Journal of the KSQC
Vol. 19, No. 1, June. 1991

1991년 6월 韓國品質管理學會誌 제19권 제1호

A Study on Sensitivity Analysis in Ridge Regression
능형 회귀에서의 민감도 분석에 관한 연구

Soon-Kwi Kim*

ABSTRACT

In this paper, we discuss and review various measures which have been presented for studying outliers, high-leverage points, and influential observations when ridge regression estimation is adopted. We derive the influence function for $\hat{\beta}_R$, the ridge regression estimator, and discuss its various finite sample approximations when ridge regression is postulated. We also study several diagnostic measures such as Welsh-Kuh's distance, Cook's distance etc.

1. Introduction

Very little is known about the effect that collinearity can have on the influence of any given case. Belsley et al.(1980, p.210) noted that the influence of the majority of the cases in biased estimation was lower than their corresponding influence in the methed of least squares(LSM). The influence of certain cases, however, actually increases when collinearity is controlled.

* Dept. of Statistics, Kangnung University.

Belsley et al.(1980) pointed out, "we provisionally conclude that reduction in collinearity should be a first step for the effective detection of unusual data components."

A large number of statistical quantities have been proposed to study outliers and influence of individual observations in regression analysis. The basic idea behind the study of influential cases is to monitor the changes in a selected phase of the analysis as individual cases or groups of cases are in turn removed from the data. Case deletion will be used throughout this paper. A more revealing development can be based on various empirical versions of the influence curve as in Cook and Weisberg(1982). Chatterjee and Hadi(1986) gave an excellent review on the study of the sensitivity of the regression results.

On the other hand, few statistical quantities have been proposed to study outliers and influence of individual observations when ridge regression (RR) is used to mitigate the effects of collinearity. The roles of leverage, and influence in RR were discussed by Walker and Birch (1988). They also proposed approximate deletion formulas for the detection of influential observation for RR.

In this paper, when RR is used, the corresponding influence measures are discussed. In Section 2, the influence curve for $\hat{\beta}_R$, the ridge estimator, and several approximations of the influence curve are discussed. Also, several measures based on the influence curve are derived and their properties are discussed when RR is used. In Section 4, an example is given to illustrate the contents discussed in Section 2 and 3. In Section 5, summary and concluding remarks are provided.

2. Influence Function Approach

An important class of measures for the influence of the ith observation on the regression results is based on the idea of the influence curve (IC) or influence function (IF). (Hampel, 1974.) In case IF is a vector, it must be normalized so that observations can be orderd in a meaningful way. Thus one may usem

$$D_i(M, c) = \frac{(IF_i)'M(IF_i)}{c}$$

to assess the influence of the ith observation on the regression coefficients relative to M and c. A large value of $D_i(M, c)$ indicates that the ith observation has strong influence on $\hat{\beta}_R$ relative to M and c.

In this section, we will derive the IF for the ridge estimator of $\hat{\beta}_R$, and provide finite sample approximations for the IF.

2.1 Influence Function for $\hat{\beta}_R$

The influence function can be used in the assessment of the influence of observations on estimators. The estimator of interest here is the ridge estimator of $\underline{\beta}$. This is obtained by solving the system of p linear equations

$$\frac{1}{n}\left[\sum_{i}^{n}\underline{x}_i(\underline{y}_i-\underline{x}_i'\underline{\beta})-kI_P\underline{\beta}\right]=0, \quad \cdots\cdots(2.1)$$

where $\underline{x}_i$ and y_i are the ith row of X and the ith element of $\underline{y}$, respectively. Assuming that the (p+1) vector $(\underline{x}', y)$ has a joint cumulative distribution function(cdf) F, we can write (2.1) as

$$E_F[\underline{x}\underline{x}'+k/nI_P]\hat{\underline{\beta}}_R(F)=E_F(\underline{x}y).$$

Now, suppose that

$$E_F\left[\begin{pmatrix}\underline{x}\\ y\end{pmatrix}^{(\underline{x}'\,y)}\right]=\begin{bmatrix}\Sigma_{xx}(F) & \Sigma_{xy}(F)\\ \Sigma'_{xy}(F) & \sigma_{yy}(F)\end{bmatrix},$$

then the functional for $\underline{\beta}$ is

$$\hat{\underline{\beta}}_R(F)=[\Sigma_{xx}(F)+k/n\ I_P]^{-1}\ \Sigma_{xy}(F). \quad \cdots\cdots(2.2)$$

Thus, the IC may be obtained by substituting $(\underline{x}', y)$ for $\underline{z}$ and replacing T by the functional $\underline{\beta}$ (F). In this case, the n vectors of influence curves are defined, for i=1, 2, ⋯, n, by

$$I(\underline{x}', y, F, T)=\lim_{\varepsilon\to 0}\frac{T\{(1-\varepsilon)F+\varepsilon\delta_{\underline{x}',y}\}-T\{F\}}{\varepsilon}. \quad \cdots\cdots(2.3)$$

Theorem 2.1 provides explicit forms of the IC for $\hat{\underline{\beta}}_R$ given by (2.3), but first we need the following result.

$$(I+\varepsilon A)^{-1}=I+\sum_{1}^{\infty}(-1)^i\varepsilon^iA^i, \quad \cdots\cdots(2.4)$$

where A is any matrix such that $(I+\varepsilon A)^{-1}$ exists.

Theorem 2.1

The influence curve for $\hat{\beta}_R$ is

$$\hat{\beta}_R^{(1)}(F)=[\Sigma_{xx}(F)+k/n\ I_P]^{-1}[\underline{x}y-\Sigma_{xy}(F)-\underline{xx}'\hat{\beta}_R(F)]$$
$$+\Sigma_{xx}(F)\hat{\beta}_R(F)]. \quad \cdots\cdots(2.5)$$

Proof The substitution of (2.2) into (2.3) yields the IC for $\hat{\beta}_R$, i.e.,

$$\hat{\beta}_R^{(1)}(F)=\lim_{\varepsilon\to 0}\frac{\hat{\beta}\{(1-\varepsilon)F+\varepsilon\delta_{\underline{x}',y}\}-\hat{\beta}(F)}{\varepsilon}. \quad \cdots\cdots(2.6)$$

The first term in the numerator of this expression is

$$\begin{aligned}\hat{\beta}\{(1-\varepsilon)F+\varepsilon\delta_{\underline{x}',y}\}&=\{\Sigma_{xx}(F\varepsilon)+k/n\ I_P]^{-1}\ \Sigma_{xy}(F\varepsilon)\\ &=[\Sigma_{xx}\{(1-\varepsilon)F+\varepsilon\delta_{\underline{x}',y}\}+k/n\ I_P]^{-1}\ \Sigma_{xy}\{(1-\varepsilon)F+\varepsilon\delta_{\underline{x}',y}\}\\ &=\{(1-\varepsilon)\Sigma_{xx}(F)+\varepsilon\underline{xx}'+k/n\ I_P\}^{-1}\{(1-\varepsilon)\Sigma_{xy}(F)+\varepsilon\underline{x}y\}\\ &=\{\Sigma_{xx}(F)+k/nI_P+\varepsilon[\underline{xx}'-\Sigma_{xx}(F)]\}^{-1}\{\Sigma_{xy}(F)\}+\varepsilon[\underline{x}y-\Sigma_{xy}(F)]\}\\ &=\{I_P+\varepsilon[\Sigma_{xx}(F)+k/nI_P]^{-1}[\underline{xx}'-\Sigma_{xx}(F)]\}^{-1}[\Sigma_{xx}(F)+k/nI_P]^{-1}\\ &\quad\cdot\{\Sigma_{xy}(F)+\varepsilon[\underline{x}y-\Sigma_{xy}(F)]\}.\end{aligned}$$

Using (2.4), we obtain

$$\hat{\beta}\{(1-\varepsilon)F+\varepsilon\delta_{\underline{x}',y}\}=\{I-\varepsilon[\Sigma_{xx}(F)+k/nI_P]^{-1}[\underline{xx}'-\Sigma_{xx}(F)]+o(\varepsilon^2)\}$$
$$\cdot\{\hat{\beta}_R(F)+\varepsilon[\Sigma_{xx}(F)+k/n\ I_P]^{-1}[\underline{x}y-\Sigma_{xy}(F)]\}. \quad \cdots\cdots(2.7)$$

Upon substituting (2.7) into (2.6) and taking the limit, (2.5) follows. ■

2.2 Approximating the Influence Curve

The IF in (2.5) measures the influence on $\hat{\beta}_R$ when one observation $(\underline{x}', y)$ is added to a very large sample. In practice, we do not always have a very large sample. Several Finite sample versions of the IC that depend on an observed sample have been suggested. We give here four of the most common ones. These are

(1) the empirical influence curve based on n observations,

(2) the sample influence curve,

(3) the sensitivity curve, and

(4) the empirical influence curve based on (n−1) observations.

(1) The Empirical Influence Curve Based on n Observations

The empirical influence curve(EIC) based on n observations is found from (2.5) by approximationg F by the empirical cdf $\hat{F}$ and substituting $(\underline{x}_i', y_i)$, $n^{-1}(X'X)$, and $\hat{\beta}_R$ for $(\underline{x}', y)$, $\Sigma_{xx}(F)$, and $\hat{\beta}_R(F)$, respectively, and obtaining

$$\begin{aligned} EIC_i &= n(X'X+kI_P)^{-1}[\underline{x}_i e_i^* + 1/n(X'X\hat{\beta}_R - X'\underline{y})] \\ &= n(X'X+kI_P)^{-1}[\underline{x}_i e_i^* - k/n\,\hat{\beta}_R],\ i=1, 2, \cdots, n \end{aligned} \qquad \cdots(2.8)$$

where $e_i^* = y_i - \underline{x}_i'\hat{\beta}_R$.

(2) The Sample Influence Curve

The sample influence curve (SIC) is obtained from (2.3) by taking $(\underline{x}', y) = (\underline{x}_i', y_i)$, $F=\hat{F}$ and $\varepsilon = -(n-1)^{-1}$ and omitting the limit. Thus, we have,

$$\begin{aligned} SIC_i &= -(n-1)\{T(\frac{n}{n-1}\hat{F} + \frac{-1}{n-1}\delta_{\underline{x}_i', y_i}) - T(\hat{F})\} \\ &= (n-1)\{T(\hat{F}) - T(\hat{F}_{(i)})\}, \end{aligned}$$

where $\hat{F}_{(i)}$ is the empirical cdf when the ith observation is omitted. Setting $T(\hat{F}) = \hat{\beta}_R$ and $T(\hat{F}_{(i)}) = \hat{\beta}_{R(i)}$, we have

$$SIC_i = (n-1)\{\hat{\beta}_R - \hat{\beta}_{R(i)}\} \qquad \cdots(2.9)$$

where $\hat{\beta}_{R(i)} = (X_{(i)}'X_{(i)} + kI_P)^{-1}X_{(i)}'\underline{y}_{(i)}$, is the ridge estimate of $\underline{\beta}$ when the ith observation is omitted. We assume that $X_{(i)}$ and $\underline{y}_{(i)}$ are centered and scaled so that $X_{(i)}'X_{(i)}$ and $X_{(i)}'\underline{y}_{(i)}$ are in correlation form. If after deleting the ith row, however, $X_{(i)}$ is not recentered and rescaled, then $X'X - \underline{x}_i\underline{x}'$ will not be in exact correlation form. Using the approximate deletion formula for RR(see Walker et al. 1988, p.225), namely,

$$\hat{\beta}_{R(i)} \cong \hat{\beta}_F - \frac{(X'X+kI)^{-1}\underline{x}_i e_i^*}{1-h_i^*} \qquad \cdots(2.10)$$

where h_i^* is the ith diagonal element of the matrix $H^* = X(X'X+kI_P)^{-1}X'$ which plays the same role as the hat matrix in least squares, we obtain

$$SIC_i \cong (n-1)(X'X+kI_P)^{-1}\underline{x}_i\frac{e_i^*}{1-h_i^*},\ i=1, 2, \cdots, n. \qquad \cdots(2.11)$$

(3) The Sensitivity Curve

The sensitivity curve (SC) is obtained in a similar way, but here we set $F=\hat{F}_{(i)}$ and $\varepsilon=n^{-1}$, and omit the limit in (2.3). This gives

$$SC_i=n\{T(\frac{n-1}{n}\hat{F}_{(i)}+\frac{1}{n}\delta_{\underline{x}_i', y_i})-T(\hat{F}_{(i)})\}$$
$$=n\{T(\hat{F})-T(\hat{F})_{(i)}\}$$

Taking $T(\hat{F})=\hat{\beta}_R$ and $T(\hat{F}_{(i)})=\hat{\beta}_{R(i)}$ and using (2.10) again, we obtain

$$SC_i=n(\hat{\beta}_R-\hat{\beta}_{R(i)})$$
$$\cong \frac{n(X'X+kI)^{-1}\underline{x}_i e_i^*}{1-h_i^*},\ i=1, 2, \cdots, n \quad \cdots\cdots (2.12)$$

(4) The Empirical Influence Curve Based on $(n-1)$ Observations

The fourth approximation of the influence curve for $\hat{\beta}_R$ is obtained from (2.5) by taking $F=\hat{F}_{(i)}$, $(\underline{x}', y)=(\underline{x}_i', y_i)$, $\Sigma_{xx}(F)=(n-1)^{-1}(X_{(i)}'X_{(i)})$, $\Sigma_{xy}(F)=(n-1)^{-1}X_{(i)}'y_{(i)}$ and $\hat{\beta}_R(F)=\hat{\beta}_{R(i)}$. This yields the EIC based on $(n-1)$ observations, i. e.,

$$EIC_{(i)}=(n-1)(X_{(i)}'X_{(i)}+kI_P)^{-1}[\underline{x}_i(y_i-\underline{x}_i'\hat{\beta}_{R(i)})]$$
$$+(X_{(i)}'X_{(i)}\hat{\beta}_{R(i)}-X_{(i)}'\underline{y}_{(i)})/(n-1) \quad \cdots\cdots (2.13)$$

Using (2.10), the quantity $(y_i-\underline{x}_i'\hat{\beta}_{R(i)})$ can be written as

$$y_i-\underline{x}_i'\hat{\beta}_{R(i)}\cong \frac{e_i^*}{1-h_i^*}. \quad \cdots\cdots (2.14)$$

Using the Sherman-morrison-Woodbury theorem (Chatterjee et al. 1988, p.21), given by

$$(A+BDC')^{-1}=A^{-1}-A^{-1}B(D^{-1}+C'A^{-1}B)^{-1}C'A^{-1}$$

where A and D are nonsingular matrices of orders k and m, respectively, B is k x m and C is k x m and substituting (2.14) into (2.13), we get

$$EIC_{(i)}\cong(n-1)(X'X+kI_P)^{-1}\underline{x}_i\frac{e_i^*}{(1-h_i^*)^2}-k[X_{(i)}'X_{(i)}+kI_P]^{-1}\hat{\beta}_{R(i)}$$
$$=(n-1)(X'X+kI_P)^{-1}\{\underline{x}_i\frac{e_i^*}{(1-h_i^*)^2}-\frac{k}{n-1}[I-\frac{\underline{x}_i\underline{x}_i'(X'X+kI_P)^{-1}}{1-h_i^*}]$$
$$\cdot[\hat{\beta}_R-\frac{(X'X+kI_P)^{-1}\underline{x}_i e_i^*}{1-h_i^*}]\},\ i=1, 2, \cdots, n \quad \cdots\cdots (2.15)$$

In comparing the four approximations of the IC for $\hat{\beta}_R$, namely EIC_i, SIC_i, and $EIC_{(i)}$, we see that the main differences among them are in the power of $(1-h_i^*)$ and in the value of $\hat{\beta}_R$ or $\hat{\beta}_{R(i)}$. Both SIC_i and SC_i are proportional to the distance between $\hat{\beta}_R$ and $\hat{\beta}_{R(i)}$. In the next section, we shall examine $D_i(M, c)$ for several suggested choices of M and c.

3. Influence in Ridge Regression

3.1 Leverage in Ridge Regression

The vector of fitted values is

$$\hat{\underline{y}}^* = X\hat{\underline{\beta}}_R = X(X'X+kI)^{-1}X'\underline{y} \qquad (3.1)$$

Therefore, the matrix $H^* = X = (X'X+kI)^{-1}X'$ plays the same role as the hat matrix in LSM. It is important to note, however, that the matrix H^* is not a projection matrix because it is not idempotent.

Using the singular value decompostition(Jolliffe, 1986, p.37), Walker and Brich(1988) showed that the ridge leverage of the ith point and the i-jth element of H^* can be written as

$$h_i^* = \sum_{h=1}^{p} \frac{\lambda_h}{\lambda_h + k} u_{ih}^2 \qquad (3.2)$$

and

$$h_{ij}^* = \sum_{h}^{p} \frac{\lambda_{ih}}{\lambda_{ih} + k} u_{ih} u_{jh} \qquad (3.3)$$

repectively, where λ_j is the jth largest eigenvalue of $X'X$ and the columms of $U=(u_{ij})$ are the p eigenvectors of $X'X$ associated with its p nonzero eigenvalues. For $k>0$, $h_i^* < h_{ii}$(see Walker & Birch(1988)), but h_{ij}^* can be shown either larger or smaller than h_{ij} for $i=1, 2, \cdots, n$ and $j=1, 2, \cdots, n$.

3.2 Measure Based on the Influence Curve

A version of Welsh-Kuh's distance(WK_i) for RR, proposed by Walker and Birch(1988), is

$$WK_{(i)} = \underline{x}_i'(\hat{\beta}_R - \hat{\beta}_{R(i)})/se(\underline{x}_i'\hat{\beta}_R),$$

where $se(\underline{x}_i', \hat{\beta}_R)$ is an estimator of the standard error(SE) of the fitted value. If the LS estimator of σ(s or $s_{(i)}$) is used, then $WK_{(i)}$ can be written as

$$WK_{(i)} = \frac{x_i'(\hat{\beta}_R - \hat{\beta}_{R(i)})}{s_{(i)}[\sum_j (h_{ij}^*)^2]^{1/2}}.$$

Usint the approximate deletion formula in (2. 10), $WK_{(i)}$ can be approximately represented as

$$WK_{(i)} \cong \frac{h_i^* e_i^*}{s_{(i)}(1-h_i^*)[\sum_j (h_{ij}^*)^2]^{1/2}} \qquad \cdots\cdots (3.4)$$

$WK_{(i)}$ is related to the IC for $\hat{\beta}_R$ given by (2. 6). For example, if we use the norm $D_i(M, C)$ and the SIC_i in (2. 9) to approximate (2. 6), then $WK_{(i)}$ is approximately written as

$$WK_{(i)} \cong \sqrt{D_i(X'X+kI,\ (n-1)^2 s^2_{(i)} \sum_j (h_{ij}^*)^2 / h_i^*)}.$$

Whereas, if the SC_i given by (2. 12) is used instead of (2. 9), then $WK_{(i)}$ is approximately expressible as

$$WK_{(i)} \cong \sqrt{D_i(X'X+kI,\ n^2 s^2_{(i)} \sum_j (h_{ij}^*)^2 / h_i^*)}.$$

Welsh's Distance

Using the SIC_i in (2. 9), as an approximation to the IC for $\hat{\beta}_R$ in (2. 5) and setting

$$M = X_{(i)}'X + kI$$

and

$$c = (n-1)^2 s^2_{(i)} \sum_j (h_{ij}^*)^2 / h_i^*,$$

$D_i(M, C)$ becomes approximately

$$W^2_{(i)} = D_i(X_{(i)}'X_{(i)} + kI,\ (n-1)^2 s^2(i) \sum (h_{ij}^*)^2 / h_i^*)$$
$$\cong \left[\frac{h_i^* e_i^*}{s_{(i)}\sqrt{(1-h^*)\sum_j (h_{ij}^*)^2}}\right]^2 \qquad \cdots\cdots (3.5a)$$

Comparing (3. 4) and (3. 5a) gives

$$W_{(i)} \cong WK_{(i)}\sqrt{1-h_i^*} \quad \cdots\cdots(3.5b)$$

We suggest using $W_{(i)}$ as a diagnostic tool. It is clear from (3. 5b), however, that $W_{(i)}$ approximately gives less emphasis to high-leverage points.

Cook's Distance

Two versions of Cook's distance, proposed by Walker and Birch(1988), were constructed for RR, namely,

$$D_i^* = (1/(ps^2))(\hat{\beta}_R - \hat{\beta}_{R(i)})'X'X(\hat{\beta}_R - \hat{\beta}_{R(i)}) \quad \cdots\cdots(3.6a)$$

and

$$D_i^{**} = (1/(ps^2))(\hat{\beta}_R - \hat{\beta}_{R(i)})'(X'X+kI)(X'X)^{-1}(X'X+kI)(\hat{\beta}_R - \hat{\beta}_{R(i)}). \quad \cdots\cdots(3.6b)$$

D_i^{**} is based on the fact that $Var(\hat{\beta}_R) = \sigma^2(X'X+kI)^{-1}X'X(X'X+kI)$.

However, if $\underline{v}$ is a px1 vector, then we note that

$$\sup_{\underline{v}} \frac{\underline{v}'(\hat{\beta}_R - \hat{\beta}_{R(i)})}{(\underline{v}'(X'X+kI)^{-1}X'X(X'X+kI)^{-1}\underline{v})^{1/2}}$$
$$= [(\hat{\beta}_R - \hat{\beta}_{R(i)})'(X'X+kI)(X'X)^{-1}(X'X+kI)(\hat{\beta}_R - \hat{\beta}_{R(i)})]^{1/2}$$
$$= (ps^2 D_i^{**})^{1/2}$$

Thus, if D_i^{**} does not declare the ith observation to be influential then the ith observation seems not influential on the prediction at any point $\underline{x}_i$, $i=1, 2, \cdots, n$ when $WK_{(i)}$ is used as a diagnostic measure.

Modified Cook's Distance

Using the SIC_i in (2. 9) and setting

$$M = X'X + kI,$$
$$c = \frac{p(n-1)^2\sum_j(h_{ij}^*)^2}{(n-p)h_i^*} s^2_{(i)}$$

and taking the square root of $D_i(M, C)$, modified Cook's distance becomes approximately,

$$MC_{(i)} = D_i\left[\left(X'X + kI, \ \frac{p(n-1)^2}{n-p} \ \frac{\sum_j (h_{ij}^*)^2 s^2_{(i)}}{h_i^*}\right)\right]^{1/2} \quad \cdots\cdots (3.7a)$$

$$\cong WK_{(i)}\left(\frac{n-p}{p}.\right)^{1/2} \quad \cdots\cdots (3.7b)$$

The various measures for the influence of the ith observation on the regression coefficients, can be obtained by using different approximations of the IC for $\hat{\beta}_R$ in combination with varying choices for M and c.

D_i^* uses $X'X$ and s^2, $WK_{(i)}$ and $MC_{(i)}$ use $X'X+kI$ and $s^2_{(i)}$ and $W_{(i)}$ uses $(X_{(i)}'X_{(i)}+kI)$ and $s^2_{(i)}$. Thus D_i^* measures the influence of the ith observation on $\hat{\beta}_R$ only, whereas $WK_{(i)}$, $MC_{(i)}$ and $W_{(i)}$ measure the influence on both $\hat{\beta}_R$ and s^2.

4. Example of Ridge Regression

The data in this example were taken from Hill(1977). Let's suppose that X is an nxp centered and standardized matrix of known constants and the value of k, shrinkage parameter, for this data set was chosen to be 0.03 as used in Walker & Birch(1988). Then, the ridge estimator of β can be written as

$$\hat{\beta}_R = (Z'Z + 0.03I^*)^{-1} Z'\underline{y},$$

where $Z=(1:X)$ is an n x (p+1) matrix and $I^*=\text{diag}\ (0, 1, \cdots, 1)$ of dimension p+1.

Table 4. 1 shows $e_i^*/(s\sqrt{1-h_i^*})$, $r_{iR}^*=e_i^*/(s_{(i)}\sqrt{1-h_i^*})$, and h_i^*. The scatter plot of r_{iR} versus $\hat{y}_i^*$(Figure 4. 1) and the normal probability plot(Figure 4. 2) do not also show any gross violation of the model assumptions. Observation #8, however, has a moderately large residual ($r_{iR}=2.01$). Only two observations (#1, #2) have $h_i^*>0.553$, and hence, they can be declared to be high-leverage points.

Figure 4. 3 shows the boxplots for r_{iR} and h_i^*. The boxplot for h_i^* shows that observations #1 and #2 are separated from the bulk of other observations.

The L−R plot, defined as the scatter plot of h_i^* versus $(a_i^*)^2=e_i^*/(\underline{e}^*)'\underline{e}^*$, for the Hill's data is shown in Figure 4. 4. Three observations are separated from the bulk of other ploints. We find the high-leverage points(#1, #2) in the upper-left corner and the outlier(#13) in the lower-right corner.

Figure 4.1 Scatter Plot of r_{iR} versus $\hat{y}_i$

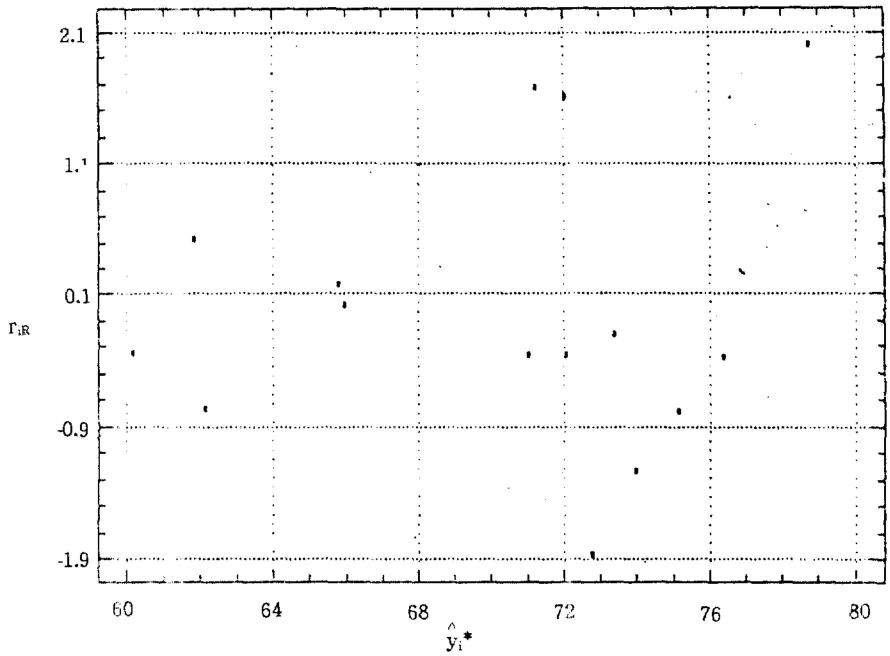


Figure 4.2 Normal Probability Plot

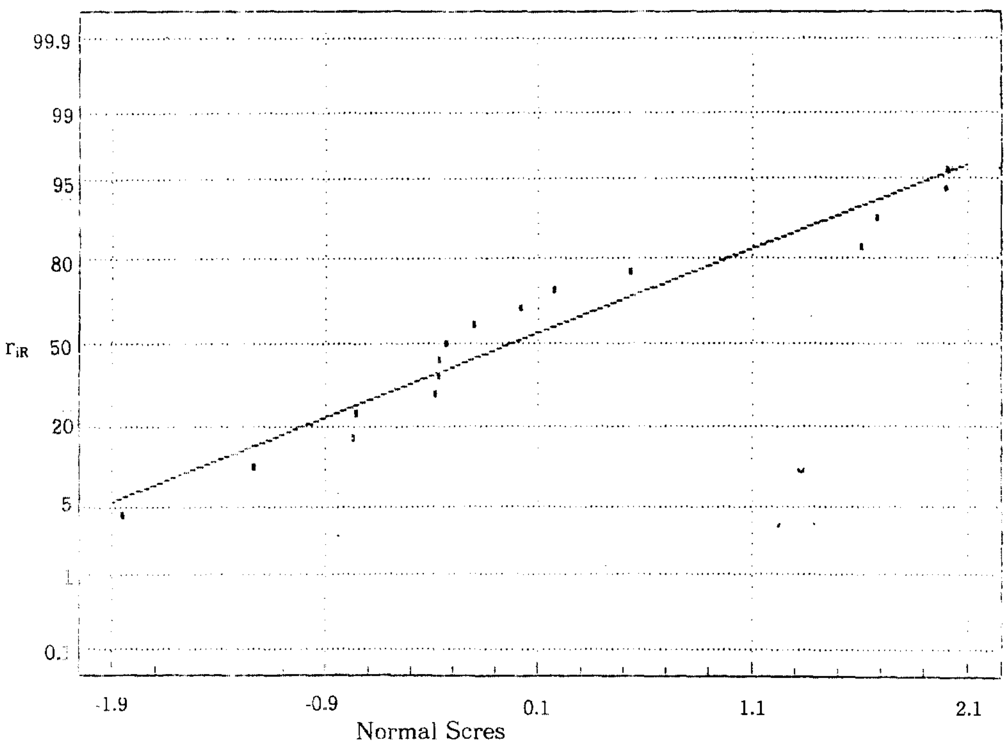


Table 4.1 e_i^*, r_{iR}, r_{iR}^*, and h_{ii}^*

Row	e_i^*	r_{iR}	r_{iR}^*	h_i^*
1	−0.9560	−0.7601	−0.8604	0.7571
2	0.4625	0.5311	0.6613	0.7476
3	−0.7791	−0.3404	−0.3195	0.4170
4	0.4126	0.1726	0.1614	0.3545
5	0.0390	0.0147	0.0138	0.2441
6	−0.6779	−0.3680	−0.3995	0.4609
7	−0.9918	−0.3843	−0.3752	0.2652
8	4.4639	2.0085	2.3077	0.4440
9	−1.9407	−0.7750	−0.7964	0.2918
10	−2.8740	−1.2367	−1.2817	0.4219
11	−0.5552	−0.2043	−0.1912	0.2226
12	3.8747	1.6039	1.5633	0.4330
13	4.7863	1.6816	1.8824	0.1544
14	−0.8279	−0.3656	−0.3436	0.4402
15	−4.1363	−1.8504	−1.9318	0.3903

Next, we examine the influence measures based on the IC. These are also in Table 4.2. The corresponding boxplots (Figure 4.5) show that observation #8 is the most influential on $\hat{\beta}_{E}$. Examination of residuals has not pointed out any peculiarities regarding observation #8.

Examination of the data for the presence of outliers, high leverage points, or influential observations has brought four observation to our attention, each of which has different characteristic. The L−R plot(Figure 4.4) explains the differences among these four observations. Observation #8 is an example of an influential observation that is not a high-leverage point but can be an outlier. Observation #1 or #2 is an example of a high-leverage point nor influential. Note again that examination of residuals is not sufficient for the detection of influential observations.

Figure 4.3 Boxplots of r_{iR} and h_i^*

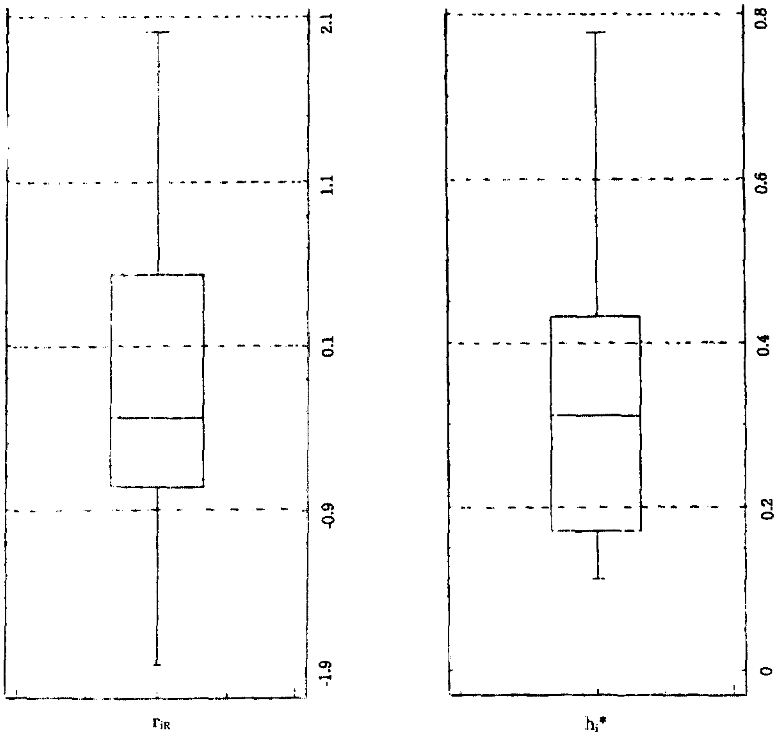


Figure 4.4 L−r Plot

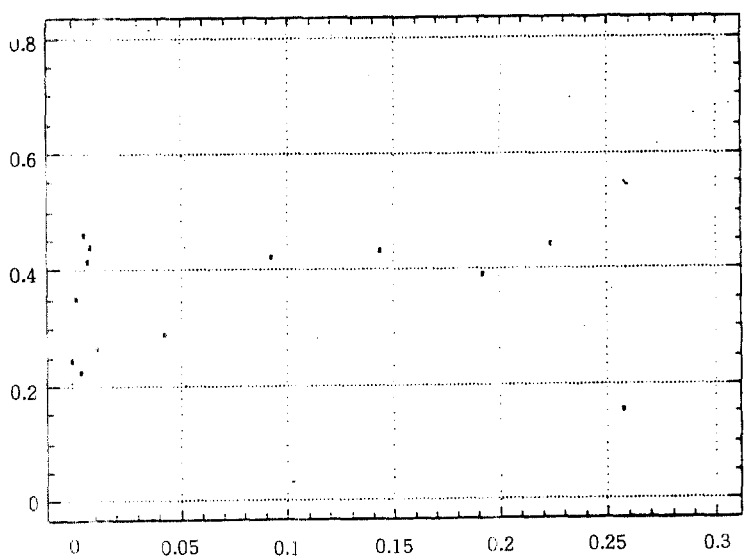
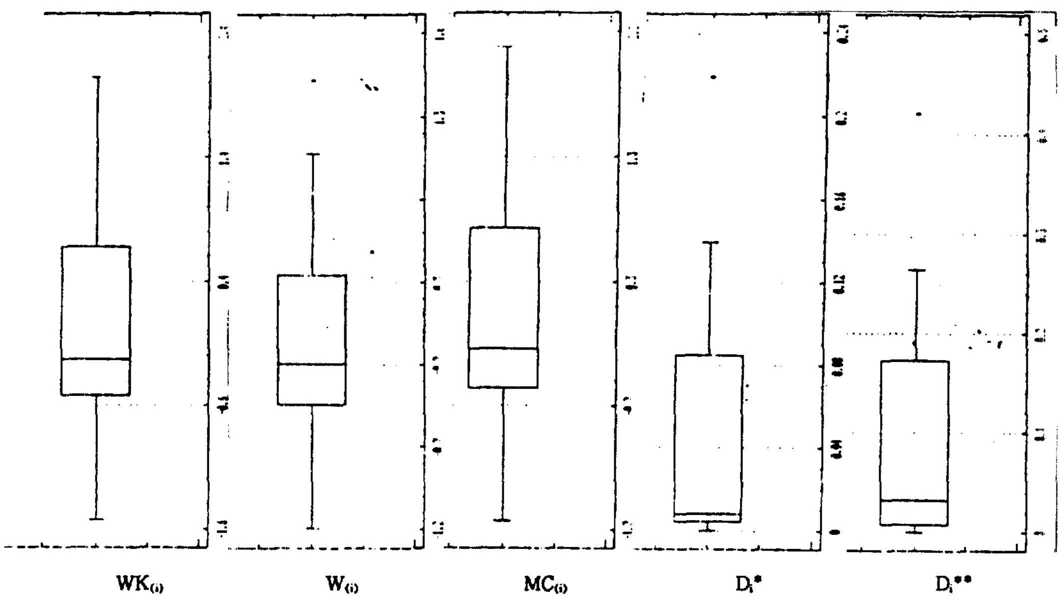


Table 4. 2 Influence Measures

Row	$WK_{(i)}$	$W_{(i)}$	$MC_{(i)}$	D_i^*	D_i^{**}
1	−1.2893	−0.6355	−1.3784	0.0391	0.1726
2	0.6833	0.3433	0.7305	0.0080	0.0370
3	−0.2704	−0.2065	−0.2891	0.0059	0.0101
4	0.1210	0.9722	0.1294	0.0012	0.0021
5	0.0080	0.0069	0.0085	0.00007	0.00001
6	−0.3009	−0.2210	−0.3217	0.0043	0.0107
7	−0.2299	−0.1917	−0.2458	0.0047	0.0071
8	0.0416	1.5223	2.1825	0.2189	0.4214
9	−0.5242	−0.4411	−0.5604	0.0198	0.3223
10	−1.0995	−0.8360	−1.1755	0.0848	0.1536
11	1.4371	1.0821	1.5363	0.1311	0.2534
12	0.8181	0.7223	0.8745	0.0588	0.0728
13	−0.3048	−0.2280	−0.3258	0.0075	0.0142
14	−1.5270	−1.1924	−1.6324	0.1295	0.2645

Gigure 4.5 Boxplots fo $WK_{(I)}$ $W_{(I)}$ $MC_{(I)}$ D_I^* and D_I



5. Some concluding Remarks

When ridge regression estimation is adopted to elliminate multicollinearities among independent variables, we have discussed and reviewed various measures which have been presented for studying outliers, high-leverage points, and influential observations.

In Section 2, 3 we suggested and reviewed some diagnostic measures when ridge regression (RR) was used. Even with a vast number of regression diagnostics, it is not easy to write down rules that can be wisely used to guide a regression data analysis. It is clear that some individual data points may be identified as outliers, high-leverage points, or influential points. Any point falling into one of these categories should be scrutinized for accuracy, relevancy, or special significance. Outliers should be always carefully examined. Points with high-leverage that are not influential should be carefully looked.

In Section 4, a numerical example was illustrated. Some indiviual data points may be flagged as outliers, high-leverage points, or influential points. Any point falling into one of these categories should be carefully examined for accuracy(transciption error, etc), relevancy (whether it belongs to the data set or not), or special significance (abnormal conditions, etc).

References

1. Belsley, D. A., Kuh, E., and Welsh, R. E.(1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinesrity, New York: John Wiley & Sons.
2. Chatterjee, S. and Hadi, A. S.(1986). Influential observations, High Leverage Points, and Outliers in Linear Regression. Statistical Science 1, 379−416.
3. Chatterjee, S. and Hadi, A. S.(1988). Sensitivity Analysis in Linear Regression, John Wiley & Sons.
4. Cook, R. D. and Weisberg, S.(1982). Residuals and Influence in Regression, Chapman and Hall.
5. hampel, F. R.(1974). The influence Curve and its role in robust estimation. J. Amer. Statist. Assoc. 69, 383−93.
6. Hill, R. W.(1977). Robust Regression When There Are Outliers in the Carriers. unpublished ph. D. dissertation, Harvard University, Dept. of Statistics.
7. Jolliffe, I. T.(1986). Principal Component Analysis. Springer Verlag.
8. Walker, E. and Birch, J. B.(1988). Influence measures in Ridge Regression. Technometrics, 30, 221−227.