Communications for Statistical Applications and Methods
2016, Vol. 23, No. 3, 231–239

http://dx.doi.org/10.5351/CSAM.2016.23.3.231
Print ISSN 2287-7843 / Online ISSN 2383-4757

Deletion diagnostics in fitting a given regression model to a new observation

Myung Geun Kim[1,a]

[a]Department of Mathematics Education, Seowon University, Korea

Abstract

A graphical diagnostic method based on multiple case deletions in a regression context is introduced by using the sampling distribution of the difference between two least squares estimators with and without multiple cases. Principal components analysis plays a key role in deriving this diagnostic method. Multiple case deletions of test statistic are also considered when a new observation is fitted to a given regression model. The result is useful for detecting influential observations in econometric data analysis, for example in checking whether the consumption pattern at a later time is the same as the one found before or not, as well as for investigating the influence of cases in the usual regression model. An illustrative example is given.

Keywords: case deletions, covariance matrix, influence, principal components analysis, test statistic

1. Introduction

Regression analysis has been used widely in econometric field. We often have a question, "Is an economic relationship at a later time the same as the one found before?". When we use a linear regression to represent an economic relationship, a statistical expression of this question is to say "Does an observation at a later time follow the regression model estimated by the data obtained before?". This question can be answered statistically in a regression context by taking two steps. The first step is to establish a regression model among economic factors of interest and the second step is to check whether an observation at hand follows this established regression model, by using an appropriate test of hypotheses given in for example Chow (1960) or Ghilagaber (2004). Specific examples are, "Is the consumption pattern at a later time the same as the one found before?", "Is the dependency of the price of a commodity at present on some economic factors the same as the one found before?", and so on.

In a regression context, it is well known that one observation or a few observations can substantially influence the least squares estimators and their relevant quantities (Chatterjee and Hadi, 1988; Cook and Weisberg, 1982). Hence in an analysis of regression data, it is very important to detect such observations and to assess their influence on diverse regression quantities. In our problem, there may exist some influential observations in each of two steps mentioned above. Even a single influential observation in either step can lead us to a wrong analysis result. The first step of our analysis is just the usual regression analysis for which a diagnostic method based on multiple case deletions is introduced in Section 3 by using the sampling distribution of the difference between two least squares

[1] Department of Mathematics Education, Seowon University, 377-3 Musimseoro, Heungdeok-gu, Cheongju 28674, Korea. E-mail: mgkim@seowon.ac.kr

Published 31 May 2016 / journal homepage: http://csam.or.kr

estimators with and without multiple cases. Principal components analysis plays a key role in deriving this diagnostic method. As in the usual regression analysis, some observations can have a large influence in testing whether an observation at hand comes from a given regression model. Hence a suitable method of detecting influential observations is needed for the second step. To this end multiple case deletions of test statistic are derived in Section 4. In Section 5, a numerical example is provided for illustration.

2. Preliminaries

We consider the multiple linear regression model defined by

$$y = X\beta + \varepsilon, \tag{2.1}$$

where $y = (y_1, \ldots, y_n)^T$ is a column vector of response variables of size n, $X = (x_1, \ldots x_n)^T$ is an $n \times p$ full column rank matrix of n measurements on p fixed regressors, β is a column vector of p unknown regression coefficients of size p, and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$ is a column vector of unobservable random errors of size n. The errors $\varepsilon_1, \ldots, \varepsilon_n$ are assumed to be independent and identically distributed as a normal distribution $N(0, \sigma^2)$ with zero mean and variance σ^2.

We denote the least squares estimator of β by $\hat{\beta} = (X^T X)^{-1} X^T y$. The residual vector is given by $e = (e_1, \ldots, e_n)^n = y - X\hat{\beta} = (I_n - H)y$, where $H = (h_{ij}) = X(X^T X)^{-1} X^T$ is the projection matrix and I_n is the identity matrix of order n. We have an alternative expression $e_i = y_i - x_i^T \hat{\beta}$. The residual sum of squares is written as $s^2 = (y - X\hat{\beta})^T (y - X\hat{\beta}) = e^T e$. An unbiased estimator of σ^2 is $\hat{\sigma}^2 = s^2/(n-p)$. The covariance matrix of $\hat{\beta}$ becomes $\mathrm{cov}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$. More details can be found in Seber (1977).

3. Influence in estimating regression coefficients

Let J be an index set of k indices among $1, \ldots, n$. We denote by y_J the column vector formed by the elements of y corresponding to J and let X_J consist of the rows of X indexed by J. When we write as $y_{(J)}$ the vector y from which y_J are removed and as $X_{(J)}$ the matrix X from which the rows of X_J are removed, we have

$$X^T X = X_{(J)}^T X_{(J)} + X_J^T X_J$$

so that

$$\left(X_{(J)}^T X_{(J)}\right)^{-1} = \left(X^T X\right)^{-1} + \left(X^T X\right)^{-1} X_J^T \left(I_k - H_J\right)^{-1} X_J \left(X^T X\right)^{-1}, \tag{3.1}$$

where $H_J = X_J (X^T X)^{-1} X_J^T$. We write as $\hat{\beta}_{(J)}$ the least squares estimator of β for the regression (2.1) computed without the cases indexed by J. Since $(I_k - H_J)^{-1} = I_k + (I_k - H_J)^{-1} H_J$ and $X_{(J)}^T y_{(J)} = X^T y - X_J^T y_J$, a little computation yields

$$\begin{aligned} \hat{\beta}_{(J)} &= \left(X_{(J)}^T X_{(J)}\right)^{-1} X_{(J)}^T y_{(J)} \\ &= \hat{\beta} - \left(X^T X\right)^{-1} X_J^T \left(I_k - H_J\right)^{-1} e_J, \end{aligned}$$

where $e_J = y_J - X_J \hat{\beta}$. Thus we have

$$\hat{\beta} - \hat{\beta}_{(J)} = \left(X^T X\right)^{-1} X_J^T \left(I_k - H_J\right)^{-1} e_J.$$

The sampling distribution of $\hat{\beta} - \hat{\beta}_{(J)}$ is seen to be determined wholly by that of e_J and we need to find it for our purpose. Since $e = (I_n - H)y$, the expectation of the residual vector e is $E(e) = 0$ and its covariance matrix is $\mathrm{cov}(e) = \sigma^2(I_n - H)$. We let Q be a subsidiary matrix of size $k \times n$ that extracts the elements indexed by J from the residual vector e. Since the rows of Q are linearly independent, the rank of Q is just k. We have $e_J = Qe$ which enables us to easily compute the covariance matrix of e_J as

$$\mathrm{cov}(e_J) = Q\,\mathrm{cov}(e)Q^T = \sigma^2(I_k - H_J).$$

The expectation of $\hat{\beta} - \hat{\beta}_{(J)}$ is zero and its covariance matrix is computed as

$$\mathrm{cov}\left(\hat{\beta} - \hat{\beta}_{(J)}\right) = \sigma^2 \left(X^T X\right)^{-1} X_J^T \left(I_k - H_J\right)^{-1} X_J \left(X^T X\right)^{-1}.$$

Since the rank of CAC^T is equal to that of C for a positive definite matrix A and a matrix C of an appropriate size, the rank of $\mathrm{cov}(\hat{\beta} - \hat{\beta}_{(J)})$ is given by that of $(X^T X)^{-1} X_J^T$. Since $X^T X$ is nonsingular, the rank of $(X^T X)^{-1} X_J^T$ is equivalent to that of X_J which is $\min\{k, p\}$. Hence the probability distribution of $\mathrm{cov}(\hat{\beta} - \hat{\beta}_{(J)})$ resides wholly in a $\min\{k, p\}$-dimensional subspace of the p-dimensional Euclidean space.

The influence of deleting the observations corresponding to the index set J on the least squares estimator $\hat{\beta}$ can be measured by the remoteness of $\hat{\beta}_{(J)}$ from $\hat{\beta}$. It is reflected in the covariance matrix $\mathrm{cov}(\hat{\beta} - \hat{\beta}_{(J)})$, not in the mean $E(\hat{\beta} - \hat{\beta}_{(J)})$ because the mean is always zero irrespective of deletions. Thus an influence analysis of deleting observations can be performed using the covariance matrix $\mathrm{cov}(\hat{\beta} - \hat{\beta}_{(J)})$ to which the principal components analysis is applied in order to remove components along redundant axes if any, which will be described in what follows. Since the rank of $\mathrm{cov}(\hat{\beta} - \hat{\beta}_{(J)})$ is $\min\{k, p\}$ (= m, say for convenience), it has m positive eigenvalues $r_1, \ldots, r_m$, and the first m standardized eigenvectors $g_1, \ldots, g_m$ associated with these eigenvalues describe the whole structure of the covariance matrix. The first m principal components of $\hat{\beta} - \hat{\beta}_{(J)}$ are the associated coordinates with respect to the first m eigenvectors $g_1, \ldots, g_m$, whose variances are the eigenvalues $r_1, \ldots, r_m$, respectively. The absolute values of the principal components of $\hat{\beta} - \hat{\beta}_{(J)}$ will become large as $\hat{\beta}_{(J)}$ stays away from $\hat{\beta}$, and so will their variances. Whenever $k < p$, there are $p - k$ redundant axes for describing the probability distribution of $\mathrm{cov}(\hat{\beta} - \hat{\beta}_{(J)})$, and projections of $\hat{\beta} - \hat{\beta}_{(J)}$ along these redundant axes make no contribution to the influence of observations in J on $\hat{\beta}$.

Since σ^2 appears commonly to all the case deletions, we can drop it for performing our principal components analysis based on $\mathrm{cov}(\hat{\beta} - \hat{\beta}_{(J)})$. Let $\hat{r}_1, \ldots, \hat{r}_m$ be the m positive eigenvalues of $(X^T X)^{-1} X_J^T (I_k - H_J)^{-1} X_J (X^T X)^{-1}$ and $\hat{g}_1, \ldots, \hat{g}_m$ be the associated standardized eigenvectors, respectively. We define

$$R_{(J)} = \sum_{i=1}^{m} \hat{r}_i,$$

$$V_{(J)} = \sum_{i=1}^{m} \left\{ \hat{g}_i^T \left(\hat{\beta} - \hat{\beta}_{(J)} \right) \right\}^2.$$

Since $R_{(J)}$ or $V_{(J)}$ or both will become large as $\hat{\beta}_{(J)}$ gets far from $\hat{\beta}$, a reasonable measure of the remoteness of $\hat{\beta}_{(J)}$ from $\hat{\beta}$ is to consider both $R_{(J)}$ and $V_{(J)}$. A graphical display of the pairs $(R_{(J)}, V_{(J)})$ in the plane can be useful for identifying subsets of influential observations for each $k = 1, 2, \ldots$, in which subsets of observations located away from the origin are potentially influential.

4. Influence in testing

4.1. Test statistic

We will review a procedure of checking whether an additional observation follows the regression model (2.1) and more details can be found in Chow (1960). Let y_{n+1} be an observation on the response variable associated with the column vector x_{n+1} of particular measurements on p regressors. In order to check whether a particular single case (y_{n+1}, x_{n+1}) comes from the regression (2.1), we will adopt the procedure described in what follows. First we assume that the case (y_{n+1}, x_{n+1}) follows the regression model

$$y_{n+1} = x_{n+1}^T \beta_* + \varepsilon_{n+1},$$

where β_* is a column vector of p unknown regression coefficients and ε_{n+1} is an unobservable random error distributed as $N(0, \sigma^2)$, independent of ε in the regression (2.1). Then a hypothesis that a particular single case (y_{n+1}, x_{n+1}) comes from the given regression (2.1) is equivalent to the following hypothesis

$$H_0 : \beta = \beta_*.$$

A test of this hypothesis can be performed using the difference between y_{n+1} and the predicted value at x_{n+1} from the regression (2.1), and this difference can be written as

$$\begin{aligned} D &= y_{n+1} - x_{n+1}^T \hat{\beta} \\ &= x_{n+1}^T (\beta_* - \beta) + \epsilon_{n+1} - x_{n+1}^T \left(X^T X\right)^{-1} X^T \varepsilon. \end{aligned}$$

Hence we can easily see that the sampling distribution of the difference D is a normal whose mean and variance are given by

$$\begin{aligned} E(D) &= x_{n+1}^T (\beta_* - \beta), \\ \text{var}(D) &= \left[1 + x_{n+1}^T \left(X^T X\right)^{-1} x_{n+1}\right] \sigma^2, \end{aligned}$$

respectively. Under the null hypothesis H_0, the sampling distribution of the ratio

$$T = \frac{n-p}{1 + x_{n+1}^T (X^T X)^{-1} x_{n+1}} \frac{\left(y_{n+1} - x_{n+1}^T \hat{\beta}\right)^2}{s^2} \tag{4.1}$$

is an F-distribution with degrees of freedom 1 and $n - p$. If the value of the test statistic T is significantly large, we would reject the null hypothesis H_0.

4.2. Multiple case deletions

Using the identity in (3.1), we have

$$x_{n+1}^T \left(X_{(J)}^T X_{(J)}\right)^{-1} x_{n+1} = x_{n+1}^T \left(X^T X\right)^{-1} x_{n+1} + x_{n+1}^T \left(X^T X\right)^{-1} X_J^T (I_k - H_J)^{-1} X_J \left(X^T X\right)^{-1} x_{n+1}. \tag{4.2}$$

The residual sum of squares computed without the cases indexed by J is computed as

$$\begin{aligned} s^2_{(J)} &= \left(y_{(J)} - X_{(J)}\hat{\beta}_{(J)}\right)^T \left(y_{(J)} - X_{(J)}\hat{\beta}_{(J)}\right) \\ &= y^T_{(J)}y_{(J)} - 2\hat{\beta}^T_{(J)}X^T_{(J)}y_{(J)} + \hat{\beta}^T_{(J)}X^T_{(J)}X_{(J)}\hat{\beta}_{(J)} \\ &= s^2 - e^T_J \left[I_k + (I_k - H_J)^{-1} H_J\right] e_J \\ &= s^2 - e^T_J (I_k - H_J)^{-1} e_J. \end{aligned} \tag{4.3}$$

The difference between y_{n+1} and the predicted value at x_{n+1} computed without the cases indexed by I is given by

$$\begin{aligned} D_{(J)} &= y_{n+1} - x^T_{n+1}\hat{\beta}_{(J)} \\ &= D + x^T_{n+1}\left(X^TX\right)^{-1} X^T_J (I_k - H_J)^{-1} e_J. \end{aligned} \tag{4.4}$$

From (4.2) and (4.3), we get

$$\begin{aligned} &\left[1 + x^T_{n+1}\left(X^T_{(J)}X_{(J)}\right)^{-1} x_{n+1}\right] s^2_{(J)} \\ &= \left[1 + x^T_{n+1}\left(X^TX\right)^{-1} x_{n+1}\right] s^2 - e^T_J (I_k - H_J)^{-1} e_J \\ &\quad + x^T_{n+1}\left(X^TX\right)^{-1} X^T_J (I_k - H_J)^{-1} X_J \left(X^TX\right)^{-1} x_{n+1}\left[s^2 - e^T_J (I_k - H_J)^{-1} e_J\right]. \end{aligned} \tag{4.5}$$

When the cases indexed by J are removed from the sample, the test statistic given in (4.1) is then computed as

$$T_{(J)} = \frac{n - p - k}{1 + x^T_{n+1}\left(X^T_{(J)}X_{(J)}\right)^{-1} x_{n+1}} \frac{D^2_{(J)}}{s^2_{(J)}} \tag{4.6}$$

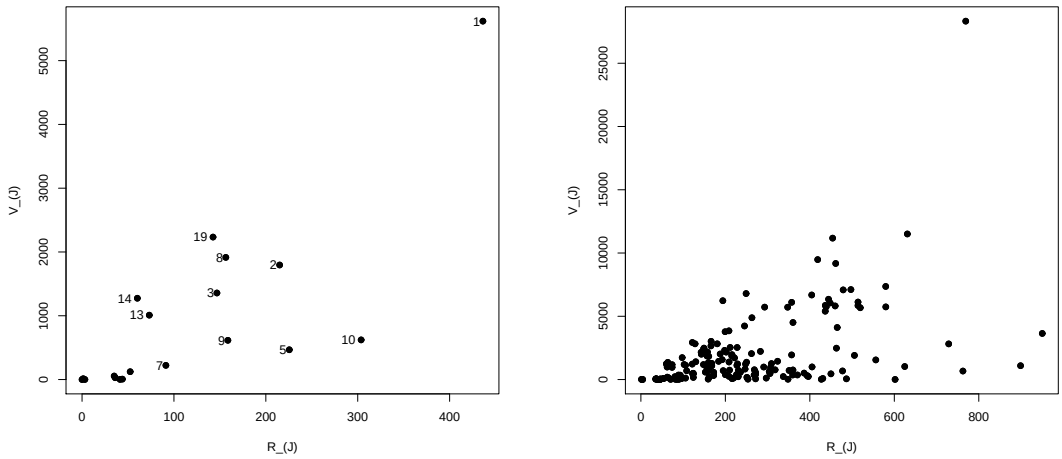
which can be evaluated using (4.4) and (4.5). The large absolute value of $T - T_{(J)}$ implies that the group effect of the cases indexed by J on the test statistic T can be high.

5. A numerical example

Single, double and triple case deletions are performed for the body fat data set (Neter *et al.*, 1996, p.261) which have 20 measurements on a single dependent variable and three independent variables. It is assumed that the intercept term is included in the regression model. For our analysis, we divide the body fat data set into two groups: the first part consists of the first 19 observations and the second part comprises the last observation only. The null hypothesis H_0 is defined as one that the last observation follows the regression estimated by the first 19 observations, and we will investigate the influence of observation belonging to the first part in testing the null hypothesis H_0. Based on the first 19 observations, the least squares estimate of $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^T$ is computed as

$$\beta_0 = 113.72, \qquad \beta_1 = 4.23, \qquad \beta_2 = -2.77, \qquad \beta_3 = -2.13.$$

The value of the test statistic T in (4.1) is computed as 0.061 and its associated p-value is 0.808. Hence we can conclude at reasonable significance levels that the last observation follows the regression formed by the first 19 observations.



(a) Single case deletions (b) Double case deletions

Figure 1: *Graphical displays of the pairs* $(R_{(J)}, V_{(J)})$.

5.1. Single case deletions

For single case deletions, the corresponding graphical display of the pairs $(R_{(J)}, V_{(J)})$ in the plane is included in Figure 1(a). Case 1 is remarkably distinct from the other cases. Cases 2, 8, 10, 19, etc are located far from the origin, lying on the outskirts of the figure. In order to confirm that the cases identified by Figure 1(a) are really influential ones, numerical single case deletions are performed and we investigated the absolute values of $\hat{\beta}_i - \hat{\beta}_{i(r)}$ $(r = 1, 2, \ldots, 19)$ as follows.

1. For $\hat{\beta}_0$, case 1 is most influential and the next ones are in this order 19, 8, 2, 3, 14, 13, etc. For each deletion of cases 1, 19, 8 and 2, $\hat{\beta}_0 - \hat{\beta}_{0(r)}$ becomes -74.9, -47.2, 43.7 and 42.3, respectively.

2. For $\hat{\beta}_1$, case 1 is most influential and the next ones are in this order 19, 8, 2, 14, 13, 3, etc. For each deletion of cases 1, 19, 8 and 2, $\hat{\beta}_1 - \hat{\beta}_{1(r)}$ becomes -2.17, -1.38, 1.36 and 1.27, respectively.

3. For $\hat{\beta}_2$, case 1 is most influential and the next ones are in this order 19, 8, 2, 3, 14, 13, etc. For each deletion of cases 1, 19, 8 and 2, $\hat{\beta}_2 - \hat{\beta}_{2(r)}$ becomes 1.94, 1.21, -1.16 and -1.10, respectively.

4. For $\hat{\beta}_3$, case 1 is most influential and the next ones are in this order 19, 8, 3, 2, 14, 13, etc. For each deletion of cases 1, 19, 3 and 8, $\hat{\beta}_3 - \hat{\beta}_{3(r)}$ becomes 1.113, 0.733, -0.678 and -0.672, respectively.

From this confirmation, we can see that Figure 1(a) provides quite accurate information about the influence of cases on $\hat{\beta}$.

Single case deletions of test statistic given in (4.6) are included in Table 1. The column with the heading T shows the value of the test statistic computed without the corresponding case and the related p-value is in the column with the heading p. Table 1 shows that the removal of each of cases 19 and 1 increases the p-value compared with the others. On the other hand, the deletion of each of cases 8 and 2 causes a great decrease in the p-value compared with the others, and that of case 9 decreases the p-value in the third place. For single case deletions these influential cases in testing H_0 are also influential in estimating the regression coefficients.

Table 1: Single case deletions of test statistic

No.	T	p	No.	T	p	No.	T	p
1	0.005	0.947	8	0.151	0.703	15	0.069	0.797
2	0.133	0.721	9	0.100	0.756	16	0.060	0.810
3	0.048	0.830	10	0.058	0.814	17	0.055	0.818
4	0.036	0.853	11	0.064	0.805	18	0.038	0.849
5	0.063	0.806	12	0.088	0.772	19	0.012	0.915
6	0.051	0.824	13	0.045	0.835			
7	0.066	0.800	14	0.110	0.745			

Table 2: $(R_{(J)}, V_{(J)})$ for some double case deletions

Cases	(1, 19)	(1, 10)	(1, 2)	(1, 8)
$(R_{(J)}, V_{(J)})$	(769.1, 28326.9)	(950.3, 3643.0)	(898.9, 1093.5)	(762.4, 671.9)
Cases	(1, 9)	(1, 7)	(2, 8)	(5, 14)
$(R_{(J)}, V_{(J)})$	(728.8, 2818.1)	(630.8872, 11506.5)	(454.0, 11176.8)	(418.6, 9479.7)

Table 3: Double case deletions of test statistic

No.	T	p	No.	T	p
(1, 4)	0.000	0.997	(2, 8)	0.343	0.568
(1, 19)	0.055	0.817	(4, 19)	0.002	0.970

5.2. Double case deletions

For double case deletions, Figure 1(b) includes the corresponding graphical display of the pairs $(R_{(J)}, V_{(J)})$ in the plane. Some double case deletions lying on the outskirts of Figure 1(b) that are far from the origin are summarized in Table 2. Deletion of double cases (1, 19) is remarkably distinct from the others. We perform numerical double case deletions to show the efficiency of Figure 1(b) based on the absolute value of $\hat{\beta}_i - \hat{\beta}_{i(J)}$ and they are summarized as follows.

1. For $\hat{\beta}_0$, deletions of double cases (1, 19), (1, 7), (2, 8) are highly influential in this order. For each deletion of double cases (1, 19), (1, 7) and (2, 8), $\hat{\beta}_0 - \hat{\beta}_{0(J)}$ becomes -168.2, -107.2 and 105.6, respectively.

2. For $\hat{\beta}_1$, deletions of double cases (1, 19), (2, 8), (5, 14), (1, 7) are highly influential in this order. For each deletion of double cases (1, 19), (2, 8), (5, 14) and (1, 7), $\hat{\beta}_1 - \hat{\beta}_{1(J)}$ becomes -4.90, 3.23, -3.14 and -3.10, respectively.

3. For $\hat{\beta}_2$, deletions of double cases (1, 19), (2, 8), (1, 7), (5, 14) are highly influential in this order. For each deletion of double cases (1, 19), (2, 8), (1, 7) and (5, 14), $\hat{\beta}_2 - \hat{\beta}_{2(J)}$ becomes 4.33, -2.77, 2.77 and 2.52, respectively.

4. For $\hat{\beta}_3$, deletions of double cases (1, 19), (5, 14), (2, 8), (1, 7), (1, 14) are highly influential in this order. For each deletion of double cases (1, 19), (5, 14), (2, 8), (1, 7) and (1, 14), $\hat{\beta}_3 - \hat{\beta}_{3(J)}$ becomes 2.54, 1.75, -1.64, 1.59 and 1.53, respectively.

Some double case deletions which yield significant changes in the value of test statistic are included in Table 3. The p-values over all double case deletions range from 0.568 to 0.997. Each removal of double cases (1, 4) or (4, 19) does not allow us to reject the null hypothesis at any reasonable significance levels, and that of double cases (2, 8) decreases the p-value. Though each single

Table 4: $(R_{(J)}, V_{(J)})$ for some triple case deletions

Cases	(1, 7, 19)	(1, 2, 10)	(1, 2, 8)	(1, 2,19)
$(R_{(J)}, V_{(J)})$	(1026, 46005)	(1631, 31.1)	(1515, 2550)	(1523, 23152)
Cases	(1, 8, 10)	(1, 2, 9)	(1, 10, 19)	(1, 9, 19)
$(R_{(J)}, V_{(J)})$	(1448, 13.4)	(1414, 52.4)	(1416, 28876)	(1248, 27560)

deletion of cases 1 and 19 greatly increases the p-value, the removal of double cases (1, 19) does not cause a noticeable change in the p-value, that is the joint influence of cases 1 and 19 on the test statistic is not severe.

5.3. Triple case deletions

For triple case deletions, a graphical display of the pairs $(R_{(J)}, V_{(J)})$ in the plane, nor provided here, shows that the deletion of triple cases (1, 7, 19) is separated from the main body of the remaining data. Some observations residing in the outskirts of the main body are listed in Table 4.

1. For $\hat{\beta}_0$, the values of $\hat{\beta}_0 - \hat{\beta}_{0(J)}$ over all triple case deletions range from -214.3 to 157.3. The first three most influential subsets are as follows: $\hat{\beta}_0 - \hat{\beta}_{0(J)}$ for each deletion of (1, 7, 19), (1, 18, 19) and (1, 14, 19) becomes -214.3, -181.7 and -180.7, respectively.

2. For $\hat{\beta}_1$, the values of $\hat{\beta}_1 - \hat{\beta}_{1(J)}$ over all triple case deletions range from -6.22 to 4.73. The first three most influential subsets are as follows: $\hat{\beta}_1 - \hat{\beta}_{1(J)}$ for each deletion of (1, 7, 19), (1, 14, 19) and (1, 18, 19) becomes -6.22, -5.38 and -5.33, respectively.

3. For $\hat{\beta}_2$, the values of $\hat{\beta}_2 - \hat{\beta}_{2(J)}$ over all triple case deletions range from -4.067 to 5.513. The first three most influential subsets are as follows: $\hat{\beta}_2 - \hat{\beta}_{2(J)}$ for each deletion of (1, 7, 19), (1, 18, 19) and (1, 14, 19) becomes 5.513, 4.664 and 4.662, respectively.

4. For $\hat{\beta}_3$, the values of $\hat{\beta}_3 - \hat{\beta}_{3(J)}$ over all triple case deletions range from -2.48 to 3.22. The first three most influential subsets are as follows: $\hat{\beta}_0 - \hat{\beta}_{0(J)}$ for each deletion of (1, 7, 19), (1, 14, 19) and (1, 18, 19) becomes 3.22, 2.82 and 2.80, respectively.

Thus we can see that the deletion of triple cases (1, 7, 19) is most influential on all of the regression coefficients.

The p-values over all triple case deletions range from 0.45983 to 0.99993. The deletion of triple cases (1, 6, 18) increases the p-value from 0.808 to 0.99993 while the removal of triple cases (2, 8, 9) decreases the p-value from 0.808 to 0.45983. The dramatic change from the acceptance of H_0 to its rejection can not occur due to any triple case deletion. The deletion of triple cases (1, 7, 19) decreases the p-value from 0.808 to 0.797, but this change is negligible. Hence the deletion of triple cases (1, 7, 19) is influential in estimating all of the regression coefficients but not in testing H_0.

6. Concluding remarks

We note that no assumption about a distributional form is in fact needed just for applying the diagnostic method introduced in Section 3. The diagnostic statistic $\hat{\beta} - \hat{\beta}_{(J)}$ is a vector. It is usually normalized so that subsets of observations can be ordered in a meaningful way (Chatterjee and Hadi, 1988). For example one popular normalized diagnostic statistic is the Cook's distance (Cook, 1977). However, the use of the Cook's distance sometimes fails to identify influential observations correctly

as Kim (2015) has shown, where some results relevant to a single case deletion can also be found. Our diagnostic method does not need any normalization.

The diagnostic method introduced in Section 3 can be used for any statistical problems where the covariance matrix of $\hat{\beta} - \hat{\beta}_{(J)}$ is available theoretically or numerically.

References

Chatterjee S and Hadi AS (1988). *Sensitivity Analysis in Linear Regression*, John Wiley & Sons, New York.

Chow GC (1960). Tests of equality between sets of coefficients in two linear regressions, *Econometrica*, **28**, 591–605.

Cook RD (1977). Detection of influential observation in linear regression, *Technometrics*, **19**, 15–18.

Cook RD and Weisberg S (1982). *Residuals and Influence in Regression*, Chapman and Hall, New York.

Ghilagaber G (2004). Another look at Chow's test for the equality of two heteroscedastic regression models, *Quality and Quantity*, **38**, 81–93.

Kim MG (2015). Influence measure based on probabilistic behavior of regression estimators, *Computational Statistics*, **30**, 97–105.

Neter J, Kutner MH, Nachtsheim CJ, and Wasserman W (1996). *Applied Linear Regression Models* (3rd ed), McGraw-Hill Higher Education/Irwin, New York.

Seber GAF (1977). *Linear Regression Analysis*, John Wiley & Sons, New York.

Received March 15, 2016; Revised April 12, 2016; Accepted April 14, 2016