REVSTAT – Statistical Journal
Volume 14, Number 1, February 2016, 75–87

ON THE BOUNDS FOR DIAGONAL AND OFF-DIAGONAL ELEMENTS OF THE HAT MATRIX IN THE LINEAR REGRESSION MODEL

Author: Mohammad Mohammadi
– Department of Statistics, University of Isfahan,
Isfahan, 81746-73441, Iran
m.mohammadi@sci.ui.ac.ir

Received: April 2014 Revised: September 2014 Accepted: November 2014

Abstract:

- In the least squares analysis, an appropriate criterion to detect potentially influential observations, either individually or jointly, deals with the values of corresponding Hat matrix elements. Hence, some conditions for which these elements give the extreme values are interesting in the model sensitivity analysis. In this article, we find a new and sharper lower bound for off-diagonal elements of the Hat matrix in the intercept model, which is shorter than those for the no-intercept model. We give necessary and sufficient conditions on the space of design matrix, under which the corresponding Hat matrix elements get desired extreme values.

Key-Words:

- *Hat matrix; high-leverage; influential observations; linear regression model.*

AMS Subject Classification:

- 62J05.

1. INTRODUCTION

In the least squares approach, any sensitivity analysis is essentially related to how points are observed, so reflected on the elements of the Hat matrix. As the most widely used concepts in regression diagnostics, influential observations and outliers are identified by the size of these quantities. Consider the general linear regression model

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i , \qquad (i = 1, 2, ..., n) , \tag{1.1}$$

where y_i is the i-th observed response, $\mathbf{x}_i$ is a $p \times 1$ deterministic vector, $\boldsymbol{\beta} \in R^p$ is an unknown $p \times 1$ vector of parameters, and the ε_i's are uncorrelated errors with mean zero and variance σ^2. Writing $\mathbf{y} = (y_1, ..., y_n)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_n)'$, and $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_n)'$, model (1.1) can be written as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} . \tag{1.2}$$

The matrix $\mathbf{X}$ is called *design matrix*, which contains the column one in the intercept model. We assume throughout that $\mathbf{X}$ is full-rank matrix, so $\mathbf{X}'\mathbf{X}$ is nonsingular. In this case the ordinary least squares estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} . \tag{1.3}$$

The $n \times 1$ vector of ordinary predicted values of the response variable is $\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$, where the $n \times n$ prediction or Hat matrix, $\mathbf{H}$, is given by

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' . \tag{1.4}$$

The residual vector is given by $\mathbf{e} = (\mathbf{I}_n - \mathbf{H})\mathbf{y}$ with the variance-covariance matrix $\mathbf{V} = (\mathbf{I}_n - \mathbf{H})\sigma^2$, where $\mathbf{I}_n$ is the identity matrix of order n. The matrix $\mathbf{H}$ plays an important role in the linear regression analysis. Let h_{ij} indicate the (i, j)-th element of $\mathbf{H}$. Hence,

$$h_{ij} = \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_j , \qquad (i, j = 1, 2, ..., n) . \tag{1.5}$$

The diagonal element h_{ii} is so-called the leverage of the i-th data point and measures how far the observation $\mathbf{x}_i$ is from the rest of points in the $\mathbf{X}$-space. Any point with large values of h_{ii} tends to be an influential observation. Such a point is called *high-leverage*. Cook and Weisberg (1982, p. 13) point to the following conditions, to h_{ii} be large:

- $\mathbf{x}_i'\mathbf{x}_i$ is large relative to the square of the norm $\mathbf{x}_j'\mathbf{x}_j$ of the vectors $\mathbf{x}_j$; i.e. $\mathbf{x}_i$ is far removed from the bulk of other points in the data set, or
- $\mathbf{x}_i'\mathbf{x}_i$ is substantially in the direction of an eigenvector corresponding to a small eigenvalue of $\mathbf{X}'\mathbf{X}$.

The various criteria are suggested for the size of h_{ii} to $\mathbf{x}_i$ being high-leverage (see Chatterjee and Hadi, 1988, p. 100–101).

On the other hand, off-diagonal elements of the Hat matrix may be regarded as another criterion in the regression analysis. Ignoring the constant σ^2, these elements are covariances of any pair of the estimated residuals, so can be useful to check the independency assumption. From theoretical point of view, there may exist situations in which observations are jointly but not individually influential (Chatterjee and Hadi, 1988, p. 185). Huber (1975) mentions that large values of h_{ij} typically correspond to outlying design points. Hadi (1990) proposed two graphical displays of the elements of $\mathbf{H}$, that are useful in the detection of potentially influential subsets of observations.

In this paper we discuss the necessary and sufficient conditions for the design matrix to have some extreme values of Hat matrix elements, in the intercept and no-intercept linear regression models. We obtain a sharper lower bound for off-diagonal elements of the Hat matrix in the with intercept linear model, which is shorter than those for no-intercept model by $1/n$.

Repeated application of the following first lemma is made. Part (a) of this lemma is due to Chipman (1964).

Lemma 1.1. *Let $\mathbf{A}$ be a matrix of $n\times p$ with rank $p-m_1$, $(m_1>0)$.*

(**a**) *If $\mathbf{B}$, of order $m_1\times p$ and full row rank, has it's rows LIN (linearly independent) of those of $\mathbf{A}$, then*

$$\mathbf{A}(\mathbf{A}'\mathbf{A}+\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}' = \mathbf{0}_{n\times m_1} \quad \textit{and} \quad \mathbf{B}(\mathbf{A}'\mathbf{A}+\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}' = \mathbf{I}_{m_1}\ .$$

(**b**) *If $\mathbf{R}$, of order $m_2\times p$; $(m_2\le m_1)$ and rank 1, has the first row $\mathbf{r}'$ of the form $\mathbf{R}=\boldsymbol{\delta}\mathbf{r}'$, where $\boldsymbol{\delta}=(1,\delta_2,...,\delta_{m_2})'$, and $\mathbf{r}$ be LIN of rows of $\mathbf{A}$, then*

$$\mathbf{R}(\mathbf{A}'\mathbf{A}+\mathbf{R}'\mathbf{R})^{-1}\mathbf{R}' = \frac{\boldsymbol{\delta}\boldsymbol{\delta}'}{\|\boldsymbol{\delta}\|^2}\ .$$

Lemma 1.2. *Let $\mathbf{A}$ and $\mathbf{B}$ be $n\times p$ matrices. Then,* $\operatorname{rank}(\mathbf{A}-\mathbf{B}) = \operatorname{rank}(\mathbf{A})-\operatorname{rank}(\mathbf{B})$, *if and only if $\mathbf{A}\mathbf{A}^-\mathbf{B}=\mathbf{B}\mathbf{A}\mathbf{A}^-=\mathbf{B}\mathbf{A}^-\mathbf{B}=\mathbf{B}$, where $\mathbf{A}^-$ is a generalized inverse of $\mathbf{A}$ satisfying $\mathbf{A}\mathbf{A}^-\mathbf{A}=\mathbf{A}$ (see Seber, 2007).*

Throughout this paper we use the notation (i) written as a subscript to a quantity to indicate the omission of the i-th observation. For example, $\mathbf{X}_{(i)}$ and $\mathbf{X}_{(ij)}$ are matrix $\mathbf{X}$ with the i-th row and (i,j)-th rows omitted, respectively. The vector $\bar{\mathbf{x}}$ denotes the mean of $\mathbf{X}$'s rows and $\mathbf{J}_p$ is a $p\times p$ matrix of ones.

2. BOUNDS FOR DIAGONAL ELEMENTS OF THE HAT MATRIX

This section is allotted to determine the lower and upper bounds of h_{ii}, along with necessary and sufficient conditions for observation matrix $\mathbf{X}$ to take those values. These conditions are fundamentally on the basis of some special forms of $\mathbf{x}_i$ and $\mathbf{X}_{(i)}$. We consider two customary full rank linear regression models; without and with intercept.

Lemma 2.1. *Let $\mathbf{X}_{n\times k}$ be full column rank matrix without column one. Then,*

(i) $0 \le h_{ii} \le 1$.

(ii) h_{ii}=0, *if and only if* $\mathbf{x}_i = \mathbf{0}$.

(iii) h_{ii}=1, *if and only if* $\operatorname{rank}(\mathbf{X}_{(i)}) = k-1$.

Proof: Part (i) is immediately proved since $\mathbf{H}$ and $\mathbf{I}_n - \mathbf{H}$ are positive semi-definite (p.s.d.) matrices. Similarly part (ii) is obtained since $(\mathbf{X}'\mathbf{X})^{-1}$ is a p.d. matrix. To verify part (iii), without loss of generality, suppose that $\mathbf{x}_i$ is the last row of $\mathbf{X}$, i.e. $\mathbf{X}' = \begin{bmatrix} \mathbf{X}'_{(i)} & \mathbf{x}_i \end{bmatrix}$. If h_{ii}=1, then

$$\mathbf{H} = \begin{bmatrix} \mathbf{X}_{(i)}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_{(i)} & \mathbf{0}_{(n-1)\times 1} \\ \mathbf{0}_{1\times(n-1)} & 1 \end{bmatrix}. \tag{2.1}$$

Since $\mathbf{H}$ is an idempotent matrix, $\mathbf{X}_{(i)}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_{(i)}$ is also idempotent. Hence,

$$\operatorname{rank}(\mathbf{X}_{(i)}) \;=\; \operatorname{rank}\big(\mathbf{X}_{(i)}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_{(i)}\big) \;=\; \operatorname{trace}\big(\mathbf{X}_{(i)}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_{(i)}\big) \;=\; k-1 \;.$$

Conversely, let $\operatorname{rank}(\mathbf{X}_{(i)}) = k-1$. Since, $\operatorname{rank}\begin{bmatrix} \mathbf{X}'_{(i)} & \mathbf{x}_i \end{bmatrix} = \operatorname{rank}(\mathbf{X}') = k$, it follows that $\mathbf{x}_i$ is LIN from the rows of $\mathbf{X}'_{(i)}$. Using part (a) of Lemma 1.1,

$$\mathbf{x}'_i \left(\mathbf{X}'_{(i)}\mathbf{X}_{(i)} + \mathbf{x}_i\mathbf{x}'_i\right)^{-1} \mathbf{x}_i (= h_{ii}) \;=\; 1 \;,$$

and proof is completed. □

Lemma 2.2. *If the full column rank matrix $\mathbf{X}_{n\times(k+1)}$ contains column one, then*

(i) $\frac{1}{n} \le h_{ii} \le 1$.

(ii) $h_{ii} = \frac{1}{n}$, *if and only if* $\mathbf{x}_i = \bar{\mathbf{x}}$.

(iii) h_{ii}=1, *if and only if* $\operatorname{rank}(\mathbf{X}_{(i)}) = k$.

Proof: In this case $\mathbf{H} - \frac{1}{n}\mathbf{J}_n$ and $\mathbf{I}_n - \mathbf{H}$ are both p.s.d. matrices, so part (i) holds. To verify part (ii) note that in the with intercept model, we have:

$$(\mathbf{X}'\mathbf{X})^{-1}\,\bar{\mathbf{x}} \;=\; \frac{1}{n}\begin{bmatrix} 1 \\ \mathbf{0}_{k\times 1} \end{bmatrix}. \tag{2.2}$$

The sufficient condition is established by noting that

$$\mathbf{x}_i'\,(\mathbf{X}'\mathbf{X})^{-1}\,\bar{\mathbf{x}} \;=\; \bar{\mathbf{x}}'(\mathbf{X}'\mathbf{X})^{-1}\,\bar{\mathbf{x}} \;=\; \frac{1}{n}\;. \tag{2.3}$$

Conversely, if $h_{ii} = 1/n$, we have:

$$(\mathbf{x}_i - \bar{\mathbf{x}})'\,(\mathbf{X}'\mathbf{X})^{-1}\,(\mathbf{x}_i - \bar{\mathbf{x}}) \;=\; 0$$

which satisfies $\mathbf{x}_i = \bar{\mathbf{x}}$. Part (iii) is verified similar to part (iii) of Lemma 2.1. □

Example 2.1. Consider the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ with usual assumption. In this case,

$$h_{ii} \;=\; \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{k=1}^{n}(x_k - \bar{x})^2}\;.$$

It is clear that $x_i = \bar{x}$ satisfies $h_{ii} = 1/n$. Also, if for all $k \neq i$, we have $x_k = c\,(\neq x_i)$, then $\bar{x} = c + (x_i - c)/n$ and $h_{ii} = 1$. Figures 1 and 2 show two examples of these situations. In Figure 1, the i-th observation gives minimum possible value h_{ii}, and the fitted slope is not affected by this observation. Conversely, Figure 2 shows an example with maximum possible value for h_{ii}. In this case, the slope of fitted line is determined by y_i, and deleting such observation changes $\mathbf{X}'_{(i)}\mathbf{X}_{(i)}$ to a singular matrix.

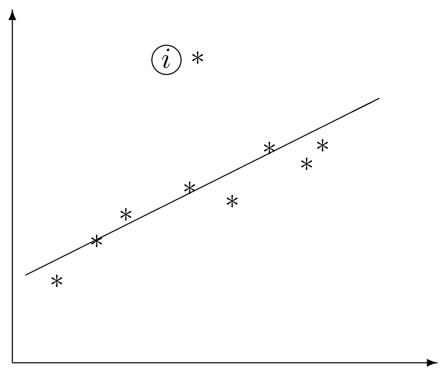

Figure 1: A simple linear regression model with intercept for which $h_{ii} = \frac{1}{n}$.

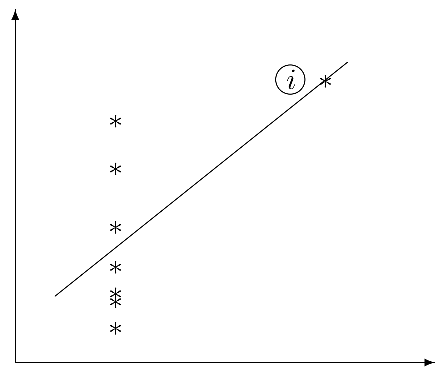

Figure 2: A simple linear regression model with intercept for which $h_{ii} = 1$.

3. BOUNDS FOR OFF-DIAGONAL ELEMENTS OF THE HAT MATRIX

In this case we assume two situations with and without intercept term in the linear regression model. Part (i) of the following lemma is shown by Chatterjee and Hadi (1988, p. 18). (They have appreciated Professor J. Brian Gray for bringing part (i) of this lemma).

Lemma 3.1. *Let* $\mathbf{X}_{n\times k}$ *be full column rank matrix without column one. Then,*

(i) $-\frac{1}{2} \le h_{ij} \le \frac{1}{2}$.

(ii) $h_{ij} = -\frac{1}{2}$, *if and only if* $\mathbf{x}_i = -\mathbf{x}_j$ *and* $\operatorname{rank}(\mathbf{X}_{(ij)}) = k-1$.

(iii) $h_{ij} = \frac{1}{2}$, *if and only if* $\mathbf{x}_i = \mathbf{x}_j$ *and* $\operatorname{rank}(\mathbf{X}_{(ij)}) = k-1$.

Proof: Since $\mathbf{H}$ is idempotent, we have:

$$h_{ii} = \sum_{i=1}^{n} h_{ik}^2 = h_{ii}^2 + h_{ij}^2 + \sum_{k\neq(i,j)} h_{ik}^2 \,, \tag{3.1}$$

which implies that $h_{ij}^2 = h_{ii}(1-h_{ii}) + \sum_{k\neq(i,j)} h_{ik}^2$. Since $0 \le h_{ii} \le 1$, part (i) is obtained by conditions $h_{ii} = h_{jj} = 1/2$ and $h_{ik} = h_{jk} = 0$ for all $k\,(\neq i,j) = 1,2,...,n$. To verify sufficient condition of part (ii), let $h_{ij} = -1/2$. From (3.1) we have $h_{ii} = h_{jj} = 1/2$, so

$$(\mathbf{x}_i+\mathbf{x}_j)'\,(\mathbf{X}'\mathbf{X})^{-1}\,(\mathbf{x}_i+\mathbf{x}_j) = 0 \,,$$

which holds only if $\mathbf{x}_i = -\mathbf{x}_j$. Again, if $\mathbf{X}' = \begin{bmatrix} \mathbf{X}'_{(ij)} & \mathbf{x}_i & \mathbf{x}_j \end{bmatrix}$, then

$$\mathbf{H} = \begin{bmatrix} \mathbf{X}_{(ij)}\,(\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'_{(ij)} & \mathbf{0}_{(n-2)\times 2} \\ \mathbf{0}_{2\times(n-2)} & \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \end{bmatrix}. \tag{3.2}$$

Since $\mathbf{H}$ is idempotent, it follows from equation (3.2) that $\mathbf{X}_{(ij)}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_{(ij)}$ is also idempotent. Hence,

$$\operatorname{rank}\big(\mathbf{X}_{(ij)}\big) = \operatorname{rank}\big(\mathbf{X}_{(ij)}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_{(ij)}\big) = \operatorname{trace}\big(\mathbf{X}_{(ij)}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_{(ij)}\big) = k-1 \,.$$

Conversely, if $\mathbf{x}_i = -\mathbf{x}_j$ and $\operatorname{rank}(\mathbf{X}_{(ij)}) = k-1$, since $\operatorname{rank}\begin{bmatrix} \mathbf{X}'_{(ij)} & \mathbf{x}_i & \mathbf{x}_j \end{bmatrix} = \operatorname{rank}(\mathbf{X}) = k$, it follows that $\mathbf{x}_i$ is LIN from the rows of $\mathbf{X}_{(ij)}$. Applying part (b) of Lemma 1.1 with replacing $\mathbf{A}$ and $\mathbf{R}$ by $\mathbf{X}_{(ij)}$ and $\begin{bmatrix} \mathbf{x}_i & -\mathbf{x}_i \end{bmatrix}$, with $\boldsymbol{\delta} = (1,-1)$ gives

$$\mathbf{R}(\mathbf{A}'\mathbf{A}+\mathbf{C}'\mathbf{C})^{-1}\mathbf{R}' = \begin{bmatrix} h_{ii} & h_{ij} \\ h_{ij} & h_{jj} \end{bmatrix} = \begin{bmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix}.$$

Part (iii) is proved similarly by multiplying $\mathbf{x}_j$ by -1. □

The following lemma gives the boundary of h_{ij} in the with intercept model. We will find its upper bound similar to the case of the no-intercept model, whereas its lower bound has sharpened by the constant $1/n$.

Lemma 3.2. *If* $\mathbf{X}_{n\times(k+1)}$ *is full column rank matrix with column one, then*

(i) $\frac{1}{n} - \frac{1}{2} \le h_{ij} \le \frac{1}{2}$.

(ii) $h_{ij} = \frac{1}{n} - \frac{1}{2}$, *if and only if* $\mathbf{x}_i + \mathbf{x}_j = 2\bar{\mathbf{x}}$ *and* $\operatorname{rank}(\mathbf{X}_{(ij)}) = k$.

(iii) $h_{ij} = \frac{1}{2}$, *if and only if* $\mathbf{x}_i = \mathbf{x}_j$ *and* $\operatorname{rank}(\mathbf{X}_{(ij)}) = k$.

Proof: In this case $\mathbf{H}$ is idempotent and has the property of a transition probability matrix, i.e. $\mathbf{H1}=\mathbf{1}$. Thus, we should minimize h_{ij} with restriction (3.1) along with

$$\sum_{i=1}^{n} h_{ik} = 1 \,. \tag{3.3}$$

Using λ as a Lagrangian multiplier, we minimize

$$h_{ij} = 1 - h_{ii} - \sum_{k\neq(i,j)} h_{ik} + \lambda \left[h_{ii}(1-h_{ii}) - h_{ij}^2 - \sum_{k\neq i,j} h_{ik}^2 \right] , \tag{3.4}$$

with respect to the λ and elements h_{ik} for $k \neq j = 1,2,...,n$. Clearly $\partial h_{ij}/\partial\lambda = 0$ gives (3.1), and $\partial h_{ij}/\partial h_{ii} = 0$ gives $h_{ii} = \frac{1}{2} - \frac{1}{2\lambda}$. On the other hand setting $\partial h_{ij}/\partial h_{ik} = 0$ results to $h_{ik} = -\frac{1}{2\lambda}$. Substituting in (3.1) gives:

$$h_{ij}^2 = \frac{1}{4}\left(1 - \frac{n-1}{\lambda^2}\right) , \tag{3.5}$$

and so (3.3) yields

$$h_{ij} = \frac{1}{2}\left(1 + \frac{n-1}{\lambda}\right) . \tag{3.6}$$

Solving equations (3.5) and (3.6) with respect to λ gives the boundary of h_{ij} as

$$\frac{1}{n} - \frac{1}{2} \leq h_{ij} \leq \frac{1}{2} \,.$$

In order to prove part (ii), note that $h_{ij} = 1/n - 1/2$ produces all h_{ik}, $(k \neq i,j)$ be equal to $1/n$, which leads to $h_{ii} = h_{jj} = 1/n + 1/2$. Hence,

$$(\mathbf{x}_i + \mathbf{x}_j - 2\bar{\mathbf{x}})' \, (\mathbf{X}'\mathbf{X})^{-1} \, (\mathbf{x}_i + \mathbf{x}_j - 2\bar{\mathbf{x}}) = 0 \,,$$

which holds only if $\mathbf{x}_i + \mathbf{x}_j = 2\bar{\mathbf{x}}$. Furthermore, we have

$$\mathbf{H} - \frac{1}{n}\mathbf{J}_n = \begin{bmatrix} \mathbf{X}_{(ij)}\,(\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'_{(ij)} - \frac{1}{n}\mathbf{J}_{n-2} & \mathbf{0}_{(n-2)\times 2} \\ \mathbf{0}_{2\times(n-2)} & \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \end{bmatrix} . \tag{3.7}$$

Since $\mathbf{H} - \frac{1}{n}\mathbf{J}_n$ is idempotent, equation (3.7) results to $\mathbf{X}_{(ij)}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_{(ij)} - \frac{1}{n}\mathbf{J}_{n-2}$ is idempotent, also. Hence,

$$\begin{aligned} k-1 &= \operatorname{trace}\left(\mathbf{X}_{(ij)}\,(\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'_{(ij)} - \frac{1}{n}\mathbf{J}_{n-2} \right) \\ &= \operatorname{rank}\left(\mathbf{X}_{(ij)}\,(\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'_{(ij)} - \frac{1}{n}\mathbf{J}_{n-2} \right) . \end{aligned}$$

We now show that the last rank of difference matrix is equal to the difference of corresponding rank of matrices. Let $\mathbf{A} = \mathbf{X}_{(ij)}\,(\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'_{(ij)}$ and $\mathbf{B} = \frac{1}{n}\mathbf{J}_{n-2}$.

Since $\mathbf{A}$ is symmetric, we have $\mathbf{A}\mathbf{A}^- = \mathbf{A}^-\mathbf{A}$, resulting $\mathbf{A}^2\mathbf{A}^- = \mathbf{A}$. Using equation (3.7) and noting that $(\mathbf{H} - \frac{1}{n}\mathbf{J})\mathbf{1} = \mathbf{0}$, we have

$$\mathbf{AB} = \mathbf{BA} = \mathbf{B}^2 = \left(\frac{n-2}{n}\right)\mathbf{B} \qquad \text{and} \qquad \mathbf{A}^2 = \mathbf{A} - \frac{2}{n}\mathbf{B}\,.$$

Therefore,

$$\left(\mathbf{A} - \frac{2}{n}\mathbf{B}\right)\mathbf{A}^- = \mathbf{A}\,. \tag{3.8}$$

Multiplying (3.8) by $\mathbf{A}$ from the left side, we find $\mathbf{ABA}^- = \mathbf{B}$. Similarly, the equality $\mathbf{A}^-\mathbf{BA} = \mathbf{B}$ is verified. It remains to show that $\mathbf{BA}^-\mathbf{B} = \mathbf{B}$. Multiplying (3.8) by $\mathbf{B}$ to the right hand side and noting that $\mathbf{A}^-\mathbf{B}$ is symmetric, we have

$$\mathbf{BA}^-\mathbf{B} = \frac{n}{2}\big(\mathbf{AA}^-\mathbf{B} - \mathbf{AB}\big) = \frac{n}{2}\left[\mathbf{B} - \left(\frac{n-2}{n}\right)\mathbf{B}\right] = \mathbf{B}\,.$$

Using Lemma 1.2, we have $\operatorname{rank}\big(\mathbf{X}_{(ij)}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'_{(ij)}\big) - \operatorname{rank}(\mathbf{J}_{n-2}) = k-1$, and thus $\operatorname{rank}\big(\mathbf{X}_{(ij)}\big) = k$.

Conversely, suppose $\mathbf{X}_{(ij)}$ of order $(n-2)\times(k+1)$ has rank k and $\mathbf{x}_i + \mathbf{x}_j = 2\bar{\mathbf{x}}$. Then $\bar{\mathbf{x}} = \bar{\mathbf{x}}_{(ij)}$, the row means of $\mathbf{X}_{(ij)}$. In this case $h_{ij} = 2/n - h_{ii}$. Now, since $\mathbf{X}_{(i)}$ is full column rank, then $\mathbf{x}_j$ is LIN from the rows of $\mathbf{X}_{(ij)}$. Using part (a) of the Lemma 1.1, we have:

$$\begin{aligned}
\mathbf{x}_i'\left(\mathbf{X}'_{(i)}\mathbf{X}_{(i)}\right)^{-1}\mathbf{x}_i &= (2\bar{\mathbf{x}} - \mathbf{x}_j)'\left(\mathbf{X}'_{(ij)}\mathbf{X}_{(ij)} + \mathbf{x}_j\mathbf{x}_j'\right)^{-1}(2\bar{\mathbf{x}} - \mathbf{x}_j) \\
&= 4\bar{\mathbf{x}}'_{(ij)}\left(\mathbf{X}'_{(ij)}\mathbf{X}_{(ij)} + \mathbf{x}_j\mathbf{x}_j'\right)^{-1}\bar{\mathbf{x}}_{(ij)} \\
&\quad - \frac{4}{n-2}\mathbf{1}'\mathbf{X}_{(ij)}\left(\mathbf{X}'_{(ij)}\mathbf{X}_{(ij)} + \mathbf{x}_j\mathbf{x}_j'\right)^{-1}\mathbf{x}_j \\
&\quad + \mathbf{x}_j'\left(\mathbf{X}'_{(ij)}\mathbf{X}_{(ij)} + \mathbf{x}_j\mathbf{x}_j'\right)^{-1}\mathbf{x}_j \\
&= 4\bar{\mathbf{x}}'_{(ij)}\left(\mathbf{X}'_{(ij)}\mathbf{X}_{(ij)} + \mathbf{x}_j\mathbf{x}_j'\right)^{-1}\bar{\mathbf{x}}_{(ij)} + 1 \\
&= 4\left(\frac{(n-1)\bar{\mathbf{x}}_{(i)} + \mathbf{x}_i}{n}\right)'\left(\mathbf{X}'_{(i)}\mathbf{X}_{(i)}\right)^{-1}\left(\frac{(n-1)\bar{\mathbf{x}}'_{(i)} + \mathbf{x}_i}{n}\right) + 1 \\
&= 4\left[\mathbf{x}_i'\left(\mathbf{X}'_{(i)}\mathbf{X}_{(i)}\right)^{-1}n + 1\right] + 1\,.
\end{aligned}$$

Hence, $\mathbf{x}_i'(\mathbf{X}'_{(i)}\mathbf{X}_{(i)})^{-1}\mathbf{x}_i = \frac{h_{ii}}{1-h_{ii}} = \frac{n+2}{n-2}$, which implies $h_{ij} = 1/n - 1/2$. Proof of part (iii) is analogous to part (iii) of Lemma 3.1. □

Example 3.1. Consider the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ with usual assumptions. In this model,

$$h_{ij} = \frac{1}{n} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{\sum_{k=1}^{n}(x_k - \bar{x})^2}\,.$$

Now if $x_i = x_j = c$ and $x_k = d \neq c$ (for every $k \neq i, j$) then $\bar{x} = d + 2(c - d)/n$. It is easy to show that $h_{ij} = 1/2$. On the other hand, if $x_i \neq x_j$, $x_k = d \neq x_i, x_j$ (for every $k \neq i, j$) and $x_i + x_j = 2\bar{x} = 2d$, then $h_{ij} = 1/n - 1/2$. Figures 3 and 4 show two examples of mentioned situations, in the case when h_{ij} gives its maximum and minimum possible values.

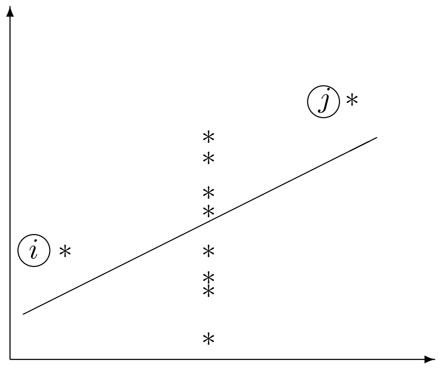

Figure 3: A simple linear regression model with intercept for which $h_{ij} = \frac{1}{n} - \frac{1}{2}$.

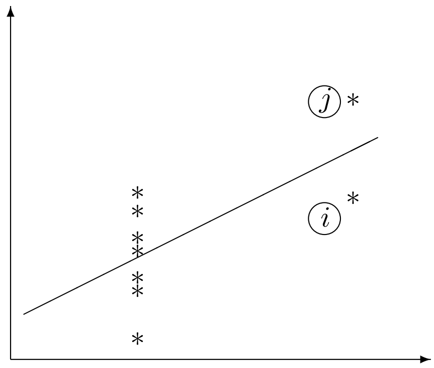

Figure 4: A simple linear regression model with intercept for which $h_{ij} = \frac{1}{n} + \frac{1}{2}$.

Example 3.2. Suppose the multiple linear regression model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i$ with design matrix $\mathbf{X}$ as

$$\mathbf{X} = \begin{bmatrix} 1 & 3 & 8 & 4 \\ 1 & 1 & 6 & 6 \\ 1 & 3 & 5 & 8 \\ 1 & 1 & 1 & 2 \\ 1 & 1 & 14 & 2 \\ 1 & 4 & 9 & 11 \\ 1 & 1 & 7 & 2 \\ 1 & 2 & 6 & 5 \end{bmatrix}.$$

Hat matrix is

$$\mathbf{H} = \begin{bmatrix} 0.625 & -0.375 & 0.125 & 0.125 & 0.125 & 0.125 & 0.125 & 0.125 \\ -0.375 & 0.625 & 0.125 & 0.125 & 0.125 & 0.125 & 0.125 & 0.125 \\ 0.125 & 0.125 & 0.298 & 0.133 & -0.161 & 0.335 & -0.003 & 0.148 \\ 0.125 & 0.125 & 0.133 & 0.618 & -0.196 & -0.235 & 0.242 & 0.188 \\ 0.125 & 0.125 & -0.161 & -0.196 & 0.791 & 0.008 & 0.296 & 0.049 \\ 0.125 & 0.125 & 0.335 & -0.235 & 0.008 & 0.658 & -0.123 & 0.106 \\ 0.125 & 0.125 & -0.003 & 0.242 & 0.260 & -0.123 & 0.250 & 0.124 \\ 0.125 & 0.125 & 0.148 & 0.188 & 0.049 & 0.106 & 0.124 & 0.126 \end{bmatrix}.$$

It is observed that $h_{12} = -0.375 = 1/n - 1/2$, and this is because of $\mathbf{x}_i + \mathbf{x}_j = 2\bar{\mathbf{x}}$ and for any $i \geq 3$: $x_{i3} = 3x_{i1} - 1$.

4. CONCLUDING REMARKS

A large number of statistical measures, such as Mahalanobis distance, weighted square standardized distance, PRESS, etc, have been proposed in the literatures of diagnosing influential observations, which are typically based on h_{ij}'s. Removing the i-th point or (i, j)-th points jointly may be useful to detect the leverage in regression diagnostics. The following outcomes are obtained from the previous lemmas in sections 2 and 3:

- $h_{ii} = 0$ (or $h_{ii} = 1/n$ in the intercept model). In this case the i-th observation potentially is an outlier, recognized by large distance between y_i and $\bar{y}$. This point has no effect on the estimation of unknown parameter $\boldsymbol{\beta}$, except constant term in the with intercept model (see Figure 1). In this situation, y_i has minimum effect to determine $\hat{y}_i$.
- $h_{ii} = 1$. Presence of such point obviates full collinearity of some columns of $\mathbf{X}$, so it is likely to be an influential observation. This point is capable to elongate the regression line itself. In other words, the fitted regression line passes through other data points to place of the i-th observation. In this case we see $e_i = 0$ (see Figure 2).
- $h_{ij} = -1/2$ (or $h_{ij} = 1/n - 1/2$ in the intercept model). This case may be declared as a competition between i-th and j-th observations. Using Lemma 3.1 and Lemma 3.2, it can be shown that if any of these points removed, then other point has the maximum value 1 of diagonal element of corresponding Hat matrix constructed based on the remaining $n - 1$ observations, so will be an influential observation. In this case, $e_i = e_j$, so $\rho(\hat{y}_i, \hat{y}_j) = -1$. This situation occurs when (i, j)-th points are at the different sides of the bulk of other points (see Figure 3).
- $h_{ij} = 1/2$. Contrary to the previous case, in this case the i-th and the j-th observations are at the same side of the bulk of other points. It can be shown that predicted values of these observations are at the same direction, i.e. $\rho(\hat{y}_i, \hat{y}_j) = 1$ (see Figure 4).

APPENDIX: Proof of Lemma 1.1

(a): Without loss of generality, let the first $p-m_1$ rows $\mathbf{A}_1$ of $\mathbf{A}$ be full row rank; then the last $(n+m_1-p)$ rows $\mathbf{A}_2$ of $\mathbf{A}$ may be written as $\mathbf{A}_2 = \mathbf{N}\mathbf{A}_1$, where $\mathbf{N}$ is $(n+m_1-p)\times(p-m_1)$. Since $\mathbf{B}$ has its rows LIN of those of $\mathbf{A}$, we may define:

$$\begin{bmatrix} \mathbf{A}_1 \\ \mathbf{B} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{D} \end{bmatrix},$$

where $\mathbf{C}_1$ and $\mathbf{D}$ are $p\times(p-m_1)$ and $p\times m_1$ matrices, respectively. Then,

$$\begin{bmatrix} \mathbf{A}_1 \\ \mathbf{B} \end{bmatrix}\begin{bmatrix} \mathbf{C}_1 & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1\mathbf{C}_1 & \mathbf{A}_1\mathbf{D} \\ \mathbf{B}\mathbf{C}_1 & \mathbf{B}\mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{p-m_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m_1} \end{bmatrix}.$$

Now define the $p\times(n+m_1-p)$ matrix $\mathbf{C}_2$ as $\mathbf{C}_2 = \mathbf{C}_1\mathbf{N}'$. So, we have

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{p-m_1} \\ \mathbf{N} \end{bmatrix}\mathbf{A}_1 ,$$

and

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 \end{bmatrix} = \mathbf{C}_1\begin{bmatrix} \mathbf{I}_{p-m_1} & \mathbf{N}' \end{bmatrix}.$$

From the solutions we obtain

$$\mathbf{A}\mathbf{C} = \begin{bmatrix} \mathbf{I}_{p-m_1} & \mathbf{N}' \\ \mathbf{N} & \mathbf{N}\mathbf{N}' \end{bmatrix}, \qquad \mathbf{A}\mathbf{D} = \mathbf{0}_{n\times m_1} = (\mathbf{B}\mathbf{C})', \qquad \mathbf{B}\mathbf{D} = \mathbf{I}_{m_1} ,$$

where $\mathrm{rank}(\mathbf{A}\mathbf{C}) = \mathrm{rank}(\mathbf{A}) = p-m_1$. Now since $\begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix}$ has rank p, $\mathbf{A}'\mathbf{A}+\mathbf{B}'\mathbf{B}$ is positive definite and therefore invertible. From above we have expressions

$$(\mathbf{A}'\mathbf{A}+\mathbf{B}'\mathbf{B})\,\mathbf{D} = \mathbf{B}' .$$

Premultiplying by $(\mathbf{A}(\mathbf{A}'\mathbf{A}+\mathbf{B}'\mathbf{B})^{-1})$, and then by $(\mathbf{B}(\mathbf{A}'\mathbf{A}+\mathbf{B}'\mathbf{B})^{-1})$ we obtain

$$\mathbf{A}\,(\mathbf{A}'\mathbf{A}+\mathbf{B}'\mathbf{B})^{-1}\,\mathbf{B}' = \mathbf{A}\mathbf{D} = \mathbf{0}_{n\times m_1}$$

and

$$\mathbf{B}\,(\mathbf{A}'\mathbf{A}+\mathbf{B}'\mathbf{B})^{-1}\,\mathbf{B}' = \mathbf{B}\mathbf{D} = \mathbf{I}_{m_1} .$$

(b): Equalizing $\mathbf{R} = \boldsymbol{\delta}\mathbf{r}'$ results to:

$$\mathbf{R}\,(\mathbf{A}'\mathbf{A}+\mathbf{R}'\mathbf{R})^{-1}\,\mathbf{R}' = \boldsymbol{\delta}\mathbf{r}'\,(\mathbf{A}'\mathbf{A}+\mathbf{r}\boldsymbol{\delta}\boldsymbol{\delta}'\mathbf{r}')^{-1}\,\mathbf{r}\boldsymbol{\delta}' .$$

Now using part (a) and substituting $\mathbf{B}$ by $\sqrt{\boldsymbol{\delta}\boldsymbol{\delta}'}\mathbf{r}'$ give

$$\mathbf{R}\,(\mathbf{A}'\mathbf{A}+\mathbf{R}'\mathbf{R})^{-1}\,\mathbf{R}' = \frac{\boldsymbol{\delta}\boldsymbol{\delta}'}{\|\boldsymbol{\delta}\|^2} .$$

REFERENCES

[1] Chatterjee, S. and Hadi, A.S. (1988). *Sensitivity Analysis in Linear Regression*, New York: John Wiley & Sons.

[2] Chipman, J.S. (1964). On least squares with insufficient observations, *Journal of the American Statistical Association*, **59**, 1078–1111.

[3] Cook, R.D. and Weisberg, S. (1982). *Residual and Influence in Regression*, London: Chapman & Hall.

[4] Hadi, A.S. (1990). Two graphical displays for the detection of potentially influential subsets in regression, *Journal of Applied Statistics*, **17**, 313–327.

[5] Hoaglin, D.C. and Welsch, R.E. (1978). The Hat matrix in regression and ANOVA *The American Statistician*, **32**, 17–22.

[6] Huber, P. (1975). *Robustness and designs.* In "A Survey of Statistical Design and Linear Models" (Jagdish N. Srivastava, Ed.), Amsterdam, North-Holland Publishing Co.

[7] Seber, S.R. (2007). *Matrix handbook for Statisticians*, New York, John Wiley & Sons.