Journal of Data Science **9**(2011), 359-372

Measuring Local Influential Observations in Modified Ridge Regression

Aboobacker Jahufer[1] and Jianbao Chen[2]
[1]*South Eastern University and* [2]*Xiamen University*

Abstract: In this paper, we use generalized influence function and generalized Cook distance to measure the local influence of minor perturbation on the modified ridge regression estimator in ridge type linear regression model. The diagnostics under the perturbation of constant variance and individual explanatory variables are obtained when multicollinearity presents among the regressors. Also we proposed a statistic that reveals the influential cases for Mallow's method which is used to choose modified ridge regression estimator biasing parameter. Two real data sets are used to illustrate our methodologies.

Key words: Local influential observations, modified ridge regression, multicollinearity, perturbation scheme.

1. Introduction

In most applications of regression analysis, the regressors are found not to be orthogonal. Sometimes, the lack of orthogonality is not a serious problem. But, when the regressors are nearly perfectly linearly related, the inferences based on the regression model can be misleading. When there exist near linear dependencies between the regressors, the problem of multicollinearity is said to be present. When the method of ordinary least squares estimator (OLSE) is applied to multicollinearity data, poor estimates of the regression coefficients could be obtained. One of the solutions to solve the problem of multicollinearity is the use of biased estimator. There are many biased estimation procedures are proposed in literature; among which the ridge type estimations are very popular.

The effect of influential observations on the parameter estimates of OLSE regression model received considerable attention in the last three decades. However, very little attention has been given to the problem of influential observations in the biased estimations. Many papers and texts address the problems of multicollinearity and influential observations in linear unbiased estimation (see Cook, 1977; Belsley *et al.*, 1980).

The local influence approach was proposed by Cook (1986) as a general method for assessing the influence of minor perturbations of a statistical model and the approach has been applied to a number of influence analysis problems. Some of the recent papers are Shi (1997) and Shi and Wang (1999) studied the local influence analysis in principal component analysis and ordinary ridge regression estimator respectively.

In this paper, we intend to analyze the local influential observations on modified ridge regression estimator (MRRE) using the minor perturbation method. This paper is composed of six sections. Section 2 gives the background of the study. Section 3 derives the local influence diagnostics of MRRE including the perturbation of constant variance and individual explanatory variables. Section 4 provides a diagnostic for detecting the local influential observations of choosing the MRRE biasing parameter. Section 5 reports examples using two real macroeconomic data sets. Comments are given in last section.

2. Background

A matrix multiple regression model using Walker and Birch (1988), takes the form

$$\mathbf{y} = \mathbf{1}\beta_0 + \mathbf{X}\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}, \tag{2.1}$$

where $\mathbf{y}$ is an $n \times 1$ vector of observable random variable, $\mathbf{1}$ is an $n \times 1$ vector of ones, β_0 is an unknown parameter, $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_m)$ is an $n \times m$ centered and standardized known matrix ($\mathbf{1}'\mathbf{x}_i = 0, \mathbf{x}_i'\mathbf{x}_i = 1, i = 1, \cdots, m$), $\boldsymbol{\beta}_1$ is a $m \times 1$ vector of unknown parameters and $\boldsymbol{\epsilon}$ is an $n \times 1$ vector of unobservable random errors with $E(\boldsymbol{\epsilon}) = 0$ and $Var(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}_n$ where $\mathbf{I}_n$ is an identity matrix of order n.

Let $\mathbf{Z} = (\mathbf{1}, \mathbf{X})$ and $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1')'$ then the OLSE of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$. The estimator of σ^2 is given by $s^2 = \mathbf{e}'\mathbf{e}/(n-p)$, where $\mathbf{e} = \mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\beta}}$ residual vector and $p = m + 1$. When there exists strong near linear relationship among the columns of matrix $\mathbf{Z}$, it can be said that multicollinearity exists in the data set.

To deal with multicollinearity problems, ridge type biased estimation techniques are often used. The ordinary ridge regression estimator (ORRE) introduced by Hoerl and Kennard (1970a), is defined as

$$\hat{\boldsymbol{\beta}}_R = (\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1}\mathbf{Z}'\mathbf{y}, \tag{2.2}$$

where $\mathbf{I}_p^* = diag(0, 1, \cdots, 1)$ and $k > 0$ is called ridge biasing parameter. $\hat{\boldsymbol{\beta}}_R$ is a biased estimator, however, the variances of its elements are less than the variances of the corresponding elements of the $\hat{\boldsymbol{\beta}}$ for suitable k.

Swindel (1976) introduced one ridge type estimator based on prior information which is called MRRE and it is defined as

$$\hat{\boldsymbol{\beta}}_{(k,b)} = (\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1}(\mathbf{Z}'\mathbf{y} + k\mathbf{b}), \tag{2.3}$$

where $k > 0$ is MRRE biasing parameter, $\mathbf{b}$ is a $m \times 1$ prior information vector. The prior information vector is same as the ORRE (see Trenkler, 1988).

A number of studies revealed the effect on MRRE. For instance, Pliskin (1987) compared the mean squared error matrix of the ORRE and MRRE; Kaçiranlar et al. (1998) and Wijekoon (1998) analyzed mean squared error comparisons of the restricted ridge regression estimator (RRRE) and MRRE and Groß (2003) used this estimator to develop a new RRRE. They proved that MRRE is superior to OLSE, ORRE and RRRE in mean squared error criteria.

The key problem in the MRRE is to choose the value of biasing parameter k. There are several methods for selecting the value of k, among which Hoerl and Kennard's an iterative procedure (Hoerl and Kennard, 1976), Mcdonald-Galarneau's method (Mcdonald and Galarneau, 1975), C_p statistic criterion (Mallows, 1973), GCV criterion (Wahba et al., 1979), PRESS procedure and VIF procedure (Marquardt, 1970) are popular.

In this paper, we use Mallows (1973) C_p statistic to choose k. The biasing parameter k is chosen by minimizing Mallows C_p and it is modified for MRRE is

$$C_{(k,b)} = \frac{SSR_{(k,b)}}{s^2} + 2tr(\mathbf{H}_{(k,b)}) - (n-2), \tag{2.4}$$

where $SSR_{(k,b)}$ is the sum of squares residual of MRRE and s^2 is the estimator of σ^2 from OLSE, $\mathbf{H}_{(k,b)}$ is hat matrix of the MRRE and $\mathbf{H}_{(k,b)} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1}[\mathbf{Z}' + k(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1}\mathbf{Z}'] = \mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1}(\mathbf{Z}' + k\mathbf{M})$ where $\mathbf{M} = (\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1}\mathbf{Z}'$ from prior information.

3. Local Influence Diagnostics

In a single case influential observation, we can use the case deletion method (see Jahufer and Chen, 2009). However, for our current case, it is almost impossible to derive an exact formula for $\hat{\boldsymbol{\beta}}_{(k,b)}(i)$ because of the scale dependency of the MRRE, where $\hat{\boldsymbol{\beta}}_{(k,b)}(i)$ is the MRRE without the i-th observation. That is, $\mathbf{Z}(-i)$ (the $\mathbf{Z}$ with the i-th row deleted) has to be rescaled to unit length before computing $\hat{\boldsymbol{\beta}}_{(k,b)}(i)$. Also, the measures induced by the case deletion method often suffer from masking effects. The main purpose of this paper then is to derive an alternative method, local influence analysis, employed to study the influence of observations on the MRRE.

Shi and Wang (1999) suggested the local influence, in which the generalized influence function (GIF) and generalized Cook (GC) statistic are defined to assess the local change of small perturbation on some key issues. The generalized influence of a concerned function of quantity $T \in R^p$ is given by

$$GIF(\mathbf{T}, l) = \lim_{a \to 0} \frac{\mathbf{T}(\omega_0 + al) - \mathbf{T}(\omega_0)}{a}, \tag{3.1}$$

where $\omega = \omega_0 + al \in R^n$ represents a perturbation, ω_0 is a null perturbation which satisfies $\mathbf{T}(\omega_0) = \mathbf{T}$ and $l \in R^n$ denotes an unit length vector. To assess the influence of the perturbations on $\mathbf{T}$, the GC statistic is defined as

$$GC(\mathbf{T}, l) = [GIF(\mathbf{T}, l)]'\mathbf{M}[GIF(\mathbf{T}, l)]/c, \tag{3.2}$$

where $\mathbf{M}$ is a $p \times p$ positive or semi positive definite matrix and c is a scalar. By minimizing the absolute value of $GC(\mathbf{T}, l)$ with respect to l, a direction $l_{max}(\mathbf{T})$ is obtained. This direction shows how to perturb the data to obtain the greatest local change in $\mathbf{T}$. Thus, it can be used as a main diagnostic. Maximum value $GC_{max}(T) = GC(\mathbf{T}, l_{max})$ indicates the serious local influence. This method removes the need of likelihood assumption.

3.1 Perturbing the constant variance

Here, the assumption of the constant variance in model (2.1) is simultaneously perturbed. This perturbation scheme is a better way to handle cases badly modeled (Lawrance, 1988). The distribution of $\boldsymbol{\epsilon}$ under the perturbation becomes

$$\boldsymbol{\epsilon}_\omega \sim N(0, \sigma^2 \mathbf{W}^{-1}), \tag{3.3}$$

where $\mathbf{W} = diag(\omega)$ is a diagonal matrix with diagonal elements of $\boldsymbol{\omega}' = (\omega_1, \cdots, \omega_n)$. Let $\boldsymbol{\omega} = \boldsymbol{\omega_0} + al$, where $\boldsymbol{\omega_0} = \mathbf{1}$, the n-vector of ones and $l' = (l_1, \cdots, l_n)$. The perturbed version of the MRRE is

$$\hat{\boldsymbol{\beta}}_{(k,b)}(\omega) = (\mathbf{Z}'\mathbf{W}\mathbf{Z} + k\mathbf{I}_p^*)^{-1}(\mathbf{Z}'\mathbf{W}\mathbf{Y} + k\mathbf{b}). \tag{3.4}$$

But we know

$$\begin{aligned}&(\mathbf{Z}'\mathbf{W}\mathbf{Z} + k\mathbf{I}_p^*)^{-1}\\ &= (\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1} - a(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1}\mathbf{Z}'\mathbf{D}(l)\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1} + 0(a^2),\end{aligned}$$

where $\mathbf{D}(l) = diag(l)$, so from (3.4) we can have

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{(k,b)}(\omega) \;=\; & \hat{\boldsymbol{\beta}}_{(k,\mathbf{b})} - a[(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1}\mathbf{Z}'\mathbf{D}(l)\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1}(\mathbf{Z}'\mathbf{Y} + k\mathbf{b})\\ & + (\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1}\mathbf{Z}'\mathbf{D}(l)\mathbf{Y}] + 0(a^2).\end{aligned}$$

Therefore the GIF of $\hat{\boldsymbol{\beta}}_{(k,b)}(\omega)$ under the perturbation is

$$GIF(\hat{\boldsymbol{\beta}}_{(k,b)}, l) = (\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1}\mathbf{Z}'\mathbf{D}(l)[\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}}_{(k,b)}].$$

The residual vector of MRRE is $\mathbf{e}_{(k,b)} = \mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}}_{(k,b)}$, then the above equation becomes

$$GIF(\hat{\boldsymbol{\beta}}_{(k,b)}, l) = (\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1}\mathbf{Z}'\mathbf{D}(e_{(k,b)})l. \tag{3.5}$$

Analogous to case deletion, two versions of the generalized Cook statistic of $\hat{\boldsymbol{\beta}}_{(k,b)}$ can be constructed:

$$GC_1(\hat{\boldsymbol{\beta}}_{(k,b)}, l) = l'\mathbf{D}(e_{(k,b)})\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1}(\mathbf{Z}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1}\mathbf{Z}'\mathbf{D}(e_{(k,b)})l/ps^2 \tag{3.6}$$

and

$$GC_2(\hat{\boldsymbol{\beta}}_{(k,b)}, l) = l'\mathbf{D}(e_{(k,b)})\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{D}(e_{(k,b)})l/ps^2. \tag{3.7}$$

In (3.6) and (3.7) $\mathbf{M}$ scaled on the OLSE and MRRE framework respectively. That is $Cov(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{Z}'\mathbf{Z})^{-1}$ and $Cov(\hat{\boldsymbol{\beta}}_{(k,b)}) = \sigma^2(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1}(\mathbf{Z}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1}$.

Hence, the associated influential diagnostics denoted by $l_{max}^{(1)}(\hat{\boldsymbol{\beta}}_{(k,b)})$ and $l_{max}^{(2)}(\hat{\boldsymbol{\beta}}_{(k,b)})$ are the eigenvectors corresponding to the largest absolute eigenvalues for matrices $\mathbf{D}(e_{(k,b)})\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1}(\mathbf{Z}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1}\mathbf{Z}'\mathbf{D}(e_{(k,b)})$ and $\mathbf{D}(e_{(k,b)})\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{D}(e_{(k,b)})$ from (3.6) and (3.7) respectively. Then, $GC_{max}^{(1)} = GC_1(\hat{\boldsymbol{\beta}}_{(k,b)}, l_{max}^{(1)})$ and $GC_{max}^{(2)} = GC_2(\hat{\boldsymbol{\beta}}_{(k,b)}, l_{max}^{(2)})$ can be used to measure the impact of the local influential observations.

3.2 Perturbing the explanatory variables

In this subsection we consider the influence perturbation of explanatory variables on the MRRE. For simplicity, we only consider the individual perturbation of m explanatory variables. The i-th column of $\mathbf{Z}$ is perturbed by

$$\mathbf{Z}_\omega = \mathbf{Z} + as_i l\mathbf{d}_i', \tag{3.8}$$

where a is limiting coefficient and $l \in R^n$ denotes an unit length vector, $\mathbf{d}_i$ is $p \times 1$ vector with a 1 in the i-th position and zeroes elsewhere. s_i denotes the scale factor and accounts for the different measurement units associated with the columns of $\mathbf{Z}, i = 2, \cdots, p$. Under this perturbation, we can easily verify that

$$\hat{\boldsymbol{\beta}}_{(k,b)}(\omega) = (\mathbf{Z}_\omega'\mathbf{Z}_\omega + k\mathbf{I}_p^*)^{-1}(\mathbf{Z}_\omega'\mathbf{Y} + k\mathbf{b}). \tag{3.9}$$

In equation (3.9), $(\mathbf{Z}'_\omega \mathbf{Z}_\omega + k\mathbf{I}_p^*)^{-1}$ can expand as

$$\begin{aligned}(\mathbf{Z}'_\omega \mathbf{Z}_\omega + k\mathbf{I}_p^*)^{-1} \;=\; & (\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1} - as_i[(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1} \\ & (\mathbf{Z}'l\mathbf{d}'_i + \mathbf{d}_i l'\mathbf{Z} + as_i\mathbf{d}_i l' l \mathbf{d}'_i)(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1}] + 0(a^2)\end{aligned}$$

and $(\mathbf{Z}'_\omega \mathbf{Y} + k\mathbf{b}) = [(\mathbf{Z}'\mathbf{Y} + k\mathbf{b}) + as_i\mathbf{d}_i l'\mathbf{Y}]$. Hence, (3.9) becomes

$$\begin{aligned}&\hat{\boldsymbol{\beta}}_{(k,b)}(\omega) \\ &= \hat{\boldsymbol{\beta}}_{(k,b)} - as_i(\mathbf{Z}'_\omega \mathbf{Z}_\omega + k\mathbf{I}_p^*)^{-1}[\mathbf{d}_i l'\mathbf{Y} - \mathbf{Z}'l\mathbf{d}'_i(\mathbf{Z}'_\omega \mathbf{Z}_\omega + k\mathbf{I}_p^*)^{-1}(\mathbf{Z}'_\omega\mathbf{Y} + k\mathbf{b}) \\ &\quad -\mathbf{d}_i l'\mathbf{Z}(\mathbf{Z}'_\omega \mathbf{Z}_\omega + k\mathbf{I}_p^*)^{-1}(\mathbf{Z}'_\omega\mathbf{Y} + k\mathbf{b})] + 0(a^2).\end{aligned}$$

Therefore GIF for individual explanatory variable perturbation is

$$GIF(\hat{\boldsymbol{\beta}}_{(k,b)}(\omega), l) = s_i(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1}[\mathbf{d}_i\mathbf{e}'_{(k,b)} - \hat{\beta}^{(i)}_{(k,b)}\mathbf{Z}']l, \tag{3.10}$$

where $\hat{\beta}^{(i)}_{(k,b)}$ is the i-th element of $\hat{\boldsymbol{\beta}}_{(k,b)}$.

Therefore, the GC statistic of $\hat{\boldsymbol{\beta}}_{(k,b)}(\omega)$ is

$$GC(\hat{\boldsymbol{\beta}}_{(k,b)}(\omega), l) = [GIF(\hat{\boldsymbol{\beta}}_{(k,b)}(\omega), l)]'\mathbf{M}[GIF(\hat{\boldsymbol{\beta}}_{(k,b)}(\omega), l)]/ps^2.$$

Using (3.10), two versions of GC statistic can be written

$$\begin{aligned}&GC_1(\hat{\boldsymbol{\beta}}_{(k,b)}(\omega), l) \qquad (3.11)\\ &= \frac{s_i^2 l'}{ps^2}[\mathbf{e}_{(k,b)}\mathbf{d}'_i - \hat{\beta}^{(i)}_{(k,b)}\mathbf{Z}](\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1}(\mathbf{Z}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1}[\mathbf{d}_i\mathbf{e}'_{(k,b)} - \hat{\beta}^{(i)}_{(k,b)}\mathbf{Z}']l\end{aligned}$$

here, $\mathbf{M}$ is based on the OLSE framework and

$$GC_2(\hat{\boldsymbol{\beta}}_{(k,b)}(\omega), l) = \frac{s_i^2 l'}{ps^2}[\mathbf{e}_{(k,b)}\mathbf{d}'_i - \hat{\beta}^{(i)}_{(k,b)}\mathbf{Z}](\mathbf{Z}'\mathbf{Z})^{-1}[\mathbf{d}_i\mathbf{e}'_{(k,b)} - \hat{\beta}^{(i)}_{(k,b)}\mathbf{Z}']l \tag{3.12}$$

in (3.12), $\mathbf{M}$ is based on MRRE framework.

The diagnostic directions l_{max} are obtained through finding the eigenvectors corresponding to the largest absolute eigenvalues of matrices in (3.11) and (3.12) respectively.

4. Assessing Influence on the Section of MRRE Parameter

The possible influential observations in the data may have serious impacts on the estimation of the MRRE parameter. The estimation of MRRE parameter k is determined from data by imposing some criteria. In this section, using local influence analysis, we give a method to study the detection of this kind of data. The selection criteria is given in (2.4) and the perturbation scheme in (3.3).

If we let $C_{(k,b)}(\omega)$, $SSR_{(k,b)}(\omega)$ and $\mathbf{H}_{(k,b)}(\omega)$ denote the perturbed versions of $C_{(k,b)}$, $SSR_{(k,b)}$ and $\mathbf{H}_{(k,b)}$ respectively, then (2.4) becomes

$$C_{(k,b)}(\omega) = \frac{SSR_{(k,b)}(\omega)}{s^2} + 2tr(\mathbf{H}_{(k,b)}(\omega)) - (n-2). \tag{4.1}$$

Let $\hat{k}(\omega)$ be the estimator of k obtained by minimizing (4.1), then the main diagnostic direction of local influence for $\hat{k}$, denoted by $l_{max}(\hat{k})$, has the form (Lawrance, 1988; Thomas and Cook, 1990; Shi, 1997; Shi and Wang, 1999):

$$l_{max}(\hat{k}) \propto \frac{\partial \hat{k}(\omega)}{\partial \omega} \tag{4.2}$$

evaluated at $\omega = \omega_0$, where $\hat{k}(\omega_0) = \hat{k}$ denotes the value by minimizing (2.4). Since $C_{(k,b)}(\omega)$ achieves the local minimum at $\hat{k}(\omega)$, we can have

$$\frac{\partial C_{(k,b)}(\omega)}{\partial k}|_{k=\hat{k}(\omega)} = 0. \tag{4.3}$$

Differentiating (4.3) with respect to ω then we get

$$\frac{\partial^2 C_{(k,b)}(\omega)}{\partial k^2}\frac{\partial \hat{k}(\omega)}{\partial \omega}|_{\omega=\omega_0,k=\hat{k}} + \frac{\partial^2 C_{(k,b)}(\omega)}{\partial k \partial \omega}|_{\omega=\omega_0,k=\hat{k}} = 0.$$

This yields

$$\frac{\partial \hat{k}(\omega)}{\partial \omega}|_{\omega=\omega_0} = -\frac{\frac{\partial^2 C_{(k,b)}(\omega)}{\partial k \partial \omega}}{\frac{\partial^2 C_{(k,b)}(\omega)}{\partial k^2}}|_{\omega=\omega_0,k=\hat{k}}. \tag{4.4}$$

Under perturbation (3.3), the sum of the squares of the residual $SSR_{k,b}(\omega)$ in MRRE becomes

$$SSR_{(k,b)}(\omega) = \mathbf{Y}'\mathbf{W}\mathbf{Y} - (\mathbf{Z}'\mathbf{W}\mathbf{Y} + k\mathbf{b})'(\mathbf{Z}'\mathbf{W}\mathbf{Z} + k\mathbf{I}_p^*)^{-1}\mathbf{Z}'\mathbf{W}\mathbf{Y}.$$

Using the knowledge of the matrix theory, we can obtain

$$\frac{\partial SSR_{(k,b)}(\omega)}{\partial \omega_i}|_{\omega=\omega_0} = (e_{(k,b)}^{(i)})^2, \tag{4.5}$$

where $e_{(k,b)}^{(i)}$ is i-th element of $\mathbf{e}_{(k,b)}$.

But, we know $\mathbf{e}_{(k,b)} = \mathbf{y} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1}(\mathbf{Z}'\mathbf{y} + k\mathbf{b})$ then

$$\begin{aligned} \frac{\partial \mathbf{e}_{(k,b)}}{\partial k} &= \mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_p^*)^{-1}(\hat{\boldsymbol{\beta}}_{(k,b)} - \mathbf{b}) \\ &= \mathbf{X}(\mathbf{X}'\mathbf{X} + k\mathbf{I}_m)^{-1}(\hat{\boldsymbol{\beta}}_{1(k,b)} - \mathbf{b}_1), \end{aligned}$$

where $\hat{\boldsymbol{\beta}}_{1(k,b)}$ is the MRRE of $\boldsymbol{\beta}_1$ and $\mathbf{b}_1$ is prior information of $\mathbf{b}$. Hence, we can have

$$\frac{\partial^2 SSR_{(k,b)}(\omega)}{\partial k \partial \omega_i}|_{\omega=\omega_0,k=\hat{k}} = 2e^{(i)}_{(k,b)}\mathbf{x}_i(\mathbf{X}'\mathbf{X} + k\mathbf{I}_m)^{-1}(\hat{\boldsymbol{\beta}}_{1(k,b)} - \mathbf{b}_1) \tag{4.6}$$

where $\mathbf{x}_i$ is the i-th row of matrix $\mathbf{X}$.

A similar matrix partial differentiation for $tr(\mathbf{H}_{(k,b)}(\omega)) = tr[\mathbf{Z}(\mathbf{Z}'\mathbf{W}\mathbf{Z} + k\mathbf{I}^*_p)^{-1}\mathbf{Z}'\mathbf{W}] + tr[k\mathbf{W}^{\frac{1}{2}}\mathbf{Z}(\mathbf{Z}'\mathbf{W}\mathbf{Z} + k\mathbf{I}^*_p)^{-1}\mathbf{M}]$ gives

$$\begin{aligned}&\tfrac{\partial tr(\mathbf{H}_{(k,b)}(\omega))}{\partial \omega}|_{\omega=\omega_0}\\ &= -\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}^*_p)^{-1}(\mathbf{Z}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}^*_p)^{-1}\mathbf{Z}' + \mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}^*_p)^{-1}\mathbf{Z}'\\ &\quad +\tfrac{k}{2}\mathbf{M}'(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}^*_p)^{-1}\mathbf{Z}' - k\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}^*_p)^{-1}(\mathbf{Z}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}^*_p)^{-1}\mathbf{M},\end{aligned}$$

the second order partial differentiation of the above equation with respect to k gives

$$\begin{aligned}\tfrac{\partial^2 tr(\mathbf{H}_{(k,b)}(\omega))}{\partial k \partial \omega}|_{\omega=\omega_0,k=\hat{k}} \;=\;& \mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}^*_p)^{-2}(\mathbf{Z}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z} + \mathbf{I}^*_p)^{-1}\mathbf{Z}'\\ &+\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}^*_p)^{-1}(\mathbf{Z}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z} + \mathbf{I}^*_p)^{-2}\mathbf{Z}'\\ &-\tfrac{1}{2}\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}^*_p)^{-2}\mathbf{Z}' - \tfrac{k}{2}\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}^*_p)^{-3}\mathbf{Z}'\\ &-\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}^*_p)^{-1}(\mathbf{Z}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z} + \mathbf{I}^*_p)^{-2}\mathbf{Z}'\\ &+k\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}^*_p)^{-2}(\mathbf{Z}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z} + \mathbf{I}^*_p)^{-2}\mathbf{Z}'\\ &+k\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}^*_p)^{-1}(\mathbf{Z}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z} + \mathbf{I}^*_p)^{-3}\mathbf{Z}'.\end{aligned}$$

But, we know matrices $(\mathbf{Z}'\mathbf{Z})$, $(\mathbf{Z}'\mathbf{Z}+k\mathbf{I}^*_p)^{-1}$, $(\mathbf{Z}'\mathbf{Z}+k\mathbf{I}^*_p)^{-2}$ and $(\mathbf{Z}'\mathbf{Z}+k\mathbf{I}^*_p)^{-3}$ are symmetric, hence the above second order partial differentiation becomes

$$\begin{aligned}\tfrac{\partial^2 tr(\mathbf{H}_{(k,b)}(\omega))}{\partial k \partial \omega}|_{\omega=\omega_0,k=\hat{k}} \;=\;& \mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}^*_p)^{-2}(\mathbf{Z}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z} + \mathbf{I}^*_p)^{-1}\mathbf{Z}'\\ &-\tfrac{1}{2}\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}^*_p)^{-2}\mathbf{Z}' - \tfrac{k}{2}\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}^*_p)^{-3}\mathbf{Z}'\\ &+2k\mathbf{Z}(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}^*_p)^{-2}(\mathbf{Z}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z} + \mathbf{I}^*_p)^{-2}\mathbf{Z}'.\end{aligned}$$

Therefore, the i-th element of $l_{max}(\hat{k})$ obtains

$$\begin{aligned}l^{(i)}_{max}(\hat{k}) \propto \tfrac{\partial \hat{k}(\omega)}{\partial \omega_i} \;=\;& e^{(i)}_{(k,b)}\mathbf{x}_i(\mathbf{X}'\mathbf{X} + k\mathbf{I}_m)^{-1}(\hat{\boldsymbol{\beta}}_{1(k,b)} - \mathbf{b}_1)/s^2\\ &+\mathbf{x}_i(\mathbf{X}\mathbf{X} + k\mathbf{I}^*_p)^{-2}(\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{X} + \mathbf{I}^*_p)^{-1}\mathbf{x}'_i\\ &-\tfrac{1}{2}\mathbf{x}_i(\mathbf{X}'\mathbf{X} + k\mathbf{I}^*_p)^{-2}\mathbf{x}'_i - \tfrac{k}{2}\mathbf{x}_i(\mathbf{X}'\mathbf{X} + k\mathbf{I}^*_p)^{-3}\mathbf{x}'_i\\ &+2k\mathbf{x}_i(\mathbf{X}'\mathbf{X} + k\mathbf{I}^*_p)^{-2}(\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{X} + \mathbf{I}^*_p)^{-2}\mathbf{x}'_i\end{aligned} \tag{4.7}$$

The index plot of $l_{max}(\hat{k})$ can reveal the influential points that have high impacts on the selection of the MRRE parameter in the linear ridge type regression.

5. Examples

Example 1. Macroeconomic Impact of Foreign Direct Investment (MIFDI) Data Sun (1998) studied MIFDI in China 1979-1996. Based on his theory, the MIFDI data were collected in Sri Lanka from 1978 to 2004 to illustrate our methodologies. The data set consists four regressors (Foreign Direct Investment, Gross Domestic Product Per Capita, Exchange Rate and Interest Rate) and one response variable (Total Domestic Investment) with 27 observations. The selected variables were tested for statistical conditions: (i) Cointegration, (ii) Constant Error Variance and (iii) Multicollinearity. The test results showed that: (i) Variables are cointegrated with a same cointegration coefficient I(1) at 1% level of significance, (ii) The estimated Durbin-Watson value for the linear model is 2.0131 so, satisfied the constant error variance condition and (iii) The scaled condition number of this data set is 31,244, this large value suggests the presence of an unusually high level of severe multicollinearity among the regressors (the proposed cutoff is 30; see Belsley et al., 1980). The MRRE biasing parameter is estimated for this data set k=0.0131.

We analyze the constant variance perturbation. The index plots of $l_{max}^{(1)}(\hat{\boldsymbol{\beta}}_{(k,b)})$ and $l_{max}^{(2)}(\hat{\boldsymbol{\beta}}_{(k,b)})$ are shown in Figure 1 and Figure 2 respectively. From these two index plots the most five influential cases are (3, 2, 15, 23, 4) and (3, 23, 15, 2, 14) using OLSE and MRRE framework respectively. In both influential measures the detected influential cases approximately same, but only the order of influence magnitude is changed.

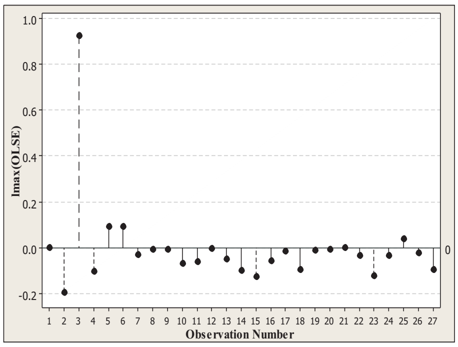

Figure 1: Index plot of $l_{max}^{(1)}(\hat{\boldsymbol{\beta}}_{(k,b)})$

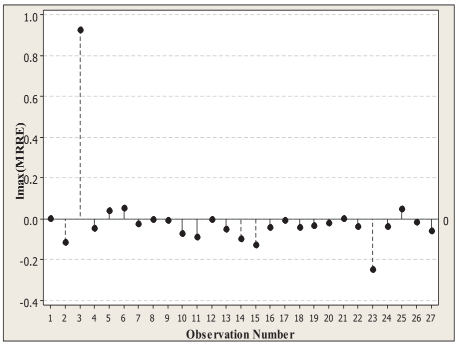

Figure 2: Index plot of $l_{max}^{(2)}(\hat{\boldsymbol{\beta}}_{(k,b)})$

Next, we consider the perturbation of individual explanatory variables. The maximum values of $l_{max}^{(2)}(\hat{\boldsymbol{\beta}}_{(k,b)})$ for separately perturbing explanatory variables X_i, $i = 1, \cdots, 4$ are 0.148787, 1.72859, 0.697375 and 0.918912 respectively using MRRE framework. Hence local change caused by perturbing X_2, X_3 and X_4 are largest among the others on MRRE. The most five influential cases when X_2, X_3 and X_4 are perturbed (3, 27, 18, 2, 4), (3, 27, 23, 14, 15) and (3, 27, 14, 22, 15) respectively.

Finally, we study the influence on the selection of MRRE biasing parameter k. An index plot of $l_{max}(\hat{k})$ is given in Figure 3. From this index plot the most five influential cases are (3, 23, 4, 27, 2) in this order.

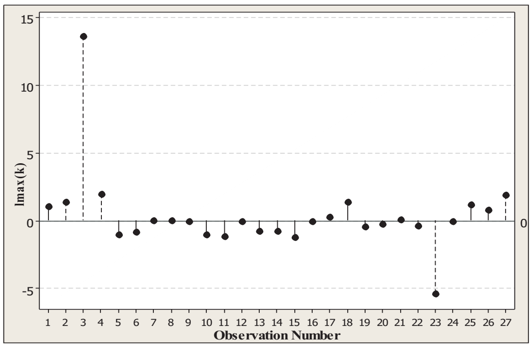

Figure 3: Index plot of $l_{max}(\hat{k})$.

For verifying these results, we contribute Table 1 of modified ridge regression estimates for the full data and the data without some influential cases detected by local influence methods. In this table, the parenthesis value indicates the percentage of change in the parameter value. The results reveals that case 3 is the most influential case while case 14 is the seventh influential point among the detected cases. At the same time parameter $\hat{\boldsymbol{\beta}}_{(k,b)0}$ has big changed and $\hat{\boldsymbol{\beta}}_{(k,b)2}$ has small changed among the others when influential cases are deleted. It is also clear from this table, that omission of single influential cases 3, 23, 15, 27, 2, 4 and 14 contribute the substantial change in the MRRE. Among all of these 3, 23, 27, 2 and 4 have a remarkable influence while cases 15 and 14 have a little influence.

Example 2. Longley Data
The second data set is Longley (1967) to explain the influential observations on the MRRE. The scaled condition number of this data set is 43,275 (see Walker and Birch, 1988). This large value suggests the presence severe multicollinearity among regressors. Cook (1977) used this data to identify the influential observations in OLSE using Cook's D_i and found that cases 5, 16, 4, 10, and 15 (in this order) were the most influential cases. Walker and Birch (1988) analyzed the same data to detect anomalous cases in ridge regression using global influence method. They observed that cases 16, 10, 4, 15 and 5 (in this order) were most influential observations. Shi and Wang (1999) also analyzed the same data to

Table 1: The most seven influence observations according to local influence analysis in MRRE

$\hat{\boldsymbol{\beta}}_{(k,b)}$	Full Data	Case Deleted						
		(3)	(23)	(15)	(27)	(2)	(4)	(14)
$\hat{\boldsymbol{\beta}}_{(k,b)0}$	0.3474	-0.1449	0.4620	0.3950	0.3518	0.2769	0.2985	0.3892
		(141.7%)	(33.0%)	(13.7%)	(1.3%)	(20.3%)	(14.1%)	(12.0%)
$\hat{\boldsymbol{\beta}}_{(k,b)1}$	0.0933	0.0697	0.0931	0.0962	0.0876	0.1049	0.0968	0.0923
		(25.3%)	(0.2%)	(3.2%)	(6.1%)	(12.5%)	(3.8%)	(1.0%)
$\hat{\boldsymbol{\beta}}_{(k,b)2}$	1.1466	1.2238	1.1267	1.1468	1.1703	1.1188	1.1515	1.1534
		(6.7%)	(1.7%)	(0%)	(2.1%)	(2.4%)	(0.4%)	(0.6%)
$\hat{\boldsymbol{\beta}}_{(k,b)3}$	-0.4588	-0.4769	-0.4227	-0.4828	-0.5459	-0.3868	-0.4809	-0.4929
		(4.0%)	(7.9%)	(5.2%)	(19.0%)	(15.7%)	(4.8%)	(7.4%)
$\hat{\boldsymbol{\beta}}_{(k,b)4}$	0.2231	0.2049	0.2047	0.2328	0.2722	0.2121	0.2434	0.2365
		(6.2%)	(8.2%)	(4.3%)	(22.0%)	(4.9%)	(9.1%)	(6.0%)

detect influential cases on the ridge regression estimator using local influence method. They detected cases 10, 4, 15, 16, and 1 (in this order) were most anomalous observations. Jahufer and Chen (2009) also used the same data to identify influential cases in modified ridge regression estimator using global influence method and they identified 16, 4, 1, 10 and 15 (in this order) were most influential cases.

The affects of MRRE biasing parameter k on the influence of observations are also studied by plotting $C_{(k,b)}$, Cook's D_i and $DFFITS_i$ against k. The value of k that minimizes (2.4) for this data set is 0.0002.

We study the influence of observations on constant variance perturbation. Analyzing the index plots of $l^{(1)}_{max}(\hat{\boldsymbol{\beta}}_{(k,b)})$ for $k = 0$ and $l^{(2)}_{max}(\hat{\boldsymbol{\beta}}_{(k,b)})$ for $k = 0.0002$, we found that cases (4, 1, 15, 16, 10) and (10, 4, 15, 16, 1) are the most five influential observations using OLSE and MRRE framework in this order respectively. The influential cases detected by these two methods are same but, only the order of magnitude is changed.

we consider the perturbation of individual explanatory variables. The maximum values of $l^{(2)}_{max}(\hat{\boldsymbol{\beta}}_{(k,b)})$ for separately perturbing explanatory variables X_i, $i = 1, \cdots, 6$ are 0.6376, 8.1347, 0.3140, 0.0494, 2.0168 and 6.6122 respectively. Hence, local change caused by perturbing X_2, X_5 and X_6 are the largest among the others on MRRE. The most five influential cases when X_2, X_5 and X_6 are perturbed (10, 4, 15, 5, 1), (10, 4, 15, 6, 1) and (5, 15, 4, 7, 11) respectively.

Finally we estimate the $l_{max}(\hat{k})$ values using (4.7). From the index plot of $l_{max}(\hat{k})$ cases 10, 16, 4, 15 and 1 in this order are the most five influential observations on MRRE parameter.

The influential observations detected by this study and the previous studies Cook (1977), Walker and Birch (1988), Shi and Wang (1999) and Jahufer and Chen (2009) are approximately same in Longley data set but, only the order of magnitude is changed.

6. Comments

In this paper, we have studied several local influence diagnostic measures that seem practical and can play a considerable part in MRRE data analysis. The local influence measures proposed focus on various outcomes of MRRE. Few of the measures introduced focus on perturbing the constant variance, other on perturbing the regressor variables and still on the selection of MRRE biasing parameter. All the proposed measures are the function of residuals, leverage points and MRRE coefficients.

Although no conventional cutoff points are introduced or developed for the MRRE local influence diagnostic quantities, it seems that index plot is an optimistic and conventional procedure to disclose influential cases. It is a bottleneck for cutoff values for the influence method. Also, the issue of accommodating influential cases has not been studied. These are additional active issues for future research study.

Acknowledgements

Much of this paper is taken from the first author's Ph.D. research work, under the supervision of the second author at the Xiamen university of China.

We would like to thank an anonymous referee for his valuable comments and suggestions. This work was supported by a research grant No: 09AZD045 by National Social Science Foundation of China.

References

Belsley, D.A., Kuh, E. and Welsch, R.E. (1980). *Regression Diagnostics: Identifying Influence Data and Source of Collinearity.* Wiley, New York.

Cook, R.D. (1977). Detection of influential observations in linear regression. *Technometrics* **19**, 15-18.

Cook, R.D. (1986). Assessment of local influence. *Journal of Royal Statistical Association B* **48**, 133-169.

Groß, J. (2003). Restricted ridge estimation. *Statistics & Probability Letters* **65**, 57-64.

Hoerl, A.E. and Kennard, R.W. (1970a). Ridge regression: biased estimation of for nonorthogonal problems. *Technometrics* **12**, 55-67.

Hoerl, A.E. and Kennard, R.W. (1976). Ridge regression: iterative estimation of the biasing parameter. *Communications in Statistics - Theory and Methods* A5, 77-88.

Jahufer, A. and Chen, J. (2009). Assessing global influential observations in modified ridge regression. *Statistics & Probability Letters* **79**, 513-518.

Lawrance, A.J. (1988). Regression transformation diagnostics using local influence. *Journal of American Statistical Association* **83**, 1067-1072.

Kaçiranlar, S., Sakallioğlu, S. and Akdeniz, F. (1998). Mean squared error comparisons of the modified ridge regression estimator and the restricted ridge regression estimator. *Communications in Statistics - Theory and Methods* **27**, 131-138.

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

McDonald, G.C. and Galarneau, D.I. (1975). A Monte Carlo evaluation of some ridge-type estimators. *Journal of American Statistical Association* **70**, 407-416.

Mallows, C.L. (1973). Some comments on C_p. *Technometrics* **15**, 661-675.

Marquardt, D.W. (1970). Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics* **12**, 591-612.

Pliskin, J.L. (1987). A ridge type estimator and good prior means. *Communications in Statistics - Theory and Methods* **16**, 3429-3437.

Shi, L. (1997). Local influence in principal component analysis. *Biometrika* **84**, 175-186.

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

Sun, H. (1988). *Macroeconomic Impact of Direct Foreign Investment in China: 1979-1996*. Blackwell Publishers Ltd, UK.

Swindel, B.F. (1976). Good ridge estimators based on prior information. *Communications in Statistics - Theory and Methods* A5, 1065-1075.

Thomas, W. and Cook, R.D. (1990). Assessing the influence on predictions from generalized linear model. *Technometrics* **32**, 59-65.

Trenkler, G. (1988). Some remarks on a ridge-type-estimator and good prior means. *Communications in Statistics - Theory and Methods* **17**, 4251-4256.

Wahba, G., Golub, G.H. and Heath, C.G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* **21**, 215-223.

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

Wijekoon, P. (1998). A comparison of the restricted ridge regression and the modified ridge regression estimators. *Ceylon Journal of Science: Physical Sciences* **5**, 75-79.

Received February 27, 2010; accepted July 31, 2010.

Aboobacker Jahufer
Department of Mathematical Sciences
Faculty of Applied Sciences
South Eastern University of Sri Lanka, Sri Lanka
jahufer@yahoo.com

Jianbao Chen
Department of Planning and Statistics
School of Economics, Xiamen University
Xiamen 361005, P. R. China
Jbjy2006@126.com