Youngnam Statistical Letters
June 1992, Vol. 3, No. 1, 1-16

Characterization of Predicted Residual Sum of Squares for Detecting Joint Influence in Regression †

Kwang Sik Oh*

ABSTRACT

In regression diagnostics, a number of joint influence measures based on various statistical tools have been discussed. We consider an alternate representation in terms of the predicted residual and g-leverage determined by the remaining points. By this approach, we choose the predicted residual sum of squares for the keypoints as joint influence measure and propose a new expression of it so that we can extend the single case form to the multiple case one. Furthermore we suggest a seach method for joint influence after investigating some properties of the new expression.

1. Introduction

It is well known that inferences based on ordinary least squares regression can be strongly influence by only a few cases in the data, and the fitted model may reflect unusual features of those cases rather than the overall relationship between the variables. An influential case is that, if removed, would substantially change certain important features of the regression analysis under consideration. For example, the deletion of a case may result in large changes in the estimated coefficients, the fitted values or the estimated variances of thier statistics. An influential subset of cases is a natural generalization of an influential case. Detailed discussion of the rational for some existing single case measures and thier extensions to diagnostics for joint influence cases can be found in Belsley, Kuh, and Welsch(1980), Cook and

† 이 논문은 1990년도 대학교수 해외파견 연구계획에 의하여 연구되었음.
* Department of Statistics, Hyosung Women's University, Kyungbuk 713-702, Korea.

Weisberg(1982), and Chatterjee and Hadi(1988). Influence measures for individual cases have been developed which are interpretable in terms of the position of the case in the orthogonal subspaces spanned by the columns of X and the residual vector e. In addition, thier single-case diagnostics can be computed efficiently. However, in contrast with the single-case diagnostics, the detection of joint influence has generally been regarded as computationally prohibitive in view of the large number of subsets involved, and there exists a situation in which observations are jointly but not individually influential, or the other way about. The situation is referred to as a masking effect which means that the influence of one observation is masked by the presence of another observation (Chatterjee and Hadi (1986)).

In this paper, we assume that a subset of k cases, to be called keypoints, has already been selected as joint influential cases. We then discuss and interpret the joint influence measures for these keypoints. In Section 2, we show notations and definitions. In Section 3 we give a review of joint influence measures and examine relationships between measures by predicted residual and g-leverage. In Section 4 we consider the predicted residual sum of squares for the keypoints to detect joint influence. We propose a new expression of it so that we can extend the single case form to the multiple case one in Section 4.1. In Section 4.2, we investigate some properties of the new expression and suggest a seach method for joint influence. Two examples are given in Section 4.3. In Section 5, we give some comments.

2. Definitions and Notations

We consider a classical multiple linear regression model $\mathbf{y} = X\beta + \varepsilon$, where $\mathbf{y}$ is an nx1 vector of response variables, X is an nxp known design matrix of rank p, β is a px1 vector of unknown parameters, and ε is an nx1 vector of random errors with mean vector, $E(\varepsilon) = \mathbf{0}$, and dispersion matrix, $Var(\varepsilon) = \sigma^2 I$. The ordinary least squares estimate of β is given by $\widehat{\beta} = (X'X)^{-1}X'\mathbf{y}$ and the vector of residuals $\mathbf{e}$ is given by $\mathbf{e} = \mathbf{y} - \widehat{\mathbf{y}}$, where $\widehat{\mathbf{y}} = X\widehat{\beta}$. We assume that a subset of k cases, Z, whose joint influence is to be examined, has already been selected. We refer to the cases of this subset as "keypoints". Without loss of generality, we assign the keypoints to the last k rows of the data matrix and remind that in a data analysis situation there are ${}_nC_k$ subsets of k data points whose joint influence may be investigated. Therefore, we partition the data matrix into a "reduced" data matrix, V, augmented by the set of keypoints, Z.

$$W = [X\,,\mathbf{y}] = \begin{bmatrix} V \\ \cdots \\ Z \end{bmatrix} = \begin{bmatrix} X_v, & \mathbf{y}_v \\ X_z, & \mathbf{y}_z \end{bmatrix}.$$

We let $\widehat{\beta}_{(z)}$ denote the OLS estimate based on the reduced data matrix, $(X_v, \mathbf{y}_v)$. That is, $\widehat{\beta}_{(z)} = (X_v'X_v)^{-1}X_v'\mathbf{y}_v$. The predicted residuals $\mathbf{d}$ is given by $\mathbf{d} = \mathbf{y} - \widehat{\mathbf{y}}_{(z)}$, where $\widehat{\mathbf{y}}_{(z)} = X\widehat{\beta}_{(z)}$. That is,

$$\begin{aligned} \mathbf{e} &= \mathbf{y} - \widehat{\mathbf{y}} = [\mathbf{e}_v', \mathbf{e}_z'], \\ \mathbf{d} &= \mathbf{y} - \widehat{\mathbf{y}}_{(z)} = [\mathbf{d}_v', \mathbf{d}_z']. \end{aligned}$$

The elements of $\mathbf{e}$ provide the difference between the y_i and the regression surface fitted to the full data. The vector $\mathbf{e}$ is orthogonal to the space spanned by the columns of X. The elements of $\mathbf{d}$ provide the distance between the y_i and the regression surface fitted to the reduced data. The vector $\mathbf{d}_v$ is orthogonal to the space spanned by the columns of X_v. The residual variance from OLS on the full data set and the reduced data set are given by $S^2 = \mathbf{e}'\mathbf{e}/(n-p)$ and $S^2_{(z)} = \mathbf{d}_v'\mathbf{d}_v/(n-k-p)$ respectively. Following the above notation, the hat matrix, $H = X(X'X)^{-1}X'$, can be partitioned as

$$H = \begin{bmatrix} X_v(X'X)^{-1}X_v' & X_v(X'X)^{-1}X_z' \\ X_z(X'X)^{-1}X_v' & X_z(X'X)^{-1}X_z' \end{bmatrix}.$$

Especially, we set $H_z = X_z(X'X)^{-1}X_z'$.

The hat matrix for the reduced data is given by $G_v = X_v(X_v'X_v)^{-1}X_v'$. Therefore we define $G = X(X_v'X_v)^{-1}X'$ as the augmented hat matrix for the reduced data and G can be partitioned as

$$G = \begin{bmatrix} X_v(X_v'X_v)^{-1}X_v' & X_v(X_v'X_v)^{-1}X_z' \\ X_z(X_v'X_v)^{-1}X_v' & X_z(X_v'X_v)^{-1}X_z' \end{bmatrix} = \begin{bmatrix} G_v & G_{vz} \\ G_{zv} & G_z \end{bmatrix}.$$

The diagonal elements of G provide the squared X-space distance between the i-th case and the centroid of the reduced data, relative to the metric induced by $(X_v'X_v)^{-1}$, for $i = 1, \cdots, n$. The off-diagonal elements of G_v and G_{zv}, that is, g_{ij}

for $i = 1, 2, \cdots, n$ and for $j = 1, 2, \cdots, n-k$, are interpretable as the rate of change in the i-th fitted value, $\widehat{y}_i = x_i\beta_{(z)}$, with respect to y_j. Following Bingham's(1977) results showing that

$$\begin{aligned}\widehat{\beta} - \widehat{\beta}_{(z)} &= (X'X)^{-1}X_z'(I - H_z)^{-1}\mathbf{e}_z \\ &= (X_v'X_v)^{-1}X_z'\mathbf{e}_z\end{aligned}$$

we obtain expressions for the predicted residuals:

$$\mathbf{d} - \mathbf{e} = \widehat{\mathbf{y}} - \widehat{\mathbf{y}}_{(z)} = X(X_v'X_v)^{-1}X_z'\mathbf{e}_z = \begin{bmatrix} G_{vz} \\ G_z \end{bmatrix} \mathbf{e}_z.$$

Therefore,

$$\begin{aligned}\mathbf{d}_v &= \mathbf{e}_v + G_{vz}\mathbf{e}_z, \\ \mathbf{d}_z &= (I + G_z)\mathbf{e}_z.\end{aligned}$$

The above equations reveals that the elements of G_{vz} and G_z, that is, g_{ij} for $i = 1, 2, \cdots, n-k$ and $j = n-k+1, \cdots, n$, are interpretable as the rate of change of the difference in fitted value for the i-th case due to adding the keypoints to the data, with respect to the j-th observed residual. Following Welsch(1982), and using the updating formula for the inverse of a crossproduct matrix,

$$(X_v'X_v)^{-1} = (X'X)^{-1} - (X'X)^{-1}X_z'\{X_z(X'X)^{-1}X_z' - I\}^{-1}X_z(X'X)^{-1}$$

and pre-multiplying by X_z and postmultiplying by X_z', we obtain the following equations;

$$\begin{aligned} G_z &= X_z(X_v'X_v)^{-1}X_z' = H_z(I - H_z)^{-1}, \\ I + G_z &= (I - H_z)^{-1} \\ \textit{and} \quad \mathbf{d}_z &= (I - H_z)^{-1}\mathbf{e}_z.\end{aligned}$$

3. An Alternate Representation of Joint Influence Measures

The diagnostic measures of jointly influential cases are usually based on the residuals $\mathbf{e}$ and the hat matrix $\mathbf{H}$ (Belsley, Kuh, and Welsch, 1980, Cook and Weisberg, 1982). Excellent summaries of the hat matrix $\mathbf{H}$ and residuals $\mathbf{e}$ are given in Rousseeuw & Leroy(1987) and Chatterjee & Hadi(1988). However, the e-residuals, the diagonal elements, h_{ii} of $\mathbf{H}_z$, and the sum of the squared residual and h-leverage for each point are bounded by

$$h_{ii} + \frac{e_i^2}{\sum e_j^2} \leq 1 \quad for \quad i = 1, \cdots, n.$$

That is, potentially influential points, those with high leverage, h_{ii}, pull the fit toward them. Therefore, an outlier may have a small e-residual. But neither the elements d_i of $\mathbf{d}_z$, the diagonal elements g_{ii} of G_z, nor their sum are bounded(Lilliam & Heiberger, 1988). Thus we favor an alternate, but fully equivalent, representation in terms of the predicted residuals for the keypoints, $\mathbf{d}_z$, and the elements of the minor of the augmented hat matrix G_z.

3.1 Outliers

The term "outlier" has been used implicitly or explicitly by many authors to describe peculiar observations, extremely deviant observations, or observations which differ from the main body of the data. These descriptions are truly informal and sometimes confusing. In the following, we shall consider the so-called mean slippage model to characterize certain peculiarity in the data. This approach to outlier identification is to consider an enlarge version of the linear model and has been used by many authors(e.g., Gentleman and Wilk(1975), Cook(1979), Cook and Weisberg(1982)).

The mean slippage model for multiple outling cases can be written as $\mathbf{y} = X\beta + A\delta + \varepsilon$ with $\mathbf{y}, X, \beta, \epsilon$ as in Section 2, δ an unknown $k \times 1$ vector, A an $n \times k$ matrix. Without loss of generality, we can assume $A = [\mathbf{o}', I_z']$. The likelihood ratio test statistic for $H_0 : \delta = 0$ versus $H_1 : \delta \neq 0$ is given by

$$F = \frac{\mathbf{d}_z'(I + G_z)^{-1}\mathbf{d}_z/p}{\mathbf{d}_v'\mathbf{d}_v/(n-k-p)} \tag{3.1}$$

(Oh, 1989). Under the null hypothesis, this statistic F is distributed as F(p,n-k-p). For further details on the detection of outliers see Barnett and Lewis(1978), Hawkins(1980), and Marasinghe(1985).

3.2 Influential Cases

An important class of measures of the influential cases is based on the idea of the influence curve or influence function introduced by Hampel(1974). We give a review of several influence measures based on the influence curve and provide alternate representations. The general form of these measures is given by

$$D(M,c) = \frac{If'\{z, f, \widehat{\beta}(F)\} \cdot M \cdot If\{z, F, \widehat{\beta}(F)\}}{c}$$

where $If\{\cdot\}$ is a generalization of the influence curve for β. A large value of $D(M,c)$ indicates that the Z cases has strong influence on β relative to M and c (Chatterjee & Hadi, 1988). We consider four common approximations for the influence curve and an appropriate choice for M and c. These are (i) the sample influence curve(SIC), (ii) the sensitivity curve(SC), (iii) the empirical influence curve based on all cases(EIC), (iv) the empirical influence curve based on reduced cases($EIC_{(z)}$). The most widely used choices of M are $(X'X)$ and $(X_v'X_v)$. Then we can obtain Table **1** and the following facts(Oh, 1989);

Table 1. Alternative Representations of some Joint Influence Diagnostics

Influence Curve	M	measures	Similarity
SIC ≈ SC	$X'X$	$SIC(F) = \frac{1}{c}\mathbf{d}_z'(I+G_z)^{-1}G_z\mathbf{d}_z$	Cook's distanc
	$X_v'X_v$	$SIC(R) = \frac{1}{c}\mathbf{d}_z'(I+G_z)^{-1}G_z(I+G_z)^{-1}\mathbf{d}_z$	MDFFIT
EIC	$X'X$	$EIC(F) = \frac{1}{c}\mathbf{d}_z'(I+G_z)G_z\mathbf{d}_z$	
	$X_v'X_v$	$EIC(R) = \frac{1}{c}\mathbf{d}_z'G_z\mathbf{d}_z$	Welsch's dist
$EIC_{(z)}$	$X'X$	$EIC_{(z)}(F) = \frac{1}{c}\mathbf{d}_z'(I+G_z)^{-1}G_z(I+G_z)^{-2}\mathbf{d}_z$	Belsley, Kuh, Wel
	$X_v'X_v$	$EIC_{(z)}(R) = \frac{1}{c}\mathbf{d}_z'(I+G_z)^{-2}G_z(I+G_z)^{-2}\mathbf{d}_z$	

(1) The main difference among these measures is in the power of $(I+G_z)^{-l}G_z(I+G_z)^{-m}$.

(2) Setting $M = X'X$ is more sensitive to G_z then setting $M = X_v'X_v$.

(3) These measures are arranged in an increasing order of sensitiving to G_z as follows; $\{EIC_{(z)}(R), EIC_{(z)}(F), SIC(R), SIC(F), EIC(R), EIC(F)\}$.

(4) Well-known influence measures, such as Cook's, Welsch's, Cook & Weisberg's and, Belsley,Kuh & Welsch's, fall into these measures.

4. Predicted Residual Sum of Squares For the Keypoints

The elements of $\mathbf{d}$ provide the distance between the y_i value and the regression surface fitted to the reduced data. We refer to the elements of $\mathbf{d}_z$ as the predicted residuals for the keypoints. In Section 3, we find that all of the joint influence measures are expressed by the quadratic forms of $\mathbf{d}_z$. Now we decompose predicted residual sum of squares for the keypoints, $\mathbf{d}_z'\mathbf{d}_z$, to sum of quadratic forms and investigate properties of the decomposed quadratic forms;

$$\mathbf{d}_z'\mathbf{d}_z = \mathbf{d}_z'(I + G_z)^{-1}\mathbf{d}_z + \mathbf{d}_z'(I + G_z)^{-1}G_z\mathbf{d}_z.$$

From the equation in Section 2, we obtain

$$\begin{aligned} \mathbf{d}_z'(I + G_z)^{-1}\mathbf{d}_z &= \mathbf{e}_z'(I - H_z)^{-1}\mathbf{e}_z \\ \mathbf{d}_z'(I + G_z)^{-1}G_z\mathbf{d}_z &= \mathbf{e}_z'(I - H_z)^{-1}H_z(I - H_z)^{-1}\mathbf{e}_z \\ &= (\mathbf{d} - \mathbf{e})'(\mathbf{d} - \mathbf{e}) \\ &= (\hat{\mathbf{y}} - \hat{\mathbf{y}}_{(z)})'(\hat{\mathbf{y}} - \hat{\mathbf{y}}_{(z)}). \end{aligned}$$

From the above equations we find the following properties. The predicted residual sum of squares for the keypoints is decomposed into two terms. One of the term is the same as "Outlier sum of square" which was proposed by Draper & John(1981) and this is the numerator part of outlier test statistic F in Section 3.1. The other term is the numerator part of Cook's distance and this is the sum of squared changes in fit to all n data points due to augmentation by keypoints. Thus, we consider a joint influence measure PD_z to detect outliers, influential cases, or both.

$$\mathbf{PD}_z = \frac{\mathbf{d}_z'\mathbf{d}_z}{c} \qquad (4.1)$$

$$where \quad c = ps^2.$$

This permits us to detect outliers and influential cases simultaneously. In spite of appealing simplicity, this formular fails to explain the interaction among cases in the subset, and has computational deficiency. In section 4.1, we propose a new expression of PD_z so that the single case can be combined with the multiple case. In section 4.2, we show some properties of the new expression for a special cases.

4.1 A New Expression of PD_z

We consider a new expression of predicted residual sum of squares for the keypoints so that the single case can be combined with the multiple case. Predicted residual sum of squares for the keypoints is written as

$$\begin{aligned}
\mathbf{d}_z'\mathbf{d}_z =& \mathbf{e}_z'(I-H_z)^{-2}\mathbf{e}_z \\
=& \mathbf{e}_z'\left[diag(I-H_z)\right]^{-1}\left[diag(I-H_z)\right](I-H_z)^{-2}\left[diag(I-H_z)\right]\left[diag(I-H_z)\right]^{-1}\mathbf{e}_z \\
=& \frac{1}{c}\mathbf{e}_z'\left[diag(I-H_z)\right]^{-1}Q\left[diag(I-H_z)\right]^{-1}\mathbf{e}_z
\end{aligned}$$

where

$$\begin{aligned}
Q =& \left[diag(I-H_z)\right](I-H_z)^{-2}\left[diag(I-H_z)\right] \\
=& \left\{\left[diag(I-H_z)\right]^{\frac{1}{2}}(I-H_z)^{-1}\left[diag(I-H_z)\right]^{\frac{1}{2}}\right\}^2 \\
=& R_z^{-2}.
\end{aligned}$$

Let $\mathbf{P}_z = \frac{1}{\sqrt{c}}\left[diag(I-H_z)\right]^{-1}\mathbf{e}_z$, then the element P_{z_i}, $z_i \in z$, of P_z is the predicted residual for single case, z_i. We derive a new expression as follows;

$$\mathbf{P}D_z = \mathbf{P}_z'R_z^{-2}\mathbf{P}_z \tag{4.2}$$

where $R_z = \left[diag(I-H_z)\right]^{-\frac{1}{2}}(I-H_z)\left[diag(I-H_z)\right]^{-\frac{1}{2}}$ is the correlation matrix of the residuals for the cases in Z and diag(A) means that zeroes are substituted for the off-diagonal elements of A and the off-diagonal elements of R_z is given by

$$\gamma_{ij} = -\frac{h_{ij}}{\sqrt{(1-h_{ii})(1-h_{jj})}}.$$

Especially, if $R_z^{-2} = I$, then (4.2) is denoted by

$$PD_z^* = P_z'P_z = \frac{1}{c}\{\mathbf{e}_z'\left[diag(I-H_z)\right]^{-1}\left[diag(I-H_z)\right]^{-1}\mathbf{e}_z\} = \sum_{z_i\in z} P_{z_i}^2.$$

which means the sum of the individual influences for the cases in Z. Note that it is theoretically impossible to obtain $R_z^{-2} = I$, but if some of the diagonal elements of H_z are close to 1 or 0, then R_z is close to I (Hadi 1988), and as the result R_z^{-2} will be close to I.

The new expression of (4.2) can be more effectively expressed than (4.1). The reasons are as follows: (i) Since the new expression is made up of P_z based on the single case and R_z based on correlation coefficients between residuals, we can find out the interrelation between the single case and the multiple case. (ii) Elements of $\mathbf{e}_z$, and R_z for all combinations will be simultaneously calculated in most regression diagnostics when the individual influence is calculated, so we only calculate R_z^{-2} when the joint influence is calculated. The problem that we need to calculate R_z^{-2} for all combinations may be improved in section 4.2.

4.2 Properties of the New Expression

We show some properties of the new expression for a case of $k = 2$, $Z = \{z_i, z_j\}$, $i < j$. We derive two eigenvalues and corresponding eigenvectors of $\mathbf{R}_z^{-2}$, where

$$\mathbf{R}_z^{-2} = \begin{bmatrix} \dfrac{(1+\gamma_{ij}^2)}{(1-\gamma_{ij}^2)^2} & \dfrac{-2\gamma_{ij}}{(1-\gamma ij^2)^2} \\ \dfrac{-2\gamma_{ij}}{(1-\gamma_{ij}^2)^2} & \dfrac{(1+\gamma_{ij}^2)}{(1-\gamma_{ij}^2)^2} \end{bmatrix}.$$

The eigenvalues of the matrix are

$$\lambda_1 = (1-\gamma_{ij})^{-2}, \quad \lambda_2 = (1+\gamma_{ij})^{-2}$$

and the coreesponding eigenvectors are

$$Z_1 = \begin{bmatrix} x_1 \\ -x_1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$$

and

$$Z_2 = \begin{bmatrix} x_1 \\ x_1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}.$$

From the above results, we have the following cases to consider the properties;

(1) when $0 < \gamma_{ij} < 1$ $\quad 0 < \lambda_2 < 1 < \lambda_1$
(2) when $\gamma_{ij} = 0$ $\quad \lambda_1 = \lambda_2 = 1$
(3) when $-1 < \gamma_{ij} < 0$ $\quad 0 < \lambda_1 < 1 < \lambda_2$.

Since the case (2) can be identified with the single case, we consider the case of (1) and (3).

Now to determine influential domain for both cases, we take a procedure to calculate the following increment from PD_z^* to PD_z.

$$PD_z - PD_z^* = P_z'(R_z^{-2} - I)P_z. \tag{4.3}$$

Since $\lambda_1 - 1 > 0 > \lambda_2 - 1$ or $\lambda_z^{-2} > 0 > \lambda_1 - 1$, $R_z^{-2} - I$ is indefinite. Hence, (4.3) becomes a hyperbola and for both cases the equations of asymptote are shown as

$$P_{z_j} = \frac{\left(\sqrt{(\lambda_1 - 1)/(1 - \lambda_2)} - 1\right)}{\left(1 + \sqrt{(\lambda_1 - 1)/(1 - \lambda_2)}\right)} P_{z_i} \tag{4.4}$$

and

$$P_{z_j} = \frac{\left(1 + \sqrt{(\lambda_1 - 1)/(1 - \lambda_2)}\right)}{\left(\sqrt{(\lambda_1 - 1)/(1 - \lambda_2)} - 1\right)} P_{z_i}. \tag{4.5}$$

by using $P_{z_i} - P_{z_j}$ coordinates. Note that the slope of (4.4) is the same sign as that of (4.5), and that as $\gamma_{ij} \to 0$ the asymptotes (4.4) and (4.5) asymtotically approach the P_{z_i} axis and the P_{z_j} axis respectively.

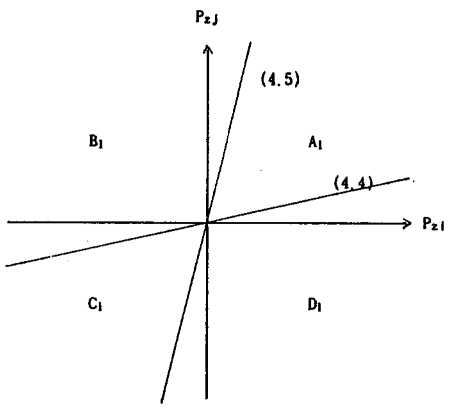

Figure 1. The Domain of the Influential Subset for the case (1)

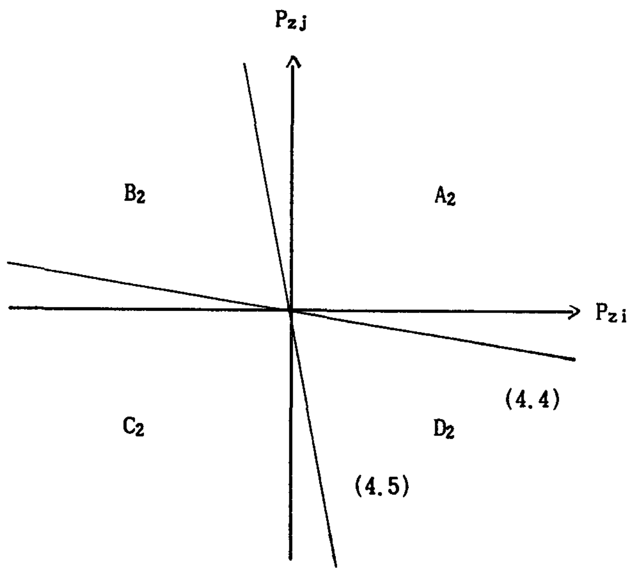

Figure 2. The Domain of the Influential Subset for the case (3)

From the above Figures, the joint influential subset, which means that the increment is positive, is determined by the domain B_1 and D_1 for the case (1), and the domains A_2 and C_2 for the case (3). Note that these domains become larger as $|\gamma_{ij}| \to 1$ from the asymptote property of the asymptote. In particular, if $\gamma_{ij} \to 1$, then (4.4) & (4.5) become $P_{z_j} = P_{z_i}$ and if $\gamma_{ij} \to -1$, then (4.4) & (4.5) become $P_{z_j} = P_{z_i}$. If $\gamma_{ij} P_{z_i} P_{z_j} < 0$, then the influence of the subset is larger than the sum of the individual influences for its elements. Hence we can use it as a rough-and-ready method to detect jointly influential observations. Even if $\gamma_{ij} P_{z_i} P_{z_j} \geq 0$, the subset may be influential, but in this case we regard the subset which is included in the influential domains determined by (4.4) and (4.5) as jointly influential cases. Now in order to decrease the number of candidate cases, we use the correlation coefficient γ_{ij} because there will not be much difference between the joint influence and the sum of the individual influence for the elements of the cases $\{z_i, z_j\}$ when $|\gamma_{ij}|$ is regarded as small. A calibration point for the correlation coefficient is given by $|\gamma_{ij}| > 1.5p/(n-p)$ (Takeuchi 1991).

By using the above properties,we suggest the following steps as a search method for jointly influential cases.

Step 1. Calculate the single case values which are elements of R_z and P_z.
Step 2. Select cases for which $|\gamma_{ij}| > 1.5p/(n-p)$.
Step 3. Calculate (4.2) for subset (i,j), if the cases is included in the influential domains determined by (4.4) and (4.5).
Step 4. Detect jointly influential cases from several viewpoints.

For Step 3, we calculate the ratio of P_{z_j}/P_{z_i} for the subset, which is $\gamma_{ij}P_{z_j}P_{z_i} \geq 0$ (Clearly the subset, which is $\gamma_{ij}P_{z_j}P_{z_i} < 0$, satisfies the condition, $PD_z > PD_z^*$.) and compare it with the slopes of (4.4) and (4.5) such as;

$$P_{z_j}/P_{z_i} < (\sqrt{\cdot}-1)/(\sqrt{\cdot}+1)$$
$$or \quad P_{z_j}/P_{z_i} < (\sqrt{\cdot}+1)/(\sqrt{\cdot}-1)$$
$$where \quad \sqrt{\cdot} = \sqrt{(\lambda_1-1)/(1-\lambda_2)}.$$

4.3 Example

(1) Artificial Data

As an illustration, consider the data in Table 2 . These data are constructed by arbitrarily choosing 20 X values and generating the first 2 responses through the model $y = 5 + x_i + (-1)^{i+1}10 + \epsilon_i$, $i = 1,2$, the last 2 responses through the model $y_i = 5 + x_i + 10 + \epsilon_i$, $i = 19, 20$, and the other responses through the model $y_i = 5 + x_i + \epsilon_i$, where the ϵ_i's are pseudo-random normal variables with mean 0 and variance 1. They are plotted in Figure 3. If case 19 or 20 is deleted, the fitted regression will change very little. If both are deleted, the estimates of parameters may be very different. Conversely, if 1 or 2 is deleted the fitted line will change but if both are deleted, the fitted line will stay about the same. Following the method, cases (19,20), (1,3), (1,4), and (1,5) can be regarded as jointly influential cases from Table 3. However case (1,2) is not detected as jointly influential. These mean that there may exist a situation in which observations are individually influential but not jointly and vice versa.

Table 2. Artificial Data

OBS	X	Y	OBS	X	Y
1	1	12.95	11	16	20.88
2	2	-3.13	12	17	21.57
3	8	14.07	13	18	22.01
4	9	13.89	14	19	23.98
5	10	15.13	15	20	23.38
6	11	15.70	16	21	23.26
7	12	15.27	17	22	27.37
8	13	17.75	18	23	28.86
9	14	19.49	19	24	38.83
10	15	18.93	20	25	41.95

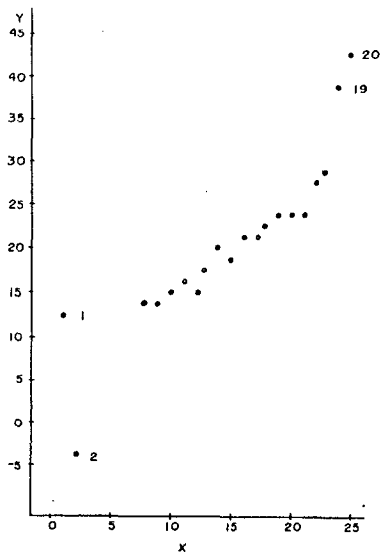

Figure 3. Plot of Artifical Data

Table 3. Summary of Jointly Influential cases in Artificial Data

No	Single	case	multiple case $PD_z \setminus \gamma_{ij}$							
	PD_z	P_{z_i}	1	2	3	4	5	18	19	20
1	4.74	2.17	—	-0.34	-0.20	-0.18	-0.16			
2	2.53	-1.59	(4.32)	—	-0.18	-0.17				
3	0.16	0.39	6.22*	(2.46)	—					
4	0.02	0.13	5.42*	(2.60)		—				
5	0.02	0.13	5.28*				—			
18	0.10	-0.32						—	-0.15	-0.16
19	1.72	1.31						(1.68)	—	-0.18
20	2.90	1.70						(2.88)	6.78*	—

· $F(.10; 2, 18) = 2.62$
· case( ) is not included in joint influence domain.

(2) Adaptive Score Data

This dataset is given by Mickey, Dunn and Clark(1967) and used by Draper & John(1981), Lilliam & Heiberger(1988), Takeuchi(1991), and many authors. According to Cook & Weisberg(1982), they point out cases 2,18, and 19 as influential, and cases (2,18), (18,19), and (11,18) as jointly influential cases by using eigenvalues of the hat matrix as a calibration point. Following the method, cases (2,18) and (11,18) can be regarede as jointly influential cases from Table 4. However case (18,19) is not detected as jointly influential.

Table 4. Summary of Jointly Influential cases in Mickey, Dunn, and Clark Data

No	Single	case	multiple case $PD_z \setminus \gamma_{ij}$				
	PD_z	P_{z_i}	2	9	11	18	19
2	0.53	-0.73	—			-0.56	
9	0.05	0.22		—		0.16	
11	0.60	0.78			—	0.20	
18	1.04	-1.02	8.19*	1.33	2.65*	—	-0.18
19	4.21	2.05				(4.20)	—

· $F(.10; 2, 19) = 2.61$
· case( ) is not included in joint influence domain.

5. Comments

We have concerned an alternate representation of joint influence measures in terms of the predicted residual and g-leverage determined by the remaining points and have found that this is useful for describing the relationships between many popular diagnostics. By this approach, the predicted residual sum of squares for the keypoint is decomposed into two quadratic forms which are used for detecting outliers and influential cases. Hence we have considerd a joint influence measure PD_z which is the predicted residual sum of squares for the keypoints, $d'_z d_z$ divided by $c = ps^2$.

Also we have proposed a new expression of PD_z and suggested a seach method for joint influence. The advantages from the theoretical point of view are that the expression is based not only on the multiple case but also on the single case, and that we can detect jointly influential cases with the masking effect by investigating the difference between the joint influence and the sum of the individual influences. The advantage from the practical point of view is that we can decrease the number of combinations of cases by using the properties as mentioned in Section 4.2.

References

1. Belsley, D., Kuh, E. and Welsch, R.(1980). *Regression Diagnostics*, New York: John Wiley.
2. Bingham, C.(1977). Some Identities Useful in the Analysis of Residual from Linear Regression, *Technical Report 300*, School of Statistics, University of Minnesota.
3. Barnett and Lewis(1978), *Outliers in Statistical Data*, New York; John Wiley & Sons.
4. Chatterjee, S., and Hadi, A.S.(1986). Influential Observations, High Leverage Points, and Outliers in Linear Regression (with discussion), *Statistical Science*, 1, 379-416.

5. Chatterjee, S., and Hadi, A.S.(1988). *Sensitivity Analysis in Linear Regression*, New York: Wiley.
6. Cook, R.D.(1979). Influential Observation in Linear Regression, *Journal of the American Statistical Association*, 74, 169-174.
7. Cook, R.D., and Weisberg, S.(1982). *Residual and Influence in Regression*, New York: Chapman and Hall.
8. Draper, N.R., and John, J.A.(1981). Influential Observations and Outliers in Regression, *Technometrics*, 23, 21-26.
9. Gentleman, J.F., and Wilk, M.B.(1975). Detecting Outliers in a two-way Table: 1. Statistical Behaviour of Residual, *Technometrics*, 17, 1-14.
10. Hampel, F.R.(1974). The Influence Curve and its Role in Robust Estimation, *Journal of the American Statistical Association*, 62, 1179-1186.
11. Hawkins, D.M.(1980), *Identification of Outliers*, London: Chapman and Hall.
12. Lilliam, K. and Heiberger, R.M.(1988). The Orthogonal Decomposition of Joint Influence Measures in Regression, *Technical Report* # 88-7, Department of Statistics, Temple University.
13. Marasinghe, M.G.(1985). A Multistage Procedure for Detecting Several Outliers in Linear Regression, *Technometrics*, 27, 395-399.
14. 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.
15. Oh, K.S.(1989). *A Study on Joint Influence Measures in Regression Diagnostics*, Ph.D. Dissertation, Department of Statistics, Korea University.
16. Rousseeuw, P.J. and Leroy, A.M.(1987). *Robust Regression and Outlier Detection*, John Wiley & Sons, New York.
17. Takeuchi, H.(1991). Dectecting Influential Observations by Using a New Expression of Cook's Distance, *Commun. Statist. -Theory Meth.*, 20(1), 261-274.
18. Welsch, R.E.(1982). Influence Functions and Regression Diagnostics, *Mordern Data Analysis*, R.L. Launert and A.F. Siegel, editors, New York: Academic Press, 149-169.