Hacettepe Journal of Mathematics and Statistics
Volume 31 (2002), 139–151

DETECTION OF INFLUENTIAL OBSERVATION VECTORS FOR MULTIVARIATE LINEAR REGRESSION

B. Altunkaynak* and M. Ekni*

Received 20. 09. 2002

Abstract

In this study, the influence on parameter estimation of observational vectors in a multivariate linear regression model is investigated. A three-stage method is proposed for this investigation. The first stage involves, with the help of a linear restriction, the transformation of the multivariate linear regression model into a restricted multivariate linear regression model. The second includes the calculation of the difference, via the projection theory, between parameter estimates of the multivariate linear regression model and that of the restricted multivariate linear regression model. The third contains the assessment of the influential observations using the generalized Cook's distance. The first two stages in the study facilitate the calculation of the difference between parameter estimates, while the third aids the easy determination of the observational vectors influential on the regression coefficients. In the final section of the study, the calculations are illustrated using a numerical example.

Key Words: Influential observation, generalized Cook's distance, multivariate linear regression, linear restriction, projection theory.

1. Introduction

As seen in Anscombe's regression samples, reaching a significant F statistic for a model in linear regression problems is not always an indicator of the fit (Anscombe, [1]). Especially in determining the regression coefficient, every observation requires a detailed examination. Even a single observation on a parameter estimate can be very influential and the removal of this from the data set can completely change the regression equation. Such an observation is called an "influential observation".

Various effectiveness measures have been developed to test whether an observation is influential or not. The most common of these is Cook's distance (Cook, [6]). This measure is fairly influential in examining the effectiveness of an observation in multiple linear regression models. This measure was later appropriately

*Gazi University, Department of Statistics, Teknikokullar, Beşevler, Ankara, Turkey

expanded by Cook and Weisberg for use in the multivariate linear regression model [7]. One of the measures suggested for the investigation the effectiveness of one or more observations used in the multivariate linear regression model is the generalised Cook's distance (GCD) (Pan and Fang, [11]). Apart from the measures of effectiveness based on Cook's distance, different measures of effectiveness have been proposed by Belsley *et al* [3]. Likewise, Jensen and Ramirez have suggested certain measures to assess the influential and outlier observations in the multiple linear regression model [10].

The model defined as a linear regression is

$$Y_{n\times 1} = X_{n\times m}\beta_{m\times 1} + \epsilon_{n\times 1}, \quad \epsilon \sim N(0, I).$$

Let ψ be the observation set whose influence on the regression model we wish to investigate, and denote by $\hat{\beta}_\psi$ the parameter estimate obtained after removing the observation set ψ from the data set. Most of the measures mentioned above require the calculation of $\hat{\beta}-\hat{\beta}_\psi$. This calculation is especially difficult for the multivariate linear regression model and involves complex matrix processes. Therefore, a study in which linear restrictions were used to easily calculate this difference in the multiple linear regression model was conducted by Pino [12]. In his study, an equation concerning $\hat{\beta}-\hat{\beta}_\psi$ is easily calculated through projection theory. However, his study was carried out for the multiple linear regression model. In parallel to this, a measure of effectiveness in which $\hat{\beta}-\hat{\beta}_\psi$ is used is required to measure the effectiveness of the observation sets.

In this paper, on the other hand, a three-stage method is suggested to assess influential observation vectors in the multivariate linear regression model. The first stage involves transforming the multivariate linear regression model into a restricted multivariate linear regression model through linear restriction. The second involves obtaining the difference between parameter estimates of the multivariate linear regression model and those of the restricted multivariate linear regression model using projection theory. These two stages represent an expansion of Pino's work to the multivariate linear regression model. The third stage involves the determination of influential observations using the GCD. The reason for using the GCD is that this measure is an influential technique for the multivariate linear regression model in investigating the effects of both one or more observation vectors on the regression coefficients. In addition, the fact that this measure is not influenced by such problems as masking makes it more attractive.

This paper is in six sections. In the second section, the calculation of the parameter estimates of the multivariate linear regression model with linear restrictions and the statement of the model equation with the help of the projection matrix is illustrated. In the third section, it is shown that $\hat{\beta}-\hat{\beta}_\psi$ is easily calculated through the use of projection theory. In the forth section, GCD is explained. In the fifth section, how the calculations have been carried out for an application is demonstrated and the numeric values obtained are discussed. In the final section, conclusions and suggestion are presented.

2. First Stage : Parameter Estimations in Linear Restriction

Consider the multivariate linear regression model

$$Y = XB + E, \tag{1}$$

where Y is an $n \times p$ matrix of responses, X is an $n \times m$, $(n > m)$ matrix of known constants of rank m, B is an $m \times p$ parameter matrix and $E \sim N(0, V \otimes I)$. Here, I denotes the identity matrix of order n and V is unknown.

The best linear unbiased estimation (BLUE) of the B parameter B occurring in (1) can be easily obtained through ordinary least squares as follows:

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

A linear restriction on the observation can be shown as:

$$T = A'Y. \tag{2}$$

Here A refers to an $n \times r$ matrix with $\text{rank}(A) = r$, when $r \leq n$ [12]. Thus, under the linear restriction (2), the multivariate linear regression model given in (1) is changed into the following restricted multivariate linear regression model:

$$T = A'XB + A'E. \tag{3}$$

It will be observed for this model that $\text{E}(T) = A'XB$ and $\text{Var}(T) = A'VA$. For the B parameter in (3) BLUE can be obtained as follows:

$$\hat{B}_\psi = (X'A(A'A)^{-1}A'X)^{-1}(X'A(A'A)^{-1}A'Y). \tag{4}$$

For any $q_1 \times q_2$ matrix M, let P_M be the matrix representing the orthogonal projection on the column space of M. It is well known that if M is of rank q_2, $(q_2 \leq q_1)$, then

$$P_M = M(M'M)^{-1}M'$$

([4], p.247). Equation (4) can be formulated as:

$$\hat{B}_\psi = (X'P_AX)^{-1}(X'P_AY), \tag{5}$$

or using the inner product $\langle a, b\rangle = a'b$, Equation (5) is expressed as:

$$\hat{B}_\psi = \langle X, P_AX\rangle^{-1}\langle X, P_AY\rangle.$$

Here

$$\langle a, P_Ab\rangle = \langle P_Aa, b\rangle = \langle P_Aa, P_Ab\rangle \tag{6}$$

is known from the general properties of a projection matrix [5].

$\hat{B}_\psi$ can be written, using Equation (6), as:

$$\hat{B}_\psi = \langle P_AX, P_AX\rangle^{-1}\langle P_AX, Y\rangle. \tag{7}$$

Then, using Equation (7), the model equation can be formulated as:

$$Y = P_AXB + E, \tag{8}$$

where $E \sim N(0, V \otimes I)$.

3. Second Stage : The Projection Theory

In this section we examine how $\hat{\beta} - \hat{\beta}_{\psi}$ is obtained with the help of projection theory. Projection theory can be formulated as:

3.1. Theorem : *A linear regression model is given as follows:*

$$Y = XB + Z\gamma + E, \tag{9}$$

where $E \sim N(0, V \otimes I)$. *If* B^* *is the BLUE in the linear regression model (9), in this case, the estimator* $\hat{B}_{\psi}$ *equals* B^* *under the condition that the matrix* Z *meets the following conditions:*

(i) $\operatorname{rank}(A) + \operatorname{rank}(Z) = n$,

(ii) $Z'A = 0$.

Proof. From (i) the following equation can be written:

$$P_A + P_Z = A(A'A)^{-1}A' + Z(Z'Z)^{-1}Z'. \tag{10}$$

When the equation (10) is multiplied with A on the right-hand side we have

$$(P_A + P_Z)A = A(A'A)^{-1}A'A + Z(Z'Z)^{-1}Z'A. \tag{11}$$

Since $(A'A)^{-1}A'A = I$ and $Z'A$ being equal to zero due to (ii), Equation (11) can be re-written as:

$$(P_A + P_z)A = A$$

If both sides of the above equation are multiplied by $A'(AA')^{-1}$, the result is:

$$P_A + P_Z = I.$$

Using this equation, the following is obtained from (8):

$$Y = (I - P_Z)XB + E = P_A XB + E. \tag{12}$$

From (12) and the above theorem the best linear unbiased estimation (BLUE) of B is seen to be B^*. Using the above theorem we obtain:

$$\hat{B} - \hat{B}_{\psi} = (X'X)^{-1}X'Z\hat{\gamma} \tag{13}$$

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

$$Y = (I - P_X)Z\gamma + E. \tag{14}$$

If $\hat{\gamma}$ is the BLUE of γ in the above model it follows from (14) that

$$\hat{\gamma} = (Z'(I - P_X)Z)^{-1}Z'(I - P_X)Y. \tag{15}$$

Thus, due to the linear restriction $T = A'Y$, the change in the regression coefficient matrix can be formulated as:

$$\hat{B} - \hat{B}_{\psi} = (X'X)^{-1}X'Z(Z'(I - P_X)Z)^{-1}Z'(I - P_X)Y. \tag{16}$$

Notation: For the matrix $G_{a\times b}$, the sub-matrix formed by choosing rows $(i_1, i_2, \ldots, i_r)$ and columns $(j_1, j_2, \ldots, j_s)$ will be denoted by G_D^F, where $D = \{i_1, i_2, \ldots, i_d\}$ and $F = \{j_1, j_2, \ldots, j_s\}$. If $D = \{1, \ldots, a\}$ or $F = \{1, \ldots, b\}$ then the corresponding subscript or superscript will be dropped [12].

If we take, $K = (I - P_X), (I - P_X)Y = (Y - X\hat{B})$ and $L = \hat{B} - \hat{B}_\psi$, then Equation (16) can be written as:

$$L = (X'X)^{-1}X'I^D(K_D^D)^{-1}I_D(Y - X\hat{B}). \tag{17}$$

Specifically, if $D = \{i\}$, then Equation (17) can be re-arranged as:

$$L = (X'X)^{-1}X'I^D I_D(Y - X\hat{B})/k_{ii}. \tag{18}$$

4. Third Stage: Generalized Cook's Distance

Those observations causing significant changes to the parameter estimates are defined as "influential observations" (Belsley *et. al.*, [4]). GCD can be employed to investigate the effect of the observation vector on the parameter estimates. GCD is defined as:

$$DC_\psi = \frac{1}{m}\text{trace}\{L'X'XLU^{-1}\} \tag{19}$$

(Barrett and Ling, [2]). Here the matrix U is a positively-defined matrix having a dimension of $m \times m$. It is possible to select the matrix U as $(E'E)/(n - m)$ ([7], p. 116) and Equation (19) can be reformulated as:

$$DC_\psi = \frac{n-m}{m}\text{trace}\{L'X'XL(E'E)^{-1}\}. \tag{20}$$

Generally speaking, in the case of $DC_\psi > 1$, the observation vector is said to have a major effect on the parameter estimates ([7], p. 118).

5. Application

A study conducted for a national firm that sells automobile tires is described by Green ([9], p. 27). One objective of the study was to determine the interest in the company's newly introduced steel-belted radial tire after respondents had viewed a television commercial advertising the tire. Among the variables measured were an interest score (0-10) for the tire advertised (Y_1), a believability score $(0 - 10)$ for the claims made about the tire in the commercial (Y_2), respondents age (X_1), family size (X_2), education level in years (X_3), and annual income to the nearest hundred dollars (X_4). The data for the first 10 respondents are shown in Table 1.

Table 1 : First 10 Respondents from an Automobile Tire Survey

Respondent	Y_1	Y_2	X_1	X_2	X_3	X_4
1	8	8	40	4	12	190
2	0	6	57	2	12	220
3	9	8	35	1	16	220
4	0	4	54	3	14	220
5	4	6	43	5	14	110
6	7	8	19	6	14	220
7	8	10	38	5	18	220
8	5	5	42	3	14	220
9	8	8	21	2	12	25
10	7	7	19	5	14	190

The value of $I - P_X$ calculated from $P_X = X(X'X)^{-1}X'$ is obtained as

$$I-P_X = \begin{bmatrix} 0.712 & -0.264 & 0.125 & -0.133 & -0.089 & -0.200 & 0.142 & 0.119 & -0.039 & -0.135 \\ & 0.501 & -0.066 & -0.303 & -0.016 & 0.069 & 0.170 & -0.206 & 0.038 & 0.077 \\ & & 0.290 & -0.079 & 0.267 & 0.065 & -0.196 & -0.190 & -0.162 & -0.055 \\ & & & 0.721 & -0.158 & 0.065 & -0.130 & -0.157 & 0.098 & 0.075 \\ & & & & 0.378 & 0.047 & -0.289 & 0.024 & -0.206 & 0.042 \\ & & & & & 0.447 & -0.063 & -0.079 & 0.069 & -0.421 \\ & & & & & & 0.329 & -0.059 & 0.149 & -0.052 \\ & & & & & & & 0.842 & 0.016 & -0.072 \\ & & & & & & & & 0.138 & -0.102 \\ & & & & & & & & & 0.642 \end{bmatrix}$$

Suppose, for instance, that $D = \{1\}$; that is, the question to explore is whether or not the first observation vector is influential on the parameter estimate for the regression coefficients. In this case, $K_D^D = 0.712$.

$$I^D = [1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]'$$

and

$$I_D = [1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]$$

are written as above. The matrix concerning the residuals can be easily calculated, using, $E = (Y - X\hat{B}) = (I - P_X)Y$, as follows:

$$E = \begin{bmatrix} 4.35 & -1.11 & 0.94 & -2.46 & -0.02 & -0.99 & 0.93 & 0.24 & -0.44 & -1.45 \\ 2.01 & 1.12 & 0.15 & -1.83 & -0.79 & 0.04 & 1.65 & -1.58 & 0.30 & -1.07 \end{bmatrix}' \quad (21)$$

These values can be placed in (18) to find the following:

$$L = \begin{bmatrix} 5.513 & 2.551 \\ 0.007 & 0.003 \\ 0.220 & 0.102 \\ -0.501 & -0.232 \\ 0.006 & 0.003 \end{bmatrix}. \tag{22}$$

According to GCD, then $DC_{\{1\}} = 0.344$ is found from (19). Thus, it is hardly possible to say that the first observation parameter or the first individual is influential on the parameter estimates. Likewise, by excluding the other observation vectors from the observation matrix, the differences in the parameter estimates and the GCD values may be found as illustrated in Table 2. The values L_{ij} in the table refer to row i and column j of the matrix L.

Table 2 : Differences in the parameter estimates and their GCD's produced by deleting observation vectors

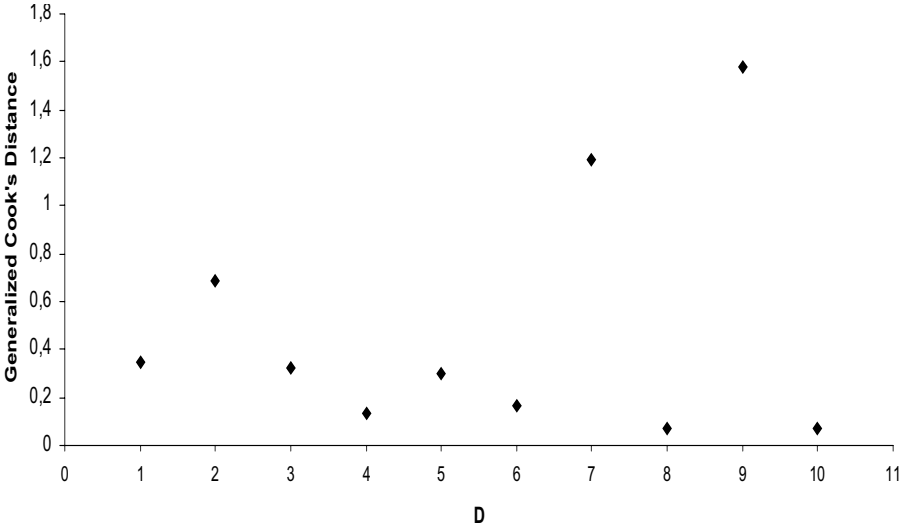
D	$\mathbf{L_{11}}$	$\mathbf{L_{21}}$	$\mathbf{L_{31}}$	$\mathbf{L_{41}}$	$\mathbf{L_{51}}$	$\mathbf{L_{12}}$	$\mathbf{L_{22}}$	$\mathbf{L_{32}}$	$\mathbf{L_{42}}$	$\mathbf{L_{52}}$	$\mathbf{DC_{\{i\}}}$
1	5.513	0.007	0.220	-0.501	0.006	2.551	0.003	0.102	-0.232	0.003	0.344
2	-1.495	-0.018	0.066	0.158	-0.003	1.509	0.018	-0.066	-0.159	0.003	0.688
3	-0.666	-0.03	-0.507	0.228	0.004	-0.103	-0.005	-0.078	0.035	0.001	0.322
4	1.165	-0.035	-0.015	-0.004	-0.001	0.865	-0.026	-0.011	-0.003	0	0.132
5	0.027	-0.001	-0.004	-0.002	0	1.463	-0.031	-0.225	-0.082	0.008	0.298
6	-1.289	0.027	-0.145	0.108	-0.005	0.051	-0.001	0.006	-0.004	0	0.166
7	-5.556	0.017	0.131	0.384	-0.003	-9.821	0.029	0.232	0.678	-0.006	1.196*
8	0.042	0	-0.008	-0.003	0	-0.276	0	0.051	0.021	-0.002	0.072
9	-4.105	0.022	0.223	0.016	0.011	2.787	-0.015	-0.151	-0.011	-0.007	1.576*
10	-1.397	0.028	-0.051	0.064	-0.003	-1.033	0.02	-0.038	0.047	-0.002	0.074

$*$ Influential cases

The GCD values in the table above can be easily examined with the help of Figure 1. As seen in Figure 1, observations 7 and 9 can be said to be influential. Similarly, thanks to GCD, the effect of more than one individual on the parameter estimates can be examined simultaneously. For example, by taking $D = \{1, 2\}$ we can ask whether or not the first and second individuals together are influential on the parameter estimate for the regression coefficients. In this case, (21) remains the same. The other calculations are performed as follows:

$$K_D^D = \begin{bmatrix} 0.712 & -0.264 \\ -0.264 & 0.501 \end{bmatrix}, \tag{23}$$

Figure 1 : The GCD values for the observations in Table 2



$$I^D = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}' \tag{24}$$

and

$$I_D = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{25}$$

If matrices (23), (24) and (25) are placed in Equation (17), the following is found:

$$L = \begin{bmatrix} 6.762 & 7.224 \\ 0.018 & 0.043 \\ 0.200 & 0.026 \\ -0.627 & -0.703 \\ 0.008 & 0.010 \end{bmatrix},$$

and $DC_{\{1,2\}} = 1.914$ is found from (20). That is, the observation values of the first and the second individuals are influential on the parameter estimates. In the same way, the GCD values for all paired combinations are shown in Table 3.

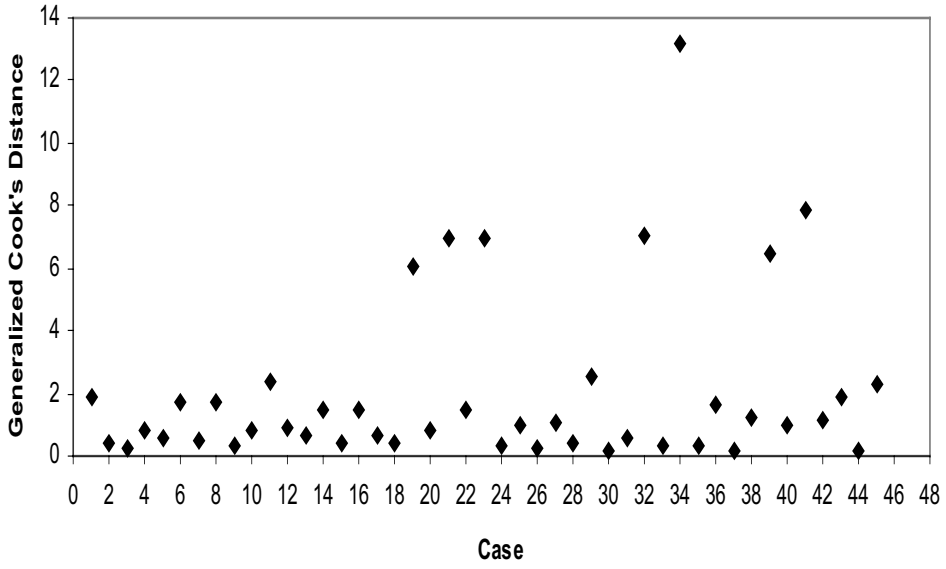
Table 3 : The GCD values where paired observation vectors are deleted

Case	D	$\mathbf{DC_{\{i,j\},i>j}}$	Case	D	$\mathbf{DC_{\{i,j\},i>j}}$	Case	D	$\mathbf{DC_{\{i,j\},i>j}}$
1	1.2	1.914*	16	2.9	1.467*	31	5.6	0.587
2	1.3	0.410	17	2.10	0.668	32	5.7	7.061*
3	1.4	0.270	18	3.4	0.380	33	5.8	0.312
4	1.5	0.793	19	3.5	6.042*	34	5.9	13.141*
5	1.6	0.539	20	3.6	0.790	35	5.10	0.344
6	1.7	1.699*	21	3.7	6.944*	36	6.7	1.660*
7	1.8	0.532	22	3.8	1.471*	37	6.8	0.162
8	1.9	1.710*	23	3.9	6.992*	38	6.9	1.264*
9	1.10	0.304	24	3.10	0.349	39	6.10	6.470*
10	2.3	0.800	25	4.5	0.970	40	7.8	1.019*
11	2.4	2.390*	26	4.6	0.280	41	7.9	7.874*
12	2.5	0.903	27	4.7	1.033*	42	7.10	1.152*
13	2.6	0.379	28	4.8	0.395	43	8.9	1.870*
14	2.7	1.491*	29	4.9	2.564*	44	8.10	0.191
15	2.8	0.379	30	4.10	0.131	45	9.10	2.331*

$*$ Influential cases

The graph of the GCD is given in Figure 2.

Figure 2: The GCD values for the cases in Table 3



As seen in Figure 2, it seems that cases numbered 1, 6, 8, 11, 14, 16, 19, 21, 22, 23, 27, 29, 32, 34, 36, 38, 39, 40, 41, 42, 43 and 45 represent paired observation vectors influential on the parameter estimate. The cases 19, 21, 23, 32, 34, 39 and 41 are especially influential on parameter estimation. When all the cases are considered, the 7 th and 9 th observations, which are influential for the case $D = \{i\}$, are also seen to be influential on parameter estimation in all positions ($i, j = 7$ or 9) for the case $D = \{i, j\}$. On examining Table 3 and Figure 2, it is important to notice that some observation vectors which are not influential in the $D = \{i\}$ case can become quite influential, when included in a paired observation vector. For example, looking in the table for $D = \{3, 5\}$ gives $DC_{\{3,5\}} = 6.042$, which shows that this paired observation vector is quite influential on the parameter estimate. However, for $D = \{i\}$, these observation vectors are not influential. In the same way the influence of the observation vectors obtained by removing various combinations of observation vectors from the observation matrix can be easily calculated. However the number of removed observations will be less than $n/2$, and with an increase in the number of observation vectors removed the effect on the parameter estimate will increase.

6. Conclusions and Suggestions

In this paper, a three-stage method to assess influential observation vectors in multivariate linear regression has been suggested. The first two stages of this method involve an expansion of Pino's work to the multivariate linear regression model. The third stage, on the other hand, proposes using the generalized Cook distance to measure the effectiveness of the observation vectors in the multivariate linear regression model. The linear restrictions and projection theory handled in this paper facilitate the calculation of $\hat{\beta} - \hat{\beta}_\psi$, and make $\hat{\beta} - \hat{\beta}_\psi$ easy to use. The exclusion of unrelated rows and columns from the matrices L and K reduces the dimensions of these matrices, facilitating the calculations. It is, then, quite possible to examine all the sub-combinations using simple computer software due to this useful form of $\hat{\beta} - \hat{\beta}_\psi$.

However, there may occur certain problems in examining all the sub-combinations in general. One of these problems is that the number of cases to be examined rapidly increases as the number of the observation goes up. For example, in the case of $n = 100$ in order to examine the effectiveness of observation vectors with ten components, the number of possible cases is $\binom{100}{10} = 17310309456440$. Considering the fact that a typical computer does one operation per microsecond, the examination of observation vectors with ten components could take up to 150 years to complete, as a rough estimate. The second problem is determining the maximum number of components to be examined in the observation groups. Using heuristic optimization methods rather than examining all possible cases could solve these two problems, ensuring the easy determination of the influential cases.

As shown in the numerical example in the fifth section, non-influential observation vectors of one component can be much more influential when taken as pairs.

This may also be true for other sub-groups of the observation vectors. This is called "masking", and forms a specific field of study.

One of the problems encountered in studies involving influential observations is how to deal with the influential observations. Most researchers maintain that the influential observations should be excluded from the data set and that the analysis should then be performed on the rest. However, it is not desirable, practically speaking, to remove an influential observation from the data set when using small samples, for this can greatly affect the results.

7. Appendices

A1. The OLS estimator of B in the multivariate linear regression model (1) is given by:

$$E'E,$$

where $E'E = (Y - XB)'(Y - XB)$ (Draper and Smith, 1980, p.86).

If a derivation of $E'E$ is taken according to B, then

$$\begin{aligned} \frac{\partial E'E}{\partial B} &= -X'(Y - X\hat{B}) \\ -X'Y + X'X\hat{B} &= 0 \\ \hat{B} &= (X'X)^{-1}X'Y. \end{aligned}$$

A2. As in A1, the OLS estimator of B in the restricted multivariate linear regression model (3) is given by:

$$E'E = (A'Y - A'XB)'(A'A)^{-1}(A'Y - A'XB)$$

If a derivation of $E'E$ is taken according to B, then

$$\begin{aligned} \frac{\partial E'E}{\partial B} &= -X'A(A'A)^{-1}(A'Y - A'X\hat{B}) = 0 \\ \hat{B}_\psi &= (X'A(A'A)^{-1}A'X)^{-1}X'A(A'A)^{-1}A'Y \end{aligned}$$

A3. Equation (12) is obtained from Equation (5) as follows:

$$(M + NDN')^{-1} = M^{-1} - M^{-1}N(N'M^{-1}N + D^{-1})^{-1}N'M^{-1},$$

where M and N are non-singular (Rao, 1973). Then

$$\begin{aligned} \hat{B}_\psi &= (X'P_AX)^{-1}(X'P_AY) \\ &= (X'(I - P_z)X)^{-1}(X'(I - P_z)Y) \\ &= (X'X - X'Z(Z'Z)^{-1}Z'X)^{-1}(X'Y - X'P_zY). \end{aligned}$$

If, $M = X'X, N = X'Z$ and $D = -(Z'Z)^{-1}$, then

$$\hat{B}_\psi = [(X'X)^{-1} - (X'X)^{-1}X'Z(Z'X(X'X)^{-1}XZ - Z'Z)^{-1}Z'X(X'X)^{-1}] \cdot (X'Y - X'P_zY).$$

This can be reformulated as:

$$\hat{B}_\psi = [(X'X)^{-1} - (X'X)^{-1}X'Z(Z'P_XZ - Z'Z)^{-1}Z'X(X'X)^{-1}] \cdot (X'Y - X'P_zY).$$

and if we multiply out and set $R = Z'(I - P_X)Z$,

$$\hat{B}_\psi = \hat{B} + (X'X)^{-1}(X'ZR^{-1}Z'P_XY - (X'X)^{-1}X'P_ZY + +(X'X)^{-1}X'ZR^{-1}Z'P_XP_zY.$$

This formulation can be rewritten, by collecting terms, as:

$$\hat{B} - \hat{B}_\psi = (X'X)^{-1}X'(P_z - ZR^{-1}Z'P_X + ZR^{-1}Z'P_XP_z)Y. \tag{26}$$

In this equation, $Q = (P_Z - ZR^{-1}Z'P_X + ZR^{-1}Z'P_XP_Z)$,??

$$R^{-1}R = I. \tag{27}$$

If the both sides of Equation (27) are multiplied by the matrix Z on the left and by $(Z'Z)^{-1}Z$ on the right, the following is obtained:

$$ZR^{-1}R(Z'Z)^{-1}Z = Z(Z'Z)^{-1}Z$$

Since, $R = Z'(I - P_X)Z$, then

$$ZR^{-1}Z'(I - P_X)P_Z = P_Z.$$

Thus,

$$\begin{aligned} Q &= (ZR^{-1}Z(I - P_X)P_Z - ZR^{-1}Z'P_X + ZR^{-1}Z'P_XP_Z \\ &= Z(R^{-1}Z'(I - P_X)) \end{aligned}$$

This equation, if substituted in (26), gives

$$\hat{B} - \hat{B}_\psi = (X'X)^{-1}X'Z\hat{\gamma}$$

with the help of Equation (15).

References

[1] Anscombe, F. J. Graphs in statistical analysis, Amer. Stat. **27**, 17–21, 1973.

[2] Barrett B. E., and Ling, R. F., General classes of influence measures for multivariate regression, Journal of the American Statistical Association **87**, No. 417, 1992.

[3] Belsley, D. A., Kuh E. and Welch, R. E. Regression diagnostics: Identifying influential data and sources of collinearity, Wiley, New York, 1980.

[4] Belsley, D. A. Conditioning diagnostics:Collinearity and weak data in regression, John Wiley and Sons, New York, 1990.

[5] Bryant, P. Geometry, Statistics, Probability:Variations on a common theme, The American Statistician **38**, No. 1, February 1984.

[6] Cook, R. D. Detection of influential observation in linear regression, Technometrics **19**, No. 1, 15–18, 1977.

[7] Cook, R. D. and Weisberg, S. Residual and influence in regression, Chapman and Hall, New York, 1982.

[8] Draper, N. R. and Smith, H. Applied regression analysis, Wiley, New York, 1980.

[9] Green, P. E. Analyzing multivariate data, Hindsdale, Ill. The Dryden Press, 1978.

[10] Jensen, D. R. and Ramirez, D. E. Bringing order to outlier diagnostics in regression models, http://www.math.virginia.edu/ der/home.html, 2001.

[11] Pan, J. and Fang, K. T. Influential observation in the growth curve model with unstructured covariance matrix, Computational Statistics and Data Analysis **22**, 71–87, 1996.

[12] Pino, G. E. Linear restrictions and two step least squares with applications, Statistics & Probability Letters **2**, 245–248, 1984.

[13] Rao, C. R. Linear statistical inference and its applications, Wiley, New York, 1973.