Journal of the KSQC
Vol. 19, No. 2, Dec.1991

1991년 12월 韓國品質管理學會誌 제19권 제2호

A case-by-case version of C_B statistic in biased estimation

Byoung Jin Ahn*

ABSTRACT

The C_B statistic, a generalization of Mallows's C_L statistic, is developed to determine the shrinkage parameter. Since not all cases in a data set play an equal role in forming C_B, a subdivision of C_B into individual components for each case is developed. This subdivision is useful both as an aid in understanding C_B and as a diagnostic procedure.

1. INTRODUCTION

A regression problem is said to be collinear if there exist approximate linear relationships among the predictor variables. The consequences are well known;in particular, coefficient estimates tend to be inflated and may even have the incorrect sign, and predicted values may be unreasonable (see, eg; Montgomery and peck, 1982).

Collinearities can occur for a variety of reasons. Mason, Gunst and Webster(1975) discuss three general source of collinearities; physical constraints on the model or population, sampling deficience and overdefinition of the model. Mason and Gunst(1985) show that collinearity can be increased without bound by increasing the leverage of a point. They also show that a k variate leverage point can produce k-1 independent collinearities.

* Department of Applied Statistic, Kon-Kuk University.

Walker and Birch(1988) show that, when ridge regression is used to reduce the level of collinearity, the influence of each point changes as a function of the shrinkage parameter. It is not unusual to have collinearity and influential points simultaneously in a data set (see, eg. Lowrence and Marsh, 19984).

Not all the cases in a set of data play an equal role in forming biased estimates. In some problem, collinearity is mainly influenced by only a few data points. For these reasons, it seems that the role of each case should be considered in biased estimation.

2. A class of biased estimators.

The standard linear model with n observations on p input variables is assumed. The data and the model are summarized in matrix as,

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

where X is $n \times p$ design matrix of rank p, and e is an $n-$vector of errors with $E(e) = 0$ and $Var(e) = \sigma^2 I$. For convenience, it is assumed that all of the variables have been standardized so that the sample means are zero and the sum of squares are one.

The singular value decomposition (SVD) of X is an useful technique for the problem of collinearity (Mandel, 1982). The $n \times p$ matrix X may be decomposed as

$$X = U\theta^{1/2} V', \tag{2.2}$$

where $U'U = V'V = I$ and $\theta^{1/2}$ is a diagonal with nonnegative diagonal elements $\theta_j^{1/2}$ $j=1, 2, \cdots, p$ called the singular values of X. The columns of U are the eigenvectors of XX' and the rows of V' are the eigenvectors of $X'X$

Introducing (2.2) into (2.1), we obtain

$$y = U\alpha + e, \tag{2.3}$$

where $\alpha = \theta^{1/2} V'\beta$.

The least squares solution for the unkknown coefficient α is given by

$$\begin{aligned} \hat{\alpha} &= (U'U)^{-1}U'y \\ &= U'y. \end{aligned} \tag{2.4}$$

In the following, we assume that we have a near collinear X matrix.

Consider a general class of biased estimators as

$$\tilde{\alpha} = B\hat{\alpha}, \tag{2.5}$$

where $B = \text{diag}(b_1, b_2, \cdots, b_p)$.

The various estimators can be described as particular cases of $\tilde{\alpha}$(Hocking et. al., 1976).

i) If $b_i=1$, $i=1, 2, \cdots, p$, $\tilde{\alpha}$ is the least squares estimator $\hat{\alpha}$.

ii) If $b_i = \begin{cases} 1 & i \leq t \\ 0 & i > t, \end{cases}$ $\tilde{\alpha}$ is the principal component estimator $\tilde{\alpha}_P$.

iii) If $b_i=c$, $i=1, 2, \cdots, p$, for $0 < c < 1$, $\tilde{\alpha}$ is the shrinkage estimator $\tilde{\alpha}_S$. (Stein, 1960)

iv) If $b_i = \frac{\theta}{\theta + k} i=1, 2, \cdots, p$, $\tilde{\alpha}$ is the ridge estimator $\tilde{\alpha}_R$.

(Hoerl and Kennard, 1970 *a*, *b*)

v) If $b_i = \frac{\theta}{\theta + k} i=1, 2, \cdots, p$, $\tilde{\alpha}$ is the generalized ridge estimator $\tilde{\alpha}_{GR}$

(Hoerl and Kennard, 1970 *a*)

3. Leverage and residuals

Using the estimator (2.5), the vector of fitted value is

$$\tilde{y} = U\tilde{\alpha} = UBU'y. \tag{3.1}$$

Therefore, the matrix $H^* = UBU'$ plays the same role as the hat matrix H in least squares estimation (LSE).

Note that $H = UU' = X(X'X)^{-1}X'$. The $i-$th fitted value can be written as

$$\tilde{y}_i = \sum_{j=1}^{n} h_{ij}^* y_j, \tag{3.2}$$

where h_{ij}^* is the (i, j) -th element of H^*, and consquently, $\partial \tilde{y}_i / \partial y_i = h_{ii}^*$.

The diagonals h_{ii}^* can be interpreted as leverage in the same sense as the hat diagonals h_{ii} of H in LSE.

Using SVD, the leverage of the $i-$th point can be written as

$$h_{ii}^* = U_i' BU_i = \sum_{j=1}^{p} b_j U_{ij}^2, \tag{3.3}$$

where U_i' is the $i-$th row of U and U_{ij} is the $(i, j)-$th element of U.

Since the matrix H^* is not idempotent, the variance of $\tilde{y}_i$ is

$$Var(\tilde{y}_i) = U_i' B^2 U_i \sigma^2. \tag{3.4}$$

Particular values of h_{ii}^* and $U_i' B^2 U_i$ are presented from Table 3.1.

Table 3.1 Particular values of h_{ii}^* and $Var(y_i)/\sigma^2$

Estimators	$h_{ii}^* = U_i' BU_i$	$Var(y_i)/\sigma^2 = U_i' B^2 U_i$
Least Square	$h_{ii} = \sum_{j=1}^{p} U_{ij}^2$	h_{ii}
Principal Component	$\sum_{j=1}^{t} U_{ij}^2$	$\sum_{j=1}^{t} U_{ij}^2$
Shrinkage	Ch_{ii}	$C^2 h_{ii}$
Ridge	$h_{R,i} = \sum_{j=1}^{p} \left(\frac{\theta_j}{\theta_j + k} \right) U_{aj}^2$	$\sum_{j=1}^{p} \left[\frac{\theta_i}{\theta_i + k} \right]^2 U_{ij}^2$
Generalized ridge	$h_{GR,i} = \sum_{j=1}^{p} \left(\frac{\theta_j}{\theta_j + k} \right) U_{aj}^2$	$\sum_{j=1}^{p} \left[\frac{\theta_i}{\theta_i + k} \right]^2 U_{ij}^2$

Several facts can be deduced from (3.3) and Table 3.1. First, $h_{ii}^* < h_{ii}$ for all $i = 1, 2, \cdots, n$; that is, when we adopt a biased estimation, leverage is smaller than the corresponding LSE leverage. Second, the leverage decreases monotonicaly as b_i decreases.

For example, the leverage decreases monotonicaly as shrinkage parameter k increases in Ridge estimation (RE).

The preceding discussion suggests that influence can be affected as b_i changes. Remember, however, that influence is not a function of leverage only but also of residual.

The i-th residual is defined as

$$e_{B,i} = y_i - \tilde{y}_i = e_i + (\hat{y}_i - \tilde{y}_i),$$

where $\hat{y}_i = U_i' \alpha$ and $e_i = y_i - \hat{y}_i$.

From (3.2), we can obtain following equation.

$$\begin{aligned} e_{B,i} &= e_i + \sum_{j=1}^{n} y_j (h_{ij} - h_{ij}^*) \\ &= e_i + \sum_{j=1}^{n} y_j \left(\sum_{m=1}^{p} (1 - b_m)\ U_{im}\, U_{jm} \right). \end{aligned} \tag{3.5}$$

Particular $e_{B,i}$ can be presented from (3.5). For example, $e_{PC,i}$ for principal component estimation (PCE) is given by

$$e_{PC,i} = e_i + \sum_{j=1}^{n} y_j \left(\sum_{m=t+1}^{p} U_{im}\, U_{jm} \right). \tag{3.6}$$

Another example is $e_{R,i}$ in RE, which is given by

$$= e_i + \sum_{j=1}^{n} y_j \sum_{m=1}^{p} \left(\frac{k}{\lambda_m + k} \right)\ U_{im}\, U_{jm}. \tag{3.7}$$

An can be seen in previous discussion, U_{ij} plays an important role. Mason and Gunst (1985) suggest a pairwise satter plot of columns of U to detect the collinearity-influential points. Such points are usually separated from the bulks of other points on the graphs.

4. A case-by-case version of C_B

One good criterion for "closeness" of $\tilde{y}$ to y is the scaled mean squared error of $\tilde{y}$, which is given by

$$\frac{MSE(\tilde{y})}{\sigma^2} = \frac{E(\tilde{y}-U\alpha)'(\tilde{y}-U\alpha)}{\sigma^2}$$

$$= tr(B^2) + \frac{\alpha'(I-B^2)\alpha}{\sigma^2}. \quad (4.1)$$

Let RSS_B and RSS represent the residual sum of squares in some biased estimation and least squares, respectively. Then

$$RSS_B = (y-y)'(y-y)$$

$$= RSS + \alpha'(I-B)^2\alpha, \quad (4.2)$$

and, it follows that

$$E(RSS_B) = \{n - 2tr(B) + tr(B^2)\}\sigma^2 + \alpha'(I-B)^2\alpha. \quad (4.3)$$

Using (4.1) and (4.3), an estimator of $MSE(y)/\sigma^2$ can be presented as follows.

$$C_B = \frac{RSS_B}{\hat{\sigma}^2} - n + 2tr(B), \quad (4.4)$$

where $\hat{\sigma}^2 = y'(I-UU')y/(n-p)$.

Paticular cases for C_B can be presented from (4.4). For example, case for principal component estimation is given by

$$C_{PC} = \frac{RSS_{PC}}{\hat{\sigma}^2} - n + 2t.$$

We may determine the number of principal components to be removed by choosing t to minimize C_{PC}. Another example is C_R in ridge estination, which is given by

$$C_R = \frac{RSS_{PC}}{\hat{\sigma}^2} - n + 2\left(\sum_{i=1}^{p}\left(\frac{\theta_i}{\theta_i + k}\right)\right).$$

And, C_R is identical to C_L statistic (Mallows, 1973), which can be used to determine k by minimizing C_L. In case of shrinkage estimation (SE), C_S is given by

$$C_S = (1 - c)^2 \hat{\alpha}' \hat{\alpha} / \sigma^2 + P(2C - 1).$$

By differentiating C_S with respect to c, we can obtain following shrinkage estimator.

$$\tilde{\alpha}_S = \max [1 - P\hat{\sigma}^2 / \hat{\alpha}' \hat{\alpha}, 0] \hat{\alpha}.$$

which is the estimator referred to as STEINM by Dempster et al. (1977).

To derive another version of C_B statistic as a sum of n components, let us consider an estimator of

$$\frac{MSE(\tilde{y}_i)}{\sigma^2} = \frac{Var(\tilde{y}_i)}{\sigma^2} + \frac{\{E(\tilde{y}_i) - E(y_i)\}^2}{\sigma^2}, \tag{4.5}$$

where $\tilde{y}_i$ is the i-th fitted value in some biased estimation.

If the model (2.3) is unbiased, then $E(y_i) = E(\tilde{y}_i)$ and for $i = 1, 2, \cdots, n$.

$$\frac{1}{\sigma^2} E(\tilde{y}_i - \hat{y}_i)^2$$

$$= \frac{1}{\sigma^2}[U_i' (I - B)^2 U_i \sigma^2 + \{E(\tilde{y}_i) - E(\hat{y}_i)\}^2]. \tag{4.6}$$

From (4.5) and (4.6), we obtain

$$\frac{1}{\sigma^2} MSE(\tilde{y}_i) = \frac{E(\tilde{y}_i - \hat{y}_i)^2}{\sigma^2} + 2U_i' BU_i - U_i' U_i \tag{4.7}$$

Replacing the expectation by its observed value and estimating $\hat{\sigma}^2$, we get

$$C_{B,i} = \frac{(\tilde{y}_i - \hat{y}_i)^2}{\hat{\sigma}^2} + 2U_i' BU_i - U'_i U_i$$

$$= \frac{(e_{B,i} - e_i)^2}{\hat{\sigma}^2} + (h_{ii}^* - h_{ii}) + h_{ii}^*. \tag{4.8}$$

Since $\sum_{i=1}^{n}(\tilde{y}_i - \hat{y}_i)^2 = RSS_B - (n-p)\hat{\sigma}^2$, $\sum_{i=1}^{n} U_i' BU_i = tr(B)$ and $\sum_{i=1}^{n} h_{ii}$

$= \sum_{i=1}^{n} U_i' U_i = tr(UU') = p$, $\sum_{i=1}^{n} C_{B,i} = C_B$.

The equations (4.8) and (3.5), together with Table 3.1, provide the paticular cases for $C_{B,i}$. The case for LSE, $C_{B,i}$ is identical to the $i-$th doagonal of H. As given in (4.8), the values of $C_{B,i}$ depend on the change in residual and the change in leverage. The $C_{B,i}$ has an analogy to $C_{P,i}$ suggested by Weisberg (1981).

The examination of $C_{B,i}$, gives information on the role of each case in determining C_B. Furthermore, it seems imformative to monitor the change in $C_{B,i}$, as the shrinkage parameter k in RE or the number of principal components to be removed vary.

5. A numerical example

We will apply the procedure discussed in the previous sections to a specific real data set presented in Walker and Birch(1988). The data are given in Table 5.1. There are 15 observations on six regressors and a response. After each of the original regressors has been standardized, the estimation procedures including LSE, SE, and RE are adopted to the data. The level of collinearity of this data is moderate, as suggested by the condition number of 119.88. Using the C_B criterion, the values of shrinkage parameters are $c = 0.895$ and $k = 0.031$ for SE and RE, respectively. Since the values of C_S and C_R are smaller than the values of $\sum_i h_{ii}$, SE and RE may be suferior to LSE for this data.

Table 5.1 Data for example

case	$X1$	$X2$	$X3$	$X4$	$X5$	$X6$	Y
1	57.0	6.40	12	293.2	41.1	45.0	61.2
2	53.0	5.00	12	354.3	51.0	31.0	62.3
3	50.3	5.75	14	293.5	24.9	29.4	59.4
4	41.2	4.50	13	299.0	19.4	20.3	66.2
5	36.7	5.15	13	286.0	18.6	17.4	66.0
6	35.5	4.25	10	254.8	17.1	14.9	71.4
7	26.4	3.35	10	270.4	17.6	14.5	75.4
8	25.0	2.50	9	239.2	13.6	13.2	83.2
9	23.5	3.45	11	270.5	14.3	11.7	73.2
10	26.7	6.00	11	298.0	12.9	10.4	71.1
11	25.8	5.70	11	247.0	11.9	15.2	72.8
12	25.7	6.75	12	260.1	12.5	19.5	75.8
13	27.0	4.95	12	228.8	10.5	18.6	75.6
14	24.5	3.65	12	179.4	8.3	19.1	70.2
15	23.1	4.05	11	176.8	8.5	15.9	68.6

Table 5.2 contains some of the caes statistics, namely, leverages, residuals, and $C_{B,i}$ for particular estimation procedures. Recall that $C_{B,i}$ for LSE is identical to h_{ii}. Figure 5.1 shows the index plots of h_{ii}, $C_{S,i}$, and $C_{R,i}$.

Table 5.2 leverages, residuals, and $C_{B,i}$

	LSE		SE			RE		
	h_{ii}	e_i	$h_{S,ii}$	$e_{S,i}$	$C_{S,i}$	$h_{R,ii}$	$e_{R,ii}$	$C_{R,i}$
1	0.7703	−0.084	0.6898	−0.114	0.6708	0.6887	−0.039	0.7386
2	0.8552	0.068	0.7659	0.264	0.7907	0.6774	0.019	0.6615
3	0.3933	−0.023	0.3523	−0.067	0.4436	0.3492	−0.033	0.3122
4	0.3443	0.002	0.3083	0.016	0.2927	0.2865	0.017	0.2454
5	0.2039	0.033	0.1826	0.011	0.1926	0.1769	0.001	0.2193
6	0.5836	0.038	0.5226	0.039	0.4618	0.3904	0.0029	0.4938
7	0.2468	−0.067	0.2211	−0.055	0.2630	0.1976	−0.0041	0.2938
8	0.4242	0.152	0.3799	0.193	0.4468	0.3764	0.187	0.4082
9	0.2870	−0.122	0.2571	−0.096	0.2720	0.2239	−0.080	0.2786
10	0.3767	−0.118	0.3373	−0.102	0.3158	0.3546	−0.120	0.3328
11	0.1721	−0.011	0.1541	−0.002	0.1466	0.1556	−0.023	0.1497
12	0.4213	0.111	0.3773	0.123	0.3431	0.3651	0.150	0.4125
13	0.0984	0.185	0.0881	0.191	0.0803	0.0875	0.200	0.0926
14	0.4047	−0.026	0.3624	−0.023	0.3207	0.3727	−0.034	0.3457
15	0.4183	−0.117	0.3746	−0.112	0.3328	0.3219	−0.173	0.4371
	$C_L = 6$		$C_S = 5.3734$			$C_R = 5.4260$		
	RSS = 0.1354		$RSS_S = 0.1449$			$RSS_R = 0.1562$		

Examination of Table 5.2 and Figure 5.1 reveals several facts.

1. Not all the casè in this data set play an equal role in forming the values of C_B. For example, $C_{S,2} = 0.7097$ and $C_{S,13} = 0.0803$. In some problems, the values of shrinkage parameter might be datermined by only a few cases while most of the cases are essentially ignored.
2. As the figure shows, case 1 and case 2 have relatively large h_{ii}, $C_{S,i}$, and $C_{R,i}$.
3. The change in $C_{B,i}$ ($= h_{ii} - C_{B,i}$) is not same. For example, $h_{22} - C_{R,2} = 0.1937$ and $h_{77} - C_{R,7}$ $= -0.047$. Note that the change in C_B ($= \sum_i h_{ij} - C_R$) is 0.574.
4. The patterns of leverages and residuals for particular estimation procedures are not same. This is mainly caused by the behavior of the leverage (3.3) and residual (3.5) as a function of b given in (2.5).

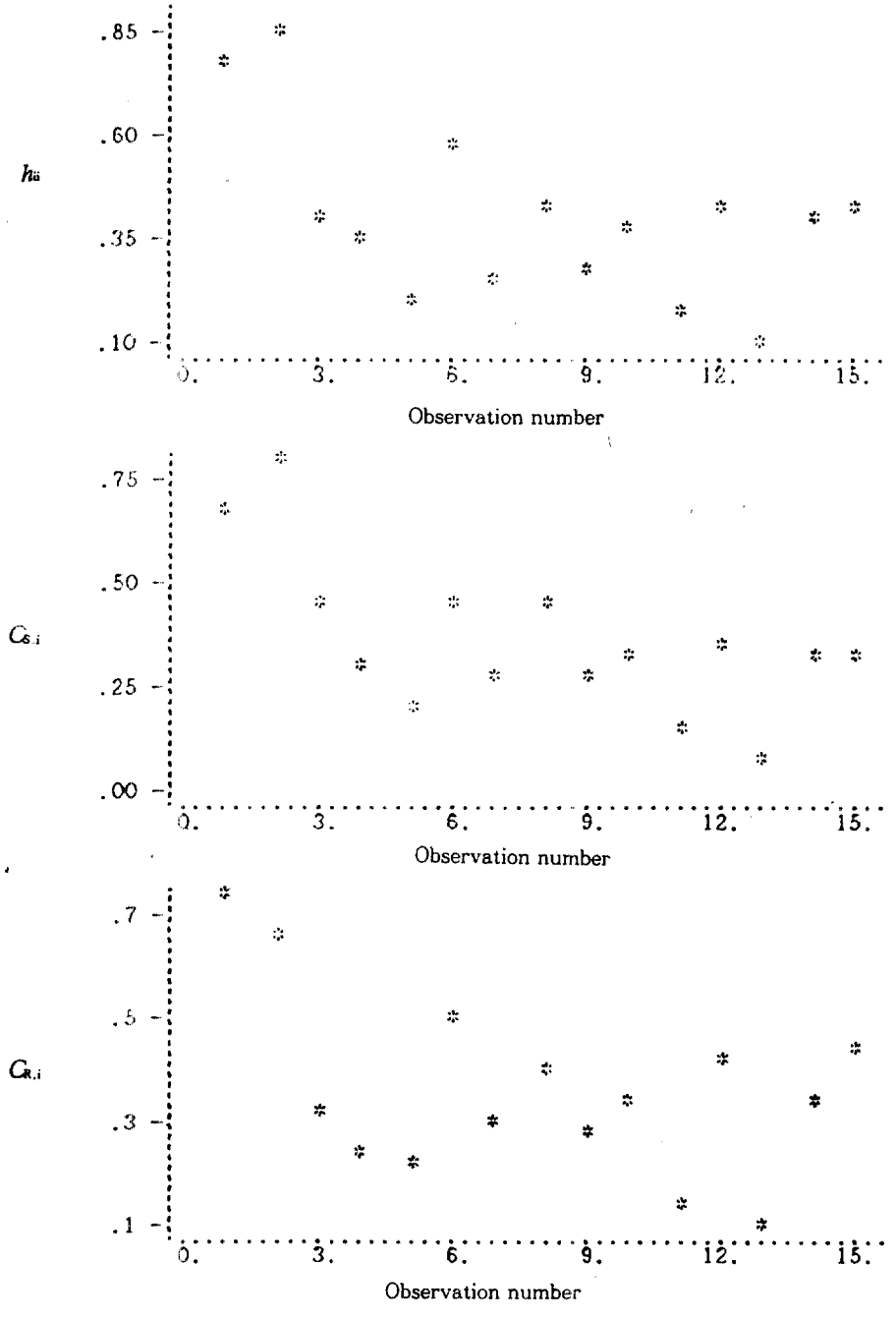

Figure 5.1 Index plots of h_{ii}, $C_{S,i}$ and $C_{R,i}$.

6. Concluding remarks

When biased estimation techniques are used to reduce the effect of collinearity, not all cases in a data set play an equal roll in determining estimtes and the influence of each case changes as a function of shrinkage parameter. The C_B statistic, a generalization of Mallows's C_L statistic, can be used in determining the shrinkage parameters for a class of biased estimators. For these reasons, it seems necessary to study the role of each case in determining C_B. In this paper, a subdivision of C_B statistic into individual components for each case is developed. These components, whose sum in the C_B statistic, are function of the change in residual and the change in leverage. Thie subdivision gives information on the role of each case in determining C_B.

REFERENCES

1) Dempster, A.P., Schatzoff, M. and Wermuth N. (1977), "A simulation study of alternatives to ordinary least squares", Journal of the American Statistical Association, 72, 77-91.

2) Hocking, R. R., Speed, F. M. and Lynn, M. J. (1976), "A class of biased estimatiors in linear regression", Tech., 18, 425-437.

3) Hoerl, A. E. and Kennard, R. W. (1970 a), "Ridge regression : biased estimation for non-orthogonal problems", Tech., 12, 55-67.

4) Hoerl, A. E. and Kennard, R. W. (1970 b), "Ridge regression : applications to non-orthogonal problems", Tech, 12, 69--82.

5) Lowrence, K. D. and Marsh, L. C. (1984), "Robust ridge regression methods for predicting U. S. coal mining fatalitier", Comm. in Stat. -Theory and Methods, 13, 139-149.

6) Mallows, C. L. (1973), "Some comments on C_P", Tech., 15, 661-675.

7) Mandel, J. (1982), "Use of the singular value decomposition in regression analysis", The American Statistician, 36, 15-24.

8) Mason, R. L. and Gunst, R. F. (1985), "Outlier-induced collinearities", Tech., 27, 401-407.

9) Mason, R. L., Webster, J. T. and Gunst, R. F. (1975), "Sources of multi-collinearity in regression analysis", Comm. in Stat., 4, 277-292.

10) Montgomery, D. C and Peck, E. A. (1982), Introduction to linear regression analysis, New York ; John Wiley.

11) Stein, C. M. (1960), "Multiple regression. Contribution to probability and statistics. Essays in honor of Harold Hotelling", ed. I. Olkin, Stanford univ. Press, 424－443.

12) Walker, E. and Birch, J. B. (1988), "Influence measures in ridge regression", Tech., 30, 221－227.

13) Weisberg, S. (1981), "A statistic for allocating C_p to individual cases", Tech., 23, 27－31.