Influence Diagnostics in Two-Parameter Ridge Regression

Yasin ASAR[1] and Murat Erişoğlu[2]

[1] Department of Mathematics-Computer Sciences, Necmettin Erbakan University [2] Departmet of Statistics, Necmettin Erbakan University

Abstract: Identifying influential observations is an important part of the model building process in linear regression. There are numerous diagnostic measures based on different approaches in linear regression analysis. However, the problem of multicollinearity and influential observations may occur simultaneously. Therefore, we propose new diagnostic measures based on the two parameter ridge estimator defined by Lipovetsky and Conklin (2005) alternative to the usual ridge regression and ordinary linear regression. We define two parameter ridge-type generalizations of DFFITS and Cook's distance. Moreover, we obtain approximate case deletion formulas and provide approximate versions of new measures. Finally, we illustrate the benefits of proposed measures in real data examples.

Key words: Influence diagnostics, two-parameter ridge estimator, DFFITS, Cook's distance, case deletion

1. Introduction and Motivation

Consider the usual multiple linear regression model with intercept which is defined as

$$y = 1\beta_0 + X\beta_1 + \varepsilon \tag{1.1}$$

where X is an $n \times t$ data matrix centered and standardized, y is an $n \times 1$ response vector, β_0 is an unknown scalar parameter, β_1 is a $t \times 1$ vector of unknown coefficients and ε is an $n \times 1$ random error vector following normal distribution $\varepsilon \sim N(0, \sigma^2 I_n)$ such that $E(\varepsilon) = 0$, $\text{var}(\varepsilon) = \sigma^2 I_n$. Now, let $Z = [1 \ X]$ is the $n \times p$ design matrix.

The ordinary least square (OLS) estimator of $\left(\beta' = \left[\beta_0, \beta_1'\right]\right)$ is $\hat{\beta} = (Z'Z)^{-1} Z'y$ and the vector of fitted responses is $\hat{y} = Z\hat{\beta}$. The unbiased estimator of σ^2 is $s^2 = e'e/(n-p)$ where $e = y - \hat{y} = (I_n - H)y$ is the residual vector and H is the hat or projection matrix having

diagonal elements $h_{ii} = z_i \left(Z'Z \right)^{-1} z_i'$ called leverages such that z_i is the ith row of Z. We refer to a couple $\left(y_i, z_i \right)$ as a case, as suggested by Cook and Weisberg (1982).

Not all data points in a data set have the same importance in determining the coefficient estimates, t-values and some other statistics. Some points may affect the analysis or estimates remarkably. Therefore it is crucial to detect this kind of points in the analysis. This process is called as influence diagnostics.

After the seminal study of Cook (1977), numerous diagnostic measures have been developed to identify influential observations in literature, for example see: Belsey et al. (1980), Atkinson (1981), Cook and Weisberg (1982), Chatterjee and Hadi (1986), and Shi (1997).

However, influential observations and the problem of multicollinearity may occur at the same time. Belsey et al. (1980) stated that using biased estimators to overcome multicollinearity may affect influence of some cases. The most common method of detecting influential observations is to use single-case deletion approach as Cook (1977) did. Therefore, Walker and Birch (1988) defined a ridge regression (Hoerl and Kennard, 1970) scheme with case deletion method and obtained the approximate case deletion formulas for the detection of influential cases and proposed ridge generalizations of Cook's distance (Cook, 1977) and DFFITS (Belsey et al., 1980) which are the most commonly used statistics based on the case deletion method in the ordinary linear regression.

After Walker and Birch (1988), generalized versions of Cook's distance and DFFITS of some biased estimators used for combating multicollinearity have been defined, for example, Jahufer and Jianbao (2009) obtained global influential observations by using a modified ridge regression scheme, Ullah et al. (2013) defined Liu versions and Ertas et al. (2013) obtained Liu and modified Liu versions of the mentioned single case diagnostics.

The purpose of this paper is to introduce new influence diagnostics based on a two-parameter ridge estimator defined by Lipovetsky and Conklin (2005) and obtain generalized versions of Cook's distance and DFFITS and approximate case deletion formulas for this estimator.

The organization of the paper is as follows: We give some quick background information and review the influence measures in ordinary linear regression in section 2. In section 3, we introduce new diagnostic measures in two-parameter ridge regression and obtain case deletion formulas. Applications of real data sets are illustrated in section 4.

2. Background Information

The main purpose of influence analysis is to measure the changes occurred in a defined aspect of the research when there is perturbation in the data. As we mentioned, one approach is to use case omission perturbation technique. We follow this technique throughout this article and we assume that the reader is familiar with the basic concepts of leverages and influence analysis in ordinary least squares.

Although there are various types of single case diagnostic methods, one popular method is the difference in fit standardized called DFFITS (Belsey et al., 1980) ,which is the standardized change in the fitted value of a case when it is deleted, can be evaluated at the ith case as

$$DFFITS_i = \frac{z_i\left(\hat{\beta}-\beta_{(i)}\right)}{se\left(z_i\hat{\beta}\right)} = \frac{e_i}{s(i)}\frac{h_{ii}^{1/2}}{1-h_{ii}} \tag{2.1}$$

where $\hat{\beta}_{(i)}$ is the OLS estimator of β when the ith case is deleted and $s(i) = \left((n-p)s^2 - \left[e_i^2/(1-h_{ii})\right]\right)/(n-p-1)$ is the OLS estimator of σ without the ith case.

Another popular and useful measure is Cook's distance (Cook and Weisberg, 1982), which is a measure of the change in the fitted values when the ith case is deleted, is defined by

$$D_i = \frac{\left(\hat{\beta}-\beta_{(i)}\right)' Z'Z\left(\hat{\beta}-\beta_{(i)}\right)}{ps^2} = \frac{e_i^2}{ps^2}\frac{h_{ii}}{\left(1-h_{ii}\right)^2}. \tag{2.2}$$

where It is observed from the above measures that the influence of a case can be interpreted as a function of residuals and leverages. Moreover, it is important to emphasize that these measures are useful for exploring the individual or single influential cases. Shi and Wang (1999) stated that measures based on the case deletion method may suffer from the masking problem which occurs in the presence of another influential cases.

D_i detects the case causing the most change in the estimates when it is deleted, moreover, $DFFITS_i$ also considers the effect on the estimates of variance s^2 (Brown and Lawrance, 2000). If the values of D_i and $DFFITS_i$ exceed some well-defined cutoff points, then it is said that the ith observation is influential. However, the cutoff points for these measures are not clear. Cook and Weisberg (1982) suggested as $D_i > 1$ and Bollen and Jackman (1990) indicated that $D_i > 4/n$ might be used. $|DFFITS_i| > 2\sqrt{p/n}$ is a commonly used cut-off point (Belsey et al., 1980). However, it is important to note that these influence measures are only useful for identifying single cases with high-influence.

3. Influence Measure in Two-Parameter Ridge Estimator

3.1 Two-Parametere Ridge Estiamator(TPR)

When the explanatory variables are correlated to each other, the variance of unbiased OLS estimator becomes inflated so that we cannot make stable estimations. Therefore, there are various studies proposing biased estimators in literature. Among them, ridge estimator (RE) $\hat{\beta}_k = \left(Z'Z + kI\right)^{-1} Z'y, \, k > 0$ defined by Hoerl and Kennard (1970) and Liu estimator (Kejian, 1993) $\hat{\beta}_d = \left(Z'Z + I\right)^{-1}\left(Z'y - d\beta\right), 0 < d < 1$ are the popular ones.

In this study, we consider a two-parameter ridge estimator (TPR) defined by Lipovetsky and Conklin (2005). Although RE is a popular estimator, its quality of fit is worse than OLS and does not satisfy the orthogonality assumptions. Therefore Lipovetsky and Conklin (2005) obtain TPR, a generalization of RE to two parameter model, considering a simultaneous minimization of the model errors, deviations from orthogonality between regressors and errors and deviations of the solutions from the pairwise regressions.

Now, let us denote the objective function of the sum of squared errors of OLS as follows:

$$S^2 = \left\|\varepsilon^2\right\| = \left(y - Z\beta\right)'\left(y - Z\beta\right) = 1 - 2\beta'r + \beta'C\beta \tag{3.1}$$

where $C = Z'Z$ and $r = Z'y$ which are in the correlation forms. The minimization of (3.1) is satisfied by the following normal system of equations

$$C\beta = r. \tag{3.2}$$

Solution of this system gives us the usual OLS estimator $\hat{\beta} = C^{-1}r$. The determination coefficient R^2 estimates the quality of the model such that

$$R^2 = 1 - S^2 = 2\beta'r - \beta'C\beta = \beta'C\beta. \tag{3.3}$$

We can describe the relation (3.2) as

$$Z'\varepsilon = 0 \tag{3.4}$$

expressing the orthogonality of each regressor (column of Z) to the error vector.

Similarly, we can obtain the objective function of RE as

$$S^2 = \|\varepsilon\|^2 + k\|\beta\|^2 = 1 - 2\beta'r + \beta'C\beta + k\beta'\beta. \tag{3.5}$$

Minimizing (3.5) gives us RE as $\hat{\beta}_k = (Z'Z + kI)^{-1} Z'y = (C + kI)^{-1} r, \, k > 0$. We can conclude that RE does not satisfy the orthogonality assumption(3.4). Therefore Lipovetsky and Conklin (2005) constructed a multi-objective least squares for a regression as follows:

$$S^2 = q_1 \|y - Z\beta\|^2 + q_2 \|Z'(y - Z\beta)\|^2 + q_3 \|Z'y - \beta\|^2 + q_4 \|y'(y - Z\beta)\|^2 + q_5 \|\beta\|^2. \tag{3.6}$$

Minimizing the equation (3.6) yields the following matrix equation

$$\left(q_1 C + q_2 C^2 + q_3 I + q_4 rr' + q_5 I\right)\beta = \left(q_1 I + q_2 C + q_3 I + q_4 I\right)r. \tag{3.7}$$

Taking $q_2 = 0$ and after some algebraic calculations (see Lipovetsky and Conklin (2005) for details), TPR is obtained as follows:

$$\hat{\beta}(k,q) = q(C + kI)^{-1} r \tag{3.8}$$

where the parameter q is chosen to maximize the function of regression fit which is given by

$$\tilde{R}^2 = 2qr'(C + kI)^{-1} r - q^2 r'(C + kI)^{-1} C(C + kI)^{-1} r \tag{3.9}$$

which can be obtained by using the equations (3.8) and (3.3). The optimal value of the parameter q is computed by

$$q = \frac{r'(C + kI)^{-1} r}{r'(C + kI)^{-1} C(C + kI)^{-1} r} \tag{3.10}$$

which is always bigger than 1. The authors also claimed that all the orthogonality assumptions hold for TPR.

3.2 Leverage and Residuals in TPR

Using TPR given in (3.8), we can obtain the vector of fitted values $\hat{y}(k,q) = Z\hat{\beta}(k,q) = H(k,q)y$ where $H(k,q) = qZ(C+kI)^{-1}Z'$ is the hat matrix for TPR, plays the same role as the hat matrix H of ordinary least squares. We can interpret the ith fitted value in terms of the elements of $H(k,q)$ as $\hat{y}_i(k,q) = \sum_{j=1}^{n} h_{ji}(k,q)y_j$; consequently, $\dfrac{\partial \hat{y}_i(k,q)}{\partial y_i} = h_{ii}(k,q)$ which is the ith diagonal element of $H(k,q)$. We can see the diagonal elements $h_{ii}(k,q)$ as the leverages for TPR regression as in the least square. Note that the matrix $H(k,q)$ is not idempotent, thus it is not a projection matrix.

Furthermore, we can consider canonical reduction by applying the singular value decomposition (SVD) (Mandel, 1982) so that $Z = U\Lambda^{1/2}V'$ where Λ is a diagonal matrix consisting of eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p$ of the matrix Z and the columns of the matrix V are the eigenvectors of Z such that $Z'Z = V\Lambda V'$. The (ij) element of the $n \times p$ matrix U is such that $u_{ij}\sqrt{\lambda_j}$ is the projection of ith row z_i onto the jth eigenvector of Z . By using the SVD theorem, i^{th} leverage of TPR can be written as follows:

$$
\begin{aligned}
h_{ii}(k,q) &= qz_i(C+kI)^{-1}C^{-1}z_i' \\
&= q\sqrt{\lambda_j}u_{ii}v_i'(V\Lambda V' + kI)^{-1}(V\Lambda V')^{-1}v_i u_{ii}\sqrt{\lambda_j} \\
&= \sum_{j=1}^{p} \frac{q\lambda_j}{\lambda_j + k}u_{ij}^2.
\end{aligned}
$$

We observe from the above result that if q approaches to one when k is fixed, then $h_{ii}(k,q)$ approaches to the ridge leverages $h_{ii}(k)$. If k approaches to zero when $q = 1$, then $h_{ii}(k,q)$ values goes to the OLS leverages h_{ii} . Moreover, we conclude that $\lambda_{\min} > k/(q-1)$ implies $h_{ii}(k,q) > h_{ii} > h_{ii}(k)$ and $\lambda_{\max} < k/(q-1)$ implies $h_{ii} > h_{ii}(k,q) > h_{ii}(k)$ where $\lambda_{\min}$ and $\lambda_{\max}$ are the minimum and maximum eigenvalues of $Z'Z$.

The ith residual of TPR is also given by

$$
\begin{aligned}
e_i(k,q) &= y_i - \hat{y}_i(k,q) \\
&= y_i - z_i \hat{\beta}_i(k,q) \\
&= \left(1 - h_{ii}(k,q)\right) y_i.
\end{aligned}
$$

We conclude similar results such that if q approaches to one when k is fixed, then $e_i(k,q)$ approaches to the ridge residuals $e_i(k)$. If k approaches to zero when $q=1$, then $e_i(k,q)$ values goes to the OLS residuals e_i.

3.3 Cook's Distance and DFFITS in TPR

We define a new version of DFFITS for TPR as

$$
DFFITS_i(k,q) = \frac{z_i\left(\hat{\beta}(k,q) - \beta_{(i)}(k,q)\right)}{se\left(z_i\hat{\beta}(k,q)\right)} = \frac{z_i\left(\hat{\beta}(k,q) - \beta_{(i)}(k,q)\right)}{s(i)\left[\sum_{j=1}^{n} h_{ij}^{2}(k,q)\right]^{1/2}} \tag{3.11}
$$

where $\hat{\beta}_{(i)}(k,q)$ is the TPR estimator (3.8) without the ith case and denominator is an estimator of the standard error of the TPR fitted value, $se\left(z_i\hat{\beta}(k,q)\right) = s\left[\sum_{j=1}^{n} h_{ij}^{2}(k,q)\right]^{1/2}$

such that k and q are assumed to be non-stochastic and $s(i) = \left[\dfrac{(n-p)s^2 - e_i^2/(1-h_{ii})}{n-p-1}\right]^{1/2}$

is the OLS estimator s of σ without the ith case. The OLS estimators s and $s(i)$ are used as measures of scale since the MSE function consists of the response and the fitted values which are both not depend on the eigenvalues of $Z'Z$ so that s and $s(i)$ are not affected by collinearity.

We also define two versions of Cook's D_i for TPR, namely,

$$
D_i^{*}(k,q) = \left(1/ps^2\right)\left(\hat{\beta}(k,q) - \beta_{(i)}(k,q)\right)' Z'Z\left(\hat{\beta}(k,q) - \beta_{(i)}(k,q)\right) \tag{3.12}
$$

and

$$D_i^{**}(k,q) = \left(q^2/ps^2\right)\left(\hat{\beta}(k,q) - \beta_{(i)}(k,q)\right)' C_k \left(Z'Z\right)^{-1} C_k \left(\hat{\beta}(k,q) - \beta_{(i)}(k,q)\right)$$

(3.13)

such that $D_i^{**}(k,q)$ is based on the fact that $\text{var}\left(\hat{\beta}(k,q)\right) = q^2\sigma^2 C_k^{-1} Z'Z C_k^{-1}$ where $C_k = C + kI$.

It is better to express these new measures as functions of leverages and residuals. However, this is not possible because of the scale dependency of the TPR estimator. Since the TPR estimator is not scale invariant, $Z(i)$, the Z matrix without the ith row, is needed to be rescaled before computing $\hat{\beta}_{(i)}(k,q)$. In the following subsection, we provide some approximate case deletion formulas to obtain the approximate versions of these measures.

3.4   Approximate Case Deletion Formulas for TPR

$\hat{\beta}_{(i)}(k,q) = q\left(Z(i)' Z(i) + kI\right)^{-1} Z(i)' y(i)$ is the TPR estimator without the ith case, $y(i)$ is the response vector without the ith element. We can write $\hat{\beta}_{(i)}(k,q)$ in the following form:

$$\hat{\beta}_{(i)}(k,q) = q\left(Z'Z + kI - z_i' z_i\right)^{-1} \left(Z'y - z_i' y_i\right).$$

Now, we apply Sherman-Morrison-Woodbury (SMW) theorem (Rao, 1973) to the matrix $\left(Z'Z + kI - z_i' z_i\right)^{-1} = \left(C_k - z_i' z_i\right)^{-1}$ and obtain $\hat{\beta}_{(i)}(k,q)$ as follows:

$$\hat{\beta}_{(i)}(k,q) = q\left(Z'Z + kI - z_i'z_i\right)^{-1}\left(Z'y - z_i'y_i\right)$$

$$\cong q\left(C_k^{-1} + \frac{C_k^{-1}z_i'z_iC_k^{-1}}{1 - z_iC_k^{-1}z_i'}\right)\left(Z'y - z_i'y_i\right)$$

$$\cong qC_k^{-1}Z'y - qC_k^{-1}z_i'y_i + \frac{qC_k^{-1}z_i'z_iC_k^{-1}Z'y}{1 - m_{ii}} - \frac{qm_{ii}C_k^{-1}z_i'y_i}{1 - m_{ii}}$$

$$\cong \hat{\beta}(k,q) - qC_k^{-1}z_i'y_i + \frac{C_k^{-1}z_i'\hat{y}_i(k,q)}{1 - m_{ii}} - \frac{qm_{ii}C_k^{-1}z_i'y_i}{1 - m_{ii}} \qquad (3.14)$$

$$\cong \hat{\beta}(k,q) - \frac{C_k^{-1}z_i'}{1 - m_{ii}}\left(qy_i - \hat{y}_i(k,q)\right)$$

$$\cong \hat{\beta}(k,q) - \frac{e_i^*C_k^{-1}z_i'}{1 - m_{ii}}$$

Where $M = ZC_k^{-1}Z'$, $m_{ii} = z_iC_k^{-1}z_i'$ and $e_i^* = \left(qy_i - \hat{y}_i(k,q)\right)$ is a vector (not residual). Thus we obtain the following difference formula:

$$\hat{\beta}(k,q) - \beta_{(i)}(k,q) \cong \frac{e_i^*C_k^{-1}z_i'}{1 - m_{ii}}. \qquad (3.15)$$

Based on the above result, we present the approximate versions of (3.11), (3.12) and (3.13) respectively as follows:

$$DFFITS_i^a(k,q) = \frac{e_i^*z_iC_k^{-1}z_i'}{(1 - m_{ii})s(i)\left[\sum_{j=1}^{n}h_{ij}^2(k,q)\right]^{1/2}}$$

$$= \frac{m_{ii}e_i^*(k,q)}{(1 - m_{ii})s(i)\left[\sum_{j=1}^{n}h_{ij}^2(k,q)\right]^{1/2}}, \qquad (3.16)$$

$$D_{ai}^{*}(k,q) = \frac{1}{ps^2} \frac{\left(e_i^*\right)^2 z_i C_k^{-1} Z'ZC_k^{-1} z_i'}{\left(1-m_{ii}\right)^2}$$

$$= \frac{1}{ps^2} \frac{\left(e_i^*\right)^2 \sum_{j=1}^{n} m_{ij}^2}{\left(1-m_{ii}\right)^2},$$

(3.17)

$$D_{ai}^{**}(k,q) = \frac{q^2}{ps^2} \frac{\left(e_i^*\right)^2 z_i C_k^{-1} C_k \left(Z'Z\right)^{-1} C_k C_k^{-1} z_i'}{\left(1-m_{ii}\right)^2}$$

$$= \frac{q^2}{ps^2} \frac{\left(e_i^*\right)^2 z_i \left(Z'Z\right)^{-1} z_i'}{\left(1-m_{ii}\right)^2}$$

(3.18)

$$= \frac{q^2}{ps^2} \frac{\left(e_i^*\right)^2 h_{ii}}{\left(1-m_{ii}\right)^2}.$$

4. Numerical Examples

In this section, we illustrate an application of new influence statistics to the widely investigated data set used by Longley (1967). There are 16 observations of response variable as total derived employment and 6 predictors namely, GNP implicit price deflator $1954 = 100\left(x_1\right)$, gross national product $\left(x_2\right)$, unemployment $\left(x_3\right)$, size of armed forces $\left(x_4\right)$, non-institutional population 14 years of age and over $\left(x_5\right)$ and the time $\left(x_6\right)$.

This data set has been used to identify influential observations by Cook (1977), Walker and Birch (1988), Jahufer and Jianbao (2009), and Ullah et al. (2013) and some other authors. To be consistent with these papers, we use the model (1.1) with the following notations: $y = 1\beta_0 + X_1\beta_1 + \varepsilon$ where 1 is a vector of 16 ones and $X_1 = \left[x_1 x_2 ... x_6\right]$ is centered and standardized so that $X_1'X_1$ is in correlation form. We use the matrix $Z = \left[1 \ X_1\right]$ as the design matrix. Thus, we use $\text{diag}\left(0,1,1,...,1\right)$ as the identity matrix as used in Walker and Birch (1988).

We used the Matlab program to compute all of the given information, so there may be some differences between our results and the literature. The condition number of the matrix Z is computed as $\kappa = \lambda_{\max} / \lambda_{\min} = 42473$ which shows that there is strong multicollinearity problem with this data set.

Cook (1977) considered this data set and found the cases $(5,16,4,10,15)$ as the most influential cases in this order. Walker and Birch (1988) obtained the cases $(16,10,4,15,1)$ as the influential cases in this order using the generalization of Cook's distance in ridge regression. Jahufer and Jianbao (2009) also used the same data set and identify the cases with the following new order $(16,4,1,10,15)$ by using the modified ridge regression generalization of the usual diagnostics. Very recently, Ullah et al. (2013) computed the influential points for different values of the parameter d used in Liu regression such that $(16,10,4,6,1)$, $(16,10,4,6,5)$, $(16,5,4,10,15)$ respectively for $d = 0.1, 0.5, 0.9$.

We obtained the values of $DFFITS_i^a(k,q)$, $D_{ai}^*(k,q)$ and $D_{ai}^{**}(k,q)$ for different values of the parameters k and q and identify the most influential observations as given in Table 1. We provide the values regarding the observations whose $DFFITS_i^a(k,q)$ value exceeds the cut-off value which is computed as 1.0445 and the five observations having largest Cook's distances $D_{ai}^*(k,q)$ and $D_{ai}^{**}(k,q)$.

We used four different estimators of the parameter k chosen from the literature and the optimal value of the parameter obtained by using (3.10) to minimize the mean squared error function as follows: $k_1 = \dfrac{ps^2}{\hat{\beta}'\beta}$, (Hoerl et al., 1975), $k_2 = \dfrac{s^2}{\hat{\beta}_{\max}^2}$, (Hoerl and Kennard, 1970), $k_3 = \text{median}\left(\dfrac{s^2}{\hat{\beta}_i^2}\right)$, (Kibria, 2003), $k_4 = \text{geomean}\left(\dfrac{s^2}{\hat{\beta}_i^2}\right)$, (Kibria, 2003).

According to Table 1, it is observed that the same observations that Cook (1977) identified as influential cases are detected as influential observations in a different order as $(16,10,5,4,15)$ and $(16,10,4,5,15)$ by D_{ai}^* and D_{ai}^{**} respectively when k_1 is used. If we use k_2 , then we obtain the same results as Cook (1977) did such that $(5,16,4,10,15)$ computed by using D_{ai}^* .

Moreover, if k_3 and k_4 are used, we obtain the influential observations in the same order as given in Ullah et al. (2013) in Table 1 by using Liu estimators. If we use k_4, we observe that the same cases given in Walker and Birch (1988) except for the sixth case instead of fifteenth are identified as influential cases by both D_{ai}^* and D_{ai}^{**} .

Table 1: The most influential cases according to DFFITS^a_i, D$^*_{ai}$ and D$^{**}_{ai}$
for different values of k and q.

k	q	Cases	D^*_{ai}	Cases	D^{**}_{ai}	Cases	$DFFITS^a_i$
k_1	9.8562×10^{-5}	1.000002					
		16	0.54984	16	0.56267	16	0.00692
		10	0.24554	10	0.26701	10	0.00632
		4	0.22988	5	0.25618	4	0.00558
		5	0.20389	4	0.24282	5	0.00553
		15	0.15142	15	0.15411	15	0.00390
k_2	2.1783×10^{-5}	1.0000004					
		16	0.48723	5	0.49321	5	0.00772
		5	0.46467	16	0.48999	16	0.00646
		4	0.24107	4	0.24452	10	0.00604
		10	0.23792	10	0.24297	4	0.00561
		15	0.16586	15	0.16653	15	0.00405
k_3	0.0018	1.000008					
		16	0.91656	16	1.13726	16	0.00968
		10	0.25723	10	0.48457	10	0.00798
		6	0.18731	6	0.21944	6	0.00395
		1	0.11507	4	0.13337	4	0.00394
		4	0.09707	1	0.12309	1	0.00317
k_4	0.0012	1.000007					
		16	0.88136	16	1.04126	16	0.00931
		10	0.26171	10	0.43866	10	0.00775
		6	0.17040	6	0.19617	4	0.00437
		1	0.12331	4	0.16011	6	0.00377
		4	0.12220	1	0.12949	1	0.00325

Values with * have smaller values than the cut-off point.

We also give some plots to summarize the results easily. In Figures 1-2, $DFFITS^a_i(k,q)$ $D^*_{ai}(k,q)$ and $D^{**}_{ai}(k,q)$ values versus observations are plotted only for the estimator in order to make it easy to observe the influential cases from these figures. According to Figure 1 and 2, the most influential cases are the 16th and 10th cases. D^*_{ai} , D^{**}_{ai} and $DFFITS^a_i(k,q)$ give similar results. These results are consistent with literature.
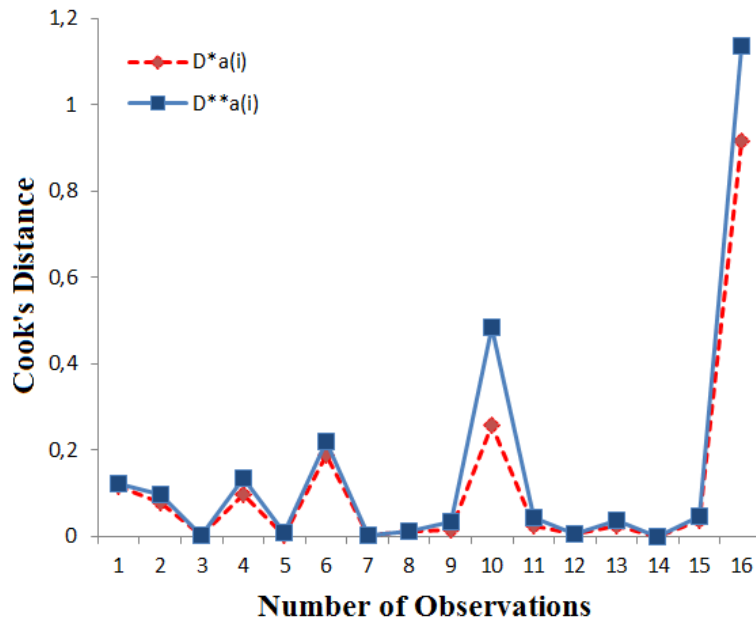
Figure 1: Plot of Cook's distance according to different two approach using k_3
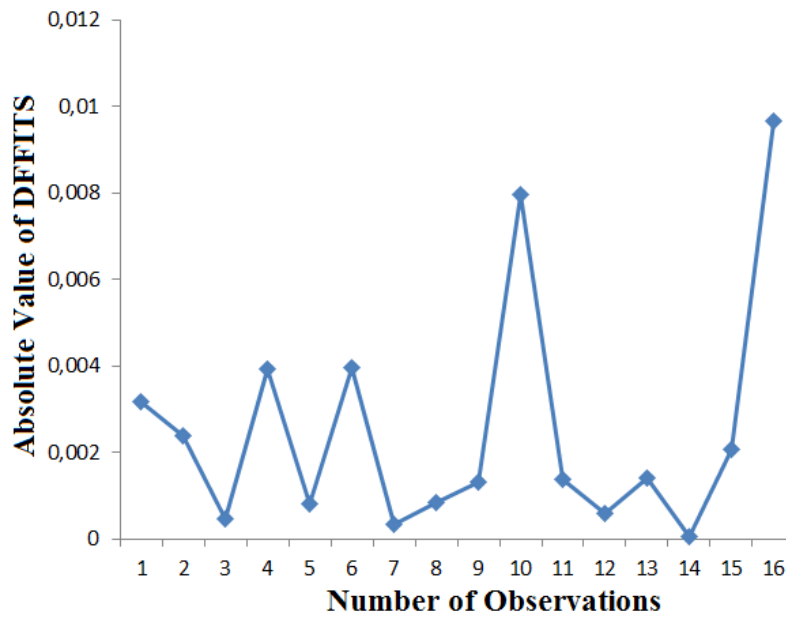
Figure 2: Plot of absolute value of DFFITS using k_3

In Figure 3, we provide the plot of hat diagonals versus observations. According to Figure 3, although the observations having the first three highest leverages are 16th, 2nd and 8th observations are not detected as influential cases. Thus, we can say that high leverage points may not be influential all the time.
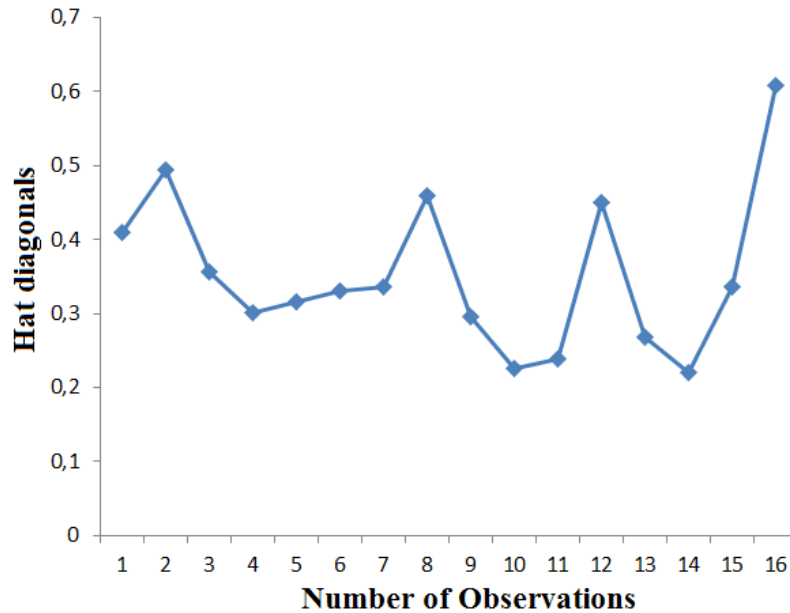

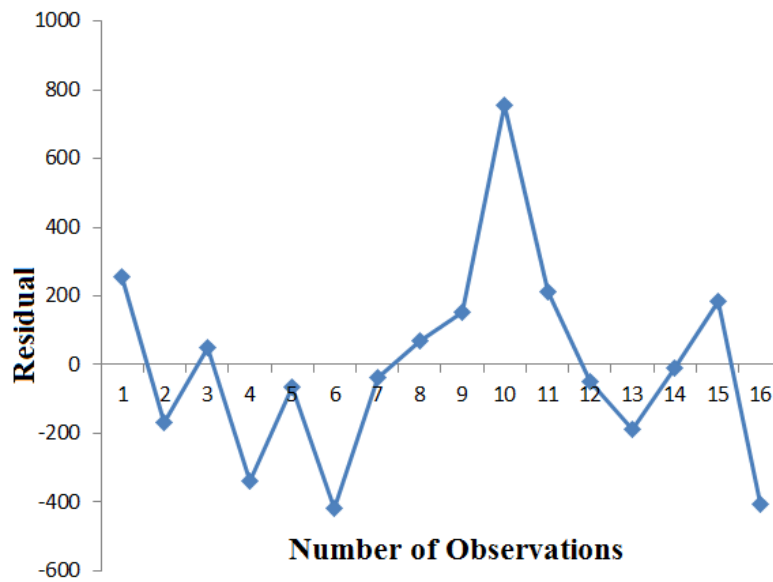Figure 3: Plot of hat diagonals $h_{ii}(k,q)$ using k_3

Figure 1. Plot of residuals $e_i(k,q)$ using k_3

The Figure 4 is the plot of residual against the observations. According to this figure, $(10, 16, 6, 4, 1)$ are the observations having largest residuals. Moreover, 10th observation has the largest residual; however it is not the most influential observation. Thus, we can say in a similar manner that having a larger residual does not guarantee to be the most influential observation.
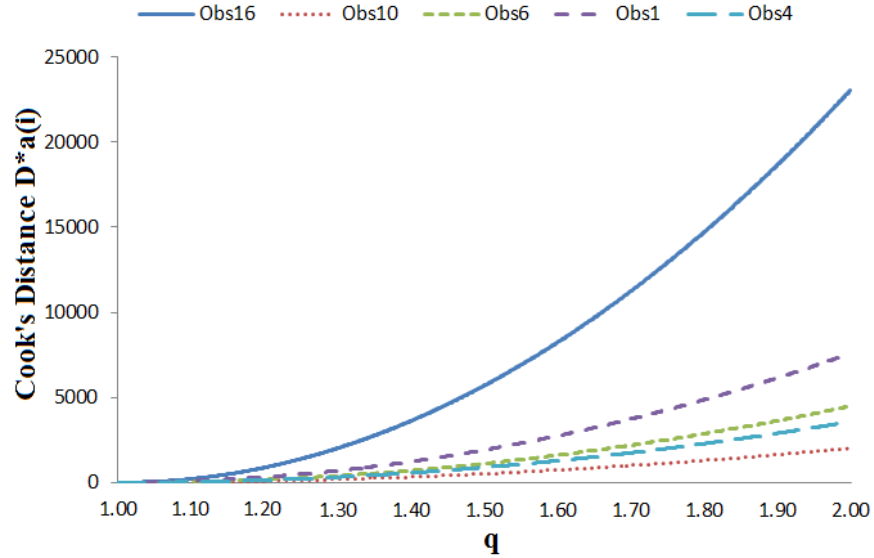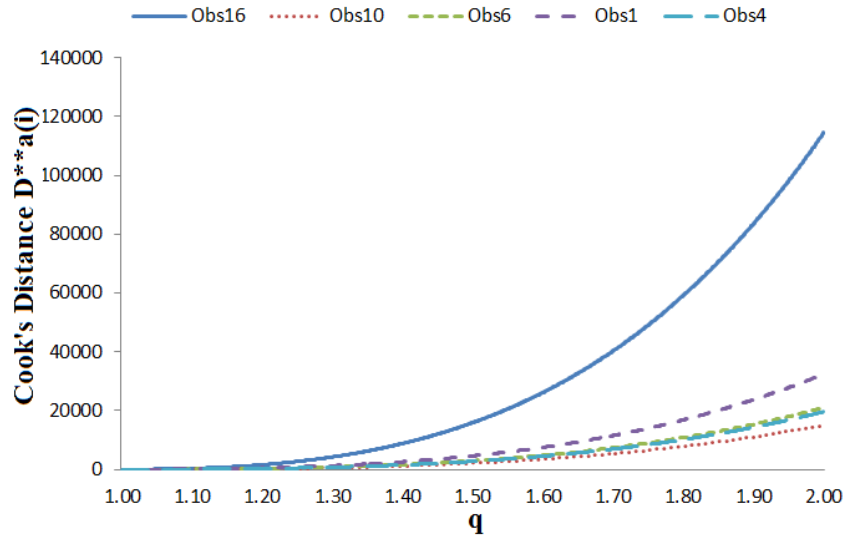

Figure 5. Plot of D^{*}_{ai} versus **q** using k_3

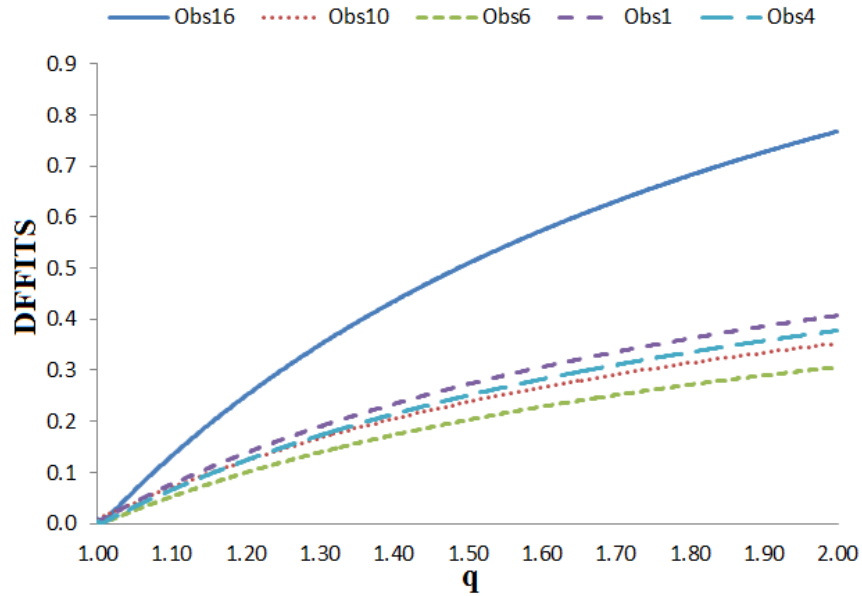Figure 6. Plot of D^{**}_{ai} versus **q** using k_3

Figure 7: Plot of DFFITS$_i$[a] versus **q** using k_3

In Figure 5, we computed the values of $D^*_{ai}(k,q)$ of the unordered influential observations for changing values of the parameter $(1,4,6,10,16)$ between 1 and 2 when k_3 is used. It can be seen from Figure 5 that increasing the value of q affects the distance values $D^*_{ai}(k,q)$ negatively. Moreover, we obtained a similar figure for $D^{**}_{ai}(k,q)$.

In the last figure, we provide the plots of the distance values of $DFFITS^a_i$ for changing values of q between 0 and 1 when k_3 is used. All distance values of influential observations are increased slowly, however, all distance values remain smaller than 1.

Finally, we consider the following data sets and obtain their influential observations and distance values using the new methods: Tobacco data (Myers, 1986), Hald data set (Hald, 1952), body fat data set (Neter et al., 1997) and crime rate data set (Agresti and Finlay, 1986).

Table 2: Influential observations and distance values of some data sets used in literature

Data	k_3	q	Cases	D_{ai}^*	Cases	D_{ai}^{**}	Cases	$DFFITS_i^a$
Tobacco	0.0038	1.0003	4	0.21568	4	0.64017	4	0.05092
			14	0.13010	14	0.27389	14	0.02510
			1	0.07646	8	0.12989	8	0.01930
			8	0.07103	1	0.08700	1	0.01424
			11	0.04500	11	0.08672	11	0.01375
Hald	0.0146	1.0002	8	0.28968	8	0.32303	8	0.68700
			11	0.15016	11	0.17892	11	0.37566
			10	0.11096	10	0.12331	6	0.35687
			6	0.07808	3	0.12133	10	0.27890
			13	0.05606	6	0.08382	13	0.23406
Body fat	0.0263	1.0007	3	0.36455	3	0.42279	3	0.51638
			13	0.15563	13	0.20096	13	0.35350
			14	0.10079	8	0.12962	14	0.31067
			8	0.06724	14	0.12739	8	0.22969
			4	0.05292	2	0.09174	4	0.19599
Crime	0.0566	1.0083	13	0.17213	13	0.26876	13	0.05055
			1	0.12978	1	0.19935	1	0.04340
			37	0.06693	37	0.08998	37	0.02892
			54	0.04821	54	0.05519	54	0.02358
			22	0.04408	22	0.05003	22	0.02156

According to Table 2, we see that our detections agree with the literature [(Cook, 1977), (Ullah, et al., 2013)]. Thus, it is showed that new diagnostics defined in two-parameter ridge estimator are successful to determine the influential observations of the data sets used in literature.

5. Conclusion

In this article, we consider the problem of multicollinearity and influential observations together and propose new diagnostic measures using a two-parameter ridge estimator. In order to obtain the approximate versions of new diagnostic measures, we present the approximate case deletion formulas in two-parameter ridge regression using SMW theorem.

Moreover, we illustrate an example of real data application using Longley (1967) data. The numerical results show that new measures are useful to identify influential observations. However, we suggest to the practitioners that it is important to use these measures along with the knowledge and expertise such that he/she needs to decide whether the identified case should be retained, removed or down weighted.

References

[1] Agresti, A., and Finlay, B. (1986). *Statistical Methods for the Social Sciences*: Pearson Prentice Hall, Inc.

[2] Atkinson, A. (1981). *Two graphical displays for outlying and influential observations in regression*. Biometrika, **68**(**1**), 13-20.

[3] Belsey, D. A., Kuh, E., and Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*: John Wiley.

[4] Bollen, K. A., and Jackman, R. W. (1990). Regression diagnostics: An expository treatment of outliers and influential cases. *Modern methods of data analysis*, 257-291.

[5] Brown, G. C., and Lawrance, A. J. (2000). Theory and illustration of regression influence diagnostics. *Communications in Statistics - Theory and Methods*, **29**(**9**-**10**), 2079-2107. doi: 10.1080/03610920008832597.

[6] Chatterjee, S., and Hadi, A. S. (1986). Influential observations, high leverage points, and outliers in linear regression. Statistical Science, 379-393.

[7] Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 15-18.

[8] Cook, R. D., and Weisberg, S. (1982). Residuals and influence in regression. *Monographs on Statistics and Applied Probability*. New York: Chapman and Hall.

[9] Ertas, H., Erisoglu, M., and Kaciranlar, S. (2013). Detecting influential observations in Liu and modified Liu estimators. *Journal of Applied Statistics*, **40**(**8**), 1735-1745. doi: 10.1080/02664763.2013.794203.

[10] Hald, A. (1952). *Statistical theory with engineering applications Statistical theory with engineering applications*: John Wiley & Sons.

[11] Hoerl, A. E., Kannard, R. W., and Baldwin, K. F. (1975). Ridge regression: some simulations. *Communications in Statistics-Theory and Methods*, **4**(**2**), 105-123.

[12] Hoerl, A. E., and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, **12**(**1**), 55-67.

[13] Jahufer, A., and Jianbao, C. (2009). Assessing global influential observations in modified ridge regression. *Statistics & Probability Letters*, **79**(**4**), 513-518.

[14] Kejian, L. (1993). A new class of blased estimate in linear regression. *Communications in Statistics-Theory and Methods*, **22**(**2**), 393-402.

[15] Kibria, B. M. G. (2003). Performance of some new ridge regression estimators. *Communications in Statistics-Simulation and Computation*, **32**(**2**), 419-435.

[16] Lipovetsky, S., and Conklin, W. M. (2005). Ridge regression in two‐parameter solution. *Applied Stochastic Models in Business and Industry*, **21**(**6**), 525-540.

[17] Longley, J. W. (1967). An appraisal of least squares programs for the electronic computer from the point of view of the user. *Journal of the American Statistical Association*, **62**(**319**), 819-841.

[18] Mandel, J. (1982). Use of the singular value decomposition in regression analysis. *The American Statistician*, **36**(**1**), 15-24.

[19] Myers, R. H. (1986). Classical and modern regression with applications. Boston: *Duxbury Press.*

[20] Neter, J., Kutner, M. H., Nachtsheim, C., and Wasserman, W. (1997). *Applied statistical models* (*4th edition*).

[21] Rao, C. R. (1973). *Linear statistical inference and its applications (Vol. 22)*: John Wiley & Sons.

[22] Shi, L. (1997). Local influence in principal components analysis. *Biometrika*, **84**(**1**), 175-186.

[23] Shi, L., and Wang, X. (1999). Local influence in ridge regression. *Computational Statistics & Data Analysis*, **31**(3), 341-353.

[24] Ullah, M. A., Pasha, G. R., and Aslam, M. (2013). Assessing Influence on the Liu Estimates in Linear Regression Models. *Communications in Statistics - Theory and Methods*, **42**(**17**), 3100-3116. doi: 10.1080/03610926.2011.620206.

[25] Walker, E., and Birch, J. B. (1988). Influence Measures in Ridge-Regression. *Technometrics*, **30**(**2**), 221-227.

Yasin Asar
Department of Mathematics-Computer Sciences
Necmettin Erbakan University
42090 Konya, Turkey
yasinasar@hotmail.com
yasar@konya.edu.tr

Murat Erisoglu
Department of Statistics
Necmettin Erbakan University
42090 Konya, Turkey
Merisoglu@konya.edu.tr