

Identifying Unusual Observations in Ridge Regression Linear Model Using Box-Cox Power Transformation Technique

Aboobacker Jahufer

Department of Mathematical Sciences, Faculty of Applied Sciences, South Eastern University of Sri Lanka, Sammanthurai, Sri Lanka Email: jahufer@yahoo.com

Received November 4, 2013; revised December 4, 2013; accepted December 11, 2013

Copyright © 2014 Aboobacker Jahufer. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. In accordance of the Creative Commons Attribution License all Copyrights © 2014 are reserved for SCIRP and the owner of the intellectual property Aboobacker Jahufer. All Copyright © 2014 are guarded by law and by SCIRP as a guardian.

ABSTRACT

The use of [1] Box-Cox power transformation in regression analysis is now common; in the last two decades there has been emphasis on diagnostics methods for Box-Cox power transformation, much of which has involved deletion of influential data cases. The pioneer work of [2] studied local influence on constant variance perturbation in the Box-Cox unbiased regression linear mode. Tsai and Wu [3] analyzed local influence method of [2] to assess the effect of the case-weights perturbation on the transformation-power estimator in the Box-Cox unbiased regression linear model. Many authors noted that the influential observations on the biased estimators are different from the unbiased estimators. In this paper I describe a diagnostic method for assessing the local influence on the constant variance perturbation on the transformation in the Box-Cox biased ridge regression linear model. Two real macroeconomic data sets are used to illustrate the methodologies.

KEYWORDS

Box-Cox Transformation; Ridge Regression; Constant Variance Perturbation; Local Influence; Influential Observations

1. Introduction

Deletion diagnostics for assessing the influential cases on the power transformation parameter estimator in the Box-Cox linear unbiased regression model has been intensively studied in the last two and half decades (see [4-6]). Rather than deleting the influential case, [7] proposed a general method for assessing the local influence of minor perturbations of a statistical model. Lawrance [2] adapted Cook's approach to obtain a diagnostic that can be used to examine the local changes of the transformation-parameter estimator caused by small perturbations on a constant-variance assumption. Tsai and Wu [8] analyzed the case-deletion model directly and obtain a more accurate and reliable transformation power estimator in weighted regression model. Also, Tsai and Wu [3] applied a case-weights perturbation scheme to obtain an alternative local influence diagnostic that takes into account the perturbation effects of the Jacobian.

In the literature, many authors noted that the influential observations on ridge type estimators are different from the corresponding least squares estimate (see [9]). The aim of this paper is to apply local influence of minor perturbation of constant variance to biased ridge regression Box-Cox power transformation model. The structure of this paper is as follows. Section 2 establishes transformation for ridge regression model for dealing with the local influence on the power transformation estimator. Section 3 gives calculation of the maximum local influence for ridge estimator. In Section 4 two real macroeconomic data sets are used to illustrate the methodologies. In the last section conclusions are given.

2. Transformation for Ridge Regression Model

A parametric family of transformations takes a column of a response vector \mathbf{y} into $\mathbf{y}^{(\lambda)}$, where λ is a scalar transformation parameter. It is assumed to be a value λ_0 of λ , ensuring that $\mathbf{y}^{(\lambda_0)}$ follows a standard regression model, having the design matrix \mathbf{X} . Thus, it is assumed that

$$\mathbf{y}^{(\lambda_0)} = \mathbf{X}\boldsymbol{\beta}_R + \boldsymbol{\varepsilon}_R \tag{1}$$

where $\hat{\boldsymbol{\beta}}_{R} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}$ is ridge regression estimator (RRE) proposed by Hoerl and Kennard (see. [10,11]), \boldsymbol{X} is a known full rank an $n \times p$ matrix and $\boldsymbol{\varepsilon}_{R}$ an $n \times 1$ random error vector of RRE, it has a multivariate normal distribution with $E(\boldsymbol{\varepsilon}_{R}) = 0$ and $\operatorname{Var}(\boldsymbol{\varepsilon}_{R}) = \sigma^{2}\boldsymbol{I}$ here \boldsymbol{I} is an identity matrix. The Equation (1) was originally proposed by Box-Cox [1]. If the Jacobian of the transformation denoted from \boldsymbol{y} to $\boldsymbol{y}^{(\lambda)}$ by

 $J(\lambda)$, defined as $z^{(\lambda)} = y^{(\lambda)} / J(\lambda)^{\frac{1}{n}}$ and $z^{(\lambda)}$ is the natural scaling of $y^{(\lambda)}$ suggested by the likelihood.

Local Influence Approach for Power Transformation on RRE

For assessing the local influence on the power transformation estimator Lawrance [2] has obtained a diagnostic by perturbing the constant variance assumption.

First, here it is to be assumed that the variance of under perturbation is

$$\operatorname{Var}(\boldsymbol{\varepsilon}_{R}(\boldsymbol{\omega})) = \sigma^{2} \boldsymbol{W}^{-1}, \qquad (2)$$

where $W = \text{diag}(\omega)$ is a diagonal matrix with diagonal elements of $\omega^{T} = (\omega_{1}, \dots, \omega_{n})$. Let $\omega = \omega_{0} + a\ell$, where $\omega_{0} = 1$, ω denotes the an $n \times 1$ vector of case-weights for the regression, $a \in R^{1}$ and ℓ is a fixed nonzero vector of unit length in R^{q} .

The distribution function for RRE linear model random error is

$$f\left(Y;\beta_{R},\sigma^{2},\lambda\right) = \frac{1}{\sqrt{2\pi\sigma^{2}}} \exp\left[-\frac{1}{2\sigma^{2}}\varepsilon_{R}^{\mathrm{T}}\varepsilon_{R}\right] J\left(\lambda\right),\tag{3}$$

where $J(\lambda)$ Jacobian of the transformation.

The perturbation of the distribution in Equation (3) becomes

$$f_{\left(Y;\beta_{R},\sigma^{2},\lambda\right)}\left(\omega\right) = \frac{1}{\sqrt{2\pi\sigma^{2}}} |\mathbf{W}|^{\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^{2}} \left(\mathbf{Y}^{(\lambda)} - \mathbf{X}\boldsymbol{\beta}_{R}\right)^{\mathrm{T}} \mathbf{W}\left(\mathbf{Y}^{(\lambda)} - \mathbf{X}\boldsymbol{\beta}_{R}\right)\right] J\left(\lambda\right)$$
(4)

The corresponding log-likelihood function for the untransformed observation y is

$$L(Y;\beta_{R},\sigma^{2},\lambda) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^{2}) + \frac{1}{2}\sum_{i=1}^{n}\log(\omega_{i}) - \frac{1}{2\sigma^{2}}\left[\mathbf{y}^{(\lambda)} - \mathbf{X}\boldsymbol{\beta}_{R}\right]^{\mathrm{T}}\mathbf{W}\left[\mathbf{y}^{(\lambda)} - \mathbf{X}\boldsymbol{\beta}_{R}\right] + \log(J(\lambda)).$$

$$(5)$$

Thus the profile likelihood function for λ is obtained when maximizing RRE β_R and σ^2 for given data set y.

The maximum likelihood estimator (MLE) of $\boldsymbol{\beta}_R$ is $\hat{\boldsymbol{\beta}}_R = (\boldsymbol{X}^T \boldsymbol{X} + k\boldsymbol{I})^{-1} \boldsymbol{X}^T \boldsymbol{y}$. Also, the MLE of σ^2 is $\hat{\sigma}^2$ this has to be estimated from the Equation (5) as

$$\frac{\partial L(Y; \boldsymbol{\beta}_{R}, \sigma^{2}, \lambda)}{\partial \sigma^{2}} = -\frac{n}{2\sigma^{2}} + \frac{1}{2\sigma^{4}} \sum_{i=1}^{n} \omega_{i} \left(\boldsymbol{y}_{i}^{\lambda} - \boldsymbol{x}_{i}^{\mathrm{T}} \boldsymbol{\beta}_{R} \right)^{2} = 0$$
$$\hat{\sigma}^{2} \Big|_{\omega=1} = \frac{1}{n} \left[\boldsymbol{y}^{(\lambda)} - \boldsymbol{X} \hat{\boldsymbol{\beta}}_{R} \right]^{\mathrm{T}} \left[\boldsymbol{y}^{(\lambda)} - \boldsymbol{X} \hat{\boldsymbol{\beta}}_{R} \right].$$

But,

$$\boldsymbol{X}\boldsymbol{\hat{\beta}}_{R} = \boldsymbol{X}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} + k\boldsymbol{I}\right)^{-1}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}^{(\lambda)}\right) = \boldsymbol{H}_{R}\boldsymbol{y}^{(\lambda)}$$

where the $\boldsymbol{H}_{R} = \boldsymbol{X} \left(\boldsymbol{X}^{\mathrm{T}} \boldsymbol{X} + k \boldsymbol{I} \right)^{-1} \boldsymbol{X}^{\mathrm{T}}$ is "hat" matrix of RRE. Therefore, the estimator of $\hat{\sigma}^{2}$ is

$$\hat{\sigma}^{2} = \frac{1}{n} \left[\mathbf{y}^{(\lambda)} - \mathbf{H}_{R} \mathbf{y}^{(\lambda)} \right]^{\mathrm{T}} \left[\mathbf{y}^{(\lambda)} - \mathbf{H}_{R} \mathbf{y}^{(\lambda)} \right] = \frac{1}{n} \mathbf{y}^{(\lambda)\mathrm{T}} \left[\mathbf{I} - \mathbf{H}_{R} \right]^{\mathrm{T}} \left[\mathbf{I} - \mathbf{H}_{R} \right] \mathbf{y}^{(\lambda)}$$

From the above results the likelihood function for transformed observation y with perturbation can be written as

$$L(\lambda_{\omega}) = -\frac{n}{2}\log(2\pi) + \frac{n}{2}\log(n) - \frac{n}{2}\log\left[z^{(\lambda)T}\left[I - H_{R}(\omega)\right]^{T}\left[I - H_{R}(\omega)\right]z^{(\lambda)}\right] + \frac{1}{2}\sum_{i=1}^{n}\log(\omega_{i}) - \frac{n}{2} + \log(J(\lambda)),$$

$$= -\frac{n}{2}\log(2\pi) + \frac{n}{2}\log(n) - \frac{n}{2}\log\left[z^{(\lambda)T}B(\omega)z^{(\lambda)}\right] + \frac{1}{2}\sum_{i=1}^{n}\log(\omega_{i}) - \frac{n}{2} + \log(J(\lambda))$$

$$= -\frac{n}{2}\log(2\pi) + \frac{n}{2}\log(n) - \frac{n}{2}\log\left[A(\lambda, \omega)\right] + \frac{1}{2}\sum_{i=1}^{n}\log(\omega_{i}) - \frac{n}{2} + \log(J(\lambda))$$

(6)

where $\boldsymbol{B}(\omega) = [\boldsymbol{I} - \boldsymbol{H}_{R}(\omega)]^{\mathrm{T}} [\boldsymbol{I} - \boldsymbol{H}_{R}(\omega)]$ is symmetric and $\boldsymbol{A}(\lambda, \omega) = \boldsymbol{z}^{(\lambda)\mathrm{T}} \boldsymbol{B}(\omega) \boldsymbol{z}^{(\lambda)}$.

The resulting MLE $\hat{\lambda}_{\omega}$ of λ can be found by minimizing Equation (6). Furthermore, the estimator $\hat{\lambda}_{\omega}$ can be regarded as a surface with Euclidean coordinates ω . A curve over this surface is mapped from a straight-line path that passes through the point of the null perturbation. The direction and location of this path are specified by $\omega = 1 + al$ passing through the null point, where the quantity "a" measures distance along the line and $l = (l_1, \dots, l_n)$ are the direction cosines of Lawrance's local influence diagnostic.

The partial differentiation of Equation (6) with respect to λ is then the result will be arrived as,

$$\frac{\partial L(\lambda_{\omega})}{\partial \lambda} = \frac{\partial A(\lambda, \omega)}{\partial \lambda} = \frac{\partial z^{(\lambda)T} B(\omega) z^{(\lambda)}}{\partial \lambda} = \left(\frac{\partial z^{(\lambda)}}{\partial \lambda}\right)^{T} B(\omega) z^{(\lambda)} + \left(\frac{\partial z^{(\lambda)}}{\partial \lambda}\right)^{T} B^{T}(\omega) z^{(\lambda)} = 0$$

But it is known that $B(\omega) = B^{T}(\omega)$, therefore, the above equation becomes

$$\left[\boldsymbol{z}^{\prime(\lambda)}\right]^{\mathrm{T}}\boldsymbol{B}(\boldsymbol{\omega})\boldsymbol{z}^{(\lambda)} = 0 \tag{7}$$

where $\left(\frac{\partial z^{(\lambda)}}{\partial \lambda}\right) = z'^{(\lambda)}$ is first derivative of $z^{(\lambda)}$. The transpose of Equation (7) can be written as

$$\left[\boldsymbol{z}^{(\lambda)}\right]^{\mathrm{T}}\boldsymbol{B}(\boldsymbol{\omega})\boldsymbol{z}^{\prime(\lambda)} = 0.$$
(8)

The local influence diagnostic is the slope of the curve on the $\hat{\lambda}_{\omega}$ surface at the point of null perturbation, at a = 0. If it is 0 small perturbations have no effect in the path points to the data cases being perturbed; the weighted set of cases that are most sensitive to local perturbations are thus specified by the direction that makes the path slope,

$$\frac{\partial \lambda(a)}{\partial a}, \hat{\lambda}(a) = \hat{\lambda}_{1+al}, \qquad (9)$$

the greatest at the null point a = 0. This is the key idea in the local-influence approach. It is not principally the value of the slope, but the direction of maximum slope that is important and that forms the main diagnostic.

This description is the basis of Cook's presentation when just one parameter is being considered; there is then no need to use a likelihood-displacement measure of the distance between $\hat{\lambda}_{\omega}$ and $\hat{\lambda}$. This also removes the need to consider curvature of the likelihood displacement, and avoids a loss of sign in connection with $\frac{\partial \hat{\lambda}(a)}{\partial a}$

at
$$a = 0$$
.

3. Calculation of the Direction of Maximum Local Influence in RRE

In this section it is tried to develop a method which is the direction that maximizes the slope of $\hat{\lambda}(a)$ at the origin of no perturbation is obtained.

Consider $\frac{\partial \hat{\lambda}(a)}{\partial a}$ for a = 0 in an arbitrary direction *l*. Let W in Equation (2) have diagonal entries

 $\omega = (1+al)$ and be denoted by W(a) for fixed l; denote $B(\omega)$ with $\omega = (1+al)$ by B(a). Let L denote the diagonal matrix with diagonal entries $l = (l_1, \dots, l_n)$. The maximum likelihood estimator $\hat{\lambda}(a)$ satis-

fies Equation (8) where each term is a function of a. Hence, differentiate the Equation (8) with respect to $\hat{\lambda}(a)$ when dealing with the variables $\mathbf{Z}^{(\lambda)}$ and $\mathbf{Z}^{(\lambda)}$.

$$\frac{\partial \left[\boldsymbol{z}^{(\lambda)} \right]^{\mathrm{T}} \boldsymbol{B}(a) \boldsymbol{z}^{\prime(\lambda)}}{\partial a} = \frac{\partial \hat{\lambda}(a)}{\partial a} \left[\boldsymbol{z}^{\prime(\lambda)} \right]^{\mathrm{T}} \boldsymbol{B}(a) \boldsymbol{z}^{\prime(\lambda)} + \left[\boldsymbol{z}^{(\lambda)} \right]^{\mathrm{T}} \frac{\partial \boldsymbol{B}(a)}{\partial a} \boldsymbol{z}^{\prime(\lambda)} + \left[\boldsymbol{z}^{\prime(\lambda)} \right] \boldsymbol{B}(a) \boldsymbol{z}^{\prime\prime(\lambda)} \frac{\partial \hat{\lambda}(a)}{\partial a} = 0$$
$$\frac{\partial \hat{\lambda}(a)}{\partial a} = -\frac{\left[\boldsymbol{z}^{(\lambda)} \right]^{\mathrm{T}} \frac{\partial \boldsymbol{B}(a)}{\partial a} \boldsymbol{z}^{\prime(\lambda)}}{\left[\left[\boldsymbol{z}^{\prime(\lambda)} \right]^{\mathrm{T}} \boldsymbol{B}(a) \boldsymbol{z}^{\prime\prime(\lambda)} + \boldsymbol{z}^{\prime(\lambda)} \boldsymbol{B}(a) \boldsymbol{z}^{\prime\prime(\lambda)} \right]} \tag{10}$$

where $z''^{(\lambda)}$ is second order derivative of $z^{(\lambda)}$ with respect to λ , and can be called the second constructed variable; all $z^{(\lambda)}$ terms in Equation (10) are used at $\lambda = \hat{\lambda}(a)$.

Let consider $\boldsymbol{B}(a) = [\boldsymbol{I} - \boldsymbol{H}_{R}(a)]^{\mathrm{T}} [\boldsymbol{I} - \boldsymbol{H}_{R}(a)]$; where $\boldsymbol{H}_{R} = \boldsymbol{X} (\boldsymbol{X}^{\mathrm{T}} \boldsymbol{X} + k\boldsymbol{I})^{-1} \boldsymbol{X}^{\mathrm{T}}$. Therefore, the perturbation matrix $\boldsymbol{B}(a)$ becomes

$$B(a) = [I - H_{R}(a)]^{\mathrm{T}} [I - H_{R}(a)] = I - 2H_{R}(a) + H_{R}(a)H_{R}(a)$$

$$= I - 2W^{\frac{1}{2}}(a)X[X^{\mathrm{T}}W(a)X + kI]^{-1}X^{\mathrm{T}}W^{\frac{1}{2}}(a)$$

$$+ W^{\frac{1}{2}}(a)X[X^{\mathrm{T}}W(a)X + kI]^{-1}X^{\mathrm{T}}W(a)X[X^{\mathrm{T}}W(a)X + kI]^{-1}X^{\mathrm{T}}W^{\frac{1}{2}}(a).$$

Now, partial differentiate B(a) with respect to a, then it gives

$$\frac{\partial \boldsymbol{B}(a)}{\partial a}\Big|_{a=0} = -\boldsymbol{L}\boldsymbol{A} + 3\boldsymbol{A}\boldsymbol{L}\boldsymbol{A} - \boldsymbol{A}\boldsymbol{L} + \frac{1}{2}\boldsymbol{L}\boldsymbol{A}^2 - \boldsymbol{A}\boldsymbol{L}\boldsymbol{A}^2 - \boldsymbol{A}^2\boldsymbol{L}\boldsymbol{A} + \frac{1}{2}\boldsymbol{A}^2\boldsymbol{L},$$

where $\boldsymbol{A} = \boldsymbol{X} \left(\boldsymbol{X}^{\mathrm{T}} \boldsymbol{X} + k \boldsymbol{I} \right)^{-1} \boldsymbol{X}^{\mathrm{T}}$.

Let $\frac{\partial \boldsymbol{B}(a)}{\partial a}\Big|_{a=0} = \boldsymbol{M}$ is a square matrix hence, the Equation (10) becomes

$$\frac{\partial \hat{\lambda}(a)}{\partial a} \bigg|_{a=0} = \frac{-\left[z^{(\lambda)}\right]^{\mathrm{T}} M z^{\prime(\lambda)}}{\left[\left[z^{\prime(\lambda)}\right]^{\mathrm{T}} B(a) z^{\prime(\lambda)} + z^{\prime(\lambda)} B(a) z^{\prime\prime(\lambda)}\right]}$$
(11)

The direction l_{max} of maximum slope is now determined by the Equation (11) and it gives the following results.

The matrix M can be written as M = ELF, where E and F are symmetric matrices. Therefore,

$$-\left[z^{(\lambda)}\right]^{\mathrm{T}} M z'^{(\lambda)} = \left[z^{(\lambda)}\right]^{\mathrm{T}} E L F z'^{(\lambda)} = \left\{\left[E z^{(\lambda)}\right]^{\mathrm{T}}\right\} L \left\{F z'^{(\lambda)}\right\}.$$

If the terms $\begin{bmatrix} Ez^{(\lambda)} \end{bmatrix}$ and $\begin{bmatrix} Fz^{\prime(\lambda)} \end{bmatrix}$ denoted by t and t' containing (t_1, \dots, t_n) and (t'_1, \dots, t'_n) , respectively. Therefore, the Equation (11) becomes

$$\frac{\partial \hat{\lambda}(a)}{\partial a} \bigg|_{a=0} = \frac{-\sum_{j=1}^{n} (t_j t'_j) l_j}{\left[\left[z'^{(\lambda)} \right]^{\mathrm{T}} \boldsymbol{B}(a) z'^{(\lambda)} + z'^{(\lambda)} \boldsymbol{B}(a) z''^{(\lambda)} \right]}$$
(12)

Therefore, the direction of l_{\max} maximum slope is now easily determined from the Equation (12) that is $l_{\max}(\lambda) = \frac{a}{\|a\|}$ where $a = \left(Ez^{(\lambda)}Fz'^{(\lambda)}\right)$ and yields the results for *i*-th l_{\max_i} is

$$U_{\max_{i}} = t_{i}t_{i}^{\prime} / \left[\sum_{j=1}^{n} (t_{j}t_{j}^{\prime})^{2}\right]^{\frac{1}{2}}, i = 1, \cdots, n.$$
(13)

Finally, consider the slope of $\hat{\lambda}(a)$ at a = 0 when just the *i*-th variance is perturbed; denoting by λ'_i , gives

$$\hat{\lambda}_{i}^{\prime} = -t_{i}t_{i}^{\prime} / \left[\left[z^{\prime(\lambda)} \right]^{\mathrm{T}} \boldsymbol{B}(a) z^{\prime(\lambda)} + z^{\prime(\lambda)} \boldsymbol{B}(a) z^{\prime\prime(\lambda)} \right], \quad i = 1, \cdots, n.$$
(14)

The result in Equation (14) is the local-influence version of computing $\hat{\lambda}$ after deleting the *i*-th data case, an operation analogous to the global perturbation $\omega_i = 0$ and $\omega_i = 1 (j \neq i)$; the global perturbation is itself algebraically intractable without approximation (see [5]).

4. Examples

4.1. Macroeconomic Impact of Foreign Direct Investment (MIFDI) in Sri Lanka Data Set

Sun [12] studied MIFDI in China 1979-1996. Based on his theory, the MIFDI data were collected in Sri Lanka form 1978 to 2004 to illustrate the methodologies derived in this paper. The data set consists four regressors (Foreign Direct Investment (FDI), Gross Domestic Product Per Capita (GDPPC), Exchange Rate (ER) and Interest Rate (IR)) and one response variable (Total Domestic Investment (TDI)) with 27 observations. The selected variables were tested for statistical conditions: 1) Cointegration, 2) Constant Error Variance and 3) Multicollinearity. The test results showed that: 1) Variables are cointegrated with a same cointegration coefficient I(1) at 1% level of significance, 2) The estimated Durbin-Watson value for the linear model is 2.0131 so, satisfied the constant error variance condition and 3) The scaled condition number of this data set is 31244, this large value suggests the presence of an unusually high level of multicollinearity among the regressors (the proposed cutoff is 30; see [9]). Hence, RRE is more preferable than ordinary least squares estimator to fit model for this data set.

The transformation parameter λ is estimated for this data set using the Box-Cox transformation model (necessary formulas for its implementations are given in the **Appendix**) is $\hat{\lambda} = 0.236$. The R^2 after the scalepreserving transformation is 96.6%, the standard deviation $\hat{\sigma}$ is 1.23374, and there is hardly any interaction; these are considerable improvements over the original values of 95.9% and 12665.3 respectively, and a large interaction. The REE biasing parameter is estimated for this data set that is k = 0.0063. The l_{max} values are estimated using the Equation (13), the values and corresponding index plot are given in below **Table 1** and **Figure 1**, respectively.

From **Table 1** and **Figure 1**, it can be observed that the most five influential cases are 1, 3, 23, 14, and 20 in this order in the Box-Cox transformation analysis for RRE in MIFDI data.

4.2. Longley Data

The second data set is [13] to explain the influential observations on the Liu estimator. The scaled condition number of this data set is 43,275 (see [14]). This large value suggests the presence severe multicollinearity among regressors. Cook [15] used this data to identify the influential observations in ordinary least squares estimator 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 [14] 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 [16] also analyzed the same data to detect influential cases on the ridge regression estimator using

Obs. No.	$l_{\rm max}$	Obs. No.	$l_{\rm max}$	Obs. No.	$l_{ m max}$
1	0.950940	10	-0.076263	19	-0.180404
2	0.067364	11	-0.066919	20	-0.182687
3	0.224807	12	-0.081785	21	-0.111494
4	-0.154101	13	-0.107358	22	-0.143976
5	-0.047689	14	-0.196588	23	-0.216221
6	-0.039656	15	-0.178989	24	-0.101125
7	-0.090456	16	-0.092639	25	0.108893
8	-0.093770	17	-0.069390	26	0.028201
9	-0.083477	18	-0.023769	27	-0.072669

Table 1. l_{max} in RRE using Box-Cox transformation.

local influence method. They detected that cases 10, 4, 15, 16, and 1 (in this order) were most anomalous observations. Researchers [17-22] also studied the same data to identify influential cases in modified ridge regression estimator and Liu estimator using global influence, local influence and Cook's minor perturbation methods and they identified 16, 4, 1, 10 and 15 were the most influential cases but the order of magnitude is changed.

For the Longley data the parameter λ is estimated using Box-Cox transformation model (necessary formulas for its implementations are given in the **Appendix**) is $\hat{\lambda} = 1.475$. The adjusted R^2 after the scale-preserving transformation is 99.3%, the standard deviation $\hat{\sigma}$ is 3.42981, and there is hardly any interaction; these are considerable improvements over the original values of 99.3% and 16.8976, respectively, and a large interaction. The RRE biasing parameter for Longley data is k = 0.00146. The l_{max} values are estimated using the Equation (13) and index plot for these values are given below.

From the index plot in **Figure 2** it can be seen that the most five influential cases are 13, 10, 14, 8, and 12 in this order in the Box-Cox transformation analysis for RRE in Longley data. Compare the influential cases detected by this method and the previous studies, there are some new influential cases were detected.



Figure 1. Index plot of *l_{max}* in RRE using Box-Cox transformation in MIFDI data.



Figure 2. Index plot of l_{max} for RRE using Box-Cox transformation in longley data.

5. Conclusions

In this paper, I have studied Box-Cox power transformation for biased RRE that seem practical and can play a considerable part in RRE data analysis. The local influence measure introduced focus on perturbing the constant variance. The influential cases detected by this method for biased RRE are different than the influential cases detected in global and local influential method for RRE.

Although no conventional cut off points are introduced or developed for the RRE Box-Cox power transformation diagnostic quantities, it seems that index plot is an optimistic and conventional procedure to disclose influential cases. It is a bottleneck for cut off 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.

REFERENCES

- [1] G. E. P. Box and D. R. Cox, "An Analysis of Transformation (with Discussion)," *Journal of the Royal Statistical Society, Series-B*, Vol. 26, 1964, pp. 211-252.
- [2] A. J. Lawrance, "Regression Transformation Diagnostics Using Local Influence," *Journal of the American Statistical Associa*tion, Vol. 83, No. 404, 1988, pp. 1067-1072. <u>http://dx.doi.org/10.1080/01621459.1988.10478702</u>
- [3] C. L. Tsai and X. Wu, "Transformation Model Diagnostics," *Technometrics*, Vol. 34, No. 2, 1992, pp. 197-202. <u>http://dx.doi.org/10.1080/00401706.1992.10484908</u>
- [4] A. C. Atkinson, "Plots, Transformations and Regression," Oxford University Press, Oxford, 1985.
- [5] R. D. Cook and P. C. Wang, "Transformations and Influential cases in Regression," *Technometrics*, Vol. 25, No. 4, 1983, pp. 337-343. <u>http://dx.doi.org/10.1080/00401706.1983.10487896</u>
- [6] D. V. Hinkley and S. Wang, "More about Transformations and Influential Cases in Regression," *Technometrics*, Vol. 30, No. 4, 1988, pp. 435-440. <u>http://dx.doi.org/10.1080/00401706.1988.10488439</u>
- [7] R. D. Cook, "Assessment of Local Influence," Journal of Royal Statistical Association, Series-B, Vol. 48, 1986, pp. 133-169.
- [8] C. L. Tsai and X. Wu, "Diagnostics in Transformation and Weighted Regression," *Technometrics*, Vol. 32, No. 3, 1990, pp. 315-322. <u>http://dx.doi.org/10.1080/00401706.1990.10484684</u>
- D. A. Belsley, E. Kuh and R. E. Welsch, "Regression Diagnostics: Identifying Influential Data and Sources of Collinearity," Wiley, New York, 1980. <u>http://dx.doi.org/10.1002/0471725153</u>
- [10] A. E. Hoerl and R. W. Kennard, "Ridge Regression: Biased Estimation for Non-Orthogonal Problems," *Technometrics*, Vol. 12, No. 1, 1970, pp. 55-67. <u>http://dx.doi.org/10.1080/00401706.1970.10488634</u>
- [11] A. E. Hoerl and R. W. Kennard, "Ridge Regression: Application to Non-Orthogonal Problems," *Technometrics*, Vol. 12, No. 1, 1970, pp. 69-82. <u>http://dx.doi.org/10.1080/00401706.1970.10488635</u>
- [12] H. Sun, "Macroeconomic Impact of Direct Foreign Investment in China 1979-1996," Blackwell Publishers Ltd., 1988.
- [13] J. W. Longley, "An Appraisal of Least Squares Programs for Electronic Computer from the Point of View of the User," *Journal of American Statistical Association*, Vol. 62, No. 319, 1967, pp. 819-841. <u>http://dx.doi.org/10.1080/01621459.1967.10500896</u>
- [14] E. Walker and J. B. Birch, "Influence Measures in Ridge Regression," *Technometrics*, Vol. 30, No. 2, 1988, pp. 221-227. <u>http://dx.doi.org/10.1080/00401706.1988.10488370</u>
- [15] R. D. Cook, "Detection of Influential Observations in Linear Regression," *Technometrics*, Vol. 19, No. 1, 1977, pp. 15-18. <u>http://dx.doi.org/10.2307/1268249</u>
- [16] L. Shi and X. Wang, "Local Influence in Ridge Regression," Computational Statistics & Data Analysis, Vol. 31, No. 3, 1999, pp. 341-353. <u>http://dx.doi.org/10.1016/S0167-9473(99)00019-5</u>
- [17] A. Jahufer and J. Chen, "Assessing Global Influential Observations in Modified Ridge Regression," *Statistics and Probability Letters*, Vol. 79, No. 4, 2009, pp. 513-518. <u>http://dx.doi.org/10.1016/j.spl.2008.09.019</u>
- [18] A. Jahufer and J. Chen, "Identifying Local Influential Observations in Liu Estimator," *Journal of Metrika*, Vol. 75, No. 3, 2012, pp. 425-438. <u>http://dx.doi.org/10.1007/s00184-010-0334-4</u>
- [19] A. Jahufer and J. Chen, "Measuring Local Influential Observations in Modified Ridge Regression," *Journal of Data Science*, Vol. 9, No. 3, 2011, pp. 359-372.
- [20] A. Jahufer, "Detecting Global Influential Observations in Liu Regression Model," Open Journal of Statistics, Vol. 3, No. 1, 2013, pp. 5-11. <u>http://dx.doi.org/10.4236/ojs.2013.31002</u>
- [21] A. Jahufer and J. Chen, "Identifying Local Influence in Modified Ridge Regression Using Cook's Method," *Sri Lankan Journal of Applied Statistics*, Vol. 9, 2008, pp. 93-108.
- [22] J. Chen and A. Jahufer, "Assessment of Anomalous Observations in Liu Estimator," *Journal of Management*, Vol. 5, 2009, 41-49.

Appendix: Constructed Variables for the Box-Cox Power Family

A useful class of transformations is the power transformation Y^{λ} , where λ is a parameter to be determined. Box-Cox [1] showed how the parameters of the regression model and λ can be estimated simultaneously using the method of maximum likelihood estimation method.

The procedure based on the [2] consists of performing a standard least squares fit using

$$z^{(\lambda)} = \begin{cases} \frac{y^{(\lambda)}}{\dot{y}^{\lambda-1}}, & \lambda \neq 0\\ \dot{y} \ln(y), & \lambda = 0 \end{cases}$$

where $y^{(\lambda)} = \frac{(y^{\lambda} - 1)}{\lambda}$

and \dot{y} is the geometric mean of (y_1, \dots, y_n) . Writing $q = \lambda \dot{y}^{\lambda-1}$ and $h = \lambda^{-1} + \ln(\dot{y})$, the two vectors of constructed variables required for variance perturbations are

$$z'^{(\lambda)} = \frac{\left\{y^{\lambda} \left[\ln\left(y\right) - h\right] + h\right\}}{q}$$
$$z''^{(\lambda)} = \frac{\left\{y^{\lambda} \left[\ln\left(y\right) - h\right]^{2} - h^{2} + \frac{\left(y^{\lambda} - 1\right)}{\lambda^{2}}\right\}}{q}.$$