable $(I - V)d_i$.

Step 2: From the regression of the residuals obtained in the preceding step on the added variable $(I - V)d_i$, the estimator of ϕ is given by:

$$\hat{\phi} = \frac{d_i^T(I - V)(I - V)X}{d_i^T(I - V)(I - V)d_i} = \frac{e_i}{1 - v_{ii}} \tag{2.17}$$

(2.17) can be found by solving for $\hat{\phi}$ in the system of normal equations from (2.16), that is

$$W^TW\hat{\beta}^* + W^T(I - V)d_i\hat{\phi} = W^TX \tag{2.16a}$$

$$d_i^T(I - V)W\hat{\beta}^* + d_i^T(I - V)(I - V)d_i\hat{\phi} = d_i^T(I - V)X \tag{2.16b}$$

If the reduced model $X = W\beta^* + \varepsilon$ can be fitted, that is $(W^TW)^{-1}$exists, the normal equation (2.16a) yields:

$$\hat{\beta}^* = (W^TW)^{-1}W^TX - (W^TW)^{-1}W^T(I-V)d_i\,\hat{\phi}$$

substitution of the above result in (2.16b) yields:

$$d_i^T(I-V)W[(W^TW)^{-1}W^TX-(W^TW)^{-1}W^T(I-V)d_i\hat{\phi}]+d_i^T(I-V)(I-V)d_i\hat{\phi} = d^T(I-V)X$$

$$\begin{aligned} d_i^T(I-V)(I-V)d_i\hat{\phi} - d_i^T(I-V)W(W^TW)^{-1}W^T(I-V)d_i\hat{\phi} &= d_i^T(I-V)X - \\ & \quad d_i^T(I-V)W(W^TW)^{-1}W^TX \end{aligned}$$

$$d_i^T(I-V)[I-W(W^TW)^{-1}W^T](I-V)d_i\hat{\phi} = d_i^T(I-V)[I-W(W^TW)^{-1}W^T]X$$

and finally expression (2.17) can be derived from this last equation, that is

$$\hat{\phi} = \frac{d_i^T(I-V)[I-W(W^TW)^{-1}W^T]X}{d_i^T(I-V)[I-W(W^TW)^{-1}W^T](I-V)d_i}$$

After arrangement

$$\hat{\phi} = \frac{d_i^T(I-V)(I-V)X}{d_i^T(I-V)(I-V)d_i} = \frac{e_i}{1-v_{ii}}$$

Test Statistic for $\phi = 0$

From step1 above, the sum of squares of regression of X on W is given by X^TVX, and from the second step the sum of squares for regression of $e = (I-V)X$ on the added variable $(I-V)d_i$ is given by

$$\begin{aligned} X^T(I-V)I(I-V)X &= X^T(I-V)(I-V)X \\ &= \hat{\phi}^2[d_i^T(I-V)(I-V)d_i] \\ &= \frac{e_i^2}{1-v_{ii}} \end{aligned}$$

Hence, the residual sum squares for the model (2.16) is

$$
\begin{aligned}
RSS &= X^T X - X^T V X - \frac{e_i^2}{1 - v_{ii}} \\
&= X^T (I - V) X - \frac{e_i^2}{1 - v_{ii}}
\end{aligned}
$$

Under the normality assumption, the statistic for the test $\phi = 0$ is:

$$
t_i = \frac{\frac{e_i}{(1-v_{ii})^{\frac{1}{2}}}}{[X^T(I-V)X - \frac{e_i^2}{1-v_{ii}}]^{\frac{1}{2}}/(n-q-1)^{\frac{1}{2}}} \tag{2.18}
$$

Under the null hypothesis, the statistic t_i follows a t-distribution with $(n-q-1)$ degrees of freedom. Thus, α being a nominal level, a two-tailed test for a single outlier will reject if $| t_i |> t(\alpha/n, n-q-1)$. If the row suspected for an outlier is unknown, an alternative rejection region (the Bonferroni inequality) is given by:

$$
max_i \quad | t_i |> t(\alpha/n, n-q-1)
$$

A similar rejection rule accommodating the $v_{ii's}$ is given by Cook and Weisberg(1980) as follows:

$$
max_j \quad | t_j |=| t_i |> t(v_{ii}\alpha/q, n-q-1)
$$

Remarks:

1 One should notice the similarity between the external studentized test statistic and the one under the mean shift outlier model; they are in fact two identical tests.

2 When $\phi \neq 0$, the distribution of t_i^2 is non-central F-distribution with noncentrality parameter $\frac{\phi^2(1-v_{ii})}{\sigma^2}$. It clearly appears that this parameter gets small as v_{ii} is close to 1, it therefore becomes difficult to identify outlying points with such values of v_{ii}, and yet these are of great interest.

Since

$$
V = W(W'W)^{-1}W'
$$

v_{ii} is increasing in q, the number of explanatory variables in the model. Thus, the larger the model the more difficult to detect outliers.

2.6 Added variable plot

In the mean shift model (2.15), replacing d_i by another explanatory variable, say Z, yields to the new model:

$$X = W\beta + \phi Z + \varepsilon \tag{2.19}$$

Multiplying both sides of (2.19) by $(I - V)$, one has:

$$(I - V)X = (I - V)W\beta + \phi(I - V)Z + (I - V)\varepsilon$$

that is

$$e = \phi(I - V)Z + (I - V)\varepsilon \tag{2.20}$$

taking expectation over both sides gives:

$$E(e) = \phi(I - V)Z \tag{2.21}$$

This last relation suggests that the new or added variable Z should enter the model linearly, that is the plot of e versus $(I - V)Z$ is linear and through the origin of axis as illustrated in figure(2.1) below. Hence, as in the previous chapter if we define P to be the projection matrix of the column space of W except W_m (the m-th column of W), one can plot $(I - P)X$ versus $(I - P)W_m$ to test wether or not case m might be suspected as outlier.

Note: Practically, the added variable plot is realized as follows: we take on the y-axis the residuals of the regression of the dependent variable X on the explanatory variables excluding the one that is being tested. On the other hand, we take on the x-axis, the residuals of the regression of this last variable on the other independent variables.

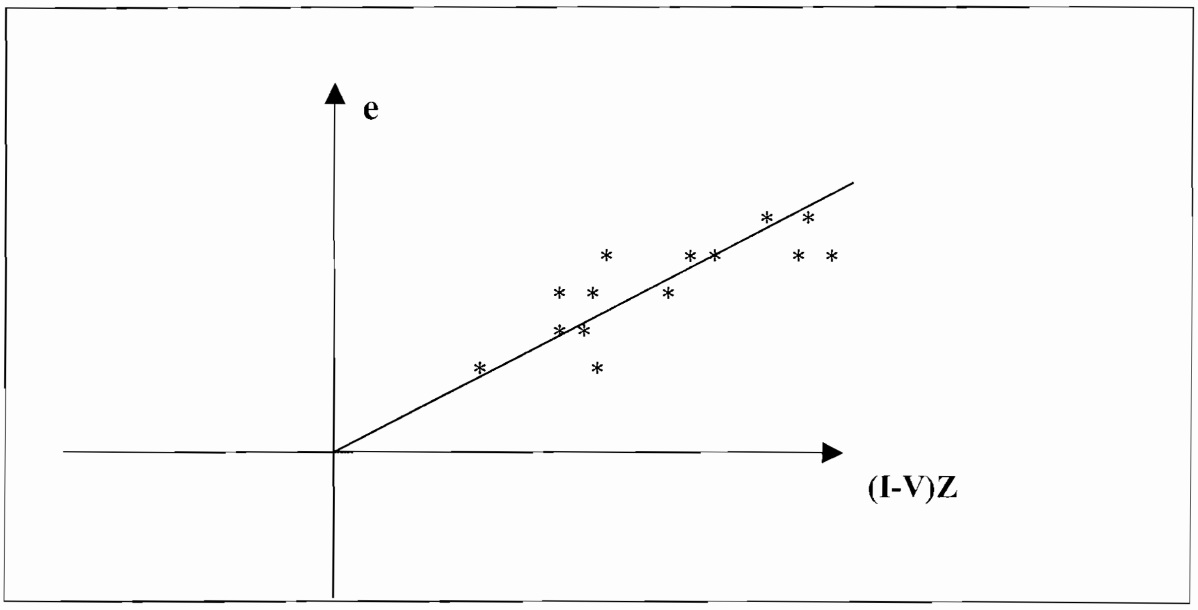


Figure 2.1 Illustration of an added Variable plot

2.7 Partial residual plots

The partial residual plot, Ezekiel (1924), is a variant of the added variable plot introduced above. Let consider the following model (full model)

$$X = W_{(j)}\beta_{(j)} + W_j\beta_j + \varepsilon \tag{2.22}$$

From the previous, the vector of residuals is given by

$$\begin{aligned}
e_X &= (I - V)\,X \\
&= \left(I - V_{(j)} - \frac{e_{W_j}e_{W_j}^T}{RSS_{W_j}}\right) X \\
&= \left(I - V_{(j)}\right) X - \frac{e_{W_j}\left(I - V_{(j)}\right)\left(I - V_{(j)}\right) X}{RSS_{W_j}} \\
&= e_{X_{(j)}} - e_{W_j}\hat{\beta}_j
\end{aligned} \tag{2.23}$$

It follows that

$$e_{X_{(j)}} = e_X + \left(I - V_{(j)}\right) W_j\hat{\beta}_j$$

Putting $V_{(j)}$ to zero, one has

$$e_{X_{(j)}} = e_X + W_j\hat{\beta}_j$$

The partial residual plot is defined as the plot of $e_X + W_j\hat{\beta}_j$ versus $\left(I - V_{(j)}\right) X_j$. It therefore appears as a special case of the added variable plot where $V_{(j)}$ has been replaced by 0. Here again, the slope is given by $\hat{\beta}_j$. This plot is restricted in detecting influential observations.

Chapter 3

MULTIPLE ROWS DETECTION

3.1 Introduction

The multiple case is analogous to the single one, but one needs to take care of some peculiarities that sometimes arise when one is trying to identify more than one outlier at a time. We have for example the problem of masking and swamping effects mentioned earlier in this study.

In this chapter we first present a non-parametric test for a multiple detection based on the mean shift outlier model. We will also look at the test procedure proposed by Gentleman and Wilk (1975); finally we will give an overview of the partial-regression leverage plots.

3.2 Mean shift model

Let $L = (i_1, i_2, ..., i_m)$ be an m-vector of case subscripts, $e_L = \{e_{i_j}; j = 1, 2, ..., m\}$ an m-vector of residuals and V_L an mxm submatrix of V with rows and columns indexed by L above. The mean shift outlier model for multiple rows detection can be written as follows:

$$X = W\beta + D\phi + \varepsilon; \quad E(\varepsilon) = \emptyset \qquad Var(\varepsilon) = \sigma^2 I \tag{3.1}$$

Where D is an nxm matrix with the $i_k - th$ element in the $k - th$ column equal to 1 while the remaining elements are all equal to 0. $\emptyset$ is the null matrix and ϕ an m-vector of unknown parameters.

Both Gentleman and Wilk (1975b) and Cook and Weisberg (1980) have shown that

$$t_L^2 = \frac{[e_L^T(I - V_L)^{-1}e_L]/m}{[(n-q)\hat{\sigma}^2 - e_L^T(I - V_L)^{-1}e_L]/(n-q-m)} \tag{3.2}$$

is the statistic for the test of $H_0 : \phi = \emptyset$ and that under the normality assumption, t_L^2 follows the F-distribution with m and $(n-q-m)$ degrees of freedom. Critical values for this multiple case outlier test are based on the Bonferroni inequality mentioned above.

Remark: As in the single case, the following relationship between t_L^2 and the multiple case analogue of the studentized residuals r_L is recommended for computation purposes :

$$t_L^2 = \frac{r_L^2\,(n-q-m)}{m(n-q-r_L^2\,)} \tag{3.3}$$

Note that

$$r_L^2 = \frac{e_L^T(I - V_L)^{-1}e_L}{\hat{\sigma}^2}$$

can easily be computed using the Choleski factorization of $(I - V_L)$ which is summarized as follows:

step1: find a triangular matrix C such that

$$C^TC = I - V_L$$

step2: solve for A in the system

$$C^TA = e_L$$

step3: compute

$$r_L^2 = \frac{A^TA}{\hat{\sigma}^2}$$

3.3 General distance measures for several cases

Let I be the set of indices for the m cases to be deleted. the empirical influence function is defined as

$$IF_I = \hat{\beta}_{(I)} - \hat{\beta} \tag{3.4}$$

and the distance function is

$$D_I(W^TW, qs^2) = \frac{(\hat{\beta}_{(I)} - \hat{\beta})^T(W^TW)(\hat{\beta}_{(I)} - \hat{\beta})}{qs^2} \tag{3.5}$$

From Bingham (1977) ,

$$\hat{\beta}_{(I)} - \hat{\beta} = -(W^TW)^{-1}W_I^T(I - V_I)_I^{-1}r_I$$

(3.5) can be written as

$$D_I(W^TW, qs^2) = \frac{r_I(I - V_I)^{-1}V_I(I - V_I)_I^{-1}r_I}{qs^2} \tag{3.6}$$

Applying the spectral decomposition to V_I, we have

$$V_I = \Gamma^T\Lambda\Gamma \tag{3.7}$$

where Γ is an orthogonal matrix and $\Lambda = diag(\lambda_1, ..., \lambda_m)$ an mxm diagonal matrix with $0 \leq \lambda_1 \leq ... \leq \lambda_m \leq 1$.

therefore (3.6) becomes

$$\begin{aligned} D_I(W^TW, qs^2) &= \frac{r_I(\Gamma^T\Gamma - \Gamma^T\Lambda\Gamma)^{-1}\Gamma^T\Lambda\Gamma(\Gamma^T\Gamma - \Gamma^T\Lambda\Gamma)^{-1}r_I}{qs^2} \\ &= \frac{(\Gamma^Tr_I)^T(I - \Lambda)^{-1}\Lambda(I - \Lambda)^{-1}(\Gamma^Tr_I)}{qs^2} \end{aligned} \tag{3.8}$$

Letting $g = \Gamma^Tr_I$, one has

$$D_I(W^TW, qs^2) = \frac{g^T(I - \Lambda)^{-1}\Lambda(I - \Lambda)^{-1}g}{qs^2}$$

that is

$$D_I(W^TW, qs^2) = \frac{\sum_{j=1}^m g_j^2\lambda_j/(1 - \lambda_j)^2}{qs^2} \tag{3.9}$$

or

$$D_I(W^TW, qs^2) = \frac{1}{q}\sum_{j=1}^m h_j^2\frac{\lambda_j}{1 - \lambda_j} \tag{3.10}$$

where

$$h_j^2 = \frac{g_j^2}{s^2(1 - \lambda_j)}$$

Following Gentleman and Wilk (1975), the likelihood ratio test statistic for the hypothesis that the m cases are not an outlying set is given by

$$F_I = \frac{(n - q - m)\sum_{j=1}^m h_j^2}{m(n - q - \sum_{j=1}^m h_j^2)} \tag{3.11}$$

$F_I \sim F(m, n - q - m)$

A comparison of (3.11) and (3.3) shows that

$$r_I^Tr_I = \sum_{j=1}^m h_j^2$$

Another interesting formulation leading to the same results as above is provided by Draper and John (1981): Writing the model (2.1) as follows

$$\begin{bmatrix} X_1 \\ X_n \end{bmatrix} = \begin{bmatrix} W_1 & \mathbf{0} \\ W_n & I_k \end{bmatrix} \begin{bmatrix} \beta \\ \lambda \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_n \end{bmatrix} \tag{3.12}$$

where X_1 and ε_1 are $(n-k)$x1 vectors, on the other hand X_n and ε_n are (kx1) vectors; W_1 is an $(n-k)$xq matrix and W_n is an (kxq) matrix . I_k is a kxk identity matrix and λ a kx1 vector of additional parameters. The least square estimate of β in this model is given by

$$\hat{\beta} = (W_1'W_1)^{-1}W_1'X_1$$

and that of λ can be derived as follows:

from (3.5) we have

$$\begin{bmatrix} \hat{\varepsilon}_1 \\ \hat{\varepsilon}_n \end{bmatrix} = \begin{bmatrix} I_{n-k} - (W_1'W_1)^{-1}W_1' & -W_1(W_1'W_1)^{-1}W_n' \\ -W_n(W'W)^{-1}W_1' & I_k - W_n(W'W)^{-1}W_n' \end{bmatrix} \begin{bmatrix} X_1 \\ X_n \end{bmatrix} \tag{3.13}$$

Hence,

$$\hat{\varepsilon}_n = [I_k - Wn(W'W)^{-1}W_n']X_n - W_n(W'W)^{-1}W_1'X_1$$

and

$$X_n = [I_k - W_n(W'W)^{-1}W_n']^{-1}[\hat{\varepsilon}_n + W_n(W'W)^{-1}W_1'X_1]$$

Since from (3.4)

$$X_n = W_n\hat{\beta} + \hat{\lambda}$$

we can write, using the above expression

$$\begin{aligned} \hat{\lambda} &= X_n - W_n\hat{\beta} \\ &= [I_k - W_n(W'W)^{-1}W_n']^{-1}[\hat{\varepsilon}_n + W_n(W'W)^{-1}W_1'X_1] - W_n(W_1'W_1)^{-1}W_1'X_1 \\ &= [I_k - W_n(W'W)^{-1}W_n']^{-1} \left\{ \begin{array}{c} \hat{\varepsilon}_n + W_n(W'W)^{-1}W_1'X_1 - \\ {[I_k - W_n(W'W)^{-1}W_n']W_n(W_1'W_1)^{-1}W_1'X_1} \end{array} \right\} \end{aligned}$$

let

$$W_n(W'W)^{-1}W_1'X_1 - [I_k - W_n(W'W)^{-1}W_n']W_n(W_1'W_1)^{-1}W_1'X_1 = A$$

one has

$$\begin{aligned} A &= W_n(W'W)^{-1}W_1'X_1 - W_n(W_1'W_1)^{-1}W_1'X_1 + W_n(W'W)^{-1}W_n'W_n(W_1'W_1)^{-1}W_1'X_1 \\ &= W_n(W'W)^{-1}[I - (W'W)(W_1'W_1)^{-1} + W_n'W_n(W_1'W_1)^{-1}]W_1'X_1 \end{aligned}$$

$$
\begin{aligned}
&= W_n(W'W)^{-1}[I - \{W'W - W_n'W_n\}(W_1'W_1)^{-1}]W_1'X_1 \\
&= W_n(W'W)^{-1}[I - \{W_1'W_1 + W_n'W_n - W_n'W_n\}(W_1'W_1)^{-1}]W_1'X_1 \\
&= W_n(W'W)^{-1}[I - \{W_1'W_1(W_1'W_1)^{-1}\}]W_1'X_1 \\
&= W_n(W'W)^{-1}[0]W_1'X_1 \\
&= 0
\end{aligned}
$$

The expression reduces to

$$\hat{\lambda} = [I_k - W_n(W'W)^{-1}W_n']^{-1}\hat{\varepsilon}_n$$

Gentleman and Wilk (1975) show that

$$Q_k = \hat{\varepsilon}_n'[I_k - W_n(W'W)^{-1}W_n']^{-1}\hat{\varepsilon}_n$$

which they termed "outlier sum of squares", the extra sum of squares due to fitting λ in the model (3.4). The statistic for the test of $H_o : \lambda = 0$ of no outliers is therefore given by

$$F = \frac{Q_k}{RSS - Q_k} \text{x} \frac{n - q - k}{k}$$

where RSS is the residual sum of squares. Under H_0, F follows the F-distribution with k and $(n - q - k)$ degrees of freedom.

3.4 The recursive residuals

The model (2.1) can also be written in partitioned form as follows,

$$\begin{bmatrix} X_1 \\ X_n \end{bmatrix} = \begin{bmatrix} W_1 \\ W_n \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_n \end{bmatrix} \tag{3.14}$$

where (X_n, W_n) is the last observation. With regard to the deletion formulae, if the n-th observation is omitted from the sample, we have the following results from the subregression

$$
\begin{aligned}
\hat{\beta}_1 &= (W_1'W_1)^{-1}W_1'X_1 \\
S_1 &= (X_1 - W_1\hat{\beta}_1)'(X_1 - W_1\hat{\beta}_1)
\end{aligned}
$$

From the work of Plackett(1950), we have

$$\hat{\beta} = \hat{\beta}_1 + \frac{(x_n - w_n\hat{\beta}_1)(W_1'W_1)^{-1}w_n'}{(1 + w_n(W_1'W_1)^{-1}w_n')}$$

$$RSS = RSS_1 + (x_n - w_n\hat{\beta}_1)^2$$

$$(x_n - w_n\hat{\beta}) = \frac{(x_n - w_n\hat{\beta}_1)}{[1 + w_n(W_1'W_1)^{-1}w_n']}$$

$$(W'W)^{-1} = (W_1'W_1)^{-1} - \frac{(W_1'W_1)^{-1}w_n'w_n(W_1'W_1)^{-1}}{[1 + w_n(W_1'W_1)^{-1}w_n']}$$

The above formulae allow us to produce $(W_1'W_1)^{-1}$, $\hat{\beta}_1$ and RSS_1 from their corresponding term in the full model, say $(W'W)^{-1}$, $\hat{\beta}$ and RSS.

Following Beckman and Trussell (1974),

$$r_n = \frac{(x_n - w_n\hat{\beta})}{\sqrt{1 + w_n(W_1'W_1)^{-1}w_n'}}$$

Under the assumption that W has full rank q, one can produce $(n - q)$ such variables, namely r_n, $r_{n-1}, \ldots, r_{q+1}$. These last are known as recursive residuals; they are independent and distributed as $N(0, \sigma^2)$.

RSS, the residual sum of squares is given by

$$RSS = \sum_{k=q+1}^{n} r_k^2$$

Now, suppose that the last k observations corresponding to the last k recursive residuals have been removed, the residual sum of squares becomes

$$RSS_k = \sum_{i=q+1}^{n-k} r_i^2$$

Following Hawkins(1980), if the observations are permuted so that the suspected outliers occupy the position n, $n-1, \ldots, n-k+1$; the following two statistics can be calculated:

$E_k = \dfrac{RSS_k}{RSS}$, a global statistic for the presence of up to k outliers

$R_k = \dfrac{r_{n-k+1}}{\sqrt{RSS_k}}$, a stepwise statistic appropriate for testing for the presence of k outliers given that the sample contains $k-1$ outliers

Small values of E_k indicate that the k observations put in the "outlier" partition are in fact contaminants. Fractiles of E_k for given n, k, ν *and* α, are provided by Hawkins (1978b) .

This formulation has the advantage of preventing several outliers from masking each others' presence. However, it is generally impossible to arrange the data as described above.

3.5 Test procedures

Procedures for detecting outliers in linear models have been suggested by many authors. e.g. Mickey, Dunn and Clark (1967), Gentleman and Wilk (1975). Named "step-up" and "step-down" procedures, they are also known under the names "forward deletion" and "backward deletion" procedures respectively; in this work we use both nomenclatures.

3.5.1 Step-up procedure

In the step-up procedure, one starts with the full data set and delete sequentially each observation that is erroneous. This procedure is recommended when the data set has at most one outlier because in this case it allows us to avoid unnecessary computation. However, in the presence of more than one outlier the forward deletion procedure possesses some weaknesses; it is insensitive to the masking effect and can cause some inliers to appear as outliers when the true outliers occur at high influence points.

The second type, the backward deletion, begins with a reasonable number of outliers k, and decreases it sequentially. This procedure has the advantage of being less affected by the masking effect. In the following section, we briefly present the backward deletion procedure based on the maximum absolute studentized residuals R_n as defined in the previous chapters.

3.5.2 Backward deletion (Gentleman and Wilk (1975))

We have two different cases; for the first one the number of possible outliers, k, is preassigned and for the second it is unknown.

In the first case the first step consists of identifying a subset of k potential outliers from the full data set. One way of doing this is to choose the subset whose deletion yields the minimum sum of squares of the residuals of the remaining observations. Another way is to select the k observations having large residuals in absolute value from the ordinary least square regression.

If k is not fixed in advance, then the procedure can be performed in the following interesting fashion:

step 1: Order the set S of residuals by absolute value, say s_1, s_2, ..., s_n with $s_i \geq s_j$ for $i > j$.

step2: Determine $s^* = \max |s_{i+1} - s_i|$

step3: Let $c = s_i$ corresponding to s^*

If we let S_* be the subset of absolute values of residuals related to the possible outliers then we can write

$$S_* = \{s_i \in S \setminus s_i > c\}$$

k is therefore the number of observations in S_*.

The interesting part of the procedure performed this way is that we can set an upper bound on k, the number of possible outliers in the data set.

step4: After the subset of k possible outliers has been selected, starting with a sample of size $(n-k)$, one reenters the possible outliers one by one and calculates its corresponding residual $^*r_i$ from $(n-k+1)$ observations.

Let $^*R = \min\{^*r_1, ^*r_2, ..., ^*r_k\}$

If $^*R$ is less than a critical value from Lund's tables of conservative values for the sample size $(n-k+1)$, the corresponding observation is entered in the data set and the process is repeated with the $(n-k+1)$ observations until none of the observations corresponding to the elements in S_*, say s_1, s_2, ..., s_p where $p \leq k$ can be entered.

3.6 Partial-regression leverage plots

Though appearing at the end of the chapter, these plots should be used as preliminary tools in applying the multiple row detection test. They display how any given group of potential outliers can work together to influence the regression parameters and most of the time, allow a clear visualization of the masking or swamping effects.

The partial-regression leverage plot is a device that can be motivated in the following way: Let $W_{(i)}$ be the nx$(q-1)$ submatrix of W obtained by deleting the i-th column w_i. If we denote by u_i and v_i respectively, the residuals of the regression of X on w_i *and* $W_{(i)}$, the i-th regression coefficient of the full model can be found by regressing u_i on v_i (simple regression). The partial-regression leverage plot is therefore the scatter plot of u_i against v_i plus their simple linear regression line. It can be shown that the residuals from this regression line are the same as the residuals from the full model (regression of X on W) and the simple correlation between u_i and v_i is equal to the partial correlation between X and W_i in the multiple regression. These plots are indicated if the rows suspected of being outliers are known.

Following Belsley et al.(1980), the added variable plots are called partial-regression leverage plots when deletion is of interest.

Chapter 4

TRANSFORMATIONS

One of the effective approaches to a variety of problems in regression analysis is the transformation of the data such that the assumptions of the linear regression model are more closely met. Generally, transformations are used for remedy to asymmetry , nonconstant variance and for linearization purpose. To achieve this, several families of transformation of independent or dependant variables are available. The following are the most used:

1) X being the variable concerned with the transformation, the power transformation family used earlier by Tukey (1949) and then by Box and Cox (1964) is defined as follows:

$$X^{(\lambda)} = \begin{cases} (x^\lambda - 1)/\lambda \ , & \lambda \neq 0 \\ \log x \ , & \lambda = 0 \end{cases} \tag{4.1}$$

where $X^{(\lambda)}$ is the transformed variable. This family is of great use when the x's are all positive.

2) The extended power transformation is defined as:

$$X^{(\lambda)} = \begin{cases} [(x + \lambda_2)^{\lambda_1} - 1]/\lambda_1 \ , & \lambda_1 \neq 0 \\ \log(x + \lambda_2) \ , & \lambda_1 = 0 \end{cases} \tag{4.2}$$

Here $(x + \lambda_2) > 0$, the transformation is applicable in the presence of negative values of the x's.

3) Modulus transformations, John and Draper(1980).

$$X^{(\lambda)} = \begin{cases} s_{(x)}[(\mid x \mid +1)^\lambda - 1]/\lambda \ . & \lambda \neq 0 \\ s_{(x)}[\log(\mid x \mid +1)] \ . & \lambda = 0 \end{cases} \tag{4.3}$$

$s_{(x)}$ is the sign of x.

4) Family of folded-power transformations, Mosteller and Tukey (1977) and Atkin-

son (1982)

$$X^{(\lambda)} = \begin{cases} [x^{\lambda} - (b-x)^{\lambda}]/\lambda \ , & \lambda \neq 0 \\ \log[x/(b-x)] & , \ \lambda = 0 \end{cases} \tag{4.4}$$

This family is useful when the x's are constrained to belong to the interval [0,b].

In this section we will focus on Atkinson and Andrews method (see Atkinson (1985) and Andrews (1971)) in addition to the maximum likelihood method for selecting a transformation because of their diagnostic ability .

4.1 TRANSFORMING THE DEPENDANT VARIABLE

4.1.1 MAXIMUM LIKELIHOOD METHOD

Andrews (1971a) demonstrates that the likelihood method for choosing a transformation is sensitive to outlying responses; that is, an outlier in the untransformed data may be brought into line by a transformation.

Let's consider the monotonic power family and suppose that the transformation is indexed by the parameter λ. The object being the normalization of the error distribution, the stabilization of the error variance and straightening the model, one can write

$$\begin{aligned} X_i^{(\lambda)} &= \beta_0 + \beta_1 w_{1i} + \ldots + \beta_q w_{qi} + \varepsilon_i \\ \varepsilon_i &\sim NID(0, \sigma^2) \end{aligned} \tag{4.5}$$

where

$$X_i^{(\lambda)} = \begin{cases} (x_i^{\lambda} - 1)/\lambda \ , & \lambda \neq 0 \\ \log x_i & , \ \lambda = 0 \end{cases} \qquad i = 1, \ldots, n$$

since, the likelihood can be regarded as a function of the transformation parameter λ and the usual regression parameters, say

$$L(\lambda, \beta_0, \beta_1, \ldots, \beta_q, \sigma^2) = (2\pi\sigma)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2}(X^{(\lambda)} - W\beta)^T (X^{(\lambda)} - W\beta) \right\} J \tag{4.6}$$

where

$$J = \Pi_{i=1}^{n} \mid \frac{\partial X_i^{(\lambda)}}{\partial X_i} \mid$$

is the Jacobian taking in account the change of scale due to the transformation.

Let's denote by $\hat{\beta}^{(\lambda)}$ the maximum likelihood estimate of β for the transformed model, we have

$$\hat{\beta}^{(\lambda)} = (W^T W)^{-1} W^T X^{(\lambda)} \tag{4.7}$$

Hence, the residual sum of squares of the $X^{(\lambda)\prime}s$ is given by

$$RSS(\lambda, X) = X^{(\lambda)^T}[I - W(W^T W)^{-1} W^T] X^{(\lambda)}$$

that is

$$RSS(\lambda, X) = X^{(\lambda)^T}[I - V] X^{(\lambda)} \tag{4.8}$$

Dividing (4.8) by n one obtains the maximum likelihood estimate of the variance denoted by $\hat{\sigma}^2_{(\lambda)}$:

$$\hat{\sigma}^2_{(\lambda)} = RSS(\lambda, X)/n \tag{4.9}$$

The log-likelihood is therefore obtained by replacing β and σ^2 in the logarithm of (4.6) by the expressions given by (4.7) and (4.9) respectively; that is

$$L_{\max}(\lambda) = -\frac{1}{2} n \log[RSS(\lambda, X)/n] + \log(J) \tag{4.10}$$

equivalently it can be written as:

$$\begin{aligned}
L_{\max}(\lambda) &= -\frac{1}{2} n \log[RSS(\lambda, X)/n] + \frac{2n}{2n} \log(J) \\
&= -\frac{1}{2} n \log[RSS(\lambda, X)/n] + \frac{1}{2} n \log(J)^{2/n} \\
&= -\frac{1}{2} n \log[\frac{X^{(\lambda)^T}[I - V] X^{(\lambda)}}{n J^{2/n}}] \\
&= -\frac{1}{2} n \log \left\{ \frac{X^{(\lambda)^T}/J^{\frac{1}{n}}[I - V] X^{(\lambda)}/J^{\frac{1}{n}}}{n} \right\} \\
&= -\frac{1}{2} n \log \left[\frac{Z^{(\lambda)^T}(I - V) Z^{(\lambda)}}{n} \right]
\end{aligned}$$

and

$$L \max(\lambda) = -\frac{1}{2} n \log [RSS(\lambda, Z)/n] \tag{4.11}$$

where the n-vector Z has elements

$$z_i^{(\lambda)} = x_i^{(\lambda)}/J^{\frac{1}{n}} \tag{4.12}$$

The usefulness of the normalized transformation (4.12) is that it allows comparisons of residual sums of squares for different values of the transformation parameter

λ.

The last step before the test consists of finding the maximum likelihood estimate of the transformation parameter λ, that is the value $\hat{\lambda}$ for which the profile log-likelihood is maximum. $\hat{\lambda}$ is obtained by finding the solution to

$$\frac{dL_{\max}(\lambda)}{d\lambda} = 0$$

Hence, the likelihood-ratio test statistic for the hypothesis $H_0 : \lambda = \lambda_0$ is given by

$$G_0^2 = -2[L_{\max}(\lambda = \lambda_0) - L_{\max}(\lambda = \hat{\lambda})] \quad (4.13)$$

and G_0^2 is distributed as a χ^2 variable with the number of degrees of freedom v equal to the number of components in λ. Since, one can built a $100(1-\alpha)\%$ confidence region for λ by considering all the values of λ satisfying

$$2[L\max(\hat{\lambda}) - L_{\max}(\lambda)] \leqslant \chi^2_{(v,\alpha)} \quad (4.14)$$

Atkinson (1985) suggests values of λ between -1 *and* 1 and considers the approximate $100(1-\alpha)\%$ confidence region for λ below

$$2[L\max(\hat{\lambda}) - L_{\max}(\lambda)] \leqslant \chi^2_{(1,\alpha)}$$

The null hypothesis of no transformation corresponds to the value λ_0 falling in the above confident region

4.1.2 THE SCORE TESTS AND CONSTRUCTED VARIABLES

Let's consider the normalized model

$$Z^{(\lambda)} = W\beta + \varepsilon \quad (4.15)$$

Expending $Z^{(\lambda)}$ in a Taylor series about λ_0 yields

$$Z^{(\lambda)} = Z^{(\lambda_0)} + (\lambda - \lambda_0)\frac{\partial\ Z^{(\lambda)}}{\partial\lambda}|_{\lambda=\lambda_0}$$

or

$$Z^{(\lambda)} = Z^{(\lambda_0)} + (\lambda - \lambda_0)v^{(\lambda_0)} \quad (4.16)$$

where

$$v^{(\lambda_0)} = \frac{\partial\ Z^{(\lambda)}}{\partial\lambda}|_{\lambda=\lambda_0}$$

is known under the name of the Constructed Variable, Box(1980).

Thus

$$Z^{(\lambda_0)} + (\lambda - \lambda_0)\psi^{(\lambda_0)} = W\beta + \varepsilon$$

and

$$Z^{(\lambda_0)} = W\beta - (\lambda - \lambda_0)\psi^{(\lambda_0)} + \varepsilon$$

$$Z^{(\lambda_0)} = W\beta + \gamma\psi^{(\lambda_0)} + \varepsilon \tag{4.17}$$

with

$$\gamma = -(\lambda - \lambda_0)$$

As in section (2.1) above, the least squares estimate of the constructed variable coefficient is given by

$$\hat{\gamma} = \frac{\psi^{(\lambda_0)T}(I - V)Z^{(\lambda_0)}}{\psi^{(\lambda_0)T}(I - V)\psi^{(\lambda_0)}}$$

$$Var(\hat{\gamma}) = \frac{\psi^{(\lambda_0)T}(I - V)(I - V)\psi^{(\lambda_0)}\sigma^2}{[\psi^{(\lambda_0)T}(I - V)\psi^{(\lambda_0)}]^2}$$

and

$$Var(\hat{\gamma}) = \frac{\sigma^2}{\psi^{(\lambda_0)T}(I - V)\psi^{(\lambda_0)}}$$

In consideration of the residual sum of squares RSS of section 2.4 , the estimate of σ^2 may be found from the expression

$$(n - q - 1)S_z^2 = Z^{(\lambda_0)T}(I - V)Z^{(\lambda_0)} - \frac{[Z^{(\lambda_0)T}(I - V)\psi^{(\lambda_0)}]^2}{\psi^{(\lambda_0)T}(I - V)\psi^{(\lambda_0)}}$$

Hence, the approximate score test for the hypothesis $\gamma = 0$ is given by the following statistic (Atkinson (1973))

$$T_q(\lambda_0) = -\frac{Z^{(\lambda_0)T}(I - V)\psi^{(\lambda_0)}}{S_Z[\psi^{(\lambda_0)T}(I - V)\psi^{(\lambda_0)}]^{1/2}}$$

The negative sign comes from $\gamma = -(\lambda - \lambda_0)$ in (4.17)

under the assumption that the linear approximation leading to (4.17) is exact, one can write

$$\hat{\gamma} = -(\lambda - \lambda_0)$$

Therefore, the test of $\gamma = 0$ is approximately equivalent to the test of the hypothesis $\lambda = \lambda_0$ of no transformation.

Defining the two following variables,

$$\psi^{(\lambda_0)*} = (I - V)\psi^{(\lambda_0)} \;\; and \;\; Z^{(\lambda_0)*} = (I - V)Z^{(\lambda_0)},$$

one can consider the plot of $Z^{(\lambda_0)*}$ against $\psi^{(\lambda_0)*}$ which is in fact the added variable plot. It tells whether the information for the transformation is concerned with all the data or depends on one or a few observations only. In this last case the observations should be given particular attention.

4.1.3 ANDREWS' METHOD

Like previously, the Andrews method is based on the test of the null hypothesis $\lambda = \lambda_0$ with the small difference that Andrews ignores the Jacobian of the transformation. The test is constructed by expending $X^{(\lambda)}$ about λ_0,

$$X^{(\lambda)} = X^{(\lambda_0)} + (\lambda - \lambda_0)\, G_x^{(\lambda_0)}$$

or equivalently

$$X^{(\lambda_0)} = X^{(\lambda)} + (\lambda_0 - \lambda)\, G_x^{(\lambda_0)} \tag{4.18}$$

where

$$G_x^{(\lambda_0)} = \frac{\partial X^{(\lambda)}}{\partial \lambda}\bigg|_{\lambda=\lambda_0}$$

is the constructed variable.

Replacing $X^{(\lambda)}$ by $W\beta + \varepsilon$, (4.18) becomes

$$X^{(\lambda_0)} = W\beta + (\lambda_0 - \lambda)\, G_x^{(\lambda_0)} + \varepsilon \tag{4.19}$$

The Andrews test statistic,which has a standard t-distribution with $(n - q - 1)$ degrees of freedom (Milliken and Graybill (1970)), is equal to the usual t-statistic for the test of $\lambda_0 - \lambda = 0$ in the model

$$X^{(\lambda_0)} = W\beta + (\lambda_0 - \lambda)\, \hat{G}_x^{(\lambda_0)} + \varepsilon$$

where $\hat{G}_x^{(\lambda_0)}$ is equal to $G_x^{(\lambda_0)}$ evaluated at the fitted values from the null model

$$X^{(\lambda_0)} = W\beta + \varepsilon.$$

Chapter 5

TECHNIQUES BASED ON MEASURES OF INFLUENCE

5.1 INTRODUCTION

In this chapter, techniques for the identification of outliers, related to different collinearity-influential diagnostics, are presented; namely the change in the volume of confidence ellipsoid, the Cook-Weisberg statistic, the COVRATIO, the Wilk's Λ , the Andrews-Pregibon statistic and the Likelihood distance. First, let us look at the Variance Inflation Factor as a measure of collinearity. This will lead us to distinguish between different categories of outliers mentioned in the following.

5.1.1 The variance inflation factor(VIF)

The variance inflation factors are diagnostic tools often used to determine at what extent regressors are involved in collinearity. By definition

$$\begin{aligned} VIF_j &= \frac{1}{1 - R^2_{W_j}} \\ &= \frac{\sum_i \tilde{w}^2_{ij}}{RSS_{W_j}} \end{aligned} \tag{5.1}$$

with $R^2_{W_j}$, the coefficient of determination from the regression of W_j on the other regressors. To see how the variance inflation factor is affected by individual observations, we can consider the jth VIF after deletion of the ith case, namely

$$\begin{aligned} VIF_{(i)j} &= \frac{\sum_k \tilde{w}^2_{kj}}{RSS_{W_{(i)j}}},\ for\ k \neq i \\ &= \frac{\sum_i \tilde{w}^2_{ij} - \left(\frac{n}{n-1}\right)\tilde{w}^2_{ij}}{RSS_{W_j} - e^2_{iW_j}/(1 - v^j_{ii})} \end{aligned}$$

$$
\begin{aligned}
&= \frac{\sum_i \tilde{w}_{ij}^2/RSS_{W_j} - \left(\frac{n}{n-1}\right)\tilde{w}_{ij}^2/RSS_{W_j}}{1 - e_{iW_j}^2/RSS_{W_j}(1 - v_{ii}^j)} \\
&= \frac{VIF_j - \left(\frac{n}{n-1}\right)\tilde{w}_{ij}^2/RSS_{W_j}}{(1 - v_{ii})/(1 - v_{ii}^j)} \\
&= VIF_j \left(\frac{1 - v_{ii}^j}{1 - v_{ii}}\right)\left(1 - \left(\frac{n}{n-1}\right)\frac{\tilde{w}_{ij}^2}{RSS_{W_j}VIF_j}\right) \\
&= VIF_j \left(\frac{1 - v_{ii}^j}{1 - v_{ii}}\right)\left(1 - \left(\frac{n}{n-1}\right)\frac{\tilde{w}_{ij}^2}{\sum_i \tilde{w}_{ij}^2}\right) \\
&= VIF_j O_{iW_j} S_{iW_j}^{-1} \qquad (5.2)
\end{aligned}
$$

where v_{ii}^j is the ith diagonal element of the hat matrix with the jth regressor omitted and where

$$
O_{iW_j} = \left(\frac{1 - v_{ii}^j}{1 - v_{ii}}\right) \qquad (5.3)
$$

and

$$
S_{iW_j} = \left(1 - \left(\frac{n}{n-1}\right)\frac{\tilde{w}_{ij}^2}{\sum_i \tilde{w}_{ij}^2}\right)^{-1} \qquad (5.4)
$$

are referred to as the **orientation away** from collinearity due to the *ith* observation and the **scale inflation** due to the *ith* observation respectively. Based on these two factors, one distinguishes between Type **A** and Type **B** observations. For the first category O_{iW_j} will generally be smaller and for the second it will be larger. Extreme type A observations have values of S_{iW_j} larger than O_{iW_j}.

Note that the expression (5.2) above can be written as

$$
\frac{VIF_{(i)j}}{VIF_j} = \frac{O_{iW_j}}{S_{iW_j}} \qquad (5.5)
$$

given that $S_{iW_j} \geq 1$ and $O_{iW_j} \geq 1; \frac{VIF_{(i)j}}{VIF_j} \geq 1$ if and only if $O_{iW_j} \geq S_{iW_j}$, that is if the *ith* observation is of type B.

In conclusion, type A observation inflates the VIF ($O_{iW_j} < S_{iW_j}$) and type B observation deflates the VIF ($O_{iW_j} > S_{iW_j}$). Observations for which $O_{iW_j} \approx S_{iW_j}$ are referred to as type **AB**.

5.1.2 Definition 5.1.1

Type A outliers refer to outliers that inflate the VIF, that is outliers for which one has either

$$
O_{iW_j} < S_{iW_j}
$$

or

$$v_{ii} - v_{ii}^{j} < (1 - v_{ii})(S_{iW_j} - 1) \tag{5.6}$$

5.1.3 Definition 5.1.2

A type B outlier deflates the VIF, this means that either

$$O_{iW_j} > S_{iW_j}$$

or

$$v_{ii} - v_{ii}^{j} > (1 - v_{ii})(S_{iW_j} - 1) \tag{5.7}$$

holds.

5.1.4 Definition 5.1.3

Type AB outliers refer to outliers for which one has either

$$O_{iW_j} \approx S_{iW_j}$$

or

$$v_{ii} - v_{ii}^{j} \approx (1 - v_{ii})(S_{iW_j} - 1) \tag{5.8}$$

5.1.5 VIF BASED PLOTS

This diagnostic plot consists of plotting in the same set of axes, the orientation away factor O_{iX_j} and the scale factor S_{iX_j}. Observations lying far above and below the 45 degrees line through the origin, can be regarded as type B and type A outliers respectively. In fact, type B outliers are such that $O_{iX_j} > S_{iX_j}$ and any observation belonging to this group lies above the 45 degrees line $O_{iX_j} = S_{iX_j}$. The same, type A outliers satisfy $O_{iX_j} < S_{iX_j}$ and observations from this group lie below the 45 degrees line $O_{iX_j} = S_{iX_j}$.

A variant of this plot, accommodating the change in the VIF due to a specific observation, is the $\varphi_{iX_j} = \frac{VIF_{(i)j} - VIF_j}{VIF_j}$ versus v_i. In other words, it is the plot of $\frac{O_{iX_j}}{S_{iX_j}} - 1$ versus v_i. Type B outliers lie far above the zero line and type A below.

5.2 Changes in the volume of confidence ellipsoid

The general equation for an ellipsoid G is given by Seber (1977) as

$$(\Psi - \Psi^*)^T L^{-1} (\Psi - \Psi^*) \leq m \tag{5.9}$$

where Ψ is the kernel of the ellipsoid and L some positive definite matrix.

The volume of this ellipsoid is proportional to the inverse of the square root of the matrix L, that is

$$Vol(G) \propto |L|^{1/2} \tag{5.10}$$

That being the case, let's write W in the partitioned form

$$W = (W_1\ W_2) \tag{5.11}$$

where W_1 is an nx$(q-p)$ and W_2 an nxp matrix. The last p components of the coefficient vector β are of interest in this case. Thus, one can define

$$C = (\mathbf{0}\ I_p) \tag{5.12}$$

where C, an pxq matrix, is such that

$$\Phi = C\beta \tag{5.13}$$

is the coefficient vector of interest, and its corresponding least squares estimator

$$\hat{\Phi} = C\hat{\beta} \tag{5.14}$$

A $(1-\alpha)$x100% confidence ellipsoid for Φ, based on the estimate $\hat{\Phi}$, is given by all Φ^* such that

$$\{E : \Phi^* \mid (\Phi^* - \hat{\Phi})^T (C(W^T W)^{-1} C^T)^{-1} (\Phi^* - \hat{\Phi}) \leq ps^2 F_{1-\alpha}(p, n-q)\} \tag{5.15}$$

or

$$\{E : \Phi^* \mid (\Phi^* - \hat{\Phi})^T [\frac{1}{ps^2 F_{1-\alpha}(p, n-q)} (C(W^T W)^{-1} C^T)^{-1}] (\Phi^* - \hat{\Phi}) \leq 1\}$$

If only one regressor is left out, this expression becomes

$$\{E : \Phi^* \mid (\Phi^* - \hat{\Phi})^T [\frac{1}{(q-1)s^2 F_{1-\alpha}(q-1, n-q)} (C(W^T W)^{-1} C^T)^{-1}] (\Phi^* - \hat{\Phi}) \leq 1\} \tag{5.16}$$

and the situation where one case is deleted follows from (5.7), that is

$$\{E_{(i)} : \Phi^* \mid (\Phi^* - \hat{\Phi}_{(i)})^T [\frac{1}{(q-1)s_{(i)}^2 F_{1-\alpha}(q-1, n-q-1)} (C(W_{(i)}^T W_{(i)})^{-1} C^T)^{-1}] (\Phi^* - \hat{\Phi}_{(i)}) \leq 1\} \tag{5.17}$$

Like in (5.2), one can write

$$vol(E) \propto [(q-1)s^2 F_{1-\alpha}(q-1, n-q)]^{(q-1)/2} \left| C(W^T W)^{-1} C^T \right|^{1/2} \tag{5.18}$$

and

$$vol(E_{(i)}) \propto [(q-1)s_{(i)}^2 F_{1-\alpha}(q-1, n-q-1)]^{(q-1)/2} \left| C(W_{(i)}^T W_{(i)})^{-1} C^T \right|^{1/2} \tag{5.19}$$

Thus, the ratio of (5.11) and (5.10) can be obtained, namely

$$\begin{aligned} \frac{vol(E_{(i)})}{vol(E)} &= \frac{[(q-1)s_{(i)}^2 F_{1-\alpha}(q-1, n-q-1)]^{(q-1)/2} \left| C(W_{(i)}^T W_{(i)})^{-1} C^T \right|^{1/2}}{[(q-1)s^2 F_{1-\alpha}(q-1, n-q)]^{(q-1)/2} \left| C(W^T W)^{-1} C^T \right|^{1/2}} \\ &= \left(\frac{s_{(i)}^2 F_{1-\alpha}(q-1, n-q-1)}{s^2 F_{1-\alpha}(q-1, n-q)} \right)^{(q-1)/2} \frac{\left| C(W_{(i)}^T W_{(i)})^{-1} C^T \right|^{1/2}}{\left| C(W^T W)^{-1} C^T \right|^{1/2}} \end{aligned} \tag{5.20}$$

Given that

$$\begin{aligned} \frac{s_{(i)}^2}{s^2} &= \frac{(n-q) - e_i^2/(1-v_{ii})s^2}{n-q-1} \\ &= \frac{n-q-r_i^2}{n-q-1} \end{aligned}$$

(5.20) becomes

$$\frac{vol(E_{(i)})}{vol(E)} = \left(\frac{F_{1-\alpha}(q-1, n-q-1)}{F_{1-\alpha}(q-1, n-q)} \text{x} \frac{(n-q-r_i^2)}{(n-q-1)} \right)^{(q-1)/2} \frac{\left| C(W_{(i)}^T W_{(i)})^{-1} C^T \right|^{1/2}}{\left| C(W^T W)^{-1} C^T \right|^{1/2}} \tag{5.21}$$

Special cases:

- m=1, p=q and C=I.

This case corresponds to the deletion of one observation from the full model and since C=I leads

$$\frac{\left|C(W_{(i)}^T W_{(i)})^{-1}C^T\right|^{1/2}}{|C(W^TW)^{-1}C^T|^{1/2}} = \frac{\left|(W_{(i)}^T W_{(i)})^{-1}\right|^{1/2}}{|(W^TW)^{-1}|^{1/2}} = \left(\frac{1}{1-v_{ii}}\right)^{1/2} \tag{5.22}$$

(5.21) reduces to

$$\frac{vol(E_{(i)})}{vol(E)} = \left(\frac{F_{1-\alpha}(q-1,n-q-1)}{F_{1-\alpha}(q-1,n-q)} \text{x} \frac{(n-q-r_i^2)}{(n-q-1)}\right)^{q/2} \left(\frac{1}{1-v_{ii}}\right)^{1/2} \tag{5.23}$$

- The j th regressor and the i th observation are omitted.

The volume ratio in this case is given by

$$\begin{aligned}\frac{vol(E_{(i)})}{vol(E)} &= \frac{\left((q-1)\, s_{(i)}^2 F_{1-\alpha}(q-1,n-q-1)\right)^{(q-1)/2} \left|C(W_{(i)}^T W_{(i)})^{-1}C^T\right|^{1/2}}{((q-1)\, s^2 F_{1-\alpha}(q-1,n-q))^{(q-1)/2} \, |C(W^TW)^{-1}C^T|^{1/2}} \\ &= \left(\frac{s_{(i)}^2 F_{1-\alpha}(q-1,n-q-1)}{s^2 F_{1-\alpha}(q-1,n-q)}\right)^{(q-1)/2} \left(\frac{1-\frac{w_{ij}^2}{\sum w_{ij}^2}}{1-v_{ii}}\right)^{1/2} \\ &= \left(\frac{F_{1-\alpha}(q-1,n-q-1)}{F_{1-\alpha}(q-1,n-q)} \text{x} \frac{(n-q-r_i^2)}{(n-q-1)}\right)^{(q-1)/2} \left(\frac{1-\frac{w_{ij}^2}{\sum w_{ij}^2}}{1-v_{ii}}\right)^{1/2} \end{aligned} \tag{5.24}$$

- An observation and the intercept are omitted

This case corresponds to C=(1 **0**). therefore one has $CW = \mathbf{1}$ and $w_{ij}^2 = 1$ and $\sum w_{ij}^2 = n$. Hence, the volume ratio can be written as

$$\frac{vol(E_{(i)})}{vol(E)} = \left(\frac{(n-q-r_i^2)F_{1-\alpha}(q-1,n-q-1)}{(n-q-1)F_{1-\alpha}(q-1,n-q)}\right)^{(q-1)/2} \left(\frac{1-\frac{1}{n}}{1-v_{ii}}\right)^{1/2} \tag{5.25}$$

These expressions have the advantage of allowing one to write down separately two factors, the first one related to outliers in the dependent space and the second one to outliers in the regressor space.

5.3 The Cook-Weisberg statistic

The Cook-Weisberg statistic is based on the change in the volume of confidence ellipsoid. It is defined as the logarithm of the ratio obtained in the previous section. For simplicity, we consider the following two confidence ellipsoids associated with (5.16) and (5.17) respectively

$$\{E : \beta^* \mid (\beta^* - \hat{\beta})^T (W^T W)(\beta^* - \hat{\beta}) < q s^2 F_{1-\alpha}(q, n-q)\} \tag{5.26}$$

and

$$\{E_{(i)} : \beta^* \mid (\beta^* - \hat{\beta}_{(i)})^T (W_{(i)}^T W_{(i)})(\beta^* - \hat{\beta}_{(i)}) < q s_{(i)}^2 F_{1-\alpha}(q, n-q-1)\} \tag{5.27}$$

Cook and Weisberg (1982) consider the ratio

$$\begin{aligned} \log \frac{Vol(E)}{Vol(E_{(i)})} &= \log\{\frac{\left|W_{(i)}^T W_{(i)}\right|^{1/2}}{|W^T W|^{1/2}} \text{x} [\frac{s^2 F_{1-\alpha}(q, n-q)}{s_{(i)}^2 F_{1-\alpha}(q, n-q-1)}]^{q/2}\} \\ &= \frac{1}{2}\log(1 - v_{ii}) + \frac{q}{2}\log\{\frac{n-q-1}{n-q-r_i^2} \text{x} \frac{F_{1-\alpha}(q, n-q)}{F_{1-\alpha}(q, n-q-1)}\} \end{aligned} \tag{5.28}$$

Following the same authors, if the i-th case is an outlier, its deletion may result in a large and positive quantity of the above ratio.

For the more general result of the above, let's write W in the following partitioned form

$$W = (W_1,\ W_2)$$

where W_1 is an nx$(q-p)$ and W_2 an nxp matrix. The last p components of the coefficient vector β are of interest in this case.

The change in the volume for estimating Φ, a subset of p components of the β vector, after deletion of m cases indexed by I, is given by

$$\log\{\frac{vol(E, \Phi)}{vol(E_{(I)}, \Phi)}\} = \frac{1}{2}\log\{\frac{|I - V_I|}{|I - U_I|}\} + \frac{p}{2}\log\{\frac{(n-q-m)}{(n-q-\sum_{j=1}^{m} h_j^2)} \text{x} \frac{F_{1-\alpha}(q, n-q)}{F_{1-\alpha}(q, n-q-m)}\} \tag{5.29}$$

where

h_j is as defined in (3.10). V_I and U_I are submatrices of V and U indexed by I.

with

$$U = W_1(W_1^T W_1)^{-1} W_1^T$$

If we let m=1. (5.29) reduces to (5.28) above.

5.4 The COVRATIO

The COVRATIO is a measure of influence based on the change of the covariance matrix $\hat{\beta}$ due to the deletion of one observation. Following Belsley, Kuh and Welsch (1980)

$$\begin{aligned} COVRATIO &= \left(\frac{s_{(i)}^2}{s^2}\right)^q \frac{\left|(W_{(i)}^T W_{(i)})^{-1}\right|}{\left|(W^T W)^{-1}\right|} \\ &= \left(\frac{s_{(i)}^2}{s^2}\right)^q \frac{\left|(W^T W)\right|}{\left|(W_{(i)}^T W_{(i)})\right|} \\ &= \left(\frac{n-q-r_i^2}{n-q-1}\right)^q \left(\frac{1}{1-v_{ii}}\right) \end{aligned} \quad (5.30)$$

Hence, cases with large values of COVRATIO should be considered as outliers in the regressor space, since high leverage observations will result in an increase in COVRATIO while cases resulting in a decrease in COVRATIO should be regarded as potential outliers in the dependent space since large values of r_i^2 decrease the factor $\left(\frac{n-q-r_i^2}{n-q-1}\right)^q$

5.5 The Wilks' Λ statistic

The Wilks'Λ statistic is another test, useful for discovering groups of outliers in the regression analysis. Let Z be the nx$(q+1)$ matrix formed by adjoining the vector X to the W matrix, that is

$$Z = [W\ X]$$

On the geometric point of view, each row of Z can be regarded as an observation in the $(q+1)$ dimensional space.

If we let $\widetilde{Z}$ denote the centered Z matrix, The Wilks'Λ statistic is defined as

$$\Lambda(\tilde{z}_i) = \frac{Det[\tilde{Z}^T\tilde{Z} - (n-1)\bar{\tilde{z}}_{(i)}^T\bar{\tilde{z}}_{(i)} - \tilde{z}_i^T\tilde{z}_i]}{Det[\tilde{Z}^T\tilde{Z}]} \quad (5.31)$$

where Det symbolizes the determinant and $\bar{\tilde{z}}_{(i)}$ the q-vector (row) of column means of $\tilde{Z}_{(i)}$. Applying the formulas for adding a column to a matrix, Rao(1973) shows that the above expression reduces to

$$\Lambda(\tilde{z}_i) = \frac{n}{n-1}(1-v_{ii})[1+\frac{r_i^2}{n-q-1}]^{-1} \quad (5.32)$$

Let D_m denote the set (of size m) of indexes of the rows to be deleted, 1_1 an nx1 vector consisting of ones for rows contained in D_m and zeros otherwise, 1 an n-vector of ones and $1_2 = 1 - 1_1$. The statistic for testing whether a group of m observations (indexed by D_m) is an outlying group is given by [Rao(1973)]

$$\Lambda(D_m) = \frac{Det[\tilde{Z}^T\tilde{Z} - \frac{1}{m}\tilde{Z}^T 1_1 1_1^T \tilde{Z} - \frac{1}{n-m}\tilde{Z}^T 1_2 1_2^T \tilde{Z}]}{Det[\tilde{Z}^T\tilde{Z}]}$$

which, following (5.32), reduces to

$$\Lambda(D_m) = 1 - \frac{n}{m(n-m)}[1_1^T\tilde{Z}(\tilde{Z}^T\tilde{Z})^{-1}1_1\tilde{Z}^T] \quad (5.33)$$

this statistic is related to the F-statistic, that is

$$\left(\frac{n-q-1}{q}\right)\left(\frac{1-\Lambda(D_m)}{\Lambda(D_m)}\right) \sim F(q, n-q-1)$$

Small values of (5.33) indicate possible discrepant observations that should be given particular attention.

5.6 The Andrews-Pregibon statistic

Adrews and Pregibon (1978) propose a similar statistic whose distribution (in the case where X is Gaussian and W fixed) has been developed: it also provides significance levels for finding sets of outliers for small values of n.

The Andrews-Pregibon Q-statistic is defined as

$$Q(D_m) = \frac{Det[Z_{(D_m)}^T Z_{(D_m)}]}{Det[\check{Z}^T\check{Z}]}$$

$$= \frac{(n-q-m)S^2_{(D_m)}Det[W^T_{(D_m)}W_{(D_m)}]}{(n-q)S^2 Det[W^TW]} \tag{5.34}$$

where D_m and Z are as in the previous section.

In the case of a single case deletion, the Andrews-Pregibon statistic is defined as

$$Q(D_1) = \frac{Det[Z^T_{(i)}Z_{(i)}]}{Det[Z^TZ]} \tag{5.35}$$

Since

$$Det[Z^TZ] = Det[W^TW]RSS_X \tag{5.36}$$

and

$$Det[Z^T_{(i)}Z_{(i)}] = Det[W^T_{(i)}W_{(i)}]RSS_{X_{(i)}} \tag{5.37}$$

where RSS_X and $RSS_{X_{(i)}}$ are the sum of squares of residuals of X and X with the i-th case deleted respectively.

Expression (5.35) becomes

$$\begin{aligned} Q(D_1) &= \frac{Det[Z^T_{(i)}Z_{(i)}]}{Det[Z^TZ]} \\ &= \frac{Det[W^T_{(i)}W_{(i)}]}{Det[W^TW]}\frac{RSS_{X_{(i)}}}{RSS_X} \\ &= (1-v_{ii})\left(1 - \frac{e_i^2}{RSS_X(1-v_{ii})}\right) \\ &= (1-v_{ii})\left(1-\frac{r_i^2}{n-q}\right) \end{aligned} \tag{5.38}$$

The Andrews-Pregibon statistic is more likely to identify outliers in the regressor space than the dependent space since r_i^2 is very small compared to its denominator $(n-q)$.

In practice, some authors make use of the following transformation

$$\begin{aligned} Q^*(D_1) &= -\frac{1}{2}\log(Q(D_1)) \\ &= -\frac{1}{2}\log(1-v_{ii}) + \frac{1}{2}\log\left(\frac{n-q}{n-q-r_i^2}\right) \end{aligned} \tag{5.39}$$

5.7 Influence measures based on likelihood distances

Let write the model (2.1) as follows

$$X = W\beta + \sigma\varepsilon \tag{5.40}$$

with $\varepsilon_i \sim iid\ N(0,1)$ and σ^2 is **known**.

Let also $L(\beta)$ and $l(\beta)$ denote respectively the likelihood function of β and its logarithm. The expression of the last is given by

$$l\left(\beta\right) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{(X - W\beta)^T(X - W\beta)}{2\sigma^2} \tag{5.41}$$

with respect to the earlier notation, the likelihood distance is defined by Cook and Weisberg (1982) as

$$ld_i = 2\left(l\left(\hat{\beta}\right) - l\left(\hat{\beta}_{(i)}\right)\right) \tag{5.42}$$

where $l\left(\hat{\beta}\right)$ is the log-likelihood evaluated at $\hat{\beta}$, namely

$$l\left(\hat{\beta}\right) = -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{RSS}{2\sigma^2} \tag{5.43}$$

the same for $l\left(\hat{\beta}_{(i)}\right)$ but with th i-th case deleted, that is

$$\begin{aligned} l\left(\hat{\beta}_{(i)}\right) &= -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{(X - \hat{X}_{(i)})^T(X - \hat{X}_{(i)})}{2\sigma^2} \\ &= -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{RSS}{2\sigma^2} - \frac{v_{ii}e_i^2}{2(1 - v_{ii})^2\sigma^2} \end{aligned} \tag{5.44}$$

the likelihood distance reduces therefore to

$$ld_i = \frac{v_{ii}e_i^2}{2(1 - v_{ii})^2\sigma^2} \tag{5.45}$$

Following Cook and Weisberg (1982) the maximum likelihood over the parameter space for θ_2 with θ_1 fixed is given by

$$l\left(\theta_1, \theta_2(\theta_1)\right) = \max_{\theta_2}\left(l(\theta_1, \theta_2)\right) \tag{5.46}$$

and the likelihood distance when the ith case is deleted is

$$ld_i(\theta_1 \mid \theta_2) = 2\left(l(\hat{\theta}) - \max_{\theta_2} l(\theta_{1(i)}, \theta_2)\right) \tag{5.47}$$

where $l(\hat{\theta}) = l(\hat{\theta}_1, \hat{\theta}_2)$.

Applying the above to the model (5.40) with $\theta_1 = \beta$ and $\theta_2 = \sigma^2$ (**unknown but not of interest**), one obtains

$$l(\hat{\beta}, \hat{\sigma}^2) = -\frac{n}{2}\log(2\pi\hat{\sigma}^2) - \frac{n}{2} \tag{5.48}$$

where

$$\hat{\sigma}^2 = \frac{(X - W\hat{\beta})^T(X - W\hat{\beta})}{n} \tag{5.49}$$

the maximum likelihood estimate of σ^2, is function of β . Hence, following (5.42) and (5.47), the likelihood distance is

$$ld_i(\beta \mid \sigma^2) = 2(l(\hat{\beta}, \hat{\sigma}^2) - \max_{\sigma^2}(l(\hat{\beta}_{(i)}, \sigma^2))) \tag{5.50}$$

with

$$\begin{aligned} l(\hat{\beta}_{(i)}, \sigma^2) &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma_*^2 - \frac{1}{2\sigma_*^2}\sum_{k=1}^{n}(x_k - w_k'\hat{\beta}_{(i)})^2 \\ &= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma_*^2 - \frac{1}{2\sigma_*^2}RSS_{X_{(i)}} \end{aligned} \tag{5.51}$$

where σ_*^2, the value of σ^2 that maximizes $l(\hat{\beta}_{(i)}, \sigma^2)$ is given by

$$\begin{aligned} \hat{\sigma}_*^2 &= \frac{1}{n}(X - \hat{X}_{(i)})^T(X - \hat{X}_{(i)}) \\ &= \frac{1}{n}\sum_{k=1}^{n}(x_k - w_k'\hat{\beta}_{(i)})^2 \\ &= \frac{1}{n}RSS_{X_{(i)}} \end{aligned} \tag{5.52}$$

Expression (5.50) becomes

$$\begin{aligned} ld_i(\beta \mid \sigma^2) &= 2(-\frac{n}{2}\log(2\pi\hat{\sigma}^2) - \frac{n}{2} - (-\frac{n}{2}\log(2\pi\sigma_*^2) - \frac{n}{2})) \\ &= n\log(\frac{\sigma_*^2}{\hat{\sigma}^2}) \end{aligned}$$

$$
\begin{aligned}
&= n\log\left(\frac{(X-\hat{X}_{(i)})^T(X-\hat{X}_{(i)})/n}{(X-W\hat{\beta})^T(X-W\hat{\beta})/n}\right) \\
&= n\log\left(\frac{RSS_X+\frac{v_{ii}e_i^2}{(1-v_{ii})^2}}{RSS_X}\right) \\
&= n\log\left(1+\frac{q}{n-q}D_i\right) \qquad (5.53)
\end{aligned}
$$

where D_i is the Cook's distance.

5.7.1 The likelihood distance when both β and σ^2 are of interest

In the former cases, interest was only on the vector β; in the following the likelihood distance is reviewed when both β and σ^2 are of interest.

As in (5.48), the log-likelihood is given by

$$
l(\hat{\beta},\hat{\sigma}^2) = -\frac{n}{2}\log(2\pi\hat{\sigma}_L^2) - \frac{n}{2} \qquad (5.54)
$$

where

$$
\hat{\sigma}_L^2 = \left(\frac{n-q}{n}\right)\hat{\sigma}_*^2 \qquad (5.55)
$$

Thus, (5.54) can be written as

$$
l(\hat{\beta},\hat{\sigma}^2) = -\frac{n}{2}\log\left(2\pi\left(\frac{n-q}{n}\right)\hat{\sigma}_*^2\right) - \frac{n}{2} \qquad (5.56)
$$

Like previously, after deletion of the ith observation this expression becomes

$$
\begin{aligned}
l(\hat{\beta}_{(i)},\hat{\sigma}_{(i)}^2) &= -\frac{n}{2}\log(2\pi\hat{\sigma}_{(i)}^2) - \frac{(X-W\hat{\beta}_{(i)})^T(X-W\hat{\beta}_{(i)})}{2\hat{\sigma}_{(i)}^2} \\
&= -\frac{n}{2}\log(2\pi\hat{\sigma}_{(i)}^2) - \frac{RSS^*_{X_{(i)}}}{2\hat{\sigma}_{(i)}^2} \qquad (5.57)
\end{aligned}
$$

where

$$
\begin{aligned}
RSS^*_{X_{(i)}} &= RSS + \frac{v_{ii}e_i^2}{1-v_{ii}} \\
&= RSS_{X_{(i)}} + \frac{e_i^2}{1-v_{ii}} + \frac{v_{ii}e_i^2}{(1-v_{ii})^2}
\end{aligned}
$$

$$
\begin{aligned}
&= RSS_{X_{(i)}} + \frac{e_i^2}{(1-v_{ii})^2} \\
&= (n-q-1)\hat{\sigma}_{(i)}^2 + \frac{\hat{\sigma}_{(i)}^2 t_i^2}{1-v_{ii}} \qquad (5.58)
\end{aligned}
$$

Replacing the likelihood estimator $\hat{\sigma}_{(i)}^2$ by $\left(\frac{n-q-1}{n-1}\right)\hat{\sigma}_{*(i)}^2$, one has

$$
\begin{aligned}
l(\hat{\beta}_{(i)}, \hat{\sigma}_{*(i)}^2) &= -\frac{n}{2}\log\left(2\pi\left(\frac{n-q-1}{n-1}\right)\hat{\sigma}_{*(i)}^2\right) - \frac{(n-q-1)\,\hat{\sigma}_{*(i)}^2}{2\left(\frac{n-q-1}{n-1}\right)\hat{\sigma}_{*(i)}^2} \\
&\quad - \frac{\hat{\sigma}_{*(i)}^2 t_i^2/(1-v_{ii})}{2\left(\frac{n-q-1}{n-1}\right)\hat{\sigma}_{*(i)}^2} \\
&= -\frac{n}{2}\log\left(2\pi\left(\frac{n-q-1}{n-1}\right)\hat{\sigma}_{*(i)}^2\right) - \left(\frac{n-1}{2}\right) \\
&\quad - \frac{(n-1)\,t_i^2}{2\,(n-q-1)\,(1-v_{ii})} \qquad (5.59)
\end{aligned}
$$

Using (5.56) and (5.59), Cook and Weisberg (1982) provide the following likelihood distance

$$
\begin{aligned}
ld_i(\beta, \sigma^2) &= 2\{-\frac{n}{2}\log\left(2\pi\left(\frac{n-q}{n}\right)\hat{\sigma}_*^2\right) - \frac{n}{2} + \frac{n}{2}\log 2\pi\left(\frac{n-q-1}{n-1}\right)\hat{\sigma}_{*(i)}^2 \\
&\quad + \left(\frac{n-1}{2}\right) + \frac{(n-1)\,t_i^2}{2\,(n-q-1)\,(1-v_{ii})}\} \\
&= n\log\left((\frac{n-1}{2})(\frac{n-q-1}{n-q})(\frac{n-q}{n-q-1+t_i^2})\right) - 1 + \frac{(n-1)\,t_i^2}{(n-q-1)\,(1-v_{ii})} \\
&= n\log\left((\frac{n}{n-1})(\frac{n-q-1}{n-q-1+t_i^2})\right) + \frac{(n-1)\,t_i^2}{(n-q-1)\,(1-v_{ii})} - 1 \qquad (5.60)
\end{aligned}
$$

For models with a constant term, the above likelihood distance is an increasing function in t_i^2. Thus, outliers in the dependent space can easily be identify by means of (5.60) . In the same expression, leverage values play an important role in identifying outliers in the independent space.

5.8 Influence measures based on the volume of confidence ellipsoids for the secondary regression

Consider the following secondary regression fit

$$
\hat{W}_j = W_{[j]}\hat{\alpha}_j \qquad (5.61)
$$

A $(1-\alpha)$x100% confidence ellipsoid, based on $\hat{\alpha}_j$, is given by

$$E(\alpha_j) = \{\alpha_j^* \mid (\alpha_j^* - \hat{\alpha}_j)^T (W_{[j]}^T W_{[j]})(\alpha_j^* - \hat{\alpha}_j) \leq (q-1)\hat{\sigma}^2_{W_j} F_{1-\alpha}(q-1, n-q+1) \quad (5.62)$$

Like previously, after the i-th observation has been deleted, the $(1-\alpha)$x100% confidence ellipsoid for α_j is given by

$$E_{(i)}(\alpha_j) = \{\alpha_j^* \mid (\alpha_j^* - \hat{\alpha}_{(i)j})^T (W_{(i)[j]}^T W_{(i)[j]})(\alpha_j^* - \hat{\alpha}_{(i)j}) \leq (q-1)\hat{\sigma}^2_{(i)W_j} F_{1-\alpha}(q-1, n-q) \quad (5.63)$$

The volume of these confident ellipsoids are proportional to the square root of the determinant of the inverses of $W_{(j)}^T W_{(j)}$ and $W_{(i)[j]}^T W_{(i)[j]}$ (in the last case, the first subscript indicates the observation that has been deleted and the second is related to the regressor that is being treated as the dependent variable in the secondary regression model). Namely we have

$$\begin{aligned} Vol[E(\alpha_j)] &\propto \left| \frac{1}{(q-1)\hat{\sigma}^2_{W_j} F_{1-\alpha}(q-1, n-q+1)} (W_{[j]}^T W_{[j]}) \right|^{-1/2} \\ &= [(q-1)\hat{\sigma}^2_{W_j} F_{1-\alpha}(q-1, n-q+1)]^{(q-1)/2} \left|(W_{[j]}^T W_{[j]})\right|^{-1/2} \quad (5.64) \end{aligned}$$

and

$$\begin{aligned} Vol[E_{(i)}(\alpha_j)] &\propto \left| \frac{1}{(q-1)\hat{\sigma}^2_{(i)W_j} F_{1-\alpha}(q-1, n-q)} (W_{(i)[j]}^T W_{(i)[j]}) \right|^{-1/2} \\ &= [(q-1)\hat{\sigma}^2_{(i)W_j} F_{1-\alpha}(q-1, n-q)]^{(q-1)/2} \left|W_{(i)[j]}^T W_{(i)[j]}\right|^{-1/2} \quad (5.65) \end{aligned}$$

The ratio of (5.63) to (5.62) gives the change related to the deletion of the ith observation in the volume of secondary confidence ellipsoid, that is

$$\begin{aligned} \frac{Vol[E_{(i)}(\alpha_j)]}{Vol[E(\alpha_j)]} &= \left(\frac{\hat{\sigma}^2_{(i)W_j}}{\hat{\sigma}^2_{W_j}} \right)^{(q-1)/2} \left(\frac{F_{1-\alpha}(q-1, n-q)}{F_{1-\alpha}(q-1, n-q+1)} \right)^{(q-1)/2} \left(\frac{\left|(W_{(i)[j]}^T W_{(i)[j]})\right|}{\left|(W_{[j]}^T W_{[j]})\right|} \right)^{-1/2} \\ &= \left(\frac{RSS_{W_{(i)j}}/(n-(q-1)-1)}{RSS_{W_j}/(n-(q-1))} \right)^{(q-1)/2} \left(\frac{F_{1-\alpha}(q-1, n-q)}{F_{1-\alpha}(q-1, n-q+1)} \right)^{(q-1)/2} \\ &\quad * \left(\frac{\left|(W_{(i)(j)}^T W_{(i)(j)})\right|}{\left|(W_{(j)}^T W_{(j)})\right|} \right)^{-1/2} \quad (5.66) \end{aligned}$$

$$= \left(\frac{n-q+1}{n-q} * \frac{RSS_{W_{(i)j}}}{RSS_{W_j}}\right)^{(q-1)/2} \left(\frac{F_{1-\alpha}(q-1,n-q)}{F_{1-\alpha}(q-1,n-q+1)}\right)^{(q-1)/2}$$
$$* \left(\frac{\left|(W^T_{(i)(j)}W_{(i)(j)})\right|}{\left|(W^T_{(j)}W_{(j)})\right|}\right)^{-1/2}$$

Since the residual sum of squares with the ith can be expressed as

$$\begin{aligned} RSS_{W_{(i)j}} &= RSS_{W_j} - \frac{e_i^2}{1-v_{ii}^j} \\ &= RSS_{W_j}\left(\frac{1-v_{ii}}{1-v_{ii}^j}\right) \\ &= RSS_{W_j}O_{iW_j}^{-1} \end{aligned} \tag{5.67}$$

the above expression becomes

$$\frac{Vol[E_{(i)}(\alpha_j)]}{Vol[E(\alpha_j)]} = \left(\frac{n-q+1}{n-q}O_{iW_j}^{-1}\frac{F_{1-\alpha}(q-1,n-q)}{F_{1-\alpha}(q-1,n-q+1)}\right)^{(q-1)/2}\left(1-v_{ii}^j\right)^{-1/2} \tag{5.68}$$

The Cook-Weisberg statistic for the secondary regression model can easily be derived from this expression by means of taking the logarithm over its right hand side, that is

$$CW_i\left(E(\alpha_j)\right) = -\frac{1}{2}\log\left(1-v_{ii}^j\right) - \frac{q-1}{2}\log O_{iW_j} - \frac{q-1}{2}\log\left(\frac{(n-q)F_{1-\alpha}(q-1,n-q+1)}{(n-q-1)F_{1-\alpha}(q-1,n-q)}\right) \tag{5.69}$$

$CW_i\left(E(\alpha_j)\right)$ is an increasing function of v_{ii}^j.

For simplicity of interpretation, some authors suggest the following adaptation

$$\Delta CW_i\left(E(\alpha_j)\right) = CW_i^*\left(E(\alpha_j)\right) - CW_i\left(E(\alpha_j)\right) \tag{5.70}$$

where $CW_i^*\left(E(\alpha_j)\right)$ is the Cook-Weisberg statistic for secondary regression model, evaluated at the following:

1 $v_{ii}^j = 1-(1-v_{ii})S_{iW_j}$

where

$$S_{iW_j} = \left(1 - \frac{n}{n-1}\frac{\tilde{w}_{ij}^2}{\sum_i \tilde{w}_{ij}^2}\right)^{-1}$$

is the scale inflation due to the *ith* observation. In this expression $\tilde{w}_{ij}$ is the *i, jth* component of the centered matrix $\tilde{W}$.

2 $v_{ii}^j = 1-(1-v_{ii})^{1/3}$ (see appendix!!!)

Let consider first the case where $v_{ii}^j = 1-(1-v_{ii})S_{iW_j}$ and $O_{iX_j} = S_{iW_j}$, expression (5.69) becomes

$$\begin{aligned} CW_i^{*(1)}\left(E(\alpha_j)\right) &= -\frac{1}{2}\log\left((1-v_{ii})S_{iW_j}\right) - \left(\frac{q-1}{2}\right)\log S_{iW_j} - K \\ &= -\frac{1}{2}\log(1-v_{ii}) - \frac{1}{2}\log(S_{iW_j}) - \left(\frac{q-1}{2}\right)\log S_{iW_j} - K \\ &= -\frac{1}{2}\log(1-v_{ii}) - \frac{q}{2}\log(S_{iW_j}) - K \qquad (5.71) \end{aligned}$$

where

$$K = \frac{q-1}{2}\log\left(\frac{(n-q)F_{1-\alpha}(q-1,n-q+1)}{(n-q-1)F_{1-\alpha}(q-1,n-q)}\right) \qquad (5.72)$$

And when $v_{ii}^j = 1-(1-v_{ii})^{1/3}$ is used, (5.69) becomes

$$\begin{aligned} CW_i^{*(2)}\left(E(\alpha_j)\right) &= -\frac{1}{2}\log\left(1-v_{ii}\right)^{1/3} - \left(\frac{q-1}{2}\right)\log(1-v_{ii})^{-2/3} - K \\ &= -\frac{1}{6}\log\left(1-v_{ii}\right) + \left(\frac{q-1}{3}\right)\log(1-v_{ii}) - K \\ &= \left(\frac{2q-3}{6}\right)\log(1-v_{ii}) - K \qquad (5.73) \end{aligned}$$

substituting respectively $CW_i^{*(1)}\left(E(\alpha_j)\right)$ and $CW_i^{*(2)}\left(E(\alpha_j)\right)$ for $CW_i^{*}\left(E(\alpha_j)\right)$ in (5.70), one has

$$\begin{aligned} \Delta CW_i^1\left(E(\alpha_j)\right) &= -\frac{1}{2}\log(1-v_{ii}) - \frac{q}{2}\log(S_{iW_j}) + \frac{1}{2}\log\left(1-v_{ii}^j\right) + \left(\frac{q-1}{2}\right)\log O_{iW_j} \\ &= -\frac{q}{2}\log(1-v_{ii}) - \frac{q}{2}\log(S_{iW_j}) + \frac{q}{2}\log\left(1-v_{ii}^j\right) \\ &= \frac{q}{2}\left(-\log(1-v_{ii}) - \log(S_{iW_j}) + \log\left(1-v_{ii}^j\right)\right) \\ &= \frac{q}{2}\left(\log(O_{iW_j}) - \log(S_{iW_j})\right) \qquad (5.74) \end{aligned}$$

and

$$\begin{aligned} \Delta CW_i^2\left(E(\alpha_j)\right) &= \left(\frac{2q-3}{6}\right)\log(1-v_{ii}) + \frac{1}{2}\log\left(1-v_{ii}^j\right) + \left(\frac{q-1}{2}\right)\log O_{iW_j} \\ &= \left(\frac{2q-3}{6}\right)\log(1-v_{ii}) + \frac{1}{2}\log\left(1-v_{ii}^j\right) + \left(\frac{q-1}{2}\right)\log\left(1-v_{ii}^j\right) \end{aligned}$$

$$- \left(\frac{q-1}{2}\right) \log\left(1 - v_{ii}\right) \tag{5.75}$$
$$= \frac{q}{2} \log\left(1 - v_{ii}^{j}\right) - \frac{q}{6} \log\left(1 - v_{ii}\right)$$

In general, type A outliers observations will have negative values of $\Delta CW_i^2\left(E(\alpha_j)\right)$ with a theoretical minimum of

$$\lim_{v_{ii}^j \to v_{ii}} \Delta CW_i^2\left(E(\alpha_j)\right) = \frac{q}{3} \log\left(1 - v_{ii}\right) \tag{5.76}$$

whereas type B outliers observations will have positive values of $\Delta CW_i^2\left(E(\alpha_j)\right)$ with a theoretical maximum of

$$\lim_{v_{ii}^j \to 0} \Delta CW_i^2\left(E(\alpha_j)\right) = -\frac{q}{6} \log\left(1 - v_{ii}\right) \tag{5.77}$$

comparison of these two limit values shows that

$$\lim_{v_{ii}^j \to v_{ii}} \left|\Delta CW_i^2\left(E(\alpha_j)\right)\right| = 2\left|\lim_{v_{ii}^j \to 0} \Delta CW_i^2\left(E(\alpha_j)\right)\right| \tag{5.78}$$

that is type A outliers can potentially have larger values of $\left|\Delta CW_i^2\left(E(\alpha_j)\right)\right|$ than type B outliers.

5.9 The likelihood distance based on the secondary regression models

From the previous results, One can easily derive the likelihood distance for secondary regression models . firstly, we make an adaptation to (5.60) above, namely

$$ld_i^{(1)} = ld_i(\alpha_j, \sigma_{W_j}^2)$$
$$= n \log\left((\frac{n}{n-1})(\frac{n-q}{n-q+t_{i_{Wj}}^2}) \right) + \frac{(n-1)t_{i_{Wj}}^2}{(n-q)(1-v_{ii}^j)} - 1 \tag{5.79}$$

where $t_{i_{Wj}}$ is the external studentized residuals from the secondary regression and is given by

$$t_{i_{Wj}} = \frac{e_{i_{W_j}}}{\hat{\sigma}_{W_{(i)j}}(1 - v_{ii}^j)^{1/2}} \tag{5.80}$$

This leads

$$\begin{aligned}
t^2_{iW_j} &= \frac{e^2_{iW_j}(n-q)}{RSS_{W_{(i)j}}(1-v^j_{ii})} \\
&= \frac{e^2_{iW_j}(n-q)}{\left(RSS_{W_j} - e^2_{iW_j}/\left(1-v^j_{ii}\right)\right)(1-v^j_{ii})} \\
&= \frac{e^2_{iW_j}(n-q)/RSS_{W_j}}{\left(1 - e^2_{iW_j}/\left(1-v^j_{ii}\right)RSS_{W_j}\right)(1-v^j_{ii})} \\
&= \frac{(v_{ii}-v^j_{ii})(n-q)}{\left(1-(v_{ii}-v^j_{ii})/\left(1-v^j_{ii}\right)\right)(1-v^j_{ii})} \\
&= \frac{(v_{ii}-v^j_{ii})(n-q)}{(1-v_{ii})} \\
&= P_{iW_j}(n-q)
\end{aligned} \tag{5.81}$$

the factor $P_{iW_j} = \frac{v_{ii}-v^j_{ii}}{1-v_{ii}}$ is known as the primary potential and relates to the orientation O_{iW_j} as follows

$$P_{iW_j} = O_{iW_j} - 1 \tag{5.82}$$

Therefore, expression (5.79) becomes

$$\begin{aligned}
ld^{(1)}_i &= n\log\left((\frac{n}{n-1})(\frac{n-q}{n-q+(n-q)P_{iW_j}})\right) + \frac{(n-q)(n-1)P_{iW_j}}{\left(1-v^j_{ii}\right)(n-q)} - 1 \\
&= n\log\left((\frac{n}{n-1})O^{-1}_{iW_j}\right) + \frac{(n-1)P_{iW_j}}{\left(1-v^j_{ii}\right)} - 1 \\
&= n\log\left((\frac{n}{n-1})\left(\frac{1-v_{ii}}{1-v^j_{ii}}\right)\right) + \frac{(n-1)\left(v_{ii}-v^j_{ii}\right)}{\left(1-v^j_{ii}\right)(1-v_{ii})} - 1
\end{aligned} \tag{5.83}$$

Secondly, another adaptation to (5.53) above leads

$$\begin{aligned}
ld^{(2)}_i &= ld_i(\alpha_j \mid \sigma^2_{W_j}) \\
&= n\log\left(1 + (\frac{q-1}{n-q+1})D^*_i\right)
\end{aligned}$$

where D^*_i, the Cook's distance for secondary regression, is given by

$$\begin{aligned}
D^*_i &= \frac{r^2_{iW_j}}{q-1}v^j_{(ii)} \\
&= \left(\frac{n-q+1}{q-1}\right)\left(\frac{v_{ii}-v^j_{ii}}{1-v^j_{ii}}\right)v^j_{(ii)} \\
&= \left(\frac{n-q+1}{q-1}\right)\left(\frac{v_{ii}-v^j_{ii}}{1-v_{ii}}\right)\left(\frac{1-v_{ii}}{1-v^j_{ii}}\right)\left(\frac{v^j_{ii}}{1-v^j_{ii}}\right) \\
&= \left(\frac{n-q+1}{q-1}\right)P_{iW_j}O^{-1}_{iW_j}P^*_{iW_j}
\end{aligned} \tag{5.84}$$

Hence,

$$
\begin{aligned}
ld_i^{(2)} &= n\log\left(1+\left(\frac{q-1}{n-q+1}\right)\left(\frac{n-q+1}{q-1}\right)P_{iw_j}O_{iw_j}^{-1}P_{iw_j}^{*}\right)\\
&= n\log\left(1+P_{iw_j}O_{iw_j}^{-1}P_{iw_j}^{*}\right)\\
&= n\log\left(1+\frac{v_{ii}-v_{ii}^{j}}{1-v_{ii}^{j}}v_{(ii)}^{j}\right)
\end{aligned}
$$

As in (5.72) and (5.73) above, substituting respectively $v_{ii}^{j*}=1-(1-v_{ii})S_{iw_j}$ and $v_{ii}^{j**}=1-(1-v_{ii})^{1/3}$ for v_{ii}^{j} in the expression of $ld_i^{(1)}$ leads

$$
\begin{aligned}
ld_i^{(1)*} &= n\log\left(\left(\frac{n}{n-1}\right)\left(\frac{1-v_{ii}}{1-1+(1-v_{ii})S_{iw_j}}\right)\right)+\frac{(n-1)\left(v_{ii}-1+(1-v_{ii})\,S_{iw_j}\right)}{\left(1-1+(1-v_{ii})\,S_{iw_j}\right)(1-v_{ii})}-1\\
&= n\log\left(\frac{n}{n-1}S_{iw_j}^{-1}\right)+(n-1)\left(\frac{S_{iw_j}-1}{(1-v_{ii})\,S_{iw_j}}\right)-1 \qquad (5.85)
\end{aligned}
$$

and

$$
\begin{aligned}
ld_i^{(1)**} &= n\log\left(\left(\frac{n}{n-1}\right)\left(\frac{1-v_{ii}}{1-1+(1-v_{ii})^{1/3}}\right)\right)+\frac{(n-1)\left(v_{ii}-1+(1-v_{ii})^{1/3}\right)}{\left(1-1+(1-v_{ii})^{1/3}\right)(1-v_{ii})}-1\\
&= n\log\left(\frac{n}{n-1}(1-v_{ii})^{2/3}\right)+(n-1)\left((1-v_{ii})^{1/3}-(1-v_{ii})\right)(1-v_{ii})^{-4/3}-1\\
&= n\log\left(\frac{n}{n-1}(1-v_{ii})^{2/3}\right)+(n-1)\left((1-v_{ii})^{-1}-(1-v_{ii})^{-1/3}\right)-1 \qquad (5.86)
\end{aligned}
$$

Hence, one can consider the following measures to identify potential collinearity-influential observations:

$$
\begin{aligned}
\Delta ld_i^{(1)*} &= ld_i^{(1)}-ld_i^{(1)*}\\
&= n\log\left(\left(\frac{n}{n-1}\right)\left(\frac{1-v_{ii}}{1-v_{ii}^{j}}\right)\right)+\frac{(n-1)\left(v_{ii}-v_{ii}^{j}\right)}{\left(1-v_{ii}^{j}\right)(1-v_{ii})}-1\\
&\quad-\left(n\log\left(\frac{n}{n-1}S_{iw_j}^{-1}\right)+(n-1)\left(\frac{S_{iw_j}-1}{(1-v_{ii})\,S_{iw_j}}\right)-1\right)\\
&= n\log\left(\frac{S_{iw_j}}{O_{iw_j}}\right)+(n-1)\left(\frac{P_{iw_j}}{\left(1-v_{ii}^{j}\right)}-\left(\frac{S_{iw_j}-1}{S_{iw_j}(1-v_{ii})}\right)\right)\\
&= n\log\left(\frac{S_{iw_j}}{O_{iw_j}}\right)+(n-1)\left(\frac{O_{iw_j}-1}{\left(1-v_{ii}^{j}\right)}-\left(\frac{S_{iw_j}-1}{S_{iw_j}(1-v_{ii})}\right)\right) \qquad (5.87)
\end{aligned}
$$

and

$$
\Delta ld_i^{(1)**} = ld_i^{(1)}-ld_i^{(1)**}
$$

$$
\begin{aligned}
&= n\log\left(\frac{(1-v_{ii})^{1/3}}{1-v_{ii}^{j}}\right)+(n-1)\left(\frac{\left(v_{ii}-v_{ii}^{j}\right)}{(1-v_{ii})\left(1-v_{ii}^{j}\right)}-\left((1-v_{ii})^{-1}+(1-v_{ii})^{-1/3}\right)\right)\\
&= n\log\left(\frac{(1-v_{ii})^{1/3}}{1-v_{ii}^{j}}\right)+(n-1)\left(\frac{v_{ii}-v_{ii}^{j}-\left(1-v_{ii}^{j}\right)+(1-v_{ii})^{2/3}(1-v_{ii})}{(1-v_{ii})\left(1-v_{ii}^{j}\right)}\right)\\
&= n\log\left(\frac{(1-v_{ii})^{1/3}}{1-v_{ii}^{j}}\right)+(n-1)\frac{(1-v_{ii})\left(\left(1-v_{ii}^{j}\right)(1-v_{ii})^{-1/3}-1\right)}{(1-v_{ii})\left(1-v_{ii}^{j}\right)}\\
&= n\log\left(\frac{(1-v_{ii})^{1/3}}{1-v_{ii}^{j}}\right)+(n-1)\left((1-v_{ii})^{-1/3}-(1-v_{ii}^{j})^{-1}\right) \qquad (5.88)
\end{aligned}
$$

$\Delta ld_i^{(1)*}$ will take on positive values in the interval $[0,\infty)$ for type B observations ($O_{i_{W_j}} \longrightarrow \infty$ *and* $S_{i_{W_j}} \longrightarrow 1$); for type A observations, for which one has $O_{i_{W_j}} \longrightarrow 1$ *and* $S_{i_{W_j}} \longrightarrow \infty$, it will take on negative values in the interval $(-\infty, 0]$.

The same, $\Delta ld_i^{(1)**}$ will take on positive values for type B observations in the interval

$$
[0, \frac{n}{3}\log(1-v_{ii})+(n-1)\left(\frac{v_{ii}}{1-v_{ii}}+(1-v_{ii})^{-1/3}-(1-v_{ii}^{j})^{-1}\right)] \qquad (5.89)
$$

since $v_{ii}^{j} \longrightarrow 0$, while for type A observations it will take on negative values in the interval

$$
[\frac{-2n}{3}\log(1-v_{ii})+(n-1)\left((1-v_{ii})^{-1/3}-(1-v_{ii})^{-1}\right), 0] \qquad (5.90)
$$

since for this category of observations $v_{ii}^{j} \longrightarrow 1$.

In the linear model, it can happen that the problem of outliers arises simultaneously with that of high correlations among the predictors. Following Marquardt (1974), the two problems must also be tackled simultaneously. In the next chapters, models with selected variables and the technique of principle components are reviewed.

Chapter 6

MODELS WITH SELECTED VARIABLES

In case we have a near collinearity problem, for example, say the i-th column of the W matrix is almost an exact linear combination of other columns of the same matrix. The variables selection technique is useful and help to eliminate from the model such a variable in order to improve the conditioning of the design matrix W^TW. Hence, more reliable estimates of the parameters of the model can be expected.

Suppose in the model (2.1) that p<q variables are retained and the rest discarded. W and β can respectively be partitioned as follows

$$W = (W_1, W_2) \ \ and \ \ \beta = (\beta_1, \beta_2)$$

with W_1 an nxp matrix of observations in the retained variables, W_2 an $n\text{x}(q-p)$ matrix of observations in the discarded variables, β_1 and β_2 are the associated vectors of parameters. We can rewrite (2.1) as below

$$X = W_1\beta_1 + W_2\beta_2 + \varepsilon \tag{6.1}$$

with $\varepsilon \sim N(\mathbf{0}, \sigma^2 I)$

The least squares estimate for β_1 is given by

$$\hat{\beta}_1 = (W_1^T W_1)^{-1} W_1^T X \tag{6.2}$$

Now

$$\begin{aligned} E(\hat{\beta}_1) &= (W_1^T W_1)^{-1} W_1^T (W_1\beta_1 + W_2\beta_2) \\ &= \beta_1 + (W_1^T W_1)^{-1} W_1^T W_2\beta_2 \end{aligned} \tag{6.3}$$

$$Var(\hat{\beta}_1) = \sigma^2 (W_1^T W_1)^{-1} \tag{6.4}$$

It can be shown that $\hat{\beta}_1$ is unbiased and $Var(\hat{\beta}) - Var(\hat{\beta}_1)$ is nonnegative definite if $\beta_2 = 0$. If $\beta_2 \neq 0$ then $\hat{\beta}_1$ is preferred to $\hat{\beta}$ under the criterion

$$\frac{\beta_2^T(W_2^T W_1 - W_2^T W_1(W_1^T W_1)^{-1} W_1^T W_2)^{-1}\beta_2}{\sigma^2} \leq q - p \tag{6.5}$$

with respect to the earlier notation, one has

$$\begin{aligned} r_1 &= X - W_1\hat{\beta}_1 \\ &= \left(I - W_1(W_1^T W_1)^{-1} W_1^T\right) X \\ &= \left(I - W_1(W_1^T W_1)^{-1} W_1^T\right)(W_1\beta_1 + W_2\beta_2 + \varepsilon) \\ &= \left(I - W_1(W_1^T W_1)^{-1} W_1^T\right)(W_2\beta_2 + \varepsilon) \\ &= (I - V_1)(W_2\beta_2 + \varepsilon) \end{aligned} \tag{6.6}$$

where

$$V_1 = W_1(W_1^T W_1)^{-1} W_1^T$$

taking expectation, one has

$$\begin{aligned} E(r_1) &= 0 \; only \; if \; \beta_2 = 0 \\ & \quad and \\ Var(r_1) &= \sigma^2(I - V_1) \end{aligned} \tag{6.7}$$

The total variance of r_1 is greater than that of the full model residuals r as one can see it from the following result

$$\begin{aligned} tr(Var(r_1)) - tr(Var(r)) &= \sigma^2(n-p) - \sigma^2(n-q) \\ &= \sigma^2(q-p) > 0 \end{aligned} \tag{6.8}$$

At this stage, the analysis will proceed as if the $(q-p)$ discarded variables had not existed, that is any of the previous diagnostic techniques can be applied to the model with p independent variables. Thus, following chapters 2 and 3,

$$(t_1)_i^2 = \frac{(r_1)_i^2}{(s_1)_{(i)}^2 (1 - (V_1)_{ii})} \qquad i = 1, 2, ..., n \tag{6.9}$$

$$(t_1)_I^2 = \frac{(r_1)_I^T (I - (V_1)_I)^{-1} (r_1)_I}{m (s_1)_{(I)}^2} \tag{6.10}$$

are statistics for testing for the single and groups of outlying observations; they follow central $F(1, n-p-1)$ and $F(m, n-p-m)$ distributions respectively. The related measures of influence are given by

$$(\triangle_1)_i = \frac{1}{p}(t_1)_i^2 (V_1)_{ii} / [1 - (V_1)_{ii}] \tag{6.11}$$

for the single case, and

$$(\triangle_1)_I = \frac{(r_1)_I^T (I - (V_1)_I)^{-1} (V_1)_I (I - (V_1)_I)^{-1} (r_1)_I}{p(s_1)_{(I)}^2} \tag{6.12}$$

for the multiple case.

In practice, it is always assumed that critical points for the statistics (6.9), (6.10), (6.11) and (6.12) can be obtained from tables of the central F distributions. However, if the assumption $\beta_2 = 0$ is not true, then the distributions are doubly noncentral (see A.K. GUPTA & V.R. GIRKO(1996)).

The reader should also be advised that the technique of subset selection of variables should be applied with care, as this approach may result in removing some of the important regression variables; it can also lead to a matrix $(W_1^T W_1)$ that is still ill-conditioned. In this last case, it is recommended to use ridge regression or the generalized inverse regression (Principle Components) which provide stable estimates. In the next chapters, Ridge and generalized inverse or principle components regressions are reviewed.

Chapter 7

RIDGE REGRESSION

The ridge regression estimate (RR) was first proposed by Hoerl and Kennard (1970). It is a biased estimation technique to be followed when the matrix (W^TW) appears to be ill-conditioned. Its procedure consists of adding a small positive constant k to the diagonal elements of (W^TW) which we will assume to be in correlation form.

The ridge estimate $\widetilde{\beta}$ is defined as the solution to

$$(W^TW + kI)\,\widetilde{\beta} = W^TX \ , \ k \geq 0 \tag{7.1}$$

that is

$$\begin{aligned} \widetilde{\beta} &= (W^TW + kI)^{-1} W^TX \\ &= UW^TX \end{aligned} \tag{7.2}$$

where

$$U = (W^TW + kI)^{-1}$$

k is an arbitrary constant chosen in such a way that the estimator $\widetilde{\beta}$ becomes stable.

$\widetilde{\beta}$ relates to β as follows

$$\begin{aligned} \widetilde{\beta} &= UW^TX \\ &= UW^TW\hat{\beta} \\ &= Z\hat{\beta} \end{aligned} \tag{7.3}$$

with

$$Z = UW^TW = I - kU \tag{7.4}$$

Properties

1 For $k = 0$, $\widetilde{\beta} = \hat{\beta}$

2 For $k \to \infty$, $UW^TW = \left(W^TW + kI\right)^{-1} W^TW$ approaches 0 and $\widetilde{\beta} \to 0$

3

$$\begin{aligned} E\left(\widetilde{\beta}\right) &= E\left(UW^TW\hat{\beta}\right) \\ &= UW^TW\beta \\ &= Z\beta \\ &= \beta - kU\beta \end{aligned} \quad (7.5)$$

4

$$\begin{aligned} Var\left(\widetilde{\beta}\right) &= Var\left(UW^TW\hat{\beta}\right) \\ &= UW^TWVar\left(\hat{\beta}\right)W^TWU \\ &= \sigma^2 UW^TWU \\ &= \sigma^2 Z\left(W^TW\right)^{-1} Z^T \end{aligned} \quad (7.6)$$

For $k > 0$, $\widetilde{\beta}$ is a biased estimator of β and the bias is given by

$$-kU\beta$$

Analyzing ridge residuals

Let

$$\widetilde{e} = X - W\widetilde{\beta} \quad (7.7)$$

be the ridge residual. Then

$$\begin{aligned} \widetilde{e} &= \left(I - WUW^T\right)X \\ &= \left(I - WUW^T\right)(W\beta + e) \\ &= W(I - Z)\beta + \left(I - WUW^T\right)e \end{aligned} \quad (7.8)$$

We have

$$\begin{aligned} E(\widetilde{e}) &\neq 0 \textit{ and} \\ Var(\widetilde{e}) &= \sigma^2\left(I - WUW^T - kWUU^TW^T\right) \\ &= \sigma^2(I - Q) \end{aligned} \quad (7.9)$$

where

$$\begin{aligned} Q &= WUW^T + kWUU^TW^T \\ &= WU(I + kU)W^T \quad \left(U = U^T\right) \end{aligned} \quad (7.10)$$

It can be shown that:

- $Var(\widetilde{e}) - Var(\hat{e})$ is a non-negative definite matrix. In fact, following Goldberger (1964), if A is an nxn positive definite matrix and P is an nxm matrix with $rank(P) = m$, then P^TAP is positive definite and if A is non-negative definite and P any matrix, then P^TAP is non-negative definite.

  Consider

$$\begin{aligned} Var(\widetilde{e}) - Var(\hat{e}) &= \sigma^2 W\{\left(W^TW\right)^{-1} - U - kUU^T\}W^T \\ rank(W) &= q \end{aligned}$$

  We need to show that

$$B = \left(W^TW\right)^{-1} - U - kUU^T \tag{7.11}$$

  is positive definite.

  Let V be an orthogonal matrix such that

$$V^TW^TWV = \Lambda = diag(\lambda_1, ..., \lambda_q)$$

  We have

$$V^TBV = V^T\left(W^TW\right)^{-1}V - V^TUV - kV^TUU^TV \tag{7.12}$$

  which is a diagonal matrix with the ith diagonal element given by

$$\frac{1}{\lambda_i} - \frac{1}{\lambda_i + k} - \frac{k}{(\lambda_i + k)^2} = \frac{k^2}{\lambda_i(\lambda_i + k)^2} \geq 0$$

  Therefore if $k > 0$ then the roots of B are positive and B is positive definite.

- $Total\ variance(\hat{e}) < Total\ variance(\widetilde{e})$

$$\begin{aligned} Total\ variance(\hat{e}) &= tr[\sigma^2\left(I - W\left(W^TW\right)^{-1}W^T\right)] \\ &= \sigma^2(n - q) \end{aligned} \tag{7.13}$$

$$Total\ variance(\widetilde{e}) = \sigma^2[n - q + \sum_{i=1}^{q}\left(\frac{k}{\lambda_i + k}\right)^2] \tag{7.14}$$

  substracting (7.12) from (7.13) proves the result.

  Thus, it seems hardly necessary to analyze ridge residuals as long as $\widetilde{e}$ is a biased estimator and more variable than the estimator $\hat{e}$, which has also the advantage of being unbiased. However, with respect to the generalized mean square error, it has been shown that under certain conditions (for e.g. when multicollinearity and erratic data points are present), ridge residuals are better estimates of e than the ordinary least square residuals.

Definition 7.1 (Theobald(1974))

Let $\hat{\theta}_1$ *and* $\hat{\theta}_2$ be two estimators of a vector parameter θ, and let

$$M_j = E\left(\hat{\theta}_j - \theta\right)\left(\hat{\theta}_j - \theta\right)^T, \ j = 1, 2$$

be the second order moment matrices. Moreover let

$$m_j = E\left(\hat{\theta}_j - \theta\right)^T B\left(\hat{\theta}_j - \theta\right), \ j = 1, 2$$

where B is a non-negative definite matrix. The m_j is called generalized mean square error. $\hat{\theta}_2$ is said to be better estimate if the generalized mean square error of $\hat{\theta}_2$ is less than the generalized mean square error of $\hat{\theta}_1$.

Lemma 7.1 (Theobald(1974))

The following two conditions are equivalent

(a) $M_1 - M_2$ is non-negative definite

(b) $m_1 - m_2 \geq 0$

for all non-negative definite matrix B.

Lemma 7.2 (Theobald(1974))

If

$$\begin{aligned} M_{LS} &= E\left(\hat{\beta} - \beta\right)\left(\hat{\beta} - \beta\right)^T \\ &= \sigma^2 V \Lambda^{-1} V^T \end{aligned}$$

and

$$\begin{aligned} M_{RR} &= E\left(\widetilde{\beta} - \beta\right)\left(\widetilde{\beta} - \beta\right)^T \\ &= \sigma^2 V \Lambda \left(\Lambda + kI\right)^{-2} V^T + k^2 V \left(\Lambda + kI\right)^{-1} V^T \beta\beta^T V \left(\Lambda + kI\right)^{-1} V^T \end{aligned}$$

Then there exists a $K > 0$ such that $M_{LS} - M_{RR}$ is positive definite whenever $0 < k < K$.

Theorem 7.1

There exists a $K > 0$ such that the generalized mean square error of $\widetilde{e}$ is less than the generalized mean square of $\hat{e}$ whenever $0 < k < K$.

<u>Proof</u> We have

$$\hat{e} - e = \left(X - W\hat{\beta}\right) - \left(X - W\beta\right)$$

$$= -X\left(\hat{\beta}-\beta\right)$$

and the second order moment matrix of $\hat{e}$ is given by

$$\begin{aligned} M_1 &= E\left(\hat{e}-e\right)\left(\hat{e}-e\right)^T \\ &= E\{X\left(\hat{\beta}-\beta\right)\left(\hat{\beta}-\beta\right)^T X^T\} \\ &= XM_{LS}X^T \end{aligned}$$

similarly

$$\widetilde{e}-e=-X\left(\widetilde{\beta}-\beta\right)$$

and the second order moment matrix of $\widetilde{e}$ is

$$\begin{aligned} M_2 &= E\left(\widetilde{e}-e\right)\left(\widetilde{e}-e\right)^T \\ &= XM_{RR}X^T \end{aligned}$$

The result follows from definition 7.1 and lemma 7.1 and 7.2.

The following statistic is suggested for testing whether the i^{th} data point is an outlier:

$$\widetilde{t}_i=\frac{\widetilde{e}_i}{s_{(i)}\sqrt{1-q_{ii}}} \quad i=1,...,n \tag{7.15}$$

Using this technique, the influence of the i^{th} observation can be measured by

$$\begin{aligned} \Delta_i &= \frac{\left(\widetilde{\beta}_{(i)}-\widetilde{\beta}\right)^T\left(W^TW+kI\right)\left(\widetilde{\beta}_{(i)}-\widetilde{\beta}\right)}{qs_{(i)}^2} \\ &= \frac{\widetilde{e}_iw_i\left(W^TW+kI\right)^{-1}\left(W^TW+kI\right)\left(W^TW+kI\right)^{-1}w_i^T\widetilde{e}_i}{qs_{(i)}^2\left(1-\widetilde{v}_{ii}\right)^2} \\ &= \frac{\widetilde{v}_{ii}\widetilde{e}_i^2}{qs_{(i)}^2\left(1-\widetilde{v}_{ii}\right)^2} \\ &= \frac{1}{q}\left(\frac{\widetilde{e}_i}{s_{(i)}\sqrt{1-q_{ii}}}\right)^2\left(\frac{\widetilde{v}_{ii}}{\left(1-\widetilde{v}_{ii}\right)}\right)\left(\frac{1-q_{ii}}{1-\widetilde{v}_{ii}}\right) \\ &= \frac{1}{q}\widetilde{t}_i^2\left(\frac{\widetilde{v}_{ii}}{\left(1-\widetilde{v}_{ii}\right)}\right)\left(\frac{1-q_{ii}}{1-\widetilde{v}_{ii}}\right) \end{aligned} \tag{7.16}$$

Comparing (7.16) to (6.11) we notice the bias

$$\left(\frac{1-q_{ii}}{1-\widetilde{v}_{ii}}\right)$$

that has been introduced by the ridge effect.

For the multiple case, (7.16) can be generalized as follows

$$\widetilde{\Delta}_I = \frac{\left(\widetilde{\beta}_{(I)} - \widetilde{\beta}\right)^T W^T W \left(\widetilde{\beta}_{(I)} - \widetilde{\beta}\right)}{q s_{(I)}^2} \tag{7.17}$$

with respect to earlier notations.

Let

$$U_{(I)} = \left(W_{(I)}^T W_{(I)} + kI\right)^{-1} \tag{7.18}$$

$$T = W_I U_{(I)} W_I^T \tag{7.19}$$

We have

$$W^T W + kI = W_{(I)}^T W_{(I)} + W_I^T W_I + kI \tag{7.20}$$

and

$$\begin{aligned} I - T(I+T)^{-1} &= (I+T)(I+T)^{-1} - T(I+T)^{-1} \\ &= (I+T)^{-1} \end{aligned} \tag{7.21}$$

let

$$U^* = U_{(I)} - U_{(I)} W_I^T (I+T)^{-1} W_I U_{(I)} \tag{7.22}$$

multiplying (7.20) and (7.22) leads

$$\begin{aligned} \left(W^T W + kI\right) U^* &= \left(W_{(I)}^T W_{(I)} + kI\right) U_{(I)} - \left(W_{(I)}^T W_{(I)} + kI\right) U_{(I)} W_I^T (I+T)^{-1} W_I U_{(I)} \\ &\quad + \left(W_I^T W_I\right) U_{(I)} - \left(W_I^T W_I\right) U_{(I)} W_I^T (I+T)^{-1} W_I U_{(I)} \\ &= I - W_I^T (I+T)^{-1} W_I U_{(I)} + \left(W_I^T W_I\right) U_{(I)} - W_I^T T (I+T)^{-1} W_I U_{(I)} \\ &= I - W_I^T \left[(I+T)^{-1} + I - T(I+T)^{-1}\right] W_I U_{(I)} \\ &= I - W_I^T \left[(I+T)^{-1} - (I+T)^{-1}\right] W_I U_{(I)} \\ &= I \end{aligned}$$

Thus

$$U^* = U = U_{(I)} - U_{(I)} W_I^T (I+T)^{-1} W_I U_{(I)} \tag{7.23}$$

It follows that

$$\begin{aligned} U W_I^T &= U_{(I)} W_I^T - U_{(I)} W_I^T (I+T)^{-1} W_I U_{(I)} W_I^T \\ &= U_{(I)} W_I^T \left[I - (I+T)^{-1} T\right] \\ &= U_{(I)} W_I^T (I+T)^{-1} \end{aligned} \tag{7.24}$$

by definition

$$\begin{aligned}\widetilde{V}_I &= W_I U W_I^T \\ &= W_I \left(W^T W + kI\right)^{-1} W_I^T \end{aligned} \tag{7.25}$$

so that

$$\begin{aligned}\left(I - \widetilde{V}_I\right) &= I - W_I U W_I^T \\ &= I - W_I U_{(I)} W_I^T (I+T)^{-1} \quad from \ (7.24) \\ &= I - T(I+T)^{-1} \\ &= (I+T)^{-1} \end{aligned} \tag{7.26}$$

The residuals vector $\widetilde{e}_I$ is given by

$$\begin{aligned}\widetilde{e}_I &= X_I - W_I \widetilde{\beta} \\ &= X_I - W_I U W_I^T X \\ &= X_I - W_I U \left(W_{(I)}^T X_{(I)} + W_I^T X_I\right) \\ &= \left(I - W_I U W_I^T\right) X_I - W_I U W_{(I)}^T X_{(I)} \\ &= \left(I - \widetilde{V}_I\right)\left(X_I - W_I \widetilde{\beta}_{(I)}\right) \end{aligned} \tag{7.27}$$

Now

$$\begin{aligned}\widetilde{\beta} &= U W^T X \\ &= U\left(W_{(I)}^T X_{(I)} + W_I^T X_I\right) \\ &= \left(U_{(I)} - U_{(I)} W_I^T (I+T)^{-1} W_I U_{(I)}\right)\left(W_{(I)}^T X_{(I)} + W_I^T X_I\right) \ from \ (7.22) \\ &= U_{(I)} W_{(I)}^T X_{(I)} - U_{(I)} W_I^T (I+T)^{-1} W_I U_{(I)} W_{(I)}^T X_{(I)} + U_{(I)} W_I^T X_I \\ &\quad - U_{(I)} W_I^T (I+T)^{-1} W_I U_{(I)} W_I^T X_I \\ &= \widetilde{\beta}_{(I)} - U_{(I)} W_I^T \left[(I+T)^{-1} W_I \widetilde{\beta}_{(I)} - X_I + (I+T)^{-1} T X_I\right] \\ &= \widetilde{\beta}_{(I)} + U_{(I)} W_I^T (I+T)^{-1}\left(X_I - W_I \widetilde{\beta}_{(I)}\right) \\ &= \widetilde{\beta}_{(I)} + U W_I^T \left(X_I - W_I \widetilde{\beta}_{(I)}\right) \\ &= \widetilde{\beta}_{(I)} + U W_I^T \left(I - \widetilde{V}_I\right)^{-1} \widetilde{e}_I^T \end{aligned}$$

this expression can be written as

$$\widetilde{\beta}_{(I)} - \widetilde{\beta} = -\left(W^T W + kI\right)^{-1} W_I^T \left(I - \widetilde{V}_I\right)^{-1} \widetilde{e}_I^T \tag{7.28}$$

Substituting (7.28) into (7.17) we obtain what follows

$$
\begin{aligned}
\widetilde{\Delta}_I &= \frac{1}{qs_{(I)}^2}\widetilde{e}_I^T\left(I-\widetilde{V}_I\right)^{-1}W_I\left(W^TW+kI\right)^{-1}W^TW\left(W^TW+kI\right)^{-1}W_I^T\left(I-\widetilde{V}_I\right)^{-1}\widetilde{e}_I \\
&= \frac{1}{qs_{(I)}^2}\widetilde{e}_I^T\left(I-\widetilde{V}_I\right)^{-1}W_IUW^TWUW_I^T\left(I-\widetilde{V}_I\right)^{-1}\widetilde{e}_I \\
&= \frac{1}{qs_{(I)}^2}\widetilde{e}_I^T\left(I-\widetilde{V}_I\right)^{-1}W_I\left(I-kU\right)UW_I^T\left(I-\widetilde{V}_I\right)^{-1}\widetilde{e}_I \\
&= \frac{1}{qs_{(I)}^2}\left[\begin{array}{l}\widetilde{e}_I^T\left(I-\widetilde{V}_I\right)^{-1}\widetilde{V}_I\left(I-\widetilde{V}_I\right)^{-1}\left(I-\widetilde{V}_I\right)^{-1}\widetilde{e}_I \\ -k\widetilde{e}_I^T\left(I-\widetilde{V}_I\right)^{-1}W_IU^TUW_I^T\left(I-\widetilde{V}_I\right)^{-1}\widetilde{e}_I\end{array}\right]
\end{aligned}
\tag{7.29}
$$

A group of observations for which the above statistic is large, given k, can be regarded as being an influential group. To test whether this group is an outlying group one can use the generalized studentized residuals

$$
\widetilde{t}_I^2 = \frac{\widetilde{e}_I^T\left(I-Q_I\right)^{-1}\widetilde{e}_I}{ms_{(I)}^2} \tag{7.30}
$$

with m the number of observations in the group and

$$
Q_I = W_IU\left(I+kU\right)W^T
$$

$\widetilde{t}_I^2$ follows a non-central F distribution with m and $(n-q-m)$ degrees of freedom; the non-centrality parameter is a function of k and the unknown parameter β. This statistic approaches the ordinary least squares statistic, that is a central F distribution, as $k \to 0$. Small values of k are indicated because the central F distribution can be used as an approximation.

Estimating the ridge constant k

An interesting way of choosing the value of k is the use of the ridge trace. The ridge trace is formed by plotting $\widetilde{\beta}$ against k, as k varies through the interval [0,1]. The value of k is then chosen at the point where the coefficients estimates stabilize. some authors prefer values of k within the interval $10^{-4} < k < 1$ (see for e.g. Marquardt(1970)). In the context of identification of outliers and influential observations, it is recommended that the analysis be performed for a number of different values of k, whereas a table or a graph should be prepared, showing the values of the statistics $\widetilde{t}_i^2$ and $\widetilde{\Delta}_i$ $\left(\text{or } \widetilde{t}_I^2 \text{ and } \widetilde{\Delta}_I \text{ depending on the case}\right)$ for different values of k; this is a sort of ridge trace, that will focus attention to observations that are potential outliers but where the outlying effect has been masked by the collinearity. Unfortunately such a ridge trace is heavy on computer time. One may therefore specify a limited number of values of k (say k=0, k=0.01, k=0.05, k=0.1. k=0.2 and k=0.3) and focus a particular attention on those $\widetilde{t}_i^2$ and $\widetilde{\Delta}_i$ that dramatically change with k.

If deletion is of consideration, then good candidates for deletion from the regression could be the variables whose estimates do not stabilize.

Chapter 8

PRINCIPAL COMPONENTS

When the matrix $W_1^T W_1$ is ill-conditioned, an alternative technique is the generalized inverse or principle components regression (Marquard (1970)).

Assume that $W_1^T W_1$ is in the form of a correlation matrix. Let $V = (V_1, V_2, ..., V_q)$ be the matrix of orthonormal latent vectors of $W_1^T W_1$ and further more let

$$V^T W^T W V = \Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_q) \tag{8.1}$$

with

$$\lambda_1 > \lambda_2 > ... > \lambda_q > 0$$

Suppose that the last $(q - r)$ roots of $W^T W$ are assumed to be close enough to zero to indicate $(q - r)$ near collinearity. Partition V and Λ as

$$V = (V_r, V_{q-r}) \quad \Lambda = \begin{pmatrix} \Lambda_r & 0 \\ 0 & \Lambda_{q-r} \end{pmatrix} \tag{8.2}$$

The principle component estimates of β are given by

$$\beta_{pc} = A^- W^T X \tag{8.3}$$

where

$$A^- = \left(W_1^T W_1\right)^- = V_r \Lambda_r^{-1} V_r^T = \sum_{j=1}^{r} \frac{V_j V_j^T}{\lambda_j} \tag{8.4}$$

is the generalized inverse of $\left(W_1^T W_1\right)$. Hence the naming β_{pc} as the generalized estimator.

If the model (2.1) is reparameterized as

$$\begin{aligned} X &= W V V^T \beta + \varepsilon \\ &= Z\delta + \varepsilon \end{aligned} \tag{8.5}$$

with

$$Z = WV \ and \ \delta = V^T\beta$$

and if the columns of Z are ordered so that Z_i corresponds to the smallest root, then the estimator in (8.3) can also be obtained by deleting the principal components $Z_{q-r}, \ldots, Z_q$ from the model (8.5), applying ordinary least squares to the retained components and making the necessary inverse linear transformations to obtain estimates for β. That is, if $Z = (Z_r, Z_{q-r})$, δ is partitioned conformably in δ_1 and δ_2, the restricted model is

$$\begin{aligned} X &= WV_r\delta_1 + e^* \\ &= Z_r\delta_1 + e^* \end{aligned} \tag{8.6}$$

Where δ_1 is now estimated by means of ordinary least squares technique, say $\hat{\delta}$. The principle component estimator is then given by

$$\beta_{pc} = V_r\hat{\delta}_1$$

Note: β_{pc} is identical to the restricted least square(RLS) estimator obtained by imposing $(q-r)$ independent linear restrictions $V_{q-r}^T\beta = 0$ on the full model.

Following Judge et al.(1980), since β_{pc} can be regarded as a RLS estimator, one has

$$\beta_{pc} = \hat{\beta} - \left(W^TW\right)^{-1} V_{q-r} \left(V_{q-r}^T \left(W^TW\right)^{-1} V_{q-r}\right)^{-1} V_{q-r}^T\beta \tag{8.7}$$

Therefore β_{pc} is a biased estimator, unless $V_{q-r}^T\beta$ is exactly zero.

$$\begin{aligned} Var\left(\beta_{pc}\right) &= \sigma^2\left(\left(W^TW\right)^{-1} - V_{q-r}\left(V_{q-r}^T\left(W^TW\right)^{-1}V_{q-r}\right)^{-1}V_{q-r}^T\left(W^TW\right)^{-1}\right) \\ &= \sigma^2\left(\left(W^TW\right)^{-1} - C\right) \end{aligned} \tag{8.8}$$

with

$$C = \left(W^TW\right)^{-1} V_{q-r} \left(V_{q-r}^T \left(W^TW\right)^{-1} V_{q-r}\right)^{-1} V_{q-r}^T \left(W^TW\right)^{-1} \tag{8.9}$$

Toro-Vizcarrondo and Wallace (1968) have shown that β_{pc} is preferred to $\hat{\beta}$ under the generalized mean square error criterion if and only if

$$\frac{\left(V_r^T\beta\right)^T \left(V_r^T \left(W^TW\right)^{-1} V_r\right)^{-1} \left(V_r^T\beta\right)}{\sigma^2} \leq 1$$

However, this condition cannot be verified exactly, as it depends upon the unknown parameters β.

Given that C is positive semi-definite and the variance of the ordinary least square estimate is

$$Var\left(\hat{\beta}\right) = \sigma^2 \left(W^T W\right)^{-1}$$

it follows that the diagonal elements of $Var\left(\beta_{pc}\right)$ are equal to or less than the diagonal elements of $Var\left(\hat{\beta}\right)$. If the restrictions mentioned above are true then β_{pc} will be unbiased and best. And if they are not true the β_{pc} will be biased but with smaller variance than $\hat{\beta}$.

Having found β_{pc}, the principal components residuals e^- are given by

$$\begin{aligned} e^- &= X - W\beta_{pc} \\ &= \left(I - WA^- W^T\right) X \\ &= \left(I - WA^- W^T\right) W\beta + \left(I - WA^- W^T\right) e \end{aligned} \tag{8.10}$$

Hence

$$Var\left(e^-\right) = \sigma^2 \left(I - WA^- W^T\right) \tag{8.11}$$

Suppose that the m observations suspected as outliers are grouped together and indexed by I. without lost of generality, let it be the first m positions. Let

$$\begin{aligned} N &= I - WA^- W^T \\ &= I - Q \end{aligned}$$

Partition the above matrix as

$$N = \begin{pmatrix} N_m & N_a \\ N_a^T & N_{n-m} \end{pmatrix}$$

where

$$N_m = I - Q_I$$

with

$$Q_I = W_I \left(W^T W\right)^- W_I^T$$

The following statistics can be used to test whether the principal components residuals are outliers

$$t_i = \frac{e_i^-}{\sqrt{s_{(i)}^2 \left(1 - q_{ii}\right)}} \tag{8.12}$$

for the single case, and for the multiple case

$$F_I = \frac{\left(e_I^-\right) N_m^{-1} e_I^-}{m s_{(I)}^2} \tag{8.13}$$

with

$$s_{(I)}^2 = e^{-T} e^- - e^{-T} \left(I - Q_I\right)^{-1} e^- / \left(n - r - q\right)$$

which is the estimate of σ^2 using the generalized inverse residuals with the m suspected residuals deleted.

Chapter 9

USING STOCHASTIC PRIOR INFORMATIONS

9.1 Introduction

In the preceding section, we have underlined the dramatic effects that outliers and influential observations can have on the least squares estimates of the regression coefficients. However, there can be a considerable improvement in the estimation of β when stochastic prior information is available and hence, an improvement also in the residuals.

Models with stochastic prior information are more realistic because in many practical situations, prior information on the regression parameters will not be exact; in other words, exact restrictions are inappropriate. The procedure of this technique consists in adding prior information to the data matrix as we will see it below.

When testing for outliers and influential observations in the presence of collinearity, related techniques presented earlier in this work can also be characterized by stochastic prior information or combined with stochastic prior information in order to improve regression estimates.

9.2 Stochastic prior information

Let consider the general linear model (2.1), say

$$X = W\beta + \varepsilon$$

with $X_{(n\times 1)}$, $W_{(n\times q)}$, $E(\varepsilon) = 0$ and $E\left(\varepsilon^T\varepsilon\right) = \sigma^2\Lambda = \sum_1$.

Assume that the prior information is given by

$$V\beta + u = h \tag{9.1}$$

where h is an $l - row$ vector of unknown constants, V is an lxq design matrix, u is an $l - row$ vector distributed as $N(\delta, \sigma^2 R)$; σ^2 is unknown but R is known (Belsey,Kuh and Welsch, 1980).

Combining (2.1) and (9.1) one can write

$$\begin{pmatrix} X \\ h \end{pmatrix} = \begin{pmatrix} W \\ V \end{pmatrix} \beta + \begin{pmatrix} \varepsilon \\ u \end{pmatrix} \tag{9.2}$$

or

$$X^* = W^* \beta + \varepsilon^* \tag{9.3}$$

with

$$E(\varepsilon^*) = E \begin{pmatrix} \varepsilon \\ u \end{pmatrix} = \begin{pmatrix} 0 \\ \delta \end{pmatrix},$$

and

$$\begin{aligned} Cov\left(\varepsilon^* \varepsilon^{*T}\right) &= \begin{pmatrix} \sum_1 & 0 \\ 0 & \sum_2 \end{pmatrix} = \begin{pmatrix} \sigma^2 \Lambda & 0 \\ 0 & \sigma^2 R \end{pmatrix} \\ &= \sigma^2 \Omega \end{aligned} \tag{9.4}$$

where

$$\Omega = \begin{pmatrix} \sigma^2 \Lambda & 0 \\ 0 & \sigma^2 R \end{pmatrix}$$

If we let $R = I$ and $\delta = 0$, model (9.2) reduces to the ordinary least squares regression. in what follows, we will assume that $R \neq I$ and $\delta \neq 0$.

From the above model, the generalized least squares estimator for β is given by

$$\beta^* = \left(W^T W + V^T R^{-1} V\right)^{-1} \left(W^T X + V^T R^{-1} h\right) \tag{9.5}$$

or again

$$\begin{aligned} \beta^* &= \left(W^{*^T} \Omega^{-1} W^*\right)^{-1} W^{*^T} \Omega^{-1} X^* \\ &= A W^{*^T} \Omega^{-1} X^* \end{aligned} \tag{9.6}$$

where

$$A = \left(W^{*^T} \Omega^{-1} W^*\right)^{-1} = \left(W^T W + V^T R^{-1} V\right)^{-1} \tag{9.7}$$

It follows that

$$E(\beta^*) = \beta + AV^T R^{-1}\delta$$

and

$$Var(\beta^*) = \sigma^2 A$$

Unless $\delta = 0$, β^* is a biased estimator. However, β^* has smaller variance than the ordinary least squares estimator $\hat{\beta}$.

Computing residuals

Let

$$\begin{aligned} r^* &= X^* - W^*\beta^* \\ &= \left(I - W^*AW^{*^T}\Omega^{-1}\right) X^* \\ &= \left(I - W^*AW^{*^T}\Omega^{-1}\right)(W^*\beta + \varepsilon^*) \\ &= \left(I - W^*AW^{*^T}\Omega^{-1}\right)\varepsilon^* \\ &= (I - B)\,\varepsilon^* \end{aligned} \tag{9.8}$$

With

$$B = W^*AW^{*^T}\Omega^{-1} \text{ and } B \text{ is idempotent}$$

Then

$$E(r^*) = (I - B)\begin{pmatrix} 0 \\ \delta \end{pmatrix}$$

and

$$Var(r^*) = \sigma^2 (I - B)\,\Omega\,(I - B)^T \tag{9.9}$$

Like previously, under the assumption that the inequality

$$\frac{\delta^T\left[V\left(W^TW\right)^{-1}V^T + R\right]\delta}{\sigma^2} \leq 1 \tag{9.10}$$

is true, the first n elements of the generalized least squares residuals r^* are better estimate of the unknown ε than e is for ε.

Searching for outliers and influential observations

Suppose that there are m suspected outliers indexed by I. The search for outliers in models with stochastic prior information is conducted on the basis of the sample information. In other words, outliers will be sought only amongst the first n observations of r^*.

Consider any subset r_I^* of m elements from the first n elements of r^*, say

$$\begin{aligned} r_I^* &= (I_m, 0)\, r^* \\ &= (I_m, 0)\,(I - B)\, e^* \end{aligned} \tag{9.11}$$

Let

$$M = I - B \quad and \quad N = (I - B)\,\Omega\,(I - B)^T = M\Omega M^T$$

Like previously, partition N as

$$N = \begin{pmatrix} N_m & N_a \\ N_a^T & N_{n+l-m} \end{pmatrix}$$

where N_m is the submatrix of N formed by the rows and columns indexed by I. We have

$$Var\,(r_I^*) = \sigma^2 N_m \tag{9.12}$$

and

$$\begin{aligned} \frac{r_I^{*T} N_m^{-1} r_I^*}{\sigma^2} &\sim \chi_m^2\,(\nu) \\ with\;\; \nu &= \frac{1}{2}\left(0, \delta^T\right)(I - B)^T \begin{pmatrix} N_m^{-1} & 0 \\ 0 & 0 \end{pmatrix} (I - B) \begin{pmatrix} 0 \\ \delta \end{pmatrix} \end{aligned} \tag{9.13}$$

A statistic for testing whether a group of m observations is an outlying group is given by

$$F_I^* = \frac{r_I^{*T} N_m^{-1} r_I^*}{m s_{(I)}^{*2}} \tag{9.14}$$

with

$$s_{(I)}^{*2} = \frac{r_I^{*T} \Omega^{-1} r_I^* - r_I^{*T} N_m^{-1} r_I^*}{n + l - q - m} \tag{9.15}$$

F_I^* is distributed as a doubly noncentral $F - variable$ with m and $n + l - q - m$ degrees of freedom and noncentrality parameters given by

$$\begin{aligned} \nu &= \frac{1}{2}\left(0, \delta^T\right)(I - B)^T \begin{pmatrix} N_m^{-1} & 0 \\ 0 & 0 \end{pmatrix} (I - B) \begin{pmatrix} 0 \\ \delta \end{pmatrix} \\ & and \\ \lambda &= \frac{1}{2}\left(0, \delta^T\right)(I - B)^T \begin{pmatrix} \Lambda^{-1} - N_m^{-1} & 0 \\ 0 & R^{-1} \end{pmatrix} (I - B) \begin{pmatrix} 0 \\ \delta \end{pmatrix} \end{aligned}$$

Given that outlying points positions are not known, it is convenient to compute the maximum of the different $F - statistics$ and use the Bonferroni-upper bound with critical level $F\left(\alpha/\begin{pmatrix} n \\ m \end{pmatrix}\right)$.

In testing whether a single observation is an outlier, we observe that for this special case $m = 1$, and (9.14) reduces to

$$t_i^* = \frac{r_i^*}{s_{(i)}^* \sqrt{n_{ii}}} \tag{9.16}$$

with n_{ii} the *ith* diagonal element of $N = M\Omega M^T$ and

$$s_{(i)}^{*2} = \frac{{r^*}^T \Omega^{-1} r^* - \frac{r_i^{*2}}{n_{ii}}}{n + l - q - 1} \tag{9.17}$$

t_i^* has a doubly noncentral student $t - distribution$

Here again, if the position i of the outlying observation is not known, attention should be focused on $\max |t_i^*|$ with upper bound for the critical point given by

$$t^{\alpha/n}(n + l - q - 1)$$

The above is the upper α/n critical value of the central $t - distribution$ with $(n + l - q - 1)$ degrees of freedom.

Measure of influence

An appropriate measure of influence would be the Cook's distance (*Cook* (1977)) for the multiple case, given by

$$\Delta_I = \frac{\left(\beta_{(I)}^* - \beta^*\right)^T A^{-1} \left(\beta_{(I)}^* - \beta^*\right)}{q s_{(I)}^{*2}} \tag{9.18}$$

where $s_{(I)}^{*2}$and A^{-1} are as defined earlier.

For $\Lambda = I$, the above expression reduces to the following form

$$\Delta_I = \frac{1}{q s_{(I)}^{*2}} (r_I^*)^T N_m^{-1} X_I A X_I^T N_m^{-1} (r_I^*) \tag{9.19}$$

where

$$N_m^{-1} = \left(I - X_I A X_I^T\right)^{-1} \tag{9.20}$$

The single case, corresponding to $m = 1$, is given by

$$\begin{aligned} \Delta_i &= \frac{(\beta_i^* - \beta^*)^T A^{-1} (\beta_i^* - \beta^*)}{q s_i^{*2}} \\ &= \frac{(r_i^*)^2 w_i A w_i^T}{q n_{ii}^2 s_{(i)}^{*2}} \end{aligned} \tag{9.21}$$

with w_i the *ith* raw of W and $n_{ii} = 1 - w_i A w_i^T$

Chapter 10

Other techniques

In this chapter, other plotting techniques based on residuals, the Biplot, some additional influence related measures, as well as the Informational Complexity Criteria for regression models are reviewed .

10.1 Diagnostics based on residuals plots

A variety of plots have been widely used in regression diagnostics for the analysis of residuals. Three types of plots can indicate inaccuracy in a proposed model, and some trends or influential points in data. The first type is a plot of some kind of residuals against the index i; the second type is a plot of residuals against the independent variable w_{ji}; the third type is a plot of residuals against the predicted value $\hat{x}_i$. Figure 1 below shows possible graph shapes which can occur in plots of residuals:

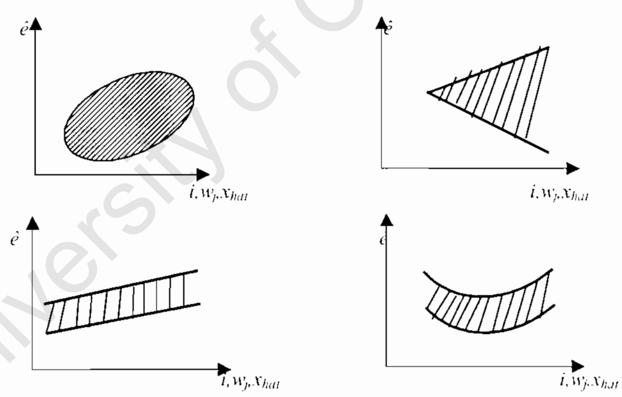


Fig.1: possible residuals' graph shapes

For the first and second type of plot, a band pattern (bottom left) indicates some error in calculation or the absence of w_j in the model: but a band pattern may also be caused by outlying points.

10.1.1 The Williams graph (William (1973))

This graph has the diagonal elements of the hat matrix v_{ii} on the x-axis and the Jackknife residuals $\hat{t}_i$ on the y-axis. Two boundary lines are drawn, the first for high-leverages, $v_{ii} = \frac{2m}{n}$ and the second for outliers, $t_i = t_{0.95}(n-m-1)$, where $t_{0.95}(n-m-1)$ is the 95% quantile of the student distribution with $(n-m-1)$ degrees of freedom; n and m are the sample size and the number of independent variables respectively. An illustration of Williams graph is given in figure 2 below:

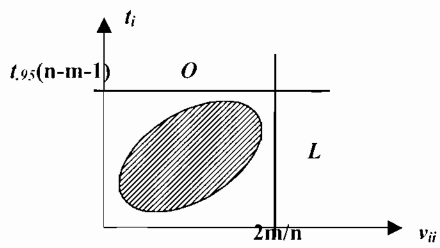


Fig.2 Williams graph: *O stands for outliers and L for leverage points.*

10.1.2 The Pregibon graph (Pregibon (1981))

Like the previous, this graph has the diagonal elements v_{ii} on the x-axis; but on the y-axis it has the square of the normalized residuals $\hat{e}_{N,i}$. Based on the validity of the expression $E\left(v_{ii} + \hat{e}_{N,i}^2\right) = \frac{m+1}{n}$, the following two constraining lines are drawn

$$\begin{aligned}(1) \quad y &= -x + \frac{2(m+1)}{n} \quad \textit{and}, \\ (2) \quad y &= -x + \frac{3(m+1)}{n}\end{aligned}$$

To distinguish among influential points, the following two rules are applied:

(a) A point is strongly influential if it is located above the upper line;

(b) A point is influential if it is located between the two lines.

In both cases (a) and (b) the influential point can be either an outlier or a high-leverage point.

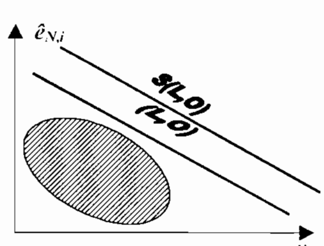


Fig.3 Pregibon graph: (L,O) are influential points and S(L.O) are strongly influential points.

Note: The normalized residuals, also known as scaled residuals, are defined as

$$\hat{e}_{N,i} = \frac{\hat{e}_i}{\hat{\sigma}} \tag{10.1}$$

It is falsely assumed that these residuals are normally distributed quantities with zero mean and variance equal to one, but in reality they have non-constant variance. When these residuals are used, it is strongly recommended to apply the classic rule of 3σ, that is quantities of $\hat{e}_{N,i}$ of magnitude greater than $\pm 3\sigma$ are classified as outliers; however. this approach is misleading at some extent and may cause wrong decision to be taken regarding data.

10.1.3 The McCulloch and Meeter graph (McCulloch and Meeter (1983))

This graph has $\ln\left[\frac{v_{ii}}{m(1-v_{ii})}\right]$ on the x-axis and the logarithm of square of the standardized residuals $\ln(\hat{r}_i^2)$ on the y-axis. In figure 4 below, the solid line drawn represents the locus of points with identical influence, with slope -1 and intercept given by $-\ln F_{0.9}(n-m,m)$. In other words, the equation of the 90% confidence line is given by

$$y = -x - \ln F_{0.9}(n-m,m)$$

The boundary line for high-leverage points is defined as

$$x = \ln\left[\frac{2}{n-m} t_{0.95}^2 (n-m-1)\right]$$

where $t_{0.95}^2(n-m-1)$ is the 95% quantile of the student distribution with $(n-m-1)$ degrees of freedom.

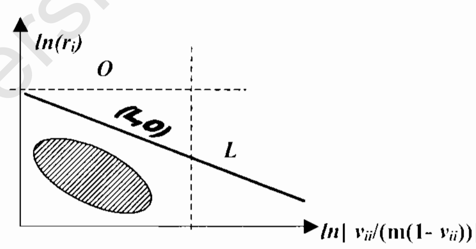


Fig.4: the McCulloch and Meeter graph

10.1.4 Gray's L-R graph (Gray (1983))

This graph has the diagonal elements v_{ii} on the x-axis and the squared normalized residuals $\hat{e}_{N,i}^2$ on the y-axis. All the points will lie under the hypotenuse of a triangle with the right angle in the origin of the two axes and the hypotenuse defined by the limiting equality

$$v_{ii} + \hat{e}_{N,i}^2 = 1 \tag{10.2}$$

Contours of the same critical influence are plotted in the Gray's L-R graph, and the locations of individual points are compared with them. It may be determined that the contours are hyperbolic as described by the equation

$$y = \frac{-(x-1)^2}{(1-k)x-1}$$

where $k = n(n-m-1)/c^2m$ and c is a constant usually equal to 2,4 or 8. Figure 5 below is an illustration of this graph:

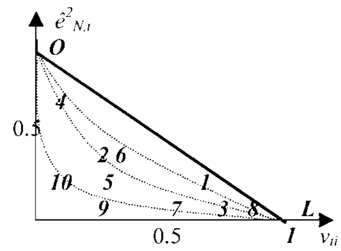


Fig 5. The Gray's L-R graph

The digit in the triangle stand for the order index i of the response variable x_i. Points towards to the upper part of the triangle are outliers, while points towards the right angle of triangle are high-leverages.

10.1.5 The predicted residuals graph

The predicted residuals, also known as the cross-validated residuals are defined as

$$\hat{e}_{p,i} = \frac{\hat{e}_i}{1 - v_{ii}} \tag{10.3}$$

The graph of the predicted residuals has the predicted residuals on the $x-$axis and the ordinary residuals $\hat{e}_i$ on the $y-$axis. High leverage points lie outside the line $y = x$ and are located quite far from this line whereas the outliers are located on the same line but far from its central pattern as shown in figure 6 below:

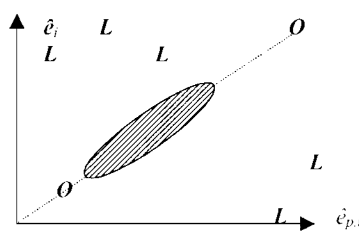


Fig. 6 The graph of predicted residuals

10.2 Biplot

A biplot is a graphical display of rows and columns of a rectangular (nxq) data matrix W, where the rows are often individuals or other sample units like cases, and the columns are variables. The biplot was introduced by Gabriel(1971) and is widely used and discussed by Greenacre and Underhill(1982), Greenacre(1984), and Gower and Hand (1996).

In almost all applications, this analysis starts with performing some transformation on W, depending on the nature of the data; to obtain a transformed matrix Z which is the one actually displayed.

Assume that the transformed matrix Z has rank r (and not q). Then Z can be factorized as

$$Z = FG^T \tag{10.4}$$

where F is (nxr) and G is (qxr).

The rows of F and rows of G provide the coordinates of n points for the rows and q points for the columns in an $r-$dimensional Euclidean space, called the *full space* since it has as many dimensions as the rank of Z. The joint plot of the two sets of points can be referred to as the exact biplot in the full space. There are an infinite number of ways to choose F and G, and certain choices favor the display of the rows, others the display of the columns. However, for any particular choice the biplot in r-dimensions has the property that the scalar product between the i^{th} row point and the j^{th} column point with respect to the origin is equal to the $(i,j)^{th}$ element z_{ij} of Z; that is if

$$F_{(nxr)} = \begin{pmatrix} f_1 \\ . \\ . \\ . \\ f_i \\ . \\ . \\ . \\ f_n \end{pmatrix} \quad \text{and } G_{(qxr)} = \begin{pmatrix} g_1 \\ . \\ . \\ g_j \\ . \\ . \\ g_p \end{pmatrix}$$

are rows and vectors then

$$z_{ij} = f_i g_j^T \tag{10.5}$$

We are mainly interested in low dimensional biplot of Z, especially in two dimensions. By means of appendix A, these can be conveniently achieved by using the singular value decomposition (SVD) of Z, namely

$$Z = UDV^T \tag{10.6}$$

where U and V are matrices of left and right singular vectors, each with r-orthonormal columns, and D is the diagonal matrix of positive singular values in decreasing order of magnitude, that is

$$d_1 \geq d_2 \geq ... \geq d_r > 0$$

The Eckart-Young theorem (Eckart and Young (1936)) states that if one calculates the $(n\text{x}r^*)$ matrix $Z_{[r^*]}$ using the first r^* singular values and corresponding singular vectors, then $Z_{[r^*]}$ is the least squares rank r^* approximation of Z. That is, over all possible matrices of rank r^*, $Z_{[r^*]}$ minimizes the fit criterion

$$\|Z - X\|^2 = \sum_i \sum_j (z_{ij} - x_{ij})^2$$

where $\|...\|$ denotes the Frobenius matrix norm. It is this approximate matrix $Z_{[r^*]}$ which is biplotted in the lower dimensional r^*-space, called the reduced space. This biplot will be as accurate as is the approximation of $Z_{[r^*]}$ to Z. The sums of squares of Z decomposes into two parts:

$$\|Z\|^2 = \left\|Z_{[r^*]}\right\|^2 + \left\|Z - Z_{[r^*]}\right\|^2$$

where

$$\begin{aligned} \left\|Z_{[r^*]}\right\|^2 &= d_1^2 + d_2^2 + ... + d_{r^*}^2 \text{ and} \\ \left\|Z - Z_{[r^*]}\right\|^2 &= d_{r^*+1}^2 + ... + d_r^2 \end{aligned} \tag{10.7}$$

The goodness-of-fit is thus measured by the proportion of explained sums of squares given by

$$\frac{d_1^2 + d_2^2 + ... + d_{r^*}^2}{d_1^2 + d_2^2 + ... + d_r^2}$$

usually expressed as a percentage.

The special case $r^* = 2$ is of interest in this work. In this case one has

$$\begin{aligned} Z_{[2]} &= (\mathbf{u}_1, \mathbf{u}_2) \begin{pmatrix} d_1 & o \\ 0 & d_2 \end{pmatrix} (\mathbf{v}_1, \mathbf{v}_2) \\ \left\| Z_{[2]} \right\|^2 &= d_1^2 + d_2^2 \\ \left\| Z - Z_{[2]} \right\|^2 &= d_3^2 + d_4^2 + ... + d_r^2 \end{aligned}$$

where $\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}_1$ *and* $\mathbf{v}_2$ are vectors.

and the goodness-of-fit in this case is given by

$$\frac{d_1^2 + d_2^2}{d_1^2 + d_2^2 + ... + d_r^2}$$

The natural choice of F and G for the biplot in two dimensions is provided by the SVD of Z, namely

$$Z_{[2]} = F_{[2]} G_{[2]}^T$$

with

$$F_{[2]} = (d_1^{\alpha} \mathbf{u}_1, d_2^{\alpha} \mathbf{u}_2) \quad , \quad G_{[2]} = \left(d_1^{1-\alpha} \mathbf{v}_1, d_2^{1-\alpha} \mathbf{v}_2\right)$$

for some constant α.

The common choices of α are 1 and 0, when the singular values are assigned entirely either to the left singular vectors of U or the right singular vectors of V, respectively. with $\alpha = 0.5$ the square roots of the singular values are split equally between left and right singular vectors. Each choice, while giving exactly the same matrix approximation, will highlight a different aspect of then data matrix.

The term "principal coordinates" refers to the singular vectors scaled by the singular values, while "standard coordinates" are the unscaled singular vectors U and V (Greenacre(1984)).

The most common biplot is of an individual(case)-by-variables data matrix W which has been transformed by centering with respect to column means $\overline{w}_j$ to give

$$\begin{aligned} Z &= [z_{ij}] \\ &= [w_{ij} - \overline{w}_j] \, , \quad i = 1, ... \quad j = 1, ... q \end{aligned} \tag{10.8}$$

Next, compute the SVD of Z to obtain

$$Z = UDV^T$$

- **Form Biplot:** $\alpha = 1$

$$F = [d_1\mathbf{u}_1, d_2\mathbf{u}_2]$$
$$G = [\mathbf{v}_1, \mathbf{v}_2] \quad (10.9)$$

where $\mathbf{u}_1, \mathbf{u}_2$, $\mathbf{v}_1$ *and* $\mathbf{v}_2$ are vectors.

The rows (cases) are the principal coordinates and the columns are in standard coordinates. This type of biplot favours the display of rows.

- **Covariance Biplot:** $\alpha = 0$

$$F = [\mathbf{u}_1, \mathbf{u}_2]$$
$$G = [d_1\mathbf{v}_1, d_2\mathbf{v}_2] \quad (10.10)$$

The rows are in standard coordinates and the columns are the principal coordinates.

- **Equal weigths Biplot:** $\alpha = 0.5$

$$F = \left[d_1^{1/2}\mathbf{u}_1, d_2^{1/2}\mathbf{u}_2\right]$$
$$G = \left[d_1^{1/2}\mathbf{v}_1, d_2^{1/2}\mathbf{v}_2\right] \quad (10.11)$$

Neither rows or columns are favoured. This biplot is often preferred when dealing with the correlation matrix as in factor analysis.

In what follows, of great interest will be the first and second type of biplot. The first will provide us with an optimal approximation of the form matrix ZZ^T by the corresponding form matrix FF^T of F; and the second, an optimal approximation of the covariance matrix $S = ZZ^T/(n-1)$ by the corresponding matrix $GG^T/(n-1)$.

Since the FORM biplot favours the display of the rows, the FORM biplot is used to identify outliers and influential observations. This can conveniently be done by mean of the SVD of the W matrix for influential observations and the SVD of the composite matrix (X, W) for outliers. A scatter plot with a confidence ellipse will immediately identify outlying and influential observations. The Biplot works very well to overcome the "masking" and/or "swamping" effects mentioned earlier in this work. The Biplot is a very good procedure for determining "multiple outliers".

10.3 Other Influence related measures

10.3.1 Hadi's influence measure (Hadi (1994))

This influence measure is based on the fact that potentially influential observations are outliers in the dependent or independent space. It is give by the expression

$$H_i^2 = \frac{m}{1 - v_{ii}} \frac{\hat{e}_{N,i}^2}{1 - \hat{e}_{N,i}^2} + \frac{v_{ii}}{1 - v_{ii}} \tag{10.12}$$

where $\hat{e}_{N,i}^2$ is the ith normalized residual. Large values of H_i^2 indicate influential points. Hadi (1992) provides extensions of H_i^2 for the multiple case, namely H_I^2; influential subsets indexed by the set I can be identified by means of the $\mathbf{J}_I$**−class** that the above author defines as follows

$$J_I = g\left(n, , q, m\right) f\left(K_I\left(u, v, w, a, b, c\right)\right), \tag{10.13}$$

where $g\left(n, , q, m\right)$ is a function that is independent of the elements of the subset whose influence we wish to monitor, $f\left(A\right)$ is a norm of the kernel matrix A, and

$$K_I\left(u, v, w, a, b, c\right) = \left(I_m - V_I\right)^{-u} V_I^a \left(I_m - Q_I\right)^{-v} Q_I^b \left(I_m - V_I - Q_I\right)^{-w} \left(V_I + Q_I\right)^c \tag{10.14}$$

is the kernel of the $J_I-class$ which depends on integer parameters u, v, w, a, b and c and on the following matrices: I_m, the identity matrix of order m; V_I as defined in the notation section and $Q_I = e_I\left(e_I^T e_I\right)^{-1} e_I^T$.

For $g\left(n, q, m\right) = \frac{q}{m}$, $\left(u, v, w, a, b, c\right) = \left(1, 1, 0, 0, 1, 0\right)$ and

$$f\left(K_I\left(u, v, w, a, b, c\right)\right) = tr\left[\left(I_m - V_I\right)^{-1}\left(I_m - Q_I\right)^{-1} Q_I\right] + tr\left[\left(I_m - V_I\right)^{-1} V_I\right]$$

Hadi's overall influence measure is given by

$$H_I^2 = \frac{q\left(X_I - W_I\hat{\beta}_{(I)}\right)^T \left(I_m - V_I\right)\left(X_m - W_I\hat{\beta}_{(I)}\right)}{e_{(I)}^T e_{(I)}} + \frac{1}{m} tr\left(W_I\left(W_{(I)}^T W_{(I)}\right)^{-1} W_I^T\right) \tag{10.15}$$

Note: Depending on the values given to the components of the above expression (10.5), different tests statistic can be derived using the $J_I - class$.

10.3.2 The Atkinson influence measure (Atkinson (1985))

This measure, known as the modified version of Cook's distance, enhances the sensitivity of distance measures to high-leverage points. It is defined as

$$A_i = |t_i| \sqrt{\frac{n-m}{m} \text{x} \frac{v_{ii}}{1-v_{ii}}} \qquad (10.16)$$

where $|t_i|$ is the absolute value of the jackknife residuals. This measure is also convenient for graphical interpretation and Atkinson (1985) recommends that absolute values of A_i be plotted in any of the ways customary for residuals.

The Atkinson influence measure is similar to the Belsey $DFFITS_i$ measure which is given by the expression

$$DFFITS_i = |t_i| \text{x} \frac{v_{ii}}{1-v_{ii}} \qquad (10.17)$$

The ith point is considered to be significantly influential if $DFFITS_i$ is larger in absolute value than $2\sqrt{\frac{m}{n}}$

10.4 Informational Complexity Criteria for regression

The Information Complexity Criteria (ICOMP), developed by Bozdogan (1994), is a new approach for detecting influential observations in linear models. It consists of two stages: in the first, Bozdogan's ICOMP criterion is used as the fitness function for variables' selection and in the second, it is used with case-deletion on the selected set of variables to detect influential cases. In what follows, we focus our attention on detecting influential observations using the ICOMP.

Consider the usual multiple regression model

$$X = W\beta + \varepsilon$$

as described in chapter 2. Without loss of generality, let assume that we have a given finite number of nested models of the form

$$m_k = \{\beta_0, \beta_1, ..., \beta_k, \sigma^2\}$$

the dimension of m_k as a model is equal to $(k+2)$. The ICOMP, as a model selection criterion, penalizes the covariance complexity of the model rather than the number of free parameters directly, as it is the case for the AIC and BIC criterion

$$AIC = -2\ln(L_M) + 2d, \tag{10.18}$$

$$BIC = -2\ln(L_M) + d\ln(n) \tag{10.19}$$

where L_M is the maximized likelihood.

The ICOMP criterion is defined by

$$ICOMP = -2\ln L\left(\hat{\theta}\right) + 2C\left(\hat{\sum}_{model}\right) \tag{10.20}$$

where L is the likelihood function, $\hat{\theta}$ is an estimator of the unknown parameter θ, C represents a complexity measure and $\hat{\sum}_{model}$ represents the estimated covariance matrix of the parameter vector estimated by the model. Originally Van Emden (1971) defined the covariance complexity as

$$C_0\left(\sum\right) = \frac{1}{2}\sum_{j=1}^{k}\ln\sigma_{jj}^2 - \frac{1}{2}\ln\left|\sum\right|, \tag{10.21}$$

where k is the dimension of $\sum$, and the σ_{jj}^2 are the diagonal elements of $\sum$. The reader should notice that $C_0\left(\sum\right) = 0$ when $\sum$ is a diagonal matrix. Based on the fact that $C_0\left(\sum\right)$ is not invariant under orthonormal transformations, Bozdogan (1990) introduces

$$C_1\left(\sum\right) = \max_T C_0\left(\sum\right) = \frac{k}{2}\ln\frac{tr\left(\sum\right)}{k} - \frac{1}{2}\ln\left|\sum\right|, \tag{10.22}$$

as a penalty functional, where T is the set of all orthonormal transformations. Van Emden (1971) and Bozdogan (1990) show that C_1 is invariant with respect to scalar multiplication and orthonormal transformation.

Formulation of ICOMP

In the context of multiple regression models with identical and independently distributed errors, the complexity of the model m_k is defined as

$$C_1(m_k) = C_1\left(\hat{\beta}_k, \hat{\varepsilon}\right)$$

where $\left(\hat{\beta}, \hat{\varepsilon}\right)$ is the vector of estimated parameters and residuals under m_k. Following Bozdogan (1990), given that the maximum likelihood or the least squares estimators $\hat{\beta}$ and $\hat{\varepsilon}$ are independent, the complexity C_1 satisfies the property

$$C_1\left(\hat{\beta}_k, \hat{\varepsilon}_k\right) = C_1\left(\hat{\beta}_k\right) + C_1\left(\hat{\varepsilon}_k\right) \tag{10.23}$$

The same author defines the complexity of $\hat{\beta}_k$ as

$$C_1\left(\hat{\sum}_{\hat{\beta}_k}\right) = C_1\left(\sigma^2\left(W_k^TW_k\right)^{-1}\right) = C_1\left(\left(W_k^TW_k\right)^{-1}\right) \tag{10.24}$$

Under the assumption that the random errors have $\sigma^2 I_n$ as their covariance matrix, the complexity of $\hat{\varepsilon}_k$ is given by

$$\begin{aligned} C_1\left(\hat{\varepsilon}_k\right) &= C_1\left(\sigma^2 I_{n-q}\right) \\ &= C_1\left(I_{n-q}\right) \\ &= 0 \end{aligned}$$

Expression (10.15) becomes

$$\begin{aligned} C_1\left(\hat{\beta}_k, \hat{\varepsilon}_k\right) &= C_1\left(\hat{\sum}_{\hat{\beta}_k}\right) \\ &= C_1\left(\left(W_k^TW_k\right)^{-1}\right) \end{aligned} \tag{10.25}$$

and the ICOMP for the multiple regression model can be written as

$$ICOMP\left(\hat{\beta}_k, \hat{\varepsilon}_k\right) = -2\ln\left(L_M\right) + 2C_1\left(\left(W_k^TW_k\right)^{-1}\right) \tag{10.26}$$

It can be shown that this expression is equivalent to

$$\begin{aligned} ICOMP\left(\hat{\beta}_k, \hat{\varepsilon}_k\right) &= n\ln\left(2\pi\right) + n\ln\left(\frac{RSS_k}{n}\right) + n \\ &\quad + q\ln\left[\frac{tr\left(W_k^TW_k\right)^{-1}}{q}\right] - \ln\left[\det\left(W_k^TW_k\right)^{-1}\right] \end{aligned} \tag{10.27}$$

with respect to the previous notation.

Another approach to the formulation of the ICOMP considers the complexity of the vector of estimated parameters $\left(\hat{\beta}_k, \hat{\sigma}_k^2\right)$ with $\hat{\sigma}_k^2 = \frac{RSS_k}{n}$. $\hat{\beta}_k$ and $\hat{\sigma}_k^2$ are independent and their covariance matrix is given by

$$Q = \begin{pmatrix} \sigma^2\left(W_k^TW_k\right)^{-1} & 0 \\ 0 & 2\sigma^4\left(\frac{n-q}{n^2}\right) \end{pmatrix} \tag{10.28}$$

Q is asymptotically equivalent, as $n \to \infty$, to the Inverse-Fisher Information Matrix (IFIM) given by

$$F^{-1} = \begin{pmatrix} \sigma^2\left(W_k^TW_k\right)^{-1} & 0 \\ 0 & 2\sigma^4/n \end{pmatrix} \tag{10.29}$$

Hence, we can use $\hat{Q}$ or $\hat{F}^{-1}$ to define the complexity of the vector $\left(\hat{\beta}_k, \hat{\sigma}_k^2\right)$, namely $C_1\left(\hat{Q}\right)$ or $C_1\left(\hat{F}^{-1}\right)$ respectively; this leads to the following two other versions of ICOMP

$$\begin{aligned} ICOMPIFIM\left(\hat{\beta}_k, \hat{\sigma}_k^2\right) &= -2\ln\left(L_M\right) + 2C_1\left(\hat{F}^{-1}\right) \\ &= n\ln\left(2\pi\right) + n\ln\left[\frac{RSS_k}{n}\right] + n \\ &\quad + (q+1)\ln\left[\frac{tr\left(W_k^T W_k\right)^{-1} + 2\hat{\sigma}_k^2/n}{q+1}\right] \\ &\quad - \ln\det\left(\left(W_k^T W_k\right)^{-1}\right) - \ln\left(\frac{2\hat{\sigma}_k^2}{n}\right) \end{aligned} \tag{10.30}$$

and

$$\begin{aligned} ICOMPCOV\left(\hat{\beta}_k, \hat{\sigma}_k^2\right) &= -2\ln\left(L_M\right) + 2C_1\left(\hat{Q}\right) \\ &= n\ln\left(2\pi\right) + n\ln\left[\frac{RSS_k}{n}\right] + n \\ &\quad + (q+1)\ln\left[\frac{tr\left(W_k^T W_k\right)^{-1} + 2\hat{\sigma}_k^2\left(\frac{n-q}{n^2}\right)}{q+1}\right] \\ &\quad - \ln\det\left(\left(W_k^T W_k\right)^{-1}\right) - \ln\left(2\hat{\sigma}_k^2\left(\frac{n-q}{n^2}\right)\right) \end{aligned} \tag{10.31}$$

ICOMP as an influence measure

The approach of detecting influential observations using ICOMP takes into account both lack-of-fit and model complexity in one criterion function. As in the previous cases, an observation is said to be influential, if deleting it from the sample causes the ICOMP score to fall, that is if

$$\Delta ICOMP = ICOMP_{full-data}\left(m_k\right) - ICOMP_{(i)}\left(m_k\right) > 0 \tag{10.32}$$

where $ICOMP_{full-data}\left(m_k\right)$ is the IMCOMP for the subset m_k when the full data set is used, and $ICOMP_{(i)}\left(m_k\right)$ is the ICOMP for the same subset m_k when observation i is deleted. This measure of influence can be extended to the multiple case.

Chapter 11

CASE STUDIES

In this chapter we present some case studies using 4 different data sets that we shortly describe below. As we mentioned it earlier, we will combine different techniques to identify outliers and influential observations (if any) in each of the data sets. In particular graphing techniques will be used when deletion of observations is not convenient. The complete data sets are displayed in appendix D.

Data set 1: The Siml4 data set

This data set of 40 observations and 5 independent variables was generated by Jacobs, M.(1982) as described below.

The first independent variable W_1 is a trend variable, namely $W_i = i$ for $i = 1, ..., n$. The variables W_2, W_3 and W_4 are independently normally distributed, with $W_2 \sim N(5,1)$ and $W_3,\ W_4 \sim N(10,1)$. An approximate linear relationship between the independent variables was introduced by making $W_5 = W_3 + W_4 + e$ where $e \sim N(0, 0.1^2)$. The dependent variable was generated as

$$X = 100 + 2W_1 + 10W_2 + 10W_3 + 10W_4 + 10W_5 + e^*$$

where $e^* \sim N(0, 25)$

The following extreme cases were introduced into the data set: in observation 1, the variable W_2 was given a value of 10 (W_2 is usually centered on 5); in observations 2 and 3, W_3 and W_4 were centred on 15 instead of 10, and corresponding adjustments were made in the values of W_5 and X. In addition a possible outlier was introduced by substracting 10 from the X value of observation 2 only. Finally, observation 4 is made to be a straightforward outlier by adding 20 to its X value.

Data set 2: Building Society Staff Requirements (Bld.soc)

The size of this data set is 48 cases x 9 variables:

- X =Clerical staff

- W_1=Banking hall transactions (in thousands)
- W_2=Number of saving cheques drawn (in thousands)
- W_3=Investor transactions (in thousands)
- W_4=Loans transactions (in thousands)
- W_5=Insurance transactions (in thousands)
- W_6=Loans information transactions (in thousands)
- W_7=Name and address capture (in thousands)
- W_8=Other general information capture (in thousands)

X is the dependent variable and all the remaining are independent variables.

Data set 3: The JSEANG data set

This is made of data that allow to model the Johannesburg Stock Exchange as a function of different indices, the following are the variables,

- X =Jse
- W_1=ang
- W_2=dmark$
- W_3=gold_r
- W_4=jelec
- W_5=jfood
- W_6=jinsr
- W_7=jplat
- W_8=yen_$

Data set 4: The ANGJSE data set

The data set is the same as above except that $\mathbf{W}_1$ becomes the dependent variable and X an independent variable.

The numerical results for each case study were found using R-codes (appendix E). For quality reasons, most graphs in this section were plotted in Eviews.

From the graph displayed below (see Jambu (1991)) , observations 1, 2 and 3 may be suspected as being outliers, since they lie far outside the 95% confidence ellipse; we therefore use this information to run the R-program and find out whether these observations are in fact outliers, both individually and as a group. The results are presented in tables showing values of some statistics computed in this work, as well as the changes in some regression parameters owing to the deletion of different combinations of the suspected outliers.

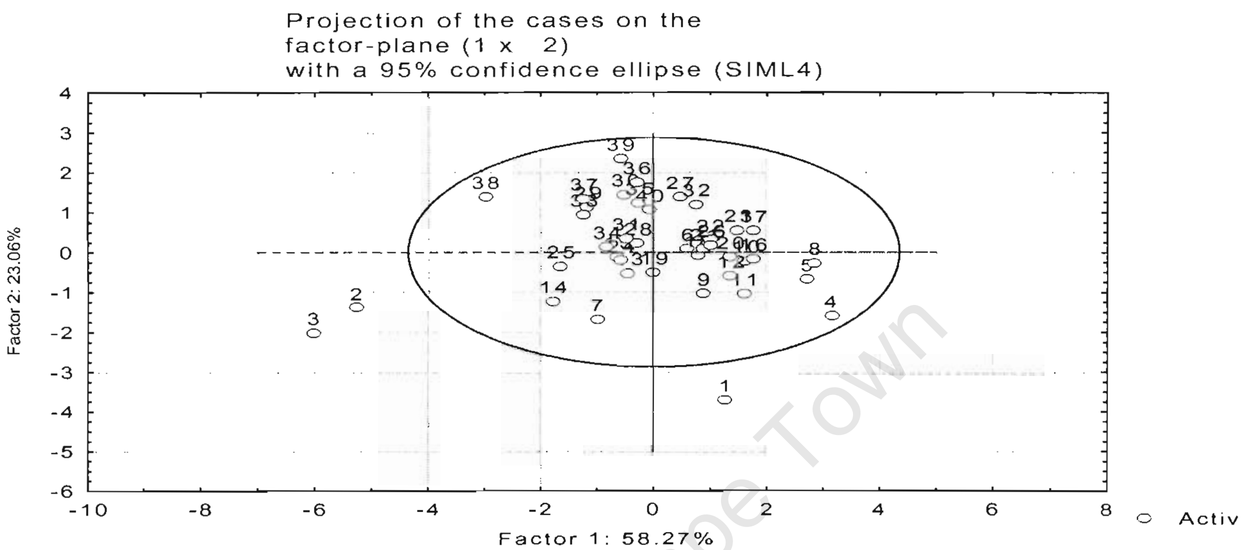


fig. 7: Projection of cases on the factor plane with a 95% confidence ellipse.

TABLE 1: SIML4 OUTPUT

	1 i	2 sdz	3 stz	4 F.i.del	5 vii	6 icomp	7 icomp.i.del	8 delta.icomp	9 F-test
1	1	-0.38	-0.37	0.14	0.46	279.76	200.84	78.92	
2	2	-2.03	-2.13	4.54	0.41	279.76	195.91	83.85	*
3	3	2.37	2.56	6.5	0.4	279.76	193.84	85.92	*
4	4	3.24	3.84	14.57	0.26	279.76	186.24	93.52	*
5	5	0.21	0.2	0.04	0.14	279.76	200.94	78.82	
6	6	-0.51	-0.51	0.25	0.24	279.76	200.7	79.06	
7	7	-0.01	-0.01	0	0.13	279.76	201.02	78.74	
8	8	0.27	0.26	0.07	0.21	279.76	200.91	78.85	
9	9	-1.47	-1.49	2.24	0.08	279.76	198.39	81.37	
10	10	-0.72	-0.72	0.51	0.14	279.76	200.39	79.37	
11	11	-0.61	-0.6	0.37	0.06	279.76	200.57	79.19	
12	12	-0.26	-0.26	0.07	0.06	279.76	200.93	78.83	
13	13	1.15	1.16	1.34	0.06	279.76	199.43	80.33	
14	14	-0.39	-0.39	0.15	0.09	279.76	200.86	78.9	
15	15	0.06	0.06	0	0.05	279.76	201.03	78.73	
16	16	-0.84	-0.83	0.7	0.08	279.76	200.18	79.58	
17	17	-0.79	-0.79	0.62	0.18	279.76	200.28	79.48	
18	18	-0.41	-0.4	0.16	0.09	279.76	200.83	78.93	
19	19	-0.66	-0.65	0.43	0.19	279.76	200.53	79.23	
20	20	-1.05	-1.05	1.11	0.1	279.76	199.71	80.05	
21	21	0.14	0.13	0.02	0.08	279.76	201.01	78.75	
22	22	1.08	1.09	1.17	0.06	279.76	199.63	80.13	
23	23	0.34	0.34	0.11	0.25	279.76	200.9	78.86	
24	24	-0.18	-0.18	0.03	0.23	279.76	201.02	78.74	
25	25	0.14	0.13	0.02	0.1	279.76	201.05	78.71	
26	26	-0.31	-0.31	0.09	0.1	279.76	200.94	78.82	
27	27	-0.29	-0.29	0.08	0.13	279.76	200.96	78.8	
28	28	0.29	0.28	0.08	0.07	279.76	200.97	78.79	
29	29	0.61	0.6	0.37	0.08	279.76	200.65	79.11	
30	30	1.83	1.9	3.6	0.1	279.76	196.92	82.84	
31	31	0.93	0.93	0.86	0.07	279.76	200.06	79.7	
32	32	0.83	0.82	0.68	0.09	279.76	200.27	79.49	
33	33	0.67	0.66	0.44	0.14	279.76	200.58	79.18	
34	34	-2.36	-2.54	6.43	0.16	279.76	193.96	85.8	*
35	35	-0.91	-0.91	0.82	0.09	279.76	200.11	79.65	
36	36	-0.46	-0.46	0.21	0.2	279.76	200.86	78.9	
37	37	0.2	0.2	0.04	0.11	279.76	201.08	78.68	
38	38	0.63	0.63	0.39	0.18	279.76	200.67	79.09	
39	39	-0.73	-0.73	0.53	0.15	279.76	200.49	79.27	
40	40	0.48	0.47	0.23	0.15	279.76	200.87	78.89	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

TABLE 2:CHANGES IN SOME REGRESSION PARAMETERS (SIML4)

	FULL MODEL	(1,2) DEL	(1,3) DEL	(2,3) DEL	(1,2,3) DEL
Intercept	113.473	104.883	122.714	117.291	116.47
b1	2.039	1.959	2.096	2.046	2.044
b2	9.934	9.732	10.016	9.675	9.831
b3	22.881	22.697	23.98	24.142	23.768
b4	24.032	24.809	25.738	26.131	25.829
b5	-4.077	-3.805	-6.019	-5.888	-5.544
R-squared	0.9893	0.9895	0.9886	0.9871	0.9871
F-statistic	626.2	604.5	553.4	490.4	474.7
DF	5 and 34	5 and 32	5 and 32	5 and 32	5 and 31
p-value	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16

Comments (SIML4 OUTPUT)

Table 1 indicates that of the 3 observations suspected as outliers, the test statistic failed for case 1 and was only significant for cases 2 and 3 for $\alpha = 5\%$. However, case 1 seems to be a high leverage point with the largest hat value of 0.46. Observation 4 should be given a sustained attention as well as a careful study, since it lies near the 95% confidence ellipse and its deletion leads the highest change in the ICOMP (delta.icomp=93.52); In fact, from the simulation of this data set, observation 4 was constructed to be a straightforward outlier. For case 34, there is not enough evidence that it is either an outlier or influential observation. Deletion of different combinations of the suspected observations does not considerably change most regression parameters (see table 2 above).

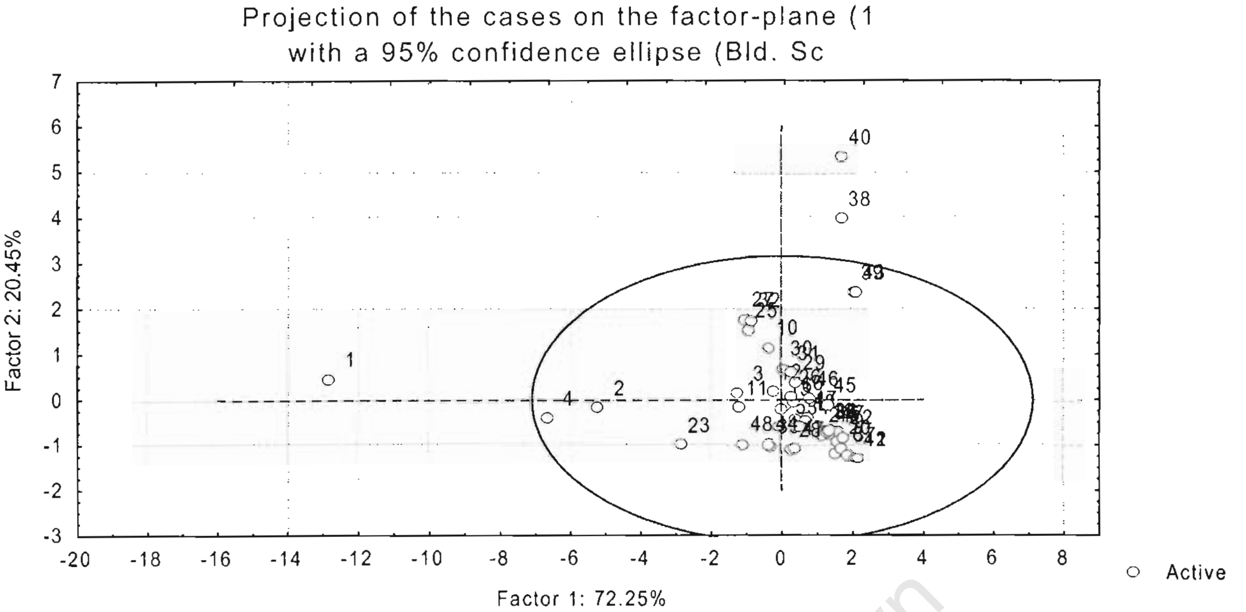


Fig 8.: Projection of cases on the factor plane with a 95% confidence ellipse.

TABLE 3: BLD.SOC OUTPUT

i	sdz	stz	F.i.del	vii	icomp	icomp.i.del	delta.icomp	F-test
1	-1.22	-1.23	1.51	0.6	390.24	268.84	121.4	
2	4.11	5.39	28.51	0.43	390.24	240.17	150.07	*
3	0.43	0.43	0.18	0.33	390.24	266.56	123.68	
4	-1.01	-1.01	1.02	0.37	390.24	266.31	123.93	
5	0.49	0.49	0.24	0.04	390.24	266.3	123.94	
6	-0.42	-0.42	0.17	0.2	390.24	266.4	123.84	
7	0.19	0.19	0.04	0.1	390.24	266.51	123.73	
8	-0.15	-0.15	0.02	0.06	390.24	266.53	123.71	
9	0.32	0.31	0.1	0.05	390.24	266.43	123.81	
10	-2.3	-2.45	5.94	0.2	390.24	259.68	130.56	*
11	0.41	0.4	0.16	0.12	390.24	266.51	123.73	
12	0.25	0.25	0.06	0.15	390.24	266.51	123.73	
13	1.44	1.46	2.13	0.14	390.24	263.97	126.27	
14	-1.35	-1.36	1.86	0.1	390.24	264.29	125.95	
15	0.49	0.49	0.24	0.05	390.24	266.26	123.98	
16	-0.19	-0.19	0.04	0.08	390.24	266.56	123.68	
17	0.05	0.05	0	0.18	390.24	266.56	123.68	
18	0.59	0.58	0.34	0.05	390.24	266.15	124.09	
19	-0.15	-0.14	0.02	0.05	390.24	266.53	123.71	
20	-0.76	-0.75	0.57	0.06	390.24	265.84	124.4	
21	-0.13	-0.13	0.02	0.06	390.24	266.54	123.7	
22	-0.12	-0.12	0.01	0.05	390.24	266.53	123.71	
23	-0.49	-0.49	0.24	0.71	390.24	266.53	123.71	
24	-0.14	-0.13	0.02	0.03	390.24	266.53	123.71	
25	-0.9	-0.9	0.81	0.16	390.24	265.62	124.62	
26	-0.37	-0.37	0.13	0.1	390.24	266.41	123.83	
27	2.07	2.17	4.68	0.24	390.24	261.03	129.21	*
28	0.58	0.57	0.33	0.14	390.24	266.18	124.06	
29	-1.96	-2.03	4.14	0.14	390.24	261.62	128.62	*
30	-0.85	-0.85	0.72	0.14	390.24	265.71	124.53	
31	0.79	0.78	0.62	0.36	390.24	265.84	124.4	
32	0.77	0.76	0.59	0.88	390.24	265.85	124.39	
33	1.49	1.52	2.29	0.23	390.24	263.79	126.45	
34	-0.52	-0.51	0.27	0.24	390.24	266.26	123.98	
35	-0.37	-0.37	0.13	0.04	390.24	266.39	123.85	
36	0.48	0.47	0.23	0.14	390.24	266.27	123.97	
37	0.02	0.02	0	0.07	390.24	266.55	123.69	
38	1.37	1.38	1.92	0.26	390.24	264.2	126.04	
39	-0.97	-0.97	0.94	0.12	390.24	265.37	124.87	
40	0.18	0.18	0.03	0.42	390.24	266.51	123.73	
41	-0.5	-0.49	0.25	0.07	390.24	266.24	124	
42	0	0	0	0.06	390.24	266.55	123.69	
43	-0.85	-0.85	0.72	0.19	390.24	265.65	124.59	
44	-1.23	-1.24	1.53	0.3	390.24	264.71	125.53	
45	1.07	1.07	1.15	0.06	390.24	265.12	125.12	
46	0.87	0.87	0.75	0.21	390.24	265.62	124.62	
47	0.49	0.48	0.24	0.1	390.24	266.3	123.94	
48	-1.24	-1.25	1.56	0.13	390.24	264.76	125.48	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

TABLE 4: CHANGES IN SOME REGRESSION PARAMETERS (BLD.SOC)

4.1 Full Model

Residuals:

Min	1Q	Median	3Q	Max
-26.8651	-7.0984	-0.7301	6.0419	40.4265

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	3.2031	3.2386	0.989	0.32875
W 1	0.5435	0.1721	3.158	0.00307**
W 2	-5.2677	2.2554	-2.336	0.02475*
W 3	0.8175	4.9872	0.164	0.87065
W 4	11.5285	4.1293	-0.752	0.45657
W 5	-3.2608	4.3362	-0.752	0.45657
W 6	-4.5805	2.896	-1.582	0.1218
W 7	-0.1839	0.8848	-0.208	0.83646
W 8	4.1591	0.6533	6.366	1.61e-07***

Signif. Codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.02 on 39 degrees of freedom

Multiple R-squared: 0.98, Adjusted R-squared: 0.9759

F-statistic: 238.6 on 8 and 39 DF, p-value: <2.2e-16

4.2 Cases 2 and 10 deleted

Residuals:

Min	1Q	Median	3Q	Max
-19.9811	-5.1025	-0.6385	5.3933	18.1298

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	3.5538	2.2704	1.565	0.12603
W 1	0.5866	0.1251	4.69	3.66e-05***
W 2	-1.4174	1.8058	-0.785	1.46e-06***
W 3	8.9323	3.7136	2.405	0.02127
W 4	17.3448	3.0272	5.73	1.46e-06***
W 5	-3.1442	3.0358	-1.036	0.30706
W 6	-5.8432	2.0377	-2.868	0.00679**
W 7	0.787	0.6406	1.229	0.22699
W 8	2.7472	0.506	5.429	3.73e-06***

Signif. Codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.116 on 37 degrees of freedom

Multiple R-squared: 0.9887, Adjusted R-squared: 0.9862

F-statistic: 404.3 on 8 and 37 DF, p-value: <2.2e-16

4.3 Cases 2 and 27 deleted

Residuals:

Min	1Q	Median	3Q	Max
-23.331	-4.7809	0.6227	4.7353	20.5441

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	3.8247	2.3725	1.612	0.115439
W1	0.5956	0.1338	4.452	7.56e-05***
W2	-1.2019	1.8739	-0.641	0.52524
W3	6.784	3.8829	1.747	0.088906
W4	13.6439	3.2769	4.164	0.00018***
W5	-0.7982	3.3206	-0.24	0.811376
W6	0.0609	0.6741	0.09	0.928501
W7	2.9065	0.5195	5.595	2.22e-06***
W8	2.7472	0.506	5.429	3.73e-06***

Signif. Codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.471 on 37 degrees of freedom

Multiple R-squared: 0.9876, Adjusted R-squared: 0.9849

F-statistic: 367.1 on 8 and 37 DF, p-value: <2.2e-16

4.4 Cases 2 and 29 deleted

Residuals:

Min	1Q	Median	3Q	Max
-22.2505	-3.8764	0.4568	5.5888	17.0623

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	3.4758	2.3258	1.494	0.14354
W 1	0.4074	0.1288	3.163	0.00312**
W 2	-0.4078	1.8203	-0.224	0.82399
W 3	6.5891	3.8338	1.719	0.09403
W 4	16.9055	3.0859	5.478	3.20e-06***
W 5	-5.1501	3.226	-1.596	0.1189
W 6	-5.0329	2.0888	-2.409	0.02107*
W 7	0.4466	0.6437	0.694	0.49213
W 8	3.2864	0.5246	6.265	2.7e-07***

Signif. Codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.341 on 37 degrees of freedom

Multiple R-squared: 0.9881 Adjusted R-squared: 0.9856

F-statistic: 385.2 on 8 and 37 DF, p-value: <2.2e-16

4.5 Cases 10 and 27 deleted

Residuals:

Min	1Q	Median	3Q	Max
-17.941	-6.445	-1.3	6.273	36.234

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	4.4036	2.9121	1.512	0.13899
W 1	0.7757	0.168	4.618	4.56e-05***
W 2	-6.7617	2.0565	-3.288	0.00222**
W 3	1.0779	4.5056	0.239	0.81224
W 4	10.6279	3.9249	2.708	0.01019*
W 5	-0.396	4.0691	-0.097	0.92299
W 6	-4.8743	2.5916	-1.881	0.06789
W 7	-0.1855	0.8289	-0.224	0.82414
W 8	3.6378	0.6022	6.041	5.52e-07***

Signif. Codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.61 on 37 degrees of freedom

Multiple R-squared: 0.9847, Adjusted R-squared: 0.9813

F-statistic: 296.7 on 8 and 37 DF, p-value: <2.2e-16

4.6 Cases 10 and 29 deleted

Residuals:

Min	1Q	Median	3Q	Max
-18.2274	-6.7554	-0.9714	5.1224	37.8208

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	3.8723	2.9375	1.318	0.195529
W 1	0.5508	0.1679	3.28	0.002265**
W 2	-5.8999	2.0717	-2.848	0.007145**
W 3	0.8368	4.595	0.182	0.856499
W 4	14.3831	3.8375	3.748	0.000608***
W 5	-5.535	4.068	-1.361	0.181861
W 6	-4.6843	2.636	-1.777	0.083784
W 7	4.1168	0.6226	6.612	9.37e-08***
W 8	4.8168	0.2662	6.102	5.52e-07***

Signif. Codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.78 on 37 degrees of freedom

Multiple R-squared: 0.9844, Adjusted R-squared: 0.981

F-statistic: 292.1 on 8 and 37 DF, p-value: <2.2e-16

4.7 Cases 27 and 29 deleted

Residuals:

Min	1Q	Median	3Q	Max
-25.93	-7.384	-1.137	5.98	37.491

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	4.1472	3.038	1.365	0.1805
W1	0.5712	0.1798	3.177	0.003**
W2	-5.6378	2.1245	-2.654	0.0117*
W3	-1.3357	4.708	-0.284	0.7782
W4	10.1875	4.1085	2.48	0.0178*
W5	-4.0345	2.7046	-1.492	0.1443
W6	-4.6843	1.363	-2.013	0.083784
W7	-0.5353	0.8488	-0.631	0.5321
W8	4.2498	0.634	6.703	7.08e-08***

Signif. Codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.12 on 37 degrees of freedom

Multiple R-squared: 0.9833, Adjusted R-squared: 0.9796

F-statistic: 271.7 on 8 and 37 DF, p-value: <2.2e-16

4.8 Cases 2, 10 and 27 deleted

Residuals:

Min	1Q	Median	3Q	Max
-16.737	-3.615	-1.109	4.546	18.747

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	4.1817	2.147	1.948	0.0593
W 1	0.6897	0.1248	5.528	2.97e-06***
W 2	-2.1441	1.7212	-1.246	0.2209
W 3	7.7153	3.5219	2.191	0.035***
W 4	15.0296	2.9957	5.017	1.43e-05***
W 5	-0.8505	3.0007	-0.283	0.7785
W 6	-5.6306	1.9151	-2.94	0.0057**
W 7	0.4166	0.6202	0.672	0.5061
W 8	2.6557	0.4766	5.573	2.59e-06***

Signif. Codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.559 on 36 degrees of freedom

Multiple R-squared: 0.9901, Adjusted R-squared: 0.9879

F-statistic: 449.6 on 8 and 36 DF, p-value: <2.2e-16

4.9 Cases 2, 10 and 29 deleted

Residuals:

Min	1Q	Median	3Q	Max
-16.6492	-4.7799	-0.2441	5.3401	15.4017

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	3.7968	2.1216	1.79	0.081927
W1	0.5005	0.1216	4.117	0.000214***
W2	-1.3042	1.6862	-0.773	0.444305
W3	7.5613	3.5083	2.155	0.037902*
W4	18.2079	2.8461	6.397	2.05e-07***
W5	-5.1231	2.9388	-1.743	0.089828
W6	-5.4547	1.9083	-2.858	0.007036**
W7	0.7897	0.5979	1.321	0.19494
W8	3.0334	0.4856	6.247	3.26e-07***

Signif. Codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.51 on 36 degrees of freedom

Multiple R-squared: 0.9904, Adjusted R-squared: 0.9883

F-statistic: 464.4 on 8 and 36 DF, p-value: <2.2e-16

4.10 Cases 2, 27 and 29 deleted

Residuals:

Min	1Q	Median	3Q	Max
-22.5096	-3.8766	0.7951	5.5228	18.8526

Coefficients:

	Estimatc	Std. Error	t value	Pr(>|t|)
Intercept	3.9603	2.2628	1.75	0.08862.
W 1	0.5034	0.1344	3.745	0.00063***
W 2	-1.0155	1.7887	-0.568	0.57374
W 3	5.7526	3.7324	1.541	0.132
W 4	14.8532	3.1736	4.68	3.98e-05***
W 5	-2.9808	3.322	-0.897	0.37553
W 6	-4.904	2.0204	-2.427	0.02035*
W 7	0.1349	0.6436	0.21	0.83515
W 8	3.1749	0.5105	6.219	3.54c-07***

Signif. Codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.03 on 36 degrees of freedom

Multiple R-squared: 0.989, Adjusted R-squared: 0.9865

F-statistic: 404.1 on 8 and 36 DF, p-value: <2.2e-16

4.11 Cases 10, 27 and 29 deleted

Residuals:

Min	1Q	Median	3Q	Max
-16.14	-6.865	-1.325	5.583	35.833

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	4.52491	2.83606	1.595	0.119345
W 1	0.67876	0.17277	3.929	0.000371***
W 2	-6.48971	2.00828	-3.231	0.002634
W 3	0.08054	4.42376	0.018	0.985576
W 4	11.84441	3.88445	3.049	0.004286**
W 5	-2.5955	4.15793	-0.624	0.536413
W 6	-4.558	2.52967	-1.802	0.079954.
W 7	-0.11698	0.80795	-0.145	0.885685
W 8	3.90588	0.60613	6.444	1.78e-07***

Signif. Codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.3 on 36 degrees of freedom

Multiple R-squared: 0.9858, Adjusted R-squared: 0.9827

F-statistic: 313.1 on 8 and 36 DF, p-value: <2.2e-16

4.12 **Cases 2, 10, 27 and 29 deleted**

Residuals:

Min	1Q	Median	3Q	Max
-15.865022	-3.739099	-0.004101	4.881405	17.216963

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	4.2973	2.0339	2.113	0.04182*
W1	0.6	0.1246	4.816	2.79e-05***
W2	-1.9378	1.6325	-1.187	0.24322
W3	6.7155	3.3643	1.996	0.05375.
W4	16.1198	2.8774	5.602	2.58e-06***
W5	-2.8993	2.9818	-0.972	0.33756
W6	-5.3275	1.8185	-2.93	0.00594**
W7	0.4743	0.5879	0.807	0.42521
W8	2.9161	0.4657	6.262	3.49e-07***

Signif. Codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.105 on 35 degrees of freedom

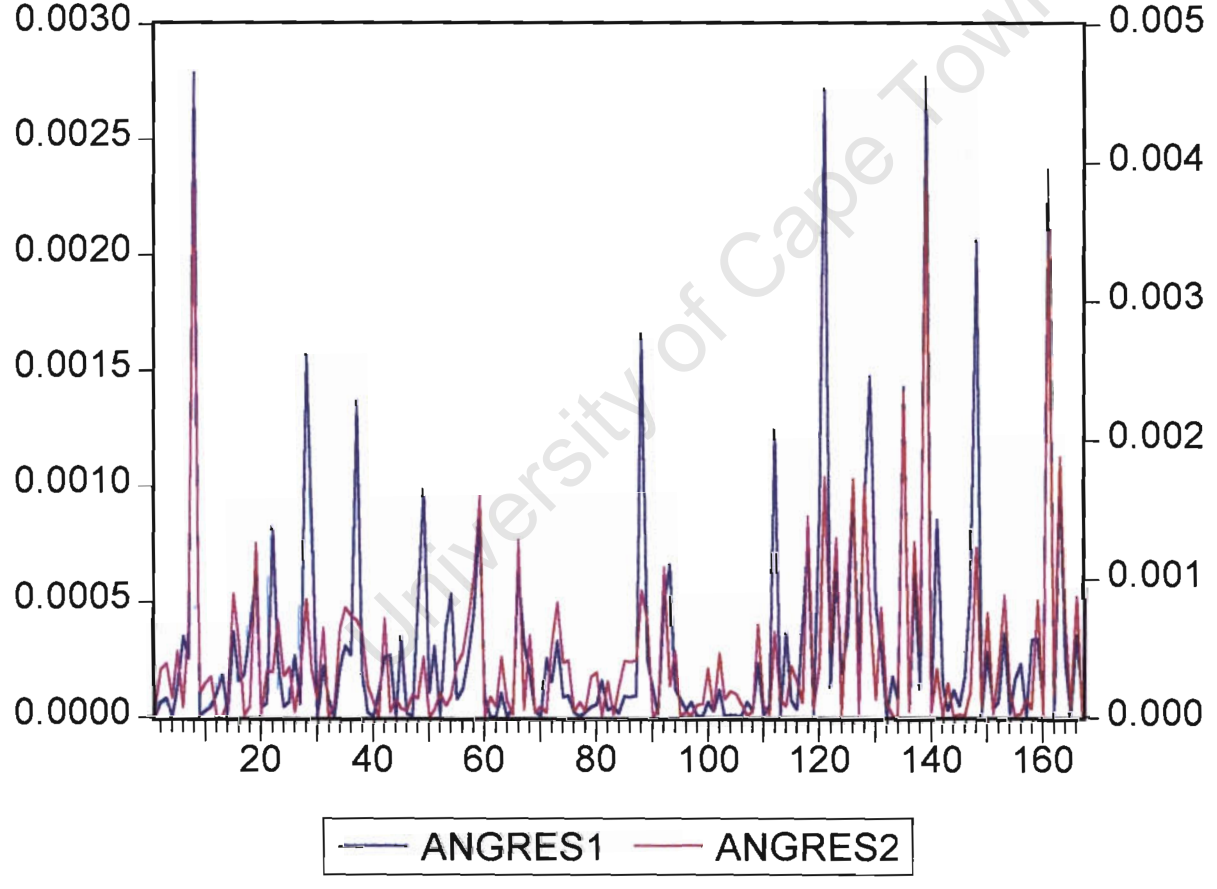
Multiple R-squared: 0.9914, Adjusted R-squared: 0.9894

F-statistic: 501.7 on 8 and 35 DF, p-value: <2.2e-16

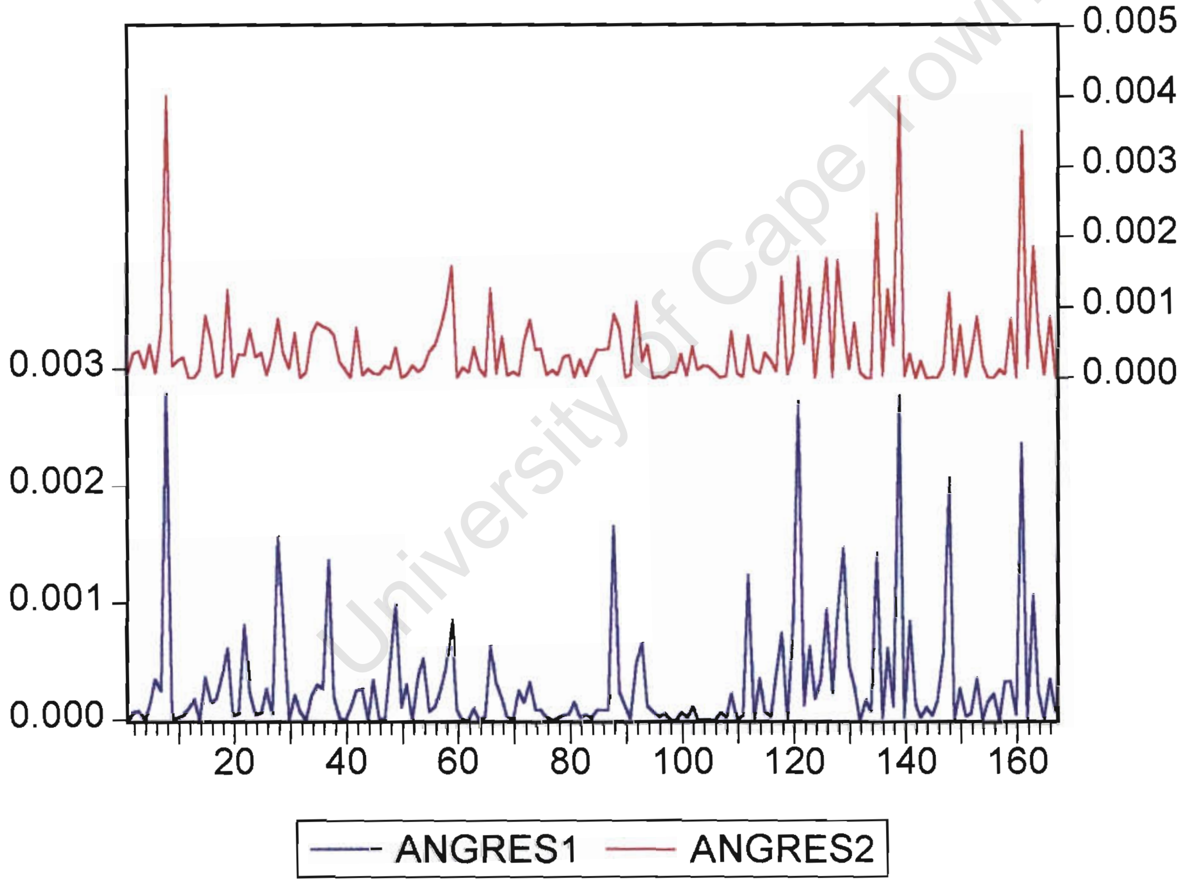
Comments (BLD.SOC OUTPUT)

For this data set the outcomes of the graphical display and the non-parametric tests introduced on pages 21 and 26 are completely different. Observation 1 has a high hat value of 0.6, and can be regarded as a high leverage point since it lies far outside the 95% confidence ellipse. However, case 1 was not identified as an outlier by the non-parametric test statistics. Like in the previous case, there is not enough evidence that observations 38 and 40 are either outliers or influential observations (see the results of table 3 above). The individual test statistics for cases 2, 10, 27 and 29 were significant at $\alpha = 5\%$ and the deletion of these observations leads in respective order the following 4 highest changes in the ICOMP: 150.07, 130.56, 129.21 and 128.62 respectively. In the same order, they also have the largest studentized residuals (stz) in absolute value . To find out whether these observations can form some outlying groups, we look at the changes of some regression parameters resulting from the deletion of different combinations of observations 2, 10, 27 and 29. In all 11 cases the changes are dramatic and significant, especially in the last case where all 4 observations have been deleted.

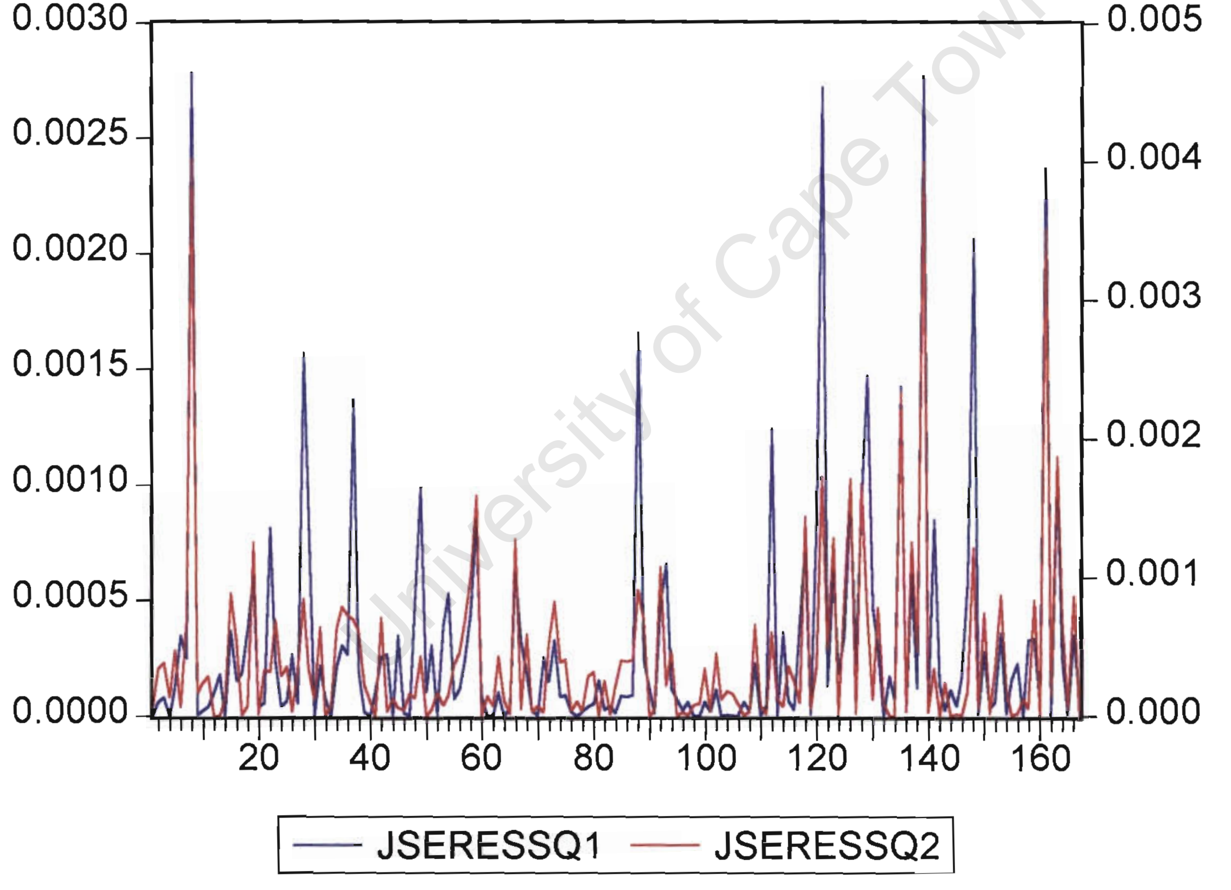
Graphs of Least Squares and Principal Component Squared Residuals of ANGLOS fit on JSE



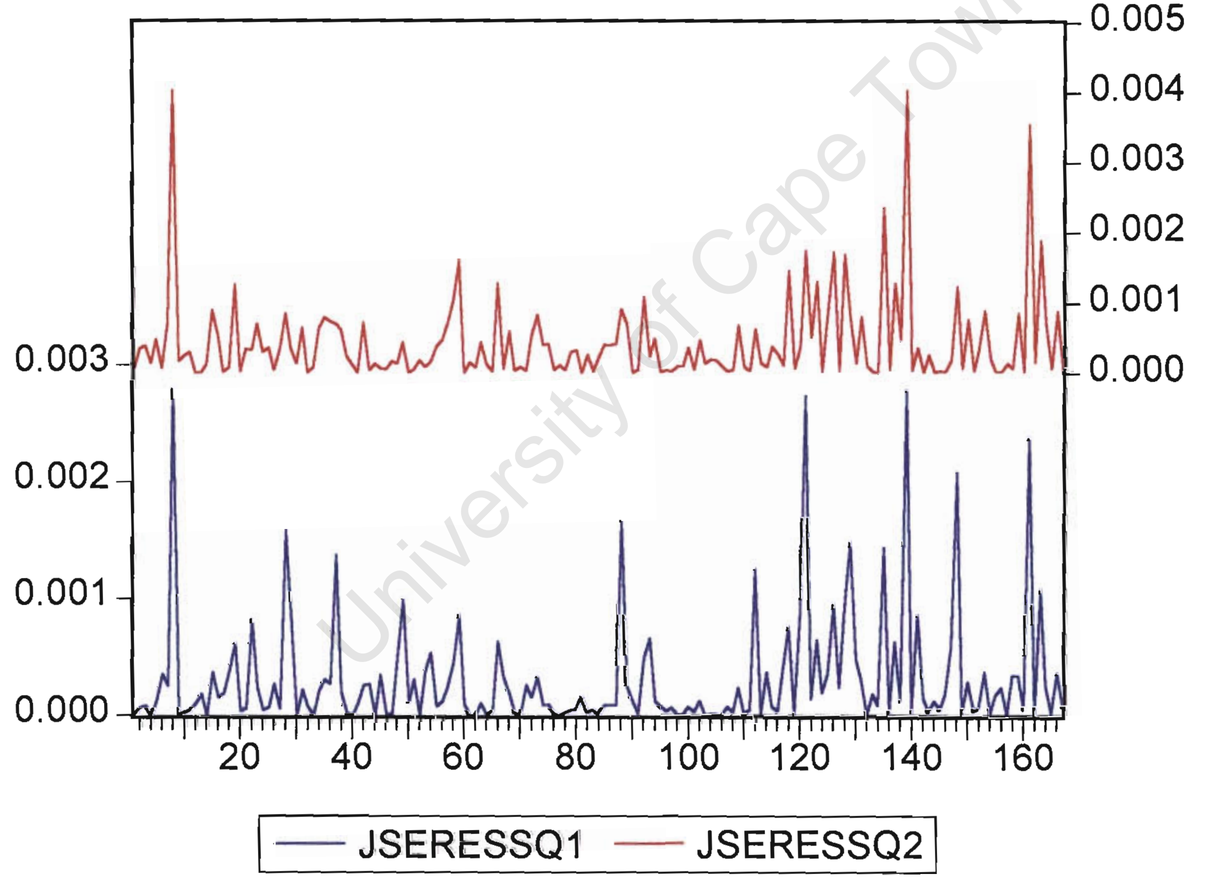
Graphs of Least Squares and Principal Component Squared Residuals of ANGLOS fit on JSE



Graphs of Least Squares and Principal Component Squared Residuals of JSE fit on ANGLOS



Graphs of Least Squares and Principal Component Squared Residuals of JSE fit on ANGLOS



Comments (ANGJSE AND JSEANG)

The analysis used for these two data sets was purely graphical. It aims to show how the principal component analysis downweights the effect of outliers on the regression parameters. In the color graphs "Graphs of Least Squares and Principal Component Squared Residuals of JSE fit on ANGLOS (respectively of ANGLOS fit on JSE)" , the ordinary least squares squared residuals (blue), and the principal component squared residuals (red) were plotted, both on a crossing dual scale and a non-crossing dual scale. The downweighting of potential outliers is remarkable.

Chapter 12

CONCLUSION

Outliers and/or influential observations arise in different contexts in general linear regression models. Whether they are harmful or not, the investigator needs to identify them. In the first case, they might be deleted from the data set or simply down-weighted using the principal component analysis if deletion is not convenient. In the second case, they should be given a sustained attention because they might be, in this case, the most important observations.

In this work, graphical and numerical methods were used for the identification of outliers and/or influential observations. Since the deletion of observations is not convenient for time series, only graphical methods were used for the ANGJSE and JSEANG data sets. For the other two data sets, we jointly analyzed the standardized and studentized residuals, the test statistics based on the mean shift outlier model, the hat values, the change in the ICOMP due to the deletion of the suspected cases, and the graphical display of the projection of observations on the factor-plane with a 95% confidence ellipse.

APPENDIX A

A1

The results of this important appendix are obtained by applying the Sherman, Morrison and Woodbury theorem [Rao(1973)] as shown in appendix A.2:

For a nonsingular matrix A and two column vectors u and z.

$$\left(A - uv^T\right)^{-1} = A^{-1} - \frac{A^{-1}uz^TA^{-1}}{1 - v^TA^{-1}u}$$

Now, letting

$$A = W^TW, \quad and \quad u = z = w_i^T,$$

one has

$$\left(W_{(i)}^TW_{(i)}\right)^{-1} = \left(W^TW\right)^{-1} + \frac{\left(W^TW\right)^{-1} w_iw_i^T \left(W^TW\right)^{-1}}{1 - v_{ii}}$$

Similarly

$$\left(W^TW\right)^{-1} = \left(W_{(i)}^TW_{(i)}\right)^{-1} - \frac{\left(W_{(i)}^TW_{(i)}\right)^{-1} w_iw_i^T \left(W_{(i)}^TW_{(i)}\right)^{-1}}{1 - v_{ii}}$$

Miller(1964) shows from the above that

$$\hat{\beta} - \hat{\beta}_{(i)} = \frac{\left(W^TW\right)^{-1} w_ie_i}{1 - v_{ii}} \tag{12.1}$$

Hence, the deletion formulae mentioned earlier can be derived:

Since

$$(n - q - 1)\, s_{(i)}^2 = \sum_{j \neq i}^{n} \left(x_j - w_j\hat{\beta}_{(i)}\right)^2$$

Using (11.1) leads

$$\begin{aligned}(n - q - 1)\, s_{(i)}^2 &= \sum_{j=1}^{n} \left(e_j + \frac{v_{ij}e_j}{1 - v_{ii}}\right)^2 - \frac{e_i^2}{(1 - v_{ii})^2} \\ &= (n - q)\, s^2 + \frac{2e_i}{1 - v_{ii}} \sum_{j=1}^{n} e_jv_{ij} + \frac{e_i^2}{(1 - v_{ii})^2} \sum_{j=1}^{n} v_{ij}^2 \\ &\quad - \frac{e_i^2}{(1 - v_{ii})^2}\end{aligned}$$

And because V annihilate the vector of residuals, this expression reduces to

$$(n - q - 1)\, s_{(i)}^2 = (n - q)\, s^2 - \frac{e_i^2}{(1 - v_{ii})^2}$$

A2

Given A, a qxq symmetric matrix of rank q, B and C, pxq matrices of rank p, and provided that their inverses exist, the Sherman, Morrison and Woodbury theorem is

$$\left(A + B^T C\right)^{-1} = A^{-1} - A^{-1}B^T\left(I + CA^{-1}B^T\right)^{-1}CA^{-1}$$

A3

$$\det\left[W_{(i)}^T W_{(i)}\right] = (1 - v_{ii})\det\left[W^T W\right]$$

Proof

Consider first $\det\left(I - uz^T\right)$, where and z are as in $A1$. Let Q be the orthonormal matrix such that $Qu = \|u\|\lambda_1$, where λ_1 is the first standard basis vector, then

$$\begin{aligned}\det\left(I - uz^T\right) &= \det Q\left[I - uz^T\right]Q^T \\ &= \det\left[I - \|u\|\lambda_1 z^T Q^T\right] \\ &= 1 - z^T Q^T \lambda_1 \|u\| \\ &= 1 - z^T Q^T Q u \\ &= 1 - z^T u\end{aligned}$$

Now,

$$\det\left[W_{(i)}^T W_{(i)}\right] = \det\left[\left(I - w_i^T w_i \left(W^T W\right)^{-1}\right) W^T W\right] \quad (12.2)$$

Let $u = w_i^T$ and $z^T = w_i\left(W^T W\right)^{-1}$, the above expression becomes

$$\det\left[W_{(i)}^T W_{(i)}\right] = (1 - v_{ii})\det\left(W^T W\right)$$

APPENDIX B

B1

$$\hat{\beta} - \hat{\beta}_{(i)} = \frac{\left(W^T W\right)^{-1} w_i e_i}{1 - v_{ii}}$$

Proof

$$\begin{aligned}
\hat{\beta} - \hat{\beta}_{(i)} &= \left(W^T W\right)^{-1} W^T X - \left(W_{(i)}^T W_{(i)}\right)^{-1} W_{(i)}^T X_{(i)} \\
&= \left(W^T W\right)^{-1} W^T X - \left(W^T W - w_i w_i^T\right)^T \left(W^T X - w_i x_i\right) \\
&= \left(W^T W\right)^{-1} W^T X - \left(\left(W^T W\right)^{-1} + \frac{1}{1 - v_{ii}} \left(W^T W\right)^{-1} w_i w_i^T \left(W^T W\right)^{-1}\right) * \\
& \quad \left(W^T X - w_i x_i\right) \\
&= \left(W^T W\right)^{-1} w_i x_i - \frac{1}{1 - v_{ii}} \left(W^T W\right)^{-1} w_i w_i^T \left(W^T W\right)^{-1} W^T X \\
& \quad + \frac{1}{1 - v_{ii}} \left(W^T W\right)^{-1} w_i w_i^T \left(W^T W\right)^{-1} w_i x_i \\
&= \left(W^T W\right)^{-1} w_i x_i - \frac{1}{1 - v_{ii}} \left(W^T W\right)^{-1} w_i w_i^T \hat{\beta} + \frac{1}{1 - v_{ii}} \left(W^T W\right)^{-1} w_i v_{ii} x_i \\
&= \left(W^T W\right)^{-1} w_i x_i + \frac{1}{1 - v_{ii}} \left(W^T W\right)^{-1} w_i \left(v_{ii} x_i - x_i^T \hat{\beta}\right) \\
&= \frac{1}{1 - v_{ii}} \left(W^T W\right)^{-1} w_i \left(\left(1 - v_{ii}\right) x_i + v_{ii} x_i - \hat{x}_i\right) \\
&= \frac{1}{1 - v_{ii}} \left(W^T W\right)^{-1} w_i e_i
\end{aligned}$$

B2 Sum of squares of residuals RSS :

$$RSS_{(i)} = RSS - \frac{e_i^2}{1 - v_{ii}}$$

Proof

In the mean shift outlier model, one has

$$E\left(X\right) = W_1 \beta_1 + \phi \beta_2$$

where ϕ is an n-vector with 1 in the ith position and 0 elsewhere. Omitting the ith case is the same as fitting this mean shift outlier model, as we mentioned earlier in chapter 2. Thus, by means of $A1$

$$V_{(i)} = V_1 + \frac{\left(I - V_1\right) \phi_i \phi_i^T \left(I - V_1\right)}{\phi_i^T \left(I - V_1\right) \phi_i}$$

$$RSS_{(i)} = X^T \left(I - V_1 - \frac{\left(I - V_1\right) \phi_i \phi_i^T \left(I - V_1\right)}{\phi_i^T \left(I - V_1\right) \phi_i}\right) X$$

$$= \quad RSS - \frac{e_i^2}{1 - v_{ii}}$$

B3

$$\left(\hat{X} - \hat{X}_{(i)}\right)^T \left(\hat{X} - \hat{X}_{(i)}\right) = \frac{e_i^2 v_{ii}}{\left(1 - v_{ii}\right)^2}$$

Proof

Since $\hat{\beta} - \hat{\beta}_{(i)} = \frac{\left(W^T W\right)^{-1} w_i e_i}{1 - v_{ii}}$

$$\hat{\beta} W - \hat{\beta}_{(i)} W = \frac{W \left(W^T W\right)^{-1} w_i e_i}{1 - v_{ii}}$$

hence

$$\begin{aligned}
\left(\hat{X} - \hat{X}_{(i)}\right)^T \left(\hat{X} - \hat{X}_{(i)}\right) &= \frac{e_i^2 w_i^T \left(W^T W\right)^{-1} W^T W \left(W^T W\right)^{-1} w_i}{\left(1 - v_{ii}\right)^2} \\
&= \frac{e_i^2 w_i^T \left(W^T W\right)^{-1} w_i}{\left(1 - v_{ii}\right)^2} \\
&= \frac{e_i^2 v_{ii}}{\left(1 - v_{ii}\right)^2}
\end{aligned}$$

APPENDIX C

C1

Let M be the following qxq nonsingular matrix

$$M = \begin{pmatrix} a & \mathbf{a} \\ \mathbf{a} & M_1 \end{pmatrix}$$

where $\mathbf{a}$ is $(q-1)$x1 and M_1 is $(q-1)$x$(q-1)$

If M_1 is nonsingular, then the determinant of M can be expressed as

$$\begin{aligned} \det(M) &= \det(M_1)\det\left(1 - \mathbf{a}^T M_1^{-1}\mathbf{a}\right) \\ &= a\det\left(M_1 - \frac{1}{a}\mathbf{a}^T\mathbf{a}\right) \end{aligned}$$

Proof

Premultiplying M by the nonsingular matrix M_1^*, where

$$M_1^* = \begin{pmatrix} \frac{1}{a} & \mathbf{0}^T \\ -\frac{1}{a}\mathbf{a} & I \end{pmatrix}$$

one has,

$$M_1^* M = \begin{pmatrix} 1 & \frac{1}{a}\mathbf{a} \\ \mathbf{0} & M_1 - \frac{1}{a}\mathbf{a}\mathbf{a}^T \end{pmatrix}$$

The reader should note that

$$\det(M_1^* M) = \det\left(M_1 - \frac{1}{a}\mathbf{a}\mathbf{a}^T\right)$$

and

$$\det(M_1^* M) = \det(M_1^*)\det(M)$$

Hence

$$\begin{aligned} \det(M_1^* M) &= \det(M_1^*)\det(M) \\ &= \det\left(\frac{1}{a}I\right)\det(M) \\ &= \frac{1}{a}\det(M) \end{aligned}$$

Finally

$$\begin{aligned} \det(M) &= a\det(M_1^* M) \\ &= a\det\left(M_1 - \frac{1}{a}\mathbf{a}\mathbf{a}^T\right) \end{aligned}$$

C2

If W^TW is nonsingular, then

$$\det\left(W_{(i)}^T W_{(i)}\right) = (1 - v_{ii}) \det\left(W^T W\right)$$

Proof

Let

$$Z = \begin{pmatrix} 1 & w_i^T \\ w_i & W^T W \end{pmatrix}$$

By appendix C1, the determinant of Z can be obtained as

$$\begin{aligned} \det(Z) &= \det\left(W^T W\right) \det\left(1 - w_i^T \left(W^T W\right)^{-1} w_i\right) \\ &= (1 - v_{ii}) \det\left(W^T W\right) \end{aligned}$$

The determinant of Z can also be written as

$$\begin{aligned} \det(Z) &= \det\left(W^T W - w_i w_i^T\right) \\ &= (1 - v_{ii}) \det\left(W^T W\right) \end{aligned}$$

Since $W_{(i)}^T W_{(i)} = W^T W - w_i w_i^T$, the result follows.

APPENDIX D: DATA SETS

D.1: SIML4 DATA SET

obs	X	W 1	W 2	W 3	W 4
1	587.231	1	10	8.808	10.448
2	715.932	2	4.37	15.011	14.151
3	766.807	3	5.573	16.714	13.355
4	527.182	4	6.776	7.417	9.058
5	512.057	5	4.98	8.12	9.261
6	556.932	6	3.183	11.323	9.464
7	629.807	7	6.386	12.026	10.667
8	508.182	8	4.589	8.23	8.87
9	569.057	9	5.792	9.933	10.073
10	541.932	10	4.495	9.136	9.776
11	558.807	11	6.198	9.339	9.48
12	562.182	12	5.401	10.042	9.183
13	616.057	13	5.105	11.245	10.386
14	658.432	14	6.308	11.948	11.589
15	614.307	15	4.511	11.151	10.792
16	549.682	16	5.214	8.355	9.995
17	545.057	17	3.917	10.558	8.198
18	579.932	18	5.12	9.261	10.401
19	612.807	19	5.823	11.964	9.105
20	571.182	20	5.526	9.667	9.308
21	584.557	21	5.23	9.37	10.011
22	586.932	22	4.933	10.073	9.214
23	568.807	23	4.636	10.276	8.417
24	638.182	24	5.839	11.48	10.12
25	670.557	25	6.042	12.183	10.823
26	589.432	26	5.745	8.386	10.526
27	585.807	27	3.448	9.089	10.73
28	630.682	28	5.651	9.792	10.933
29	643.557	29	3.855	10.995	11.136
30	628.932	30	3.558	10.198	10.839
31	645.307	31	5.761	10.401	10.542
32	602.682	32	4.464	10.105	9.245
33	661.057	33	4.667	12.308	9.948
34	651.932	34	6.37	11.511	10.151
35	623.807	35	4.573	9.714	10.855
36	619.682	36	3.776	8.917	11.558
37	659.057	37	4.48	10.62	11.261
38	706.432	38	4.183	11.823	12.464
39	627.807	39	2.886	11.026	10.167
40	642.182	40	5.589	10.23	9.87

D.2: BLD.SOC DATA SET

Obs	X	W1	W2	W3	W4	W5	W6	W7	W8
1	497	379.604	14.984	8.107	0.099	0	22.406	35.008	114.871
2	329	196.433	5.302	3.456	0.133	0	11.232	15.875	61.681
3	146	105.975	3.837	1.686	0.666	0.86	3.838	8.509	27.333
4	249	209.451	10.621	4.913	0.052	0	14.099	21.368	63.052
5	82	46.641	2.31	1.554	0.322	0.397	3.701	5.601	17.713
6	98	57.725	1.226	1.615	0.648	0.822	4.408	0.77	21.41
7	32	14.87	0.237	1.021	0.219	0.622	1.164	0.022	5.687
8	23	15.601	0.393	0.703	0.05	0.069	1.33	1.592	4.98
9	36	16.826	0.405	0.441	0.276	0.208	1.012	1.844	5.739
10	109	80.64	1.512	1.369	1.84	1.252	3.799	8.726	23.455
11	124	86.852	3.345	2.269	0.732	0.398	5.088	8.405	24.547
12	97	40.114	2.652	1.644	1.099	0.634	3.73	7.199	21.496
13	87	43.165	1.544	1.56	0.603	0.661	4.225	6.647	15.769
14	65	33.714	1.463	1.196	0.434	0.501	2.58	5.468	18.354
15	46	24.735	0.484	0.659	0.4	0.275	1.359	1.763	6.728
16	78	48.252	0.995	0.917	0.662	0.771	3.12	7.005	15.839
17	64	26.543	0.734	1.583	0.585	0.499	1.906	4.181	12.678
18	62	35.609	2.026	0.943	0.423	0.579	2.419	3.052	12.15
19	43	23.206	0.697	0.902	0.311	0.507	1.832	1.468	9.305
20	25	18.481	0.266	0.323	0.183	0.119	0.632	1.678	5.753
21	37	21.287	0.651	0.818	0.435	0.379	1.508	2.751	7.275
22	26	12.469	0.766	0.395	0.278	0.406	0.806	0.038	5.547
23	175	110.42	5.54	5.03	0.011	0	7.923	0	42.429
24	42	20.472	1.306	0.722	0.255	0.362	1.646	2.765	10.097
25	126	54.363	3.494	2.233	1.455	2.172	4.777	7.638	32.259
26	84	37.196	1.392	0.878	0.7	0.891	3.499	5.996	20.131
27	160	50.162	4.086	1.69	2.133	1.644	4.233	9.987	30.808
28	60	41.921	3.457	0.851	0.023	0	2.061	5.406	13.163
29	71	31.187	2.411	0.756	1.278	0.826	2.792	5.404	21.218
30	89	50.144	2.006	0.995	1.31	1.238	4.554	5.475	21.487
31	96	50.299	3.856	0.99	1.202	1.351	2.987	0.287	19.488
32	64	25.574	4.057	1.099	0	3.907	5.74	18.624	25.574
33	90	48.554	4.412	1.515	0.055	0	2.282	5.543	18.321
34	93	42.784	1.117	0.5	0.281	0.288	1.567	5.06	20.122
35	45	23.72	0.741	0.492	0.38	0.307	1.252	2.909	9.628
36	37	19.295	0.674	0.38	0.383	0.389	2.579	2.378	7.179
37	18	11.038	0.297	0.16	0.064	0.036	0.985	0.952	3.369
38	130	5.607	0.071	0.025	3.762	3.559	0.248	2.165	18.882
39	54	4.035	0.092	0	2.613	2.597	0.194	2.612	9.783
40	129	5.38	0.094	0.058	5.068	4.226	0.509	3.842	19.208
41	7	7.88	0.12	0.074	0.021	0	0.138	0	1.624
42	8	4.729	0.153	0.119	0.027	0	0.195	0	0.831
43	55	2.874	0.016	0.012	2.24	3.017	0.297	1.175	11.029
44	86	49.364	4.121	0.792	0.017	0	3.057	6.665	25.349
45	57	15.977	1.523	0.429	0.887	0.652	0.984	3.149	8.706
46	61	22.383	1.97	0.546	0.877	0.846	3.522	3.417	13.195
47	78	42.041	1.979	0.881	0.002	0	1.755	5.374	15.549
48	95	83.133	4.326	1.363	0.006	0	5.348	7.616	26.262

D.3 ANGJSE DATA SET

Obs	X	W1	W2	W3	W4	W5	W6	W7	W8
1	-0.06416	-0.04978	0.003741	-0.02012	-0.06347	-0.00953	-0.03489	-0.07677	0.0282442
2	0.090939	0.069998	-0.00133	-0.05589	-0.01319	0.045622	0.095848	0.123836	-0.018865
3	0.058741	0.062373	-0.05614	0.037007	0.066777	0.132602	-0.01081	0.145244	-0.0703808
4	0.032261	-0.00356	-0.01842	-0.05328	0.026962	0.033	-0.07914	0.03931	-0.0222817
5	-0.03226	0.01166	0.02079	0.009626	0.039128	0.022516	0.017493	-0.00639	0.0238829
6	0.141044	0.084067	0.051096	-0.03575	0.150823	0.141963	0.121757	0.015588	0.0365248
7	0.04564	0.057184	-0.02506	0.017039	0.052592	-0.02371	0.048936	0.098805	-0.0285566
8	0.17431	0.103038	0.04047	0.018457	0.030829	0.03393	-0.00195	0.051007	0.0503007
9	-0.02031	0.023797	-0.00874	-0.01368	0.040564	0.010925	-0.00098	0.231802	0.0024122
10	-0.13406	-0.07857	0.06698	0.04111	-0.118	0.030125	-0.04499	-0.13723	0.0757361
11	0.190731	0.091588	-0.03093	0.024434	0.0342	0.092075	0.056647	0.055303	0.0012556
12	0.064015	0.017726	-0.04299	-0.05754	-0.00578	0.009217	0.025752	0.030637	-0.0473203
13	0.032937	0.039939	0.048624	0.020969	0.047179	0.105078	0.1079	0.031047	0.0549593
14	-0.00481	-0.00649	-0.04782	-0.01013	-0.04332	-0.03623	0.00674	0.005483	-0.0369956
15	-0.03601	-0.02563	-0.01701	0.002502	-0.11514	0.014064	-0.06863	0.022706	0.0212373
16	0.089612	0.058703	-0.02788	0.073994	-0.01523	-0.02499	0.044843	0.157535	0.0023165
17	0.028171	0.040108	-0.05466	-0.03919	0.016314	0.070701	0.112899	0.030224	0.0082397
18	0.230524	0.070694	-0.00593	0.036792	0.013926	0.126604	0.050179	0.037563	0.0030551
19	-0.12084	-0.03505	0.007054	-0.01944	-0.07051	0.008168	-0.01324	-0.12549	0.0304523
20	0.102436	0.054579	0.00289	-0.0544	0.0181	-0.02082	0.111265	0.072033	0.0597144
21	-0.12153	-0.07158	-0.0106	-0.00713	-0.09643	-0.07829	-0.0893	0.021905	0.0060003
22	0.094818	0.050171	-0.00095	-0.00782	0.015922	0.011017	0.150939	0.095198	-0.0495052
23	-0.10947	-0.03544	-0.00796	-0.03708	-0.01223	0.001184	-0.06762	0.002399	0.0057401
24	-0.00447	0.024082	-0.04559	0.029404	-0.02491	0.041425	0.047489	0.021027	-0.0386998
25	-0.11886	-0.05176	-0.00808	0.033881	-0.03595	-0.06295	-0.01601	-0.13801	-0.0158537
26	-0.06498	-0.08686	-0.00827	0.049158	-0.05508	-0.05179	-0.08344	-0.17716	-0.0424358
27	-0.0184	-0.02846	-0.03082	-0.08181	-0.02801	-0.02644	-0.06076	-0.07731	-0.0609735
28	-0.02956	-0.02506	-0.00907	0.004918	0.071263	0.11181	0.044484	-0.12493	0.0270595
29	0.032088	0.044736	-0.0034	0.019829	0.087342	0.148037	0.125834	0.028462	0.0115565
30	-0.10232	-0.06219	-0.00918	-0.06254	-0.06899	-0.07508	-0.01393	-0.13616	-0.0228689
31	0.123158	0.092247	0.022637	0.000935	0.06778	0.138781	0.001911	0.276142	0.0039361
32	0.005347	0.026058	0.119026	0.042697	0.054323	0.039639	0.096409	0.063165	0.0631589
33	0.037436	0.052804	0.005246	0.02583	0.002296	0.091377	0.097934	0.0926	-0.0342685
34	0.048207	0.026677	-0.00279	0.016876	0.043753	-0.0077	0.017148	-0.10391	0.0097141
35	0.075049	0.05951	0.055453	0.048317	0.046121	-0.00084	0.069165	0.020013	-0.0009454
36	0.038374	0.054449	-0.0376	-0.02278	-0.03412	0.057519	0.099671	-0.01602	0.0005819
37	-0.01251	-0.04153	0.002517	-0.03942	0.026754	-0.01914	0.043846	-0.14777	-0.0075175
38	-0.02377	-0.01565	-0.05146	-0.00408	-0.03766	-0.05313	-0.04211	-0.03162	-0.0267276
39	0.059209	0.067151	0.006474	0.016559	0.020619	0.10835	0.071654	0.163535	-0.0135603
40	0.038349	0.004527	-0.02959	0.011016	0.034832	-0.01227	-0.0283	-0.01828	-0.0081177
41	-0.01401	-0.02922	-0.06826	-0.05487	-0.00416	0.026455	-0.01763	-0.17376	-0.0404856
42	0.031558	0.04685	0.059605	0.023972	0.069409	0.069118	0.047539	-0.02298	0.0035169
43	-0.02964	-0.00222	0.013137	0.011305	-0.00487	-0.01199	0.044224	0.160059	0.0293272
44	-0.04282	-0.01315	0.007326	-0.02082	-0.01377	0.040673	-0.00387	0.000675	0.0323998
45	-0.02743	-0.02741	0.006669	-0.01573	-0.03015	0.003231	0.014245	0.01629	0.00045
46	0.050211	0.077411	-0.02887	-0.01235	0.090661	0.084603	0.118462	0.116454	-0.042426
47	-0.02915	-0.02085	-0.05352	-0.00417	-0.01977	-0.03412	-0.06271	0.072376	-0.0148163
48	-0.02347	-0.06324	-0.03193	0.038714	-0.06886	-0.07768	-0.01898	-0.12974	0.0101115
49	-0.18095	-0.08481	-0.05064	-0.05875	-0.07067	-0.00753	0.031196	-0.07044	-0.0331619
50	-0.06914	0.018545	0.006181	0.050573	0.052167	0.050053	0.097904	0.06856	-0.0268439

Obs	X	W1	W2	W3	W4	W5	W6	W7	W8
51	-0.07475	-0.06232	0.090017	0.021529	0.079723	-0.03367	-0.03594	-0.16188	0.0274125
52	0.043789	0.056385	0.034667	0.006454	0.075612	0.07349	0.091672	0.023559	0.0121067
53	-0.01146	0.019237	-0.00564	-0.00394	0.089967	0.018781	0.085939	0.028631	-0.007408
54	0.127114	0.05355	0.011254	0.012156	0.133531	0.053719	0.092034	-0.06083	0.0080496
55	-0.03638	-0.00438	0.020857	0.011767	0.006372	-0.01794	0.009948	-0.03738	-0.0562298
56	0.137471	0.040705	-0.01868	0.038539	0.02095	-0.09371	-0.04305	0.143804	-0.0259465
57	0.146837	0.047184	-0.01494	0.04729	-0.01884	-0.00536	-0.04415	0.111648	-0.0319256
58	0.081796	0.065342	0.017416	0.067314	-0.00353	0.026031	0.043621	0.117584	-0.0296406
59	0.007273	0.023075	0.048908	0.043418	-0.05521	0.059257	0.032602	-0.10589	-0.0191586
60	0.103184	0.023992	0.026459	0.079025	-0.00224	-0.02673	0.044909	0.004325	-0.0128201
61	-0.07461	-0.0346	-0.03682	-0.07766	0.014121	-0.01661	-0.03928	-0.06524	-9.558E-05
62	-0.11164	-0.06768	-0.0301	-0.02565	-0.01113	-0.021	-0.0744	-0.12876	0.0079022
63	0.097455	0.037996	0.027477	0.014436	-0.03184	0.039512	0.016628	0.044938	0.0264836
64	0.108214	0.061405	0.022957	0.005803	0.059083	0.073012	0.068182	0.016036	0.0080026
65	0.341587	0.161125	0.01212	0.061466	0.149471	0.150385	0.092782	0.195934	0.0230936
66	-0.12606	-0.02861	0.003682	-0.02411	-0.07801	-0.04739	0.018922	-0.08592	-0.0258459
67	0.017823	0.018961	-0.01767	0.022256	0.034555	0.039904	0.045257	0.103633	-0.0441386
68	0.025064	0.019216	-0.02288	0.021025	-0.00196	-0.04829	-0.00758	0.069379	-0.0133387
69	0.151318	0.081615	-0.00968	-0.01992	0.110778	0.124329	0.085016	-0.05216	-0.0129274
70	-0.07399	0.006881	-0.00563	0.054597	0.082069	-0.02511	0.059519	0.014831	0.0305775
71	0.037238	0.001481	-0.03315	0.013302	-0.00487	-0.0404	-0.00604	0.139625	-0.0580586
72	0.09462	0.044693	-0.0022	-0.00745	-0.0131	-0.05056	0.019316	0.137088	0.0150832
73	0.039375	0.031699	-0.00713	-0.01782	0.023352	-0.03784	-0.00335	0.072151	-0.0011983
74	-0.07619	-0.02728	-0.0159	0.017318	-0.03218	-0.05492	-0.02567	0.013997	-0.0103451
75	-0.01681	0.008246	-0.02827	-0.04648	0.002768	0.04732	0.002292	-0.03133	-0.0187512
76	-0.02141	0.00575	0.04043	0.012527	0.055376	0.031289	0.046589	-0.04345	0.0194577
77	0.021414	0.01893	-0.01474	-0.00297	-0.0265	0.026286	0.048339	0.026578	0.0059347
78	-0.22995	-0.14899	-0.02397	-0.02741	-0.0875	-0.09088	-0.12475	-0.29384	-0.0088648
79	0.032735	0.018234	-0.03416	0.024884	0.002927	0.005723	-0.02085	0.077584	-0.0206509
80	0.016583	0.025701	-0.05628	0.043833	-0.00293	0.029798	0.051815	-0.09525	-0.1055786
81	0.039806	0.036807	-0.00123	-0.00312	0.067976	0.036842	0.065812	0.079236	-0.0361168
82	-0.04735	0.000365	0.02363	0.002459	0.053837	-0.02473	0.037813	-0.06278	0.0080607
83	0.001227	-0.01119	-0.02045	-0.00315	-0.01253	-0.0262	0.003772	-0.00586	0.0011799
84	0.001225	0.003316	-0.0026	-0.0116	0.02183	-0.00056	-0.01657	0.036158	0.0378364
85	-0.00245	0.019125	0.060825	0.004093	0.08521	0.028538	0.010386	0.002664	0.1032577
86	0.035176	0.020358	-0.0315	0.000651	0.058679	0.012891	-0.013	-0.05222	0.0101886
87	0.024389	0.023066	-0.00777	-0.0025	0.054583	0.062708	0.04869	-0.21321	0.0364292
88	0.037826	0.031122	0.017563	0.017538	0.129148	0.042032	0.130223	0.03557	-0.0098223
89	0.020574	0.041973	0.001739	-0.00916	0.034951	0.031081	0.073118	-0.06572	0.0156697
90	0.178526	0.098109	0.035511	0.050764	0.080731	0.10716	0.108596	0.154374	0.0385066
91	-0.03876	-0.02431	-0.01283	0.044119	-0.01966	-0.03607	-0.00733	-0.01875	-0.0169735
92	0.01957	0.006393	0.011487	0.021869	-0.02143	-0.05128	-0.01681	-0.04796	0.0094653
93	0.123794	0.03323	0.030436	0.081927	0.062322	-0.0184	0.017785	0.193686	-0.0132766
94	-0.03485	-0.02291	-0.00261	-0.00738	-0.03084	-0.06057	-0.02166	-0.0578	0.0319426
95	-0.02688	0.008762	-0.00959	-0.02596	0.072607	0.030378	0.090591	-0.15887	0.0095703
96	-0.09854	-0.04035	-0.02891	0.037941	-0.02574	-0.04878	-0.02464	0.050644	-0.023537
97	0.091223	0.012486	0.003795	0.005309	-0.00191	-0.0051	0.005955	-0.0086	0.018119
98	0.019909	0.027864	0.031167	-0.01065	0.01643	0.077699	0.080168	-0.04267	0.0252764
99	0.015137	0.014004	-0.00783	0.038285	0.123927	-0.01682	-0.02008	0.007933	0.0222853
100	-0.07243	-0.03829	0.013978	-0.0413	0.021365	0.007311	-0.06816	-0.09184	-0.0021095

Figure 12-1:

(Continue)

Obs	X	W1	W2	W3	W4	W5	W6	W7	W8
101	-0.01923	-0.00838	0.011018	0.010905	0.057143	-0.01722	-0.00482	-0.01516	0.0214145
102	-0.01961	0.00285	0.052726	-0.08917	-0.0585	0.05622	0.047891	-0.04812	0.0467744
103	0.112259	0.067894	0.032957	0.016059	-0.09263	0.052137	0.107438	0.238279	-0.0096627
104	-0.03237	-0.00716	-0.00451	-0.04119	0.035102	0.009224	0.007543	-0.08149	0.0272599
105	0.037599	0.005062	0.026391	-0.02067	0.000287	0.012238	0.001402	-0.03335	0.0239379
106	-0.08246	-0.01541	-0.01669	0.021599	-0.00029	-0.00922	0.012825	0.028981	-0.085901
107	0.04384	0.055139	0.024476	-0.0187	0.122912	0.020887	0.07487	0.14349	-0.016896
108	-0.04109	0.008723	0.051925	-0.00874	0.078203	-0.00021	0.075789	-0.03362	0.030634
109	-0.08546	-0.02407	-0.02038	0.014826	-0.02114	-0.0791	-0.02147	0.006146	0.0139698
110	-0.00935	-0.02537	-0.01706	0.016973	-0.04167	-0.03178	-0.04093	0.014012	0.0124576
111	-0.13692	-0.07793	-0.02528	-0.03143	-0.04505	-0.05412	-0.05312	-0.08918	-0.0081214
112	-0.02927	-0.04073	0.023013	-0.04147	-0.03031	-0.01428	0.047677	0.003796	0.0575559
113	-0.0271	-0.0198	0.017156	-0.03051	-0.01715	-0.03519	0.028595	-0.1211	0.0218337
114	0.107901	0.054593	0.020643	0.074125	0.121863	0.058269	0.072181	0.029714	-0.021127
115	-0.05635	0.079925	-0.00939	-0.02631	0.190432	-0.00718	0.189857	0.03689	-0.005589
116	0.108833	0.065859	0.01955	0.030127	0.097007	0.002357	0.04136	0.106181	0.0452957
117	0.258892	0.083144	-0.0307	0.03931	0.181896	0.063492	-0.01393	0.065946	-0.0037036
118	-0.19106	-0.07644	-0.00475	-0.03543	0.008628	-0.08289	-0.08243	-0.2329	0.0472529
119	-0.21107	-0.11931	0.012528	0.135915	-0.06379	-0.10123	-0.14394	0.009354	0.0149117
120	0.086178	0.036114	-0.01477	0.019368	0.12573	0.020068	-0.00357	0.267586	0.0252947
121	-0.2969	-0.35485	-0.01203	0.007438	-0.4326	-0.23268	-0.44908	-0.05671	-0.0171447
122	0.036368	0.03493	-0.05047	-0.03582	-0.03127	-0.05236	0.049753	0.08415	-0.0469638
123	0.083334	0.133826	-0.01106	-0.03613	0.153217	0.121017	0.187357	-0.04835	-0.1521553
124	-0.0011	-0.03634	0.02854	0.013584	-0.09885	0.047452	-0.05729	0.115529	0.0596818
125	-0.09542	-0.03439	-0.01533	0.002343	0.023198	-0.00819	-0.08291	-0.09326	-0.0760073
126	0.036712	0.065746	0.024045	0.023895	0.109106	-0.04509	0.022481	0.117802	0.0171004
127	0.116231	0.019637	0.03468	0.026103	-0.01298	-0.00652	0.001851	0.140789	0.0308473
128	0.10225	0.076159	0.024996	-0.02323	0.111515	0.069145	0.011947	-0.03758	0.0015843
129	0.383417	0.10153	0.012538	0.004655	0.050692	0.008203	0.076585	0.163747	-0.0046768
130	-0.10962	-0.08506	0.012918	-0.03422	-0.09591	-0.08539	-0.04113	0.04718	0.0175086
131	0.00213	0.082648	0.012542	-0.06573	0.062843	0.091122	0.126647	0.14398	-0.0056092
132	0.121401	0.006788	-0.0363	-0.00017	-0.1473	-0.03661	-0.02179	0.094184	-0.0464767
133	0.050225	-0.02238	0.016457	-0.01623	-0.04355	0.010863	-0.0335	-0.0397	-0.0473787
134	0.030011	-0.01014	-0.00216	0.1706	-0.06512	0.026652	-0.05495	0.144363	-0.031192
135	-0.05358	0.040659	0.005975	0.003247	-0.07358	0.077612	0.059661	0.056677	-0.0193078
136	0.102236	0.054281	0.038996	-0.02415	0.091052	0.065399	0.044549	0.048098	-0.0233007
137	0.089218	0.123183	0.001135	-0.00846	0.070541	0.149282	0.166137	0.042527	0.0007822
138	-0.10761	-0.00787	0.028512	0.003276	0.088126	-0.00958	-0.01165	0.045953	0.043325
139	-0.22342	-0.05868	0.006393	0.042338	-0.06536	-0.08335	-0.08396	-0.09944	0.0282385
140	0.060707	-0.00439	0.016345	-0.02629	-0.06376	-0.01764	0.008792	-0.02592	-0.0546032
141	-0.09432	-0.07257	0.039471	0.022459	-0.0245	0.010967	-0.01189	-0.08554	0.0200739
142	0.096301	-0.00488	-0.01094	0.019893	-0.05132	-0.0406	-0.06382	0.158048	0.0072263
143	0.0621	0.045785	-0.0258	0.032836	0.058436	-0.02061	0.044363	0.048011	-0.0159295
144	0.019669	0.003626	0.032817	-0.01821	0.06033	-0.00727	0.001861	0.066896	0.0394034
145	0.171306	0.092757	0.034609	0.00112	0.117739	0.003313	0.075262	0.219557	-0.0270337
146	-0.02073	-0.02565	0.015675	0.029904	0.004183	-0.00424	-0.01739	0.010096	0.013423
147	0.073173	-0.0199	0.042505	0.010687	0.055311	-0.02898	-0.06449	0.047679	0.0093084
148	-0.03315	-0.03858	-0.02517	0.037542	0.000329	0.038895	0.012921	0.082349	0.0177302
149	0.028269	0.064747	-0.07416	-0.01364	0.041407	0.103267	0.111906	0.094105	0.0331502
150	0.164229	0.085699	0.001094	0.000136	0.063443	0.022639	0.049305	0.037288	0.0128207

(Continue)

Obs	X	W1	W2	W3	W4	W5	W6	W7	W8
151	0.055682	-0.00641	0.009413	-0.00347	0.015579	0.006244	-0.05593	0.019243	0.0045515
152	-0.11333	-0.09967	0.043846	0.009027	-0.07965	-0.07514	-0.08408	-0.30101	0.0637296
153	0.112936	0.095667	-0.00864	0.01713	0.067918	0.059923	0.031859	0.254835	-0.0057242
154	0.017514	0.044873	0.043884	0.01827	0.068983	0.063477	0.036127	0.155947	-0.0336284
155	-0.0803	-0.01795	0.005163	0.017135	0.079905	0.052523	0.043963	-0.13462	0.0388701
156	-0.10202	-0.07496	-0.03398	0.004546	-0.06718	0.033912	-0.02439	-0.22342	0.0049743
157	0.118606	0.048924	-0.04295	0.045872	0.024182	0.018069	-0.00966	0.134955	-0.0482861
158	-0.12418	-0.10049	-0.00042	0.140294	-0.0321	-0.03248	-0.12001	-0.08468	-0.001513
159	0.124177	0.050043	0.010743	9.09E-05	0.001917	0.033564	-0.00653	0.022447	0.026562
160	0.242688	0.099844	0.017107	0.073252	0.095061	0.07196	0.028819	0.194891	0.0128593
161	0.167153	0.100784	0.002861	0.16021	-0.00087	-0.01184	-0.02863	0.225988	0.0592632
162	0.009777	-0.01234	0.028254	-0.02763	-0.03897	-0.00879	-0.00047	-0.04473	0.0137004
163	0.071943	0.047436	-0.00313	0.045524	-0.00597	-0.00433	-0.03252	0.095756	0.0074149
164	-0.05373	0.012315	-0.02154	0.014112	-0.00653	0.042819	0.010636	0.039845	-0.0111054
165	-0.1146	0.007371	0.001174	-0.05031	0.033005	0.046766	0.099862	0.034993	-0.0332175
166	0.066211	0.017081	0	-0.03074	0.022075	0.003442	-0.05238	-0.05806	-0.0348434
167	-0.0428	-0.05139	0	0.038832	-0.00473	-0.03266	-0.0659	-0.08914	-0.036604

APPENDIX E: R-CODES

```
# Main function #
OUTLIERS<-function (ydata,xdata)
{
n<-nrow(as.matrix(xdata))
q<-ncol(as.matrix(xdata))
k<-q+1
xdata<-as.matrix(xdata)
ydata<-as.matrix(ydata)
a<-lm(as.matrix(ydata)~as.matrix(xdata))
rstand<-round(rstandard(a),2)
rstud<-round(rstudent(a),2)
ei<-residuals(a)
RSS<-crossprod(as.matrix(ei))
sigma2.full<-RSS/(n-q-1)
hat.full<-xdata%*%solve(crossprod(xdata))%*%t(xdata)
hii<-round(lm.influence(a)$hat,2)
# Orientation away factor (section 5.1)
sec.reg=0
ow=summ=0
sw=cw.delta=0
crossp=0
for(j in 1:ncol(xdata))
{varindex=j
 y=xdata[,j]
 xdata.sec=arrange1(varindex,xdata)
 sec.reg=lm(y~xdata.sec)
 hat.sec=lm.influence(sec.reg)$hat
 ow[j]=((1-hat.sec[j])/(1-hii[j]))}
# Scale factor (section 5.1)
 center=apply(xdata.sec,2,mean)
 xdata.sec.cent=sweep(xdata.sec,2,center,"-")
for(i in 1:ncol(xdata.sec.cent))
```

```
{summ[i]=crossprod(xdata.sec.cent[,i])
 sw[i]=1-(n/(n-1))*crossprod(xdata.sec.cent[,i])/summ[i]}
#single outlier detection (from section 2.2 to 2.4)
#1.
Rn<-max(rstand)
#2.
Fn<-round((n-q-1)*(Rn^2)/(n-q-(Rn^2)),2)
#3.a standard f-test for outlier
Fi<-round((n-q-1)*(rstand^2)/(n-q-rstand^2),2)
#3.b estimator of the mean shift outlier model parameter
phii<-ei/(1-hii)
#4. The Andrew-Pregibon statistic (section 5.6)
Qdi<-(1-hii)*(1-(rstand^2)/(n-q))
#5. The likelihood distance (section 5.7)
ldi<-n*log(n*(n-q-1)/((n-1)*(n-q-1+rstud^2)))+(n-1)*
(rstud^2)/((n-q-1)*(1-hii))-1
#6. The Cook-Weisberg statistic (section 5.3)
 k=(q-1)/2*log((n-q)*qf(0,975,q-1,n-q-1)/(n-q-1)/
qf(0.975,q-1,n-q)) cw=((2*q-3)/6)*log(1-hii)-k
# influence measure
 for(i in 1:ncol(xdata))
 {varindex=i
 y=xdata[,i]
 xdata.sec=arrange1(varindex,xdata)
 sec.reg=lm(y~xdata.sec)
 hat.sec=lm.influence(sec.reg)$hat
 cw.delta[i]=q/2*log(1-hat.sec)-k}
#7. Icomp (section 10.4)
 icomp=n*(log(2*pi*RSS/n)+1)
 inv=solve(crossprod(xdata))
 icomp=icomp+q*log(tr(inv)/q)
 icomp.full=round(icomp-log(det(inv)),2)
# Icomp after deletion of i-th case
```

```
RSS.i=0
icomp.i=0
yxdat.del=arrange2(ydata,xdata)
# yxdat.del=as.matrix(yxdat.del)
for(i in 1:nrow(xdata))
{z1=as.matrix(yxdat.del$ydatresult[[i]])
z2=as.matrix(yxdat.del$xdatresult[[i]])
reg=lm(z1~z2)
RSS.i[i]=crossprod(as.matrix(residuals(reg)))
icomp.del=n*(log(2*pi*RSS.i[i]/n)+1)
inv.del=solve(crossprod(as.numeric(as.matrix
(yxdat.del$xdatresult[[i]]))))
icomp.del=icomp.del+q*log(tr(inv.del)/q)
icomp.i[i]=round(icomp.del-log(det(inv.del)),2)}
# influence measure using icomp
delta.icomp=icomp.full-icomp.i
# the multiple outlier case
m=readline("how many potential outliers are suspected?
i.e observations outside the 95% confidence ellipse")
m=as.numeric(m)
if(m==0) stop("there should be at least one suspected
outlier: the program was designed for cases with at least one
potential outlier known")
m1=m
m=m-1
cat("please enter their respective case numbers in numeri-
cal",
"\n","order one at a time using the ENTER button","\n")
pos=readline("next case number ? ")
pos=as.numeric(pos)
while(m>0) {s=readline("next case number ? ")
s=as.numeric(s)
pos=as.numeric(c(pos,s))
```

```
 m=as.numeric(m)-1}
xdata.m.out<-as.matrix(arrange3(ydata,xdata,pos)$xdata.I)
ydata.m.out<-as.matrix(arrange3(ydata,xdata,pos)$ydata.I)
xdata.m.left<-as.matrix(arrange3(ydata,xdata,pos)$xdata.I.del)
ydata.m.left<-as.matrix(arrange3(ydata,xdata,pos)$ydata.I.del)
del.reg<-lm(ydata.m.left~xdata.m.left)
out.reg<-lm(ydata.m.out~xdata.m.out)
RSS.out.reg<-crossprod(as.matrix(del.reg$residuals))
sigma2.out.reg<-RSS.out.reg/(n-length(pos)-q-1)
e.L<-del.reg$residuals
I.L<-diag(n-length(pos))
hat.mult.outs<-as.matrix(xdata.m.left%*%solve
(crossprod(xdata.m.left))%*%t(xdata.m.left))
#8. (section 3.2)
t2.L.num<-(t(e.L)%*%solve(I.L-hat.mult.outs)%*%e.L)/m1
t2.L.den<-((n-q)*sigma2.full-t(e.L)%*%solve((I.L-hat.mult.outs))
%*%e.L)/(n-q-m1)
t2.L<-t2.L.num/t2.L.den
#9. Outlier sum of squares (section 3.3)
Q.k<-t(e.L)%*%solve(I.L-xdata.m.left%*%
solve(crossprod(xdata))%*%t(xdata.m.left))%*%e.L
F<-round(Q.k*(n-q-m1)/((RSS-Q.k)*m1),2)
#10.The Andrews-Pregibon statistic (section 5.6)
Q.dm.num<-(n-q-m1)*sigma2.out.reg*
det(crossprod(xdata.m.left) )
Q.dm.den<-(n-q)*sigma2.full*
det(crossprod(xdata) )
Q.dm<-Q.dm.num/Q.dm.den
#11 Output table
sink("outputfile",append=F,type="output")
sp=" "
sp1=" "
cat(sp,"i",sp,"sdz",sp,"stz",sp,"F.i.del",sp,"vii",sp,"icomp",sp,
```

```
"icomp.i.del",sp1,"delta.icomp",sp,"F-test","\n")
for(i in 1:nrow(xdata))
{cat(sp,i,sp,rstand[i],sp,rstud[i],sp,Fi[i],sp,hii[i],sp,icomp.full,
sp,icomp.i[i],sp,delta.icomp[i],sp,if(Fi[i]>qf(0.95,1,n-k-1)) {"*"}
else{if(Fi[i]>qf(0.99,1,n-k-1)){"**"}},"\n")}
sink()
plot(rstand,type="l",main="Standardized(blue) and Stu-
dentized(red) residuals",xlab="case",ylab="sta & stu",col="blue")
lines(rstud,type="l",col="red")
grid(50,50,col="lightgray",lty="dotted",lwd=.3)
}
# End of Main Function#
###############################
# deletion of variable indexed by varindex
arrange1=function(varindex,xdata1)
{xdat=xdata1[,varindex]
 for(i in 1:ncol(xdata1))
 {if(i!=varindex)
 {xdat=cbind(xdat,xdata1[,i])}}
 xdat=as.matrix(xdat[,2:ncol(xdat)])
 return(xdat)}
###############################
arrange2=function(ydata,xdata1)
 {xdatresult=vector(mode="list",length=nrow(xdata1))
 ydatresult=vector(mode="list",length=length(ydata))
 for(i in 1:nrow(xdata1))
 {xdat=xdata1[i,]
 ydat=ydata[i]
 for(j in 1:nrow(xdata1))
 {if(i!=j)
 {xdat=rbind(xdat,xdata1[j,])
 ydat=rbind(ydat,ydata[j])}}
 xdatresult[[i]]=xdat[2:nrow(xdat),]
```

```
 ydatresult[[i]]=ydat[2:length(ydat)]}
 return(ydatresult,xdatresult)}
##############################
# deletion of observations indexed by pos
arrange3=function(ydata,xdata,pos)
{ydat=ydata[pos[1]]
xdat=xdata[pos[1],]
for(i in 2:length(pos))
 {ydat=rbind(ydat,ydata[pos[i]])
 xdat=rbind(xdat,xdata[pos[i],])}
for(i in 1:length(ydata))
 {if(any(pos==i))
 {cat()}
 else {ydat=rbind(ydat,ydata[i])
 xdat=rbind(xdat,xdata[i,])}}
ydata.I=ydat[1:length(pos)]
ydata.I.del=ydat[(length(pos)+1):length(ydata)]
xdata.I=xdat[1:length(pos),]
xdata.I.del=xdat[(length(pos)+1):length(ydata),]
return(ydata.I,ydata.I.del,xdata.I,xdata.I.del)}
##############################
# trace function (section 10.4)
 tr=function(M)
 {result=sum(diag(M))
 return(result)}
# END()
```

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *In B.N. petrov and F. Czaki, (eds), 2nd International symposium on information theory. Akademiai Kiado, Budapest, 267-281.*

Andrews, D.F. (1971a). A note on the selection of data transformations. *Biometrika,* **58**, 249-254.

Andrews, D.F., and Pregibon, D. (1978). Finding the outliers that matter. *Journal of the Royal Statistical Sosiety,* ***40****(1): 85-93.*

Atkinson, A.C. (1981). Two graphical displays for outlying and influential observations in regression. *Biometrika,* ***68****(1): 13-20.*

Atkinson, A.C. (1985). Plots, transformations, and regression: an introduction to graphical methods of diagnostic regression analysis. *Oxford: Claredon Press New York: Oxford University Press.*

Atkinson, A.C. (1986). Diagnostic tests for transformations. *Technometrics,* ***28****(1): 29-37.*

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

Barnett, V., and Lewis, T. (1978). Outliers in Statistical data. *John Wiley & Sons.*

Barret, B.E., and Ling, R.F. (1992). General classes of influence measures for multivariate regression. *Journal of the American Statistical Association,* **87**, 184-191.

Beckman, R.J., and Trussell, H.J. (1974). The distribution of an arbitrary studentized residual and the effects of updating in multiple regression. *Journal of the American Statistical Association*, **69**, 199-201.

Belsey, D.A. (1991). Conditioning diagnostics: collinearity and weak Data in Regression. *wiley, New-York.*

Belsey, D.A., Kuh, E., and Welsch, R.E.(1980) Regression Diagnostics: Identifying influential data and sources of collinearity. *John Wiley and Sons.*

Bhenken, D.W., and Draper, N.R. (1972). Residuals and their variance patterns. *Technometrics.* **14**(1)**:** 101-110.

Bingham, C. (1977). Some identities useful in the analysis of residuals from linear regression. *University of Minnesota, shool of statistics, technical report no. 300.*

Box, G.E.P., and Tiao, G.C (1968). A Bayesian approach to some outliers problems. *Biometrika.* **55,** 119-129.

Box, G.E.P., and Tiao, G.C. (1975). Intervention analysis with applications to economic and enviromental problems. *Journal of the American Statistical Association*, **70**, 70-79.

Bozdogan, H. (1988). ICOMP.: A new model selection criterion. *In: H.H. Bock (ed.). Classification and related Methods of Data Analysis, North-Holland, Amsterdam, 599-608.*

Bozdogan, H. (1990). On the information-based measure of covariancecomplexity and its application on the evaluation of multivariate linear models. *Communication in statistics. Theory methods,* **19(1),** 221-278.

Bozdogan, H. and Haughton, D.M.A. (1998). Informational complexity criteria for regression models. *Computational Statistics and Datat analysis*, **28**, 51-76.

Bozdogan, H. and Bearse, P.M. (1999). Model selection using information complexity with applications to vector autoregressive (VAR) models. *Invited contributed paper for JSPI on Model Selection.*

Bradu, D., and Kass, G.V. (1977). Detecting outliers in multiple regression, *Report WISK,* **249**, C.S.I.R., Pretoria.

Bradu, D., and Gabriel, K.R.(1978). The biplot as a diagnostic tool for models of two way tables. *Technometrics, **20** , pp. 4768.*

Brockwell, P.J., and Richard, A.D. (1991). Time series: theory and methods, *2*nd *ed., New York, Berlin: Springer-Verlag.*

Carrol, R.J., and Ruppert, D. (1988). Transformation and weighting in regression, *Chapman & Hall, New York.*

Chalton, D.O., and Troskie, C.A. (1992). Identification of outlying and influential data with biased estimation: a simulation study. *Communication in statistics-simulations,***21**(3)**:**607-627.

Chalton, D.O., and Troskie, C.A. (1996). A Bayesian estimate for the constants in ridge regression. *A South African Statistical Journal (1996), **30**, 119-137.*

Chatterjee, S., and Hadi, A.S. (1988). Sensitivity analysis in Linear Regression. *John Wiley. New York.*

Chatterjee, S., and Hadi, A.S. (1986). Influential observations, high leverage points, and outliers in linear regression (with discussion). *Statistical Sciences*, **1**(3), 379-416.

Chen, J.J (1982). Testing for outliers in linear models. *Ann Arbor, Mich.: Uni-*

versity Mirofilms International.

Colett. D. and Lewis, T. (1976). The subjective nature of outliers rejection procedures, *Applied statistics*, **25**, 228-237.

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

Cook, R.D. (1979). Influential observations in linear regression. *Journal of the American Statistical Association*, **74**, 169-174.

Cook, R.D., and Prescott, P. (1981). On the accuracy of Bonferroni significance levels for detecting outliers in linear models. *Technometrics*, **23**, 59-63.

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

Cooley, W.W., and Lohnes, P.R. (1971). Multivariate data analysis. *John Wiley & Sons.*

David, H.A., and Paulson, A.S. (1965). The performance of several tests for outliers, *Biometrika*, **52**, 429-436.

Dixon, W.J. (1950). Analysis of extreme values, *Annals of Mathematical Statistics*, **21**, 488-506.

Dixon, W.J. (1953). Processing data for outliers, *Biometrics*, **9**, 74-89.

Doornbos, R. (1981). Testing for a single outlier in a linear model. *Biometrics*, **37**, 705-711.

Draper, N.R. (1998). Applied regression analysis. *New York: John Wiley & Sons.*

Draper, N.R., Guttman, I., and Kanemash, H. (1971). The distribution of certain regression statistics, *Biometrika*, **58**, 295-298.

Draper, N.R., and Smith, H. (1981). Applied regression analysis.*2nd Edition, New York: John Wiley & Sons.*

Eckart, C., and Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika,* **1**, 211-218.

Elashoff, J.D. (1972). A model for quadratic outliers in linear regression, *Journal of the American Statistical Association*, ***67***. *478-485.*

Ellemberg, J.H. (1973). The joint distribution of the standardized least squares residuals from a general linear regression, *Journal of the American Statistical Association*. **68**. 941-943.

Ellemberg, J.H. (1976). Testing for a single outlier from a general linear regression.

Biometrics, **32**, 637-645.

Ferguson, T.S. (1961a). On the rejection of outliers, *Proceeding of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*. 253-287.

Ferguson. T.S. (1961b). Rules for rejection of outliers, *Review of the International Statistical Institute*, **29**, 29-43.

Fox, A.J. (1972). Outliers in time series, *Journal of the Royal Statistical Society, Series* **B, 34**, 350-363.

Freund, J.E., and Ronald, E.W. (1980). Mathematical statistics. *Englewood Cliffs, N.J.: Prentice-Hall.*

Gebhardt, F. (1964). On the risk of some strategies for outlying observations, *Annals of Mathematical Statistics*, **35**, 1524-1536.

Gabriel, K. R. (1971). The biplot graphical display of matrices with application to principal component analysis. *Biometrika* , **58**, 453-467.

Gabriel, K.R., and Odoroff, CH.L. (1990). Biplot in biomedical research. *Statistics in Medicine* , **9**, pp. 469485.

Gentleman, J.F., and Wilk, M.B. (1975a). Detecting outliers in a two way table I. statistical behaviour of residuals, *Technometrics*, **17**, 1-14.

Gentleman, J.F., and Wilk, M.B. (1975b). Detecting outliers II. Supplementing the direct analysis of residuals. *Biometrics*, **31**, 387-410.

Gnanadesikan, R., and Kattenring, J.R. (1972). Robust estimates, residuals and outliers detection with multiresponse data, *Biometrics*, **28**, 81-124.

Gnanadesikan, R. (1977). Methods for statistical data analysis of multivariate observations. *John Wiley & Son, New York.*

Goldberger, A.S. (1964). Econometric theory. *John Wiley & Sons.*

Gray, J.B. (1983). The L-R plot: a graphical tool for assessing influence, *In: Proceedings of the statistical computing section, vols. 159-164, American statistical association.*

Gray, J.B. (1985). Graphics for regression diagnostics, *Proceedings of the Statistical Computing Section, vols 102-107, American Statistical Association.*

Graybill, F.A. (1961). An introduction to statistical models. **vol. 1**, *McGraw-Hill.*

Graybill, F.A. (1976). Theory and application of the linear model. *Duxnury press.*

Green, R.F. (1976). Outliers-prone and outliers-resistant distribution, *Journal of the American Statistical Association*, **71**, 502-505.

Greenacre, M.J. (1984). Theory and Applications of Correspondence Analysis. *Academic Press, London.*

Greenacre, M.J., and Underhill, L.G. (1982). Scaling a data matrix in low-dimensional Euclidean space. *In "Topics in Applied Multivariate Analysis" (Hawkins, D.M., ed.), 183-268. Cambridge University Press. Cambridge. UK.*

Gower, J.C., and Hand, D.J. (1996). Biplots, *Chapman & Hall, London.*

Grubbs, F.E. (1969). Procedures for detecting outlying observations in samples, *Technometrics*, **11**. 1-21.

Gunst, R.F., and Mason, R.L. (1976). Generalized mean square error properties of regression estimators. *Communication in statistics*, **A5**, 1501-1508.

Gupta, A.K., and Girko, V.L. (1996). Multidimensional Statistical Analysis of Random Matrices. *Proceedings of the sixth Eugene Lukacs Symposium Bowling Green, OH, USA.*

Gupta, A.K., and Girko, V.L. (1996). Asymptotic behaviour of spectral functions of empirical covariance matrices. *Random Operator and Stochastic Equations*, **2**, 43-61.

Hadi, A.S. (1988). Diagnosing collinearity-influential observations. *Computational statistics and Data Analysis*, **7**, 143-159.

Hadi, A.S. (1990). Two graphical displays for the detection of potentially influential subsets in regression. *Journal of Applied Statistics*, ***17**(3), 313-327.*

Hadi, A.S. (1992). A new measure of overall potential influence in linear regression. *Computational Statistical Data Analysis, **14**, 1-27.*

Hadi, A.S. (1992). Identifying multiple outliers in multivariate data. *Journal of the Royal Statistical Society*, **54**(3), 761-771.

Hadi, A.S. (1994). A modification of a method for the detection of outliers in multivariate samples. *Journal of the Royal Statistical Society B*, **56**, 393-396.

Hadi, A.S., and Wells, M.T. (1990). Assessing the effects of multiple rows on the condition number of a matrix. *Journal of the American Statistical Association*, **85**, 786-792.

Harvey, A.C. (1993). Time series models. *Harvester wheatsheaf savage, Md.: Barnes & Noble Books.*

Hawkins, D.M. (1976). A general approach to outlier testing for the exponential family. institute of statistics mimeo series. **1090**. *Consolidated University of North*

Carolina.

Hawkins, D.M. (1980). Identification of outliers. *Monographs on Applied Probability and Statistics*, **5**, 51-73.

Hawkins, D.M. (1969a). On the distribution and power of a test for a single outlier, *South African Statistical Journal*, **3**, 9-15.

Hawkins, D.M. (1969b). Mathematical analysis of optimal outliers tests, *PhD thesis, University of the Witwaterstand.*

Hawkins, D.M. (1976). A general approach to outlier testing for the exponential family, institute of statistics mimeo series, **1090**, *Consolidated University of North Carolina.*

Hawkins, D.M. (1978a). Analysis of three tests for one or two outliers, *Statistica Neerlandica*, **32**, 137-148.

Hawkins, D.M. (1978b). Fractiles of an extended multiple outlier test, *Journal of Statistical Computation and Simulation, **8**, 227-236.*

Hawkins, D.M. (1980). Identification of outliers. *Monographs on Applied Probability and Statistics*, **5**, 51-73.

Hawkins, D.M. (1991). Diagnostics for use with regression recursive residuals. *Technometrics*, **33**, 221-234.

Healy, M.J.R. (2000). Matrices for statistics. *Oxford: Claredon Press; New York: Oxford University Press.*

Hoaglin, D.C, and Welsch, R.E. (1978). The hat matrix in regression and ANOVA. *The American Statistician*, **32**, 17-22.

Hocking, R.R. (1983). Developments in linear regression methodology: 1959-1952. *Technometrics, **25**,* 219-230.

Hoerl, A.E., and Kennard, R.W. (1970). Ridge regression: Biased Estimation for Non-Orthogonal Problems. *Technometrics*, **12**, 55-67.

Irwin, J.O. (1925). On a criterion for the rejection of outlying observations, *Biometrika*, **17**, 238-250.

Jacobs, M. (1982). Influential observations in linear regression. *Graduate school of Business, University of Cape Town, Unpublished Technical Report.*

Jambu, M. (1991). Exploratory and multivariate data analysis. *Boston: Academic press.* c1991.

Johnson, S.R., Reimer, S.C., and Rothrock, T.P. (1972). Principal components and the problem of multicollinearity. *Metroeconomica*, **25**, 306-314.

Jouan-rimbaud, D., Bouveresse, E., Massart, D.L., and de Noord, O.E. (1999). Detection of prediction outliers and inliers in multivariate calibration, *An. Chim. Acta 388, 282-301.*

Krause, A., and Melvin, O. (1977). The basics of S and S-Plus. Statistics and Computing. *New York: Springer.*

Krzanowski, W.J. (1979). Between-groups comparison of principal components. Journal of the American Statistical Association, **74** (367), 703707. Correction (1981) 76, 1022.

Kullback, S., Leibler, R. (1951). On information and sufficiency. *Annual Mathematical Statistics, **22**, 79-86.*

Langford, I.H., and Lewis, T. (1998). Outliers in multilevel data, *Journal of the Royal Statistical Sosiety, **161**(2), 121-153.*

Lawrence, A.J. (1995). Deletion influence and masking in regression. *Journal of the Royal Statistical Society*, **57**(1), 181-189.

Lingappaiah, G.S. (1976). Effects of outliers on the estimation of parameters, *Metrika*, **23**, 27-30.

Lund, R.E. (1975). Tables of an appropriate test for outliers in linear models. *Technometrics*, **17**, 473-476.

Marquardt, D.W. (1974). Generalized inverse, ridge regression, biased linear estimation and nonlinear estimation. *Technometrics*, **12**, 591-612.

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.

Milliken, G.A., and Graybill, F.A. (1970). Extensions of the general linear hypothesis. *Journal of the American Statistical Association*, **65**, 797-807.

Mosteller, F., and Tukey, J.W. (1977). Data Analysis and Regression. *Reading, Mass: Addison-Wesley.*

Nair, K.R. (1948). The distribution of the extreme deviate from the sample mean and its studentized forms, *Biometrika*, **35**, 118-144.

Nishii, R. (1984). Asymptotic properties of criteria for selection of variables in multiple regression. *Annual Mathematical Statistics*, **12**, 758-765.

Plackett, R.L. (1950). Some theorems in least squares. *Biometrika*, **37**, 149-157.

Pötscher, B.M. (1989). Model selection under nonstationarity: autoregressive models and stochastic linear regression models. *Annual Mathematical Statistics*, **17**, 1257-1274.

Pregibon, D. (1981). Logistic regression, diagnostics, *Annual Mathematical Statistics*, **9**, 45-52.

Prescott, P. (1975). An approximate test for outliers in the linear models, *Technometrics*, **17**, 129-132.

Prescott, P. (1978). Examination of the bihaviour of tests for outliers when more than one outlier is present, *Applied Statistics*, **27**, 10-25.

Rancel, M.M.S., and Sierra, M.A.G. (1999). Measures and procedures for the identification of locally influential observations in linear regression, *Communication in Statistics, Theory Methods*, **28**, 343-366.

Rao, C.R. (1964). The use and interpretation of principal component analysis in applied research, *Sankhya*, **A26**, 392-358.

Rao, C.R. (1973). Linear statistical inference and its applications. *New York, John Wiley & Sons.*

Rao, C.R., and Sujit, K.M. (1971). Generalized inverse of matrices and its applications. *New York, John Wiley & Sons.*

Rawling, J.O., Pantula, S.G., and Dickey, D.A. (1998). Applied regression analysis, *A research tool, 2nd Edition, Springer, New York.*

Sall, J. (1990). Leverage plots for general linear hypotheses. *The American Statistician*, **44**, 308-315.

Schall, R., and Dunne, T.T. (1990). Influential variables in linear regression. *Technometrics*, **32**, 323-330.

Schwager, S.J. (1979). Detecting of multivariate outliers. *Ann Arbor, Mich.: University Microfilms International.*

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

Searle, S.R. (1982). Matrix algebra useful for statistics. *New York: John Wiley & Sons.*

Seber, G.A.F. (1977). Linear Regression Analysis. *John Wiley & Sons, New York.*

Sheesley, J.H. (1977). Tests for outlying observations. *Journal of Quality Technology*, **9**, 38-41.

Sherman, J., and Morrison, W.J. (1949). Adjustment of an inverse matrix corresponding to changes in the elements of a given column or row of the original matrix.

Annual Mathematical Statistics, **20**, 621.

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

Stefansky, W. (1971). Rejecting outliers by maximum normed residual, *Annals of Mathematical Statistics*, **42**, 35-45.

Stone, E.J. (1868). On the rejection of discordant observations, *Monthly Notices of the Royal Astronomical Society*, **28**, 165-168.

Theobald, C.M. (1974). Generalization of mean square error applied to ridge regression. *Journal of the Royal Statistical Society,* **B, 36**, 103-106.

Tiao, G.C., and Guttman, I. (1967). Analysis of outliers with adjusted residuals, *Technometrics*, **9**, 541-568.

Toro-Vizcarrondo, C., and Wallace, T.D. (1968). A test of the mean square error criterion for restrictions in linear regression. *Journal of the American Statistical Association*, **63**, 558-571.

Troskie, C.G., Coutsourides, D., and Jacobs, M. (1980). Detection of outliers in the presence of multicollinearity. *Technical Report No.* **7**, Department of Mathematical Statistics, UCT.

Tukey, J.W. (1949). One degree of freedom for non-additivity. *Biometrics*, **5**, 232-242.

Van Emden, M. (1971). An analysis of complexity. *Mathematisch Centrum Tracts*, **35**, *Mathematisch Centrum, Amsterdam.*

Vinod, H.D. (1976). Application of ridge regression methods to a study of bell System scale economies. *Journal of the American Statistical Association, **71**, 835-841.*

Welsch, R.E., and Kuh, E. (1977). Linear regression diagnostics. *In: R. Launer and A. Siegel, Eds.. Modern Data Analysis. Academic Press. New York.*

Welsch, R.E. (1982). Influence functions and regression diagnostics in modern data analysis. *Academic. New-York.*

Wetherill, G.B. (1986), Regression Analysis with Applications

Wichern, D.W., and Churchill, G.A. (1978). A comparison of ridge estimators. *Technometrics*, **30**, 221-227.

Wilks, S.S. (1963). Multivariate Statistical Outliers. *Sankhya.* ***25**, 407-426.*

Yager, R.M. (1998). Detecting influential observation in nonlinear regression modelling of groundwater flow, *Water Resource Res.*, **34**, 1623-1633.