

# Detecting Global Influential Observations in Liu Regression Model

**Aboobacker Jahufer**

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

Received May 14, 2012; revised June 30, 2012; accepted July 10, 2012

## ABSTRACT

In linear regression analysis, detecting anomalous observations is an important step for model building process. Various influential measures based on different motivational arguments and designed to measure the influence of observations on different aspects of various regression results are elucidated and critiqued. The presence of influential observations in the data is complicated by the presence of multicollinearity. In this paper, when Liu estimator is used to mitigate the effects of multicollinearity the influence of some observations can be drastically modified. Approximate deletion formulas for the detection of influential points are proposed for Liu estimator. Two real macroeconomic data sets are used to illustrate the methodologies proposed in this paper.

**Keywords:** Liu Estimator; Global Influential Observations; Diagnostics; Multicollinearity; Case Deletion; Approximate Deletion Formulas

## 1. Introduction

The presence of multicollinearity in the regressors seriously affects the parameter estimation and prediction. Therefore, mixed estimation and ridge type regression are suggested to mitigate this effect. In addition to multicollinearity, the presence of influential observations in the observed data is seriously affected by the estimators as those estimators are not unbiased [1,2].

In literature, many authors [1-3] noted that the influential observations on ridge type estimators are different from the corresponding least squares estimate and that multicollinearity can even disguise anomalous data. Belsey [3] investigated the leverage in ridge regression. Walker and Bitch [2] studied the influence of observations in ordinary ridge regression estimator (ORRE) based on case deletion method. Shi [4] proposed the local influence in principal component analysis by defining a generalized Cook statistic and showed that his method is equivalent to Cook's approach under the likelihood framework. Shi and Wang [2] analyzed the influential cases in ORRE using local influence method. Jahufer and Chen [5] studied global influential observations in modified ridge regression estimator (MRRE) moreover, Jahufer and Chen [6] studied local influential observations in MRRE. Besides, Jahufer and Chen [7] analyzed local influential observations in Liu estimator.

The main aim of this paper is therefore to assess the global influence of observations in the linear Liu estima-

tor using the method of case deletion. This method has been extensively studied and it is very powerful for detecting influential cases because of its intuitive appeal and its direct connection to the sample influence curves. Also, it is widely accepted as the foundation of many other statistical methods approached. The methodology proposed in this paper is illustrated using two real macroeconomic data sets. The first Data set is macro impact of foreign direct investment in Sri Lanka. This data set contains four regressors and a response variable with 27 observations. The second data set is Longley [8] data set. It consists of six regressors and a response variable with 16 observations. This paper is composed of six sections: Section 2 gives the background and definition of influential measures in least squares; Section 3 derives the influence measures in Liu estimator; Section 4 describes approximate deletion formulas for Liu estimator; Section 5 reports the examples using two real data sets. Discussion is given in the last section.

## 2. Background and Definition

### 2.1. Background

A matrix multiple linear regression model can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (1)$$

where  $\mathbf{y}$  is an  $n \times 1$  response vector,  $\mathbf{X}$  is an  $n \times p$  centered and standardized known matrix (*i.e.* the length of

the column of  $X$  is standardized to one),  $\beta$  is a  $p \times 1$  vector of an unknown parameter,  $\varepsilon$  is an  $n \times 1$  error vector with  $E(\varepsilon) = \mathbf{0}$  and  $Var(\varepsilon) = \sigma^2 I_n$  and  $I_n$  is an identity matrix of order  $n$ . Then the ordinary least squares estimator (OLSE) of  $\beta$  is  $\hat{\beta} = (X'X)^{-1} X'y$ . The estimator of  $\sigma^2$  is  $s^2 = e'e/(n-p)$ , where residual vector  $e = y - X\hat{\beta}$ .

This article assumes that the reader is acquainted with the basic ideas of leverage and influence in regression analysis, as presented, for instance, in the works of [1] and [9].

## 2.2. Definition of Influential Measures in Least Squares

The general purpose of influential analysis is to measure the changes induced in a given aspect of the analysis when the data set is perturbed. A particularly appealing perturbation scheme is case deletion. Note that this scheme is used throughout this article.

In general, the influence of a case can be viewed as the product of two factors: the first a function of the residual and the second a function of the position of the point in the  $X$  space. The position or leverage of the  $i$ -th point is measured by  $h_i$ , which is the  $i$ -th diagonal element of the "hat" matrix  $H = X(X'X)^{-1} X'$ .

Among the most popular single-case influential measure is the difference in fit standardized DFFITS [1], which evaluated at the  $i$ -th case is given by

$$DFFITS(i) = x_i [\hat{\beta} - \hat{\beta}(i)] / SE(x_i \hat{\beta}), \quad (2)$$

where  $\hat{\beta}(i)$  is the least squares estimator of  $\hat{\beta}$  without the  $i$ -th case and  $SE(x_i \hat{\beta})$  is an estimator of the standard error (SE) of the fitted values. DFFITS is the standardized change in the fitted value of a case when it is deleted. Thus it can be considered a measure of influence on individual fitted values.

Another useful measure of influence is Cook's  $D$  [9], which evaluated at the  $i$ -th case is given by

$$D_i = \frac{1}{ps^2} [\hat{\beta} - \hat{\beta}(i)]' (X'X) [\hat{\beta} - \hat{\beta}(i)], \quad (3)$$

where  $D_i$  is a measure of the change in all of the fitted values when a case is deleted. Even though  $D_i$  is based on different theoretical consideration, it is closely related to DFFITS.

Points with large values of  $D_i$  have considerable influence on the least squares estimate  $\hat{\beta}$ . In general, points for which  $D_i > F_{(\alpha, p, n-p)}$  to be influential. In DFFITS measures any observation for which

$$|DFFITS_i| > 2\sqrt{\frac{p}{n}} \text{ warrants attention.}$$

It is important to mention that these measures are use-

ful for detecting *single* cases with an unduly high influence. For generalizations of Equations (2) and (3) for detecting influential sets (see [1,9]).

## 3. Influence Measures in Liu Estimator

### 3.1. Liu Estimator

The Liu estimator  $\hat{\beta}_d$  was introduced by Liu [10] and is defined as

$$\hat{\beta}_d = (X'X + I)^{-1} (X'y + d\hat{\beta}), \quad (4)$$

where  $I$  is an identity matrix,  $\hat{\beta}$  is OLSE and  $d$  is Liu estimator biasing parameter and it is  $0 < d < 1$ . The Liu estimator combines the ORRE [11,12] estimator. The ORRE is effective in practice, but it is complicated function of its biasing parameter. Thus we often meet some complicated equations when we use some popular methods, such as ([13],  $C_k$  criterion [14], GCV criterion [15]) and etc. to choose ridge regression biasing parameter  $k$ . The advantage of Liu estimator over ORRE is that Liu estimator is a linear function of its biasing parameter  $d$ . Therefore, it is convenient to choose Liu estimator biasing parameter  $d$ .

The Liu estimator is very useful to mitigate the effect of near multicollinearity. Also, the recent literature, particularly in the area of econometrics, engineering and other statistical areas, the Liu estimator has produced a number of new techniques and ideas for example [5-7, 16-20].

### 3.2. Leverage and Residual Measures in Liu Estimator

Using Equation (4), the vector of fitted value of Liu estimator is

$$\begin{aligned} \hat{y}_d &= X\hat{\beta}_d = X(X'X + I)^{-1} (X'y + d\hat{\beta}) \\ &= X(X'X + I)^{-1} (X'X + dI)(X'X)^{-1} X'y = H_d y, \end{aligned}$$

where  $H_d = X(X'X + I)^{-1} (X'X + dI)(X'X)^{-1} X'$  is Liu estimator "hat matrix" (see [2,10]) and plays the same role as the hat matrix in OLSE. It is important to note that the matrix  $H_d$  is not a projection matrix because it is not idempotent and  $H_d$  is called quasi-projection matrix (see [2]). The  $i$ -th fitted value can be written in terms of

elements of  $H_d$  as  $\hat{y}_{di} = \sum_{j=1}^n h_{dij} y_j$ ; consequently,

$$\frac{\partial \hat{y}_i}{\partial y_i} = h_{dii} \equiv h_{di}. \text{ The "Liu hat diagonals" } h_{di} \text{ can be inter-}$$

preted as leverage in the same sense as the hat diagonals in OLSE. It is important to note that the  $H_d$  is not idempotent and it is called a quasi-projection matrix (see [2]).

The single value decomposition (SVD) (see [21]) al-

lows  $\mathbf{X}$  to be decomposed as  $\mathbf{X} = \mathbf{UDV}'$ , where  $\mathbf{D}$  is a  $p \times p$  diagonal matrix with  $i$ -th diagonal elements  $\lambda_i^{\frac{1}{2}}$  ( $\lambda_i$  is the  $i$ -th eigenvalue of  $\mathbf{X}'\mathbf{X}$ ), the column of  $\mathbf{V}$  are the eigenvectors of  $\mathbf{X}'\mathbf{X}$ . The  $(ij)$ -th element of the  $n \times p$  matrix  $\mathbf{U}(u_{ij})$  is such that  $u_{ij}\lambda_i^{\frac{1}{2}}$  is the projection of the  $i$ -th row,  $x_i$ , onto the  $j$ -th principal axis (eigenvector) of  $\mathbf{X}$ . Using the SVD, the Liu estimator leverage of the  $i$ -th point can be written as  $h_{di} = \sum_{j=1}^p \frac{(\lambda_j + d)}{(\lambda_j + 1)} u_{ij}^2$  (see [22]).

Several important facts can be deduced from the preceding expression. First, for  $d > 0$ ,  $h_{di} < h_i$  for  $i = 1, 2, \dots, n$ ; that is, for every observation the Liu estimator leverage is smaller than the corresponding OLSE leverage. It can be confirmed from the above equation that

$$\frac{(\lambda_p + d)}{(\lambda_p + 1)} h_i \leq h_{di} \leq \frac{(\lambda_1 + d)}{(\lambda_1 + 1)} h_i.$$

Second, the leverage decreases monotonically as  $d$  increases. Finally, the rate of decrement of leverage depends on the position of the particular row of  $\mathbf{X}$  along the principal axes. More specifically, the leverage of a row that lies in the direction of principal axes associated with large eigenvalues will be reduced less than the leverage of a row that lies in the direction of principal axes associated with small eigenvalues (see [2]).

The influence can be differentially affected as  $d$  increases. Remember, that influence is not a function of leverage but also of the residual. Although the leverage of every point decreases monotonically as  $d$  increases, the effect of this increment on the residuals is far less clear.

The  $i$ -th Liu estimator residual is defined as

$$e_{di} = y_i - \hat{y}_{di} = y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}}_d.$$

### 3.3. DFFITS and Cook's Measures in Liu Estimator

The DFFITS for Liu estimator can be written as

$$\text{DFFITS}_d(i) = \mathbf{x}_i [\hat{\boldsymbol{\beta}}_d - \hat{\boldsymbol{\beta}}_d(i)] / SE(\mathbf{x}_i \hat{\boldsymbol{\beta}}_d), \quad (5)$$

where  $\hat{\boldsymbol{\beta}}_d(i)$  is the Liu estimator in (4) computed with the  $i$ -th case deleted and the denominator is an estimator of the standard error of the Liu estimator fitted value. If Liu estimator biasing parameter  $d$  is assumed non-stochastic, then

$$SE(\mathbf{x}_i \hat{\boldsymbol{\beta}}_d) = s \left[ \mathbf{x}_i (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1} (\mathbf{X}'\mathbf{X} + d\mathbf{I}) \times (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\mathbf{X} + d\mathbf{I})' (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1} \mathbf{x}_i' \right]^{1/2}.$$

Hence the mean squared error is a function of the fitted values and the response, neither of which depends on individual eigenvalues of  $\mathbf{X}'\mathbf{X}$ ; it is not affected by multicollinearity. For this reason, the OLSE of  $\sigma[s]$  will be used as measures of scale.

At least two versions of Cook's  $D_i$  can be constructed for Liu estimator, they are

$$D_i^* = \frac{1}{ps^2} [\hat{\boldsymbol{\beta}}_d - \hat{\boldsymbol{\beta}}_d(i)]' (\mathbf{X}'\mathbf{X}) [\hat{\boldsymbol{\beta}}_d - \hat{\boldsymbol{\beta}}_d(i)] \quad (6)$$

and

$$D_i^{**} = \frac{1}{ps^2} [\hat{\boldsymbol{\beta}}_d - \hat{\boldsymbol{\beta}}_d(i)]' (\mathbf{X}'\mathbf{X} + \mathbf{I}) (\mathbf{X}'\mathbf{X} + d\mathbf{I})^{-1} \times (\mathbf{X}'\mathbf{X}) (\mathbf{X}'\mathbf{X} + d\mathbf{I})^{-1} (\mathbf{X}'\mathbf{X} + \mathbf{I}) [\hat{\boldsymbol{\beta}}_d - \hat{\boldsymbol{\beta}}_d(i)] \quad (7)$$

where  $D_i^*$  is the direct generalization of Cook's  $D$  in (3) and  $D_i^{**}$  is based on the fact that

$$\text{Var}(\hat{\boldsymbol{\beta}}_d) = \sigma^2 (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1} (\mathbf{X}'\mathbf{X} + d\mathbf{I}) \times (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\mathbf{X} + d\mathbf{I}) (\mathbf{X}'\mathbf{X} + \mathbf{I})^{-1}.$$

Note that both  $D_i^*$  and  $D_i^{**}$  simplify to  $D_i$  in (3) when  $d = 1$ .

It would be desirable to be able to write these measures as functions of leverage and residual, as was done in (2) and (3). This is not possible, however, because of the scale dependency of the Liu estimator. Since the Liu estimator is not scale invariant,  $\mathbf{X}(i)$  (the  $\mathbf{X}$  matrix with the  $i$ -th row deleted) has to be rescaled to unit-column length before computing  $\hat{\boldsymbol{\beta}}_d(i)$ . In the following section some approximate deletion formulas are proposed.

## 4. Deletion Formulas for Liu Estimator

In the analysis of influential observations to quantify the impact of the  $i$ -th case, the most common approach is to compute single-case diagnostics with the  $i$ -th case deleted. In Liu regression, it is impossible to derive an exact formula using case deletion because of the scale dependency of the Liu estimator. Hence, approximate deletion formulas are derived for influential measures DFFITS and two versions of Cook's statistics.

The scale dependence of Liu estimator precludes the development of deletion formulas of the types (2) and (3). The main problem resides in the computation of  $\hat{\boldsymbol{\beta}}_d(i)$  because the matrix  $\mathbf{X}(i)$  has to be standardized same as in Section 2.1, for the small values of  $d$  and/or cases with low leverage. However, approximate deletion formulas can be obtained using the Sherman-Morrison-Woodbury (SMW) theorem (see [1]).

When  $i$ -th row is deleted from  $\hat{\boldsymbol{\beta}}_d$  then  $\hat{\boldsymbol{\beta}}_d(i)$  can be written as

$$\hat{\boldsymbol{\beta}}_d(i) = [\mathbf{X}'(i)\mathbf{X}(i) + \mathbf{I}]^{-1} [\mathbf{X}'(i)\mathbf{y}(i) + d\hat{\boldsymbol{\beta}}(i)]$$

where  $X(i)$  is the matrix  $X$  without the  $i$ -th row and  $y(i)$  is the vector of response without the  $i$ -th entry. We assume that  $X(i)$  is centered and scaled so that  $X'(i)X(i)$  is in correlation form.

If after deletion the  $i$ -th row  $X$  is not recentered and rescaled. Then,  $X'X - x'_i x_i$  will not be in exact correlation form. Thus,

$$\hat{\beta}_d(i) \cong (X'X - x'_i x_i + I)^{-1} (X'y - x'_i y_i + d\hat{\beta}(i))$$

which uses the SMW theorem with  $K = (X'X + I)$  (see the Appendix),  $\hat{\beta}_d(i)$  can be approximated as:

$$\begin{aligned} \hat{\beta}_d(i) &\cong (X'X + I - x'_i x_i)^{-1} (X'y + d\hat{\beta}(i) - x'_i y_i) \\ \hat{\beta}_d(i) &\cong \left[ K^{-1} + \frac{K^{-1} x'_i x_i K^{-1}}{1 - x'_i K^{-1} x_i} \right] [X'y + d\hat{\beta}(i) - x'_i y_i] \\ &\cong \hat{\beta}_d - \frac{K^{-1} x'_i}{(1 - x'_i K^{-1} x_i)} [y_i - x_i \hat{\beta}_d] \\ &\cong \hat{\beta}_d - \frac{K^{-1} x'_i}{(1 - m_i)} (y_i - \hat{y}_{di}); \text{ where } m_i = x'_i K^{-1} x_i \\ &\cong \hat{\beta}_d - \frac{e_{di}}{(1 - m_i)} K^{-1} x'_i. \end{aligned}$$

$$\text{Therefore, } \hat{\beta}_d(i) \cong \hat{\beta}_d - \frac{K^{-1} x'_i e_{di}}{(1 - m_i)}.$$

Based on the above result, approximate version of (5), (6) and (7) can written as

$$\text{DFFITS}_{(a)d}(i) = \left[ \frac{m_i}{1 - m_i} \right] \frac{e_{di}}{\text{SE}(x_i \hat{\beta}_d)}, \quad (8)$$

$$\begin{aligned} D_{(a)i}^* &= \left[ \frac{1}{ps^2} \right] \left[ \frac{e_{di}}{1 - m_i} \right]^2 x_i (X'X + I)^{-1} \\ &\quad \times (X'X)(X'X + I)^{-1} x'_i, \end{aligned} \quad (9)$$

$$\begin{aligned} D_{(a)i}^{**} &= \left[ \frac{1}{ps^2} \right] \left[ \frac{e_{di}}{1 - m_i} \right]^2 x_i (X'X + dI)^{-1} \\ &\quad \times (X'X)(X'X + dI)^{-1} x'_i. \end{aligned} \quad (10)$$

## 5. Examples

### 5.1. Example 1: Macroeconomic Impact of Foreign Direct Investment (MIFDI) Data

Sun [23] 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 proposed in this paper. The data set consists of 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: Integration and Multicollinearity. The test results showed that variables are integrated with a same order of integration  $I(1)$  at 1% level of significance. The scaled condition number of this data set is 31,244, this large value suggests the presence of an unusually high level of multicollinearity among the regressors (the proposed cutoff is 30; see [1]). The Liu estimator biasing parameter for this data set is estimated  $d = 0.692$ .

Global influence measures Leverage, Residual, DFFITS and two versions of Cook's  $D_i$  were computed and the results for the most seven influential cases are given in **Table 1**. The influential cases detected by these methods are same except case 26 in leverage and case 14 in residual measures, but only the order of magnitude is changed.

Using the approximate case deletion formulas in Equations (8)-(10) the influential cases are estimated and it is given in **Table 2**. From this table, it can be seen that the influential cases detected by approximate case deletion formulas DFFITS and two versions of Cook's  $D_i$  are same but only the order of magnitude is changed. Moreover, influential cases detected by case deletion one by

**Table 1. The most seven influential observations using leverage, residual, DFFITS and two versions of Cook's.**

Leverage		Standardized Residual		DFFITS		$D_i'$		$D_i''$	
Case	Value	Case	Value	Case	Value	Case	Value	Case	Value
1	0.527	3	2.609	3	0.767	3	0.181	3	0.217
23	0.360	23	1.757	1	-0.494	27	0.065	27	0.071
18	0.330	27	1.483	27	0.473	1	0.059	1	0.070
3	0.268	2	-1.401	2	-0.472	2	0.057	2	0.066
2	0.218	18	1.379	18	0.387	23	0.052	23	0.056
27	0.200	22	1.373	22	0.379	18	0.040	18	0.050
26	0.186	14	-1.133	23	0.279	22	0.035	22	0.037

one formulas in Equations (5)-(7) and approximate deletion formulas in (8)-(10) are exactly same but only the order of magnitude is changed.

For verifying these results, it is plane to contribute **Table 3** of Liu estimates for the full data and the data without some influential cases. In this table, the parenthesis value indicates the percentage of change in the parameter value. The result reveals that case 3 is the most influential case while case 22 is the seventh influential point among the detected cases. It is also clear from this table, that omission of single influential cases 3, 23, 1, 27, 2, 18 and 22 contribute the substantial change in the Liu estimator. Among all of these 3, 23, 1, 2 and 18 have a remarkable influence while cases 27 and 22 have a little influence.

**5.2. Example 2: Longley Data**

The Longley data set [8] has also been used to explain the effect of extreme multicollinearity on the OLSE. The scaled condition number (see [1]) of this data set is 43,275. This large value suggests the presence of high level of multicollinearity among regressors. Cook [24] used this data to identify the influential observations in OLSE using the method of Cook’s  $D_i$  and found that cases 5, 16, 4, 10 and 15 (in this order) were the most influential observations (see **Table 4**). Walker and Birch [2] analyzed the same data to detect anomalous observations in ORRE using case deletion method. They found that cases 16, 10, 4, 15 and 5 (in this order) were most influential observations (see **Table 4**). Shi and Wang [25] also analyzed the same data to detect influential observations on the ridge regression estimator using local influence method. They detected cases 10, 4, 15, 16 and 1 (in this order) were the most five anomalous observations.

In this paper, I used the same data set to assess the influential observations in Liu estimator using global influence methods: Cook’s  $D_i$ , DFFITS, Leverage and Residual. The estimated results, the most five influential

cases and the corresponding values are given in **Table 4**.

The approximate deletion formulas in Equations (8)-(10) are used to detect influential cases for Longley data and detected influential cases are given in **Table 5**. According to this table, it can be confirmed that identified influential cases using approximate deletion formulas DFFITS and two versions of Cook’s  $D_i$  are exactly same. Besides, influential cases identified by case deletion one by one formulas in Equations (5)-(7) and approximate deletion formulas in (8)-(10) are precisely same for Longly data.

The influential cases in MIFDI and Longly data were identified using one by one deletion formulas in Equations (5)-(7) and approximate deletion formulas in Equations (8)-(10) respectively. The identified influential cases for MIFDI and Longly data are same for both measures but only the order of magnitude is changed for MIFDI data. Hence, instead of using the one by one case deletion formulas to identify influential cases in Liu estimator the approximate case deletion formulas are more suitable and appropriate to identify influential cases in Liu estimator.

**Table 2. The most seven influence observations according to approximate case deletion formulas.**

DFFITS <sub>(a)d</sub> (i)		D <sub>(a)d</sub> *		D <sub>(a)d</sub> **	
Case	Value	Case	Value	Case	Value
3	2.020	3	0.275	3	0.312
23	1.878	1	0.126	1	0.150
18	1.314	23	0.108	23	0.115
1	-1.226	27	0.079	27	0.087
2	-0.896	2	0.071	18	0.084
27	0.883	18	0.068	2	0.082
22	-0.468	22	0.035	22	0.038

**Table 3. Impact of influential cases on Liu estimator parameter.**

$\hat{\beta}_d$	Full Data	Case Deleted						
		(3)	(23)	(1)	(27)	(2)	(18)	(22)
$\hat{\beta}_{d0}$	392.13	394.80 (0.68%)	407.33 (3.88%)	393.36 (0.31%)	392.15 (0.00%)	393.12 (0.25%)	394.80 (0.68%)	392.41 (0.07%)
$\hat{\beta}_{d1}$	-1.65	-1.53 (-7.5%)	-1.88 (13.78%)	-2.03 (22.97%)	-1.65 (-0.18%)	-1.76 (6.17%)	-1.98 (19.59%)	-1.68 (1.33%)
$\hat{\beta}_{d2}$	-26.05	-26.44 (1.52%)	-27.26 (4.65%)	-25.92 (-0.48%)	-26.05 (0.01%)	-25.95 (-0.38%)	-25.81 (-0.91%)	-26.07 (0.08%)
$\hat{\beta}_{d3}$	-14.24	-14.25 (0.08%)	-13.99 (-1.74%)	-14.04 (-1.38%)	-14.23 (-0.08%)	-14.43 (1.31%)	-14.71 (3.28%)	-14.24 (0.03%)
$\hat{\beta}_{d4}$	-10.23	-10.04 (-1.79%)	-10.82 (5.83%)	-10.26 (0.33%)	-10.25 (0.22%)	-10.37 (1.38%)	-10.33 (1.06%)	-10.19 (-0.34%)

**Table 4. The five most influence observations in Longley data.**

OLSE		ORRE		Liu Estimator							
Case	$D_i^*$	Case	$D_i^{**}$	Case	$D_i^{**}$	Case	DFFITS	Case	Leverage	Case	Standardized Residual
5	0.614	16	0.582	15	0.352	15	0.667	16	0.613	10	1.981
16	0.467	10	0.251	4	0.315	4	-0.574	5	0.540	4	-1.734
4	0.244	4	0.219	1	0.231	1	0.493	2	0.493	15	1.424
10	0.235	15	0.145	6	0.134	6	-0.445	1	0.357	5	1.223
15	0.170	5	0.142	16	0.106	16	-0.179	15	0.306	1	1.133

**Table 5. The most five influential cases for Longly data using deletion formulas.**

DFFITS <sub>(a)d</sub> (i)		$D_{(a)}^*$		$D_{(a)}^{**}$	
Case	Value	Case	Value	Case	Value
15	0.642	15	0.335	15	0.323
4	-0.554	4	0.300	4	0.302
1	0.476	1	0.223	1	0.211
6	-0.423	6	0.126	6	0.112
16	-0.154	16	0.101	16	0.093

**6. Discussion**

In this article, I show that the Liu estimator user not rely on influence measures obtained for OLSE. Once the value of  $d$  is determined, influence measures should be computed for that  $d$ . If, after analyzing these indexes, it is decided to delete one or more cases from the analysis, the whole situation has to be reassessed in terms of both influence and multicollinearity.

In this research study, the Liu estimator shrinkage parameter  $d$  is estimated first and for that  $d$  value the Liu estimator co-efficients are estimated. Using these parameter quantities the influential observations are identified. But, the value of shrinkage parameter  $d$  depends on the every observation. Hence, for every influential case the value of  $d$  should be estimated. This is very difficult task so this issue will be studied in future research study.

The main advantage of the deletion formulas in Section 4 is that, as in least squares, the estimator does not have to be computed every time a case is deleted. For a value of  $d$  all of the elements in (8), (9) and (10) are readily available from a single run of Liu estimator. Moreover, these measures, based on deletion formulas are particularly helpful for large data sets. Furthermore, the deletion formulas provide computationally inexpensive approximate influence measures for Liu estimator.

Although no conventional cutoff points are introduced or developed for the Liu estimator global influence di-

agnostics: *Cook's measures, DFFITS, Leverage* and *Liu Estimator standardized residual*, 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. These are additional active issues for future research study.

**REFERENCES**

- [1] D. A. Belsley, E. Kuh and R. E. Welsch, "Regression Diagnostics: Identifying Influence Data and Source of Collinearity," Wiley, New York, 1980. [doi:10.1002/0471725153](https://doi.org/10.1002/0471725153)
- [2] E. Walker and J. B. Birch, "Influence Measures in Ridge Regression," *Technometrics*, Vol. 30, No. 2, 1988, pp. 221-227. [doi:10.1080/00401706.1988.10488370](https://doi.org/10.1080/00401706.1988.10488370)
- [3] D. A. Belsley, "Conditioning Diagnostics: Collinearity and Weak Data in Regression," Wiley, New York, 1991.
- [4] L. Shi, "Local Influence in Principal Component Analysis," *Biometrika*, Vol. 84, No. 1, 1997, pp. 175-186. [doi:10.1093/biomet/84.1.175](https://doi.org/10.1093/biomet/84.1.175)
- [5] A. Jahufer and J. B. Chen, "Assessing Global Influential Observations in Modified Ridge Regression," *Statistics and Probability Letters*, Vol. 79, No. 4, 2009, pp. 513-518. [doi:10.1016/j.spl.2008.09.019](https://doi.org/10.1016/j.spl.2008.09.019)
- [6] 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.
- [7] A. Jahufer and J. B. Chen, "Identifying Local Influential Observations in Liu Estimator," *Journal of Metrika*, Vol. 75, No. 3, 2012, pp. 425-438. [doi:10.1007/s00184-010-0334-4](https://doi.org/10.1007/s00184-010-0334-4)
- [8] J. W. Longley, "An Appraisal of Least Squares Programs for Electronic Computer for the Point of View of the User," *Journal of American Statistical Association*, Vol. 62, No. 319, 1967, pp. 819-841. [doi:10.1080/01621459.1967.10500896](https://doi.org/10.1080/01621459.1967.10500896)
- [9] R. D. Cook and S. Weisberg, "Residuals and Influence in Regression," Chapman & Hall, London, 1982.
- [10] K. Liu, "A New Class of Biased Estimate in Linear Regression," *Communications in Statistics—Theory and Methods*, Vol. 22, No. 2, 1993, pp. 393-402.
- [11] A. E. Hoerl and R. W. Kennard, "Ridge Regression: Biased

- Estimation for Nonorthogonal Problems,” *Technometrics*, Vol. 12, No. 1, 1970, pp. 55-67.  
[doi:10.1080/00401706.1970.10488634](https://doi.org/10.1080/00401706.1970.10488634)
- [12] C. Stein, “Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution,” *Proceeding of the third Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, December 1954 and July-August 1955, pp. 197-206.
- [13] G. C. McDonald and D. I. Galarnau, “A Monte Carlo Evaluation of Some Ridge-Type Estimators,” *Journal of American Statistical Association*, Vol. 70, No. 350, 1975, pp. 407-416. [doi:10.1080/01621459.1975.10479882](https://doi.org/10.1080/01621459.1975.10479882)
- [14] C. L. Mallows, “Some comments on Cp,” *Technometrics*, Vol. 15, No. 4, 1973, pp. 661-675.
- [15] G. Wahba, G. H. Golub and C. G. Heath, “Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter,” *Technometrics*, Vol. 24, No. 2, 1979, pp. 215-223. [doi:10.1080/00401706.1979.10489751](https://doi.org/10.1080/00401706.1979.10489751)
- [16] F. Akdeniz and S. Kaçiranlar, “More on the New Biased Estimator in Linear Regression,” *The Indian Journal of Statistics*, Vol. 63, No. 3, 2001, pp. 321-325.
- [17] S. Kaçiranlar and S. Sakallioğın, “Combining the Liu Estimator and the Principal Component Regression Estimator,” *Communications in Statistics—Theory and Methods*, Vol. 30, No. 12, 2001, pp. 2699-2705.
- [18] S. Kaçiranlar, G. P. H. Styan and H. J. Werner, “A New Biased Estimator In Linear Regression and a Detailed Analysis of the Widely Analyzed Dataset on Portland Cement,” *The Indian Journal of Statistics*, Vol. 61, No. B3, 1999, pp. 443-459.
- [19] M. H. Hubert and P. Wijekoon, “Improvement of the Liu Estimator in Linear Regression Model,” *Journal of Statistical Papers*, Vol. 47, No. 3, 2006, pp. 471-479.  
[doi:10.1007/s00362-006-0300-4](https://doi.org/10.1007/s00362-006-0300-4)
- [20] N. Torigoe and K. Ujiie, “On the Restricted Liu Estimator in the Gauss-Markov Model,” *Communications in Statistics—Theory and Methods*, Vol. 35, No. 9, 2006, pp. 1713-1722.
- [21] J. Mandel, “Use of the Singular Value Decomposition in Regression Analysis,” *The American Statistician*, Vol. 36, No. 1, 1982, pp. 15-24.
- [22] A. S. Topçubaşı and N. Billor, “A Class of Biased Estimators and Their Diagnostic Measures,” 2001.  
<http://idari.cu.edu.tr/sempozyum/bil26.htm>
- [23] H. Sun, “Macroeconomic Impact of Direct Foreign Investment in China: 1979-1996,” Blackwell Publishers Ltd., Oxford, 1998.
- [24] R. D. Cook, “Detection of Influential Observations in Linear Regression,” *Technometrics*, Vol. 19, No. 1, 1977, pp. 15-18. [doi:10.2307/1268249](https://doi.org/10.2307/1268249)
- [25] L. Shi and X. Wang, “Local Influence in Ridge Regression,” *Computational Statistics & Data Analysis*, Vol. 31, No. 3, 1999, pp. 341-353.  
[doi:10.1016/S0167-9473\(99\)00019-5](https://doi.org/10.1016/S0167-9473(99)00019-5)

## Appendix

### Sherman Morrison-Woodbury (SMW) Theorem

Consider the  $p \times p$  matrix  $X'X$  and let  $x_i$  be the  $i$ -th row of  $X$ . Note that  $X'(i)X(i) = X'X - x_i'x_i$  is the  $X'X$  matrix with the  $i$ -th row removed. The inverse matrix of  $(X'X - x_i'x_i)$  is:

$$(X'X - x_i'x_i)^{-1} = (X'X)^{-1} + \frac{(X'X)^{-1} x_i'x_i (X'X)^{-1}}{1 - x_i (X'X)^{-1} x_i'}$$