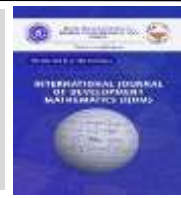




INTERNATIONAL JOURNAL OF DEVELOPMENT MATHEMATICS

ISSN: 3026-8656 (Print) | 3026-8699 (Online)

journal homepage: <https://ijdm.org.ng/index.php/Journals>



Robust Cook's Distance and DFFITS Diagnostics Based on the Robust M Kibria-Lukman Estimator for Detecting Influential Observations

Grace O. Oluwarohunbi^{a*}, Timothy O. Olatayo^a, and Abass I. Taiwo^a

^a*Department of Statistics, Olabisi Onanbanjo University Ago-Iwoye, Nigeria*

ARTICLE INFO

Article history:

Received 20 April 2026

Received in revised form 10 July 2026

Accepted 10 August 2026

Keywords:

Influential points, Robust estimation, Multicollinearity, Outliers, Regression diagnostics.

MSC 2020 Subject classification:

62P05

ABSTRACT

Influential observations can seriously distort parameter estimates, prediction accuracy, and statistical inference, particularly when multicollinearity and outliers occur simultaneously. Classical diagnostics such as Cook's Distance and DFFITS, developed under the Ordinary Least Squares (OLS) framework, often perform poorly under these conditions. This study proposes two new influential diagnostics, Cook's KL_M and DFFITS KL_M , based on the Robust M-Kibria-Lukman estimator. The methodology involved deriving the proposed diagnostics by incorporating the Robust M-KL residuals and hat matrix into the conventional influence measures and developing approximate case-deletion formulas using the Sherman-Morrison-Woodbury identity. The proposed methods were evaluated using the Manufacturing and Longley datasets and compared with existing OLS-, robust-, ridge-, Liu-, and Kibria-Lukman-based diagnostics. Results showed that Cook's KL_M consistently identified the same major influential observations as established methods, confirming its reliability. In contrast, DFFITS KL_M demonstrated greater sensitivity by detecting additional influential observations overlooked by existing diagnostics, particularly in the Longley dataset. Overall, the proposed diagnostics provide a more robust and computationally efficient approach for detecting influential observations in regression models affected by multicollinearity and outliers, making them suitable for practical statistical applications.

1 Introduction

Linear regression is a fundamental statistical modeling technique widely applied in various fields, including economics, engineering, medicine, and the social sciences. Its primary objective is to describe the relationship between a dependent variable and one or more independent variables through parameter estimation (Montgomery *et al.*, 2021). The Ordinary Least Squares (OLS) method remains the most commonly used estimation approach due to its simplicity and efficiency under ideal conditions specifically, when multicollinearity and outliers are absent (Seber and Lee, 2012). However, real-world data often violate these assumptions. Multicollinearity, which arises when independent variables are highly correlated, inflates the variance of OLS estimates, making them unstable and unreliable (Kutner *et al.*, 2004). To overcome these limitations, numerous estimators have been developed to improve parameter estimation in the presence of multicollinearity and outliers. Jegede *et al.* (2022) proposed a robust Jackknife Kibria-Lukman M-estimator based on the M-estimation framework. Arumairajan and Kayathiri (2022) introduced a stochastic restricted two-parameter estimator capable of addressing strong correlations when prior stochastic restrictions are available. Abidoye *et al.* (2022) developed an unbiased modified two-parameter estimator that incorporates prior information to reduce the effects of multicollinearity. Similarly, Majeed *et al.* (2022) proposed a robust version of the Kibria-Lukman estimator designed to manage outlying cases in the y-direction, whereas

*Corresponding author. Tel.: +2348133296458

E-mail address: golufunke@yahoo.com (Oluwarohunbi G. O.)

<https://doi.org/10.62054/ijdm/0303.20>

Adejumo *et al.* (2024) extended this work by examining the robust KL-M estimator for handling outliers occurring in the x-direction.

An influential point is an observation that significantly affects regression outcomes, including predicted responses, estimated coefficients, and hypothesis testing results (Cook and Weisberg, 1982). Such points can distort model accuracy, inflate or deflate the coefficient of determination, and lead to misleading inferences (Belsley *et al.*, 1980). Detecting influential points is therefore essential to ensure reliable regression analysis. Several diagnostic techniques have been developed for this purpose. Classical approaches include Cook's Distance (Cook, 1977), DFFITs (Welsch and Kuh, 1977), and Welsch's Distance (Welsch, 1982). These measures assess whether extreme values or leverage points exert significant influence on regression coefficients, while DFBETAS quantifies the effect of individual observations on parameter estimates.

Meloun and Militky (2001) provided a comprehensive overview of influential statistics for single-case detection in OLS regression and examined the joint influence of all variables. Leiva *et al.* (2006) investigated influential diagnostics in log-Birnbaum–Saunders regression models with censored data, introducing two deviance-type residuals for detecting outliers. Jahufer and Jianbao (2008) proposed a Modified Ridge Regression (MRR) deletion algorithm for identifying influential points. Recognizing that high-leverage observations may induce multicollinearity, Bagheri and Midi (2009) developed robust estimation approaches and proposed the Diagnostic Robust Generalized Potentials (DRGP) framework. Their findings revealed that the DRGP-MM estimator outperformed DRGP-L (L-estimator), DRGP-LTS (LTS-estimator), and DRGP-M (M-estimator). Bagheri *et al.* (2010) later employed Monte Carlo simulations to analyze the combined effects of collinearity and influential observations in multicollinear datasets. Building on these contributions, Jahufer (2013) proposed a deletion formula for influential detection under the Liu estimator.

Ayinde *et al.* (2015) compared several robust diagnostic methods including Cook's D, Welsch–Kuh distance, DFFITs, DFBETAS, MM-estimator, LTS, and S-estimators and demonstrated their relative effectiveness in identifying influential data. Further developments include the work of Asar and Erisoglu (2016), who introduced a two-parameter ridge-type diagnostic tool, extending Cook's D and DFFITs with corresponding deletion formulas. Emami and Emami (2016) applied Peña's Statistic to evaluate influential observations in Ridge regression, while Lukman and Ayinde (2016) proposed an approximate deletion formula for influential detection under the two-parameter Liu–Ridge estimator. Peng *et al.* (2016) presented the Generalized Standardized Corrected Pearson Residuals (GSCPR) for detecting multiple leverage points in generalized linear models. Lukman *et al.* (2018) applied a modified Peña's statistic to biased estimators, and Kashif *et al.* (2018) extended Peña's concept to develop an influential measure for the Liu regression estimator. Alkasadi *et al.* (2019) introduced a DFFIT-based diagnostic for identifying outliers in the Multiple Circular Regression Model (MCRM). Hussein (2021) later proposed a Liu-type diagnostic framework, deriving Cook's D and DFFITs versions alongside their deletion approximations for situations characterized by severe multicollinearity.

Recent advancements include the work of Adejumo *et al.* (2024), who developed diagnostic measures for the New Two-Parameter (NTP) estimator. They derived Cook's D, DFFITs (DFT), Peña's Statistic (PEN), Atkinson's Statistic (ATK), Hadi's Measure (HAD), and COVRATIO (CVR) for the NTP estimator and assessed their performance under various levels of multicollinearity. Their results identified CVR-NTP as the most robust measure for detecting influential points in data affected by multicollinearity and outliers. Similarly, Adejumo *et al.* (2025) extended this diagnostic framework to the Kibria–Lukman (KL) estimator, developing generalized versions of Cook's D, DFFITs, PEN, ATK, HAD, and CVR. Their simulation findings demonstrated that the CVR-KL measure performed best in detecting influential points under conditions of multicollinearity and outlier contamination.

In line with these developments, the present study proposes the Cook's D and DFFITs versions under the robust M-KL estimator, along with their approximate case-deletion formulas. These newly developed diagnostic measures are subsequently applied to real-life datasets to evaluate their effectiveness in detecting influential observations.

2 Methodology

One technique for determining the relationship between a dependent variable and one or more independent variables is linear regression analysis. Regression analysis, according to Gujarati (2003), is the study of the relationship between one or more regressors and a response variable Y . The linear regression model's standard form is

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_p x_{ij} + e_{ij} \quad , i = 1, 2, \dots, n, j = 0, 1, 2, \dots, p \quad (1)$$

where n is the number of observations, p is the number of explanatory variables such that $n > p$. The y_i denotes the i^{th} observed response, the x_{ij} represents the i^{th} observation of the regressor x_j , β_j , $j = 0, 1, \dots, p$, denotes the regression coefficients, and e_i represents the random error where $E(e_i) = 0$ and $V(e_i) = \sigma^2 I_n$ and I_n is an n -identity matrix (Birkes and Dodge, 1993).

The matrix is;

$$Y = X\beta + e_i \quad (2)$$

Such that β is estimated as:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (3)$$

2.1 Review of Some Conventional Influential Measures

The review of some of the already existing influential measures are as follows:

2.1.1 Cook's Distance Measures based on Classical approach

Cook (1977) proposed Cook's D , an aggregate influence metric that illustrates how the i^{th} case affects all fitted values. This instrument measures the impact of every data point on the regression coefficients by assessing how the fitted values change when the point is eliminated from the study. An impactful observation is indicated by larger Cook's Distance values. Typically, D_i indicates it as:

$$D_i = \frac{(\hat{\beta}_i - \hat{\beta})' (X'X) (\hat{\beta}_i - \hat{\beta})}{p \hat{\sigma}^2} \quad (4)$$

where $\hat{\beta}_i$ and $\hat{\beta}$ respectively provide estimate on all n data points and the estimate obtained after the i^{th} observation is deleted. Cook's distance measure has been observed to relate to $f(p, n - p)$ distribution and hence its percentile value can be ascertained. If the percentile value is less than about 10 or 20 percent, the i^{th} case has little apparent influence on the fitted values. If on the other hand, the percentile value is near 50 percent or more, the fitted values obtained with and without the i^{th} case should be considered to differ substantially, implying that the i^{th} case has a major influence on the fit of the regression function. According to Cook (1977), an equivalent algebraic expression is given as:

$$D_i = \frac{r_i^2}{p} \left(\frac{h_{ii}}{1 - h_{ii}} \right) \quad (5)$$

where h_{ii} is the i^{th} diagonal elements of the hat matrix and r_i is i^{th} internally studentized residual which is given as follows:

$$r_i = \frac{e_i}{\sqrt{\hat{\sigma}^2 (1 - h_{ii})}} \quad (6)$$

such that, $e_i = y_i - \hat{y}$ and $\sigma^2 = \frac{\sum_{i=1}^n e_i^2}{n-p}$.

It was suggested that observations for which $Di > 1$ warrants attention.

2.1.2 DFFITs

According to Welsch and Kuh (1977), DFFITs are a diagnostic measure that indicates a point's level of influence in a statistical regression. It can also be found by dividing the estimated standard deviation of the fit at a point by the change in the projected value for that point that is obtained when that point is excluded from the regression.

A useful measure of the influence that case i has on the fitted value $\hat{Y}$ is defined as:

$$DFFITs_i = \frac{\hat{Y} - \hat{Y}_{(-i)}}{\sqrt{S_{(-i)}^2 h_{ii}}} \quad (7)$$

where $\hat{Y}$ is the predicted value for all cases, $\hat{Y}_{(-i)}$ for the i th case obtained when the i th case is omitted in fitting the regression function, $S_{(-i)}^2$ is the estimated mean square error MSE of $\hat{Y}_{(-i)}$. Thus, DFFITs is the standardized change in the fitted value of a case when it is deleted. It can also be expressed as:

$$DFFITs_i = \frac{x_i [\hat{\beta} - \hat{\beta}_{(-i)}]}{S(x_i \hat{\beta})} \quad (8)$$

where $S(x_i \hat{\beta})$ is an estimator of standard error of the fitted values, x_i is the i th row of the X matrix, $\hat{\beta}_{(-i)}$ is the least squares estimator of β when the i th case is omitted in fitting the regression function.

Algebraically, it can further be expressed as:

$$(DFFITs)_i = \left(\frac{h_{ii}}{1-h_{ii}} \right)^{1/2} \frac{e_i}{\sqrt{\hat{\sigma}^2 (1-h_{ii})}} = \left(\frac{h_{ii}}{1-h_{ii}} \right)^{1/2} t_i \quad (9)$$

where $\hat{\sigma}^2$ is the estimate of σ , h_{ii} is the diagonal elements of the hat matrix and

$t_i = \frac{e_i}{\sqrt{\hat{\sigma}^2 (1-h_{ii})}}$ is the studentized residual (also called the external studentized residual). Lukman and Ayinde (2020).

It was suggested by Jahufer (2013) that observations for which $|DFFITs_i| > 2\sqrt{\frac{p}{n}}$ warrants attention for large data sets and if the absolute value of $DFFITs_i$ exceeds 1 for small to medium data sets.

3. The Proposed Influential Measures in Robust M- Kibria-Lukman (KL_M) Estimator

Kibria and Lukman (2020) proposed a one-parameter New Ridge-Type estimator, known as the KL estimator, as an alternative to existing one-parameter estimators for addressing the problem of multicollinearity in linear regression models. However, the presence of extreme observations in a dataset is often unavoidable, and such observations have been found to affect the performance of the KL estimator. To address this, Majeed *et al.* (2022) developed a robust M-version of the KL estimator to handle both multicollinearity and outliers occurring in the y-direction

simultaneously. Meanwhile Adejumo *et al.* (2024) extended this idea to cases involving outliers in the x-direction. Furthermore, Adejumo *et al.* (2025) introduced influential diagnostic tools for the KL estimator under the assumption that only multicollinearity is curbed. Given that outliers are frequently inevitable in practical data analysis, it becomes essential to develop influential diagnostic measures within the framework of the robust M-KL estimator. Kibria and Lukman (2020) developed a new one-parameter estimator to address multicollinearity in linear regression models, in addition to the existing ones. They defined the estimator as follows:

$$\hat{\beta}_{KL} = (X^T X + kI_p)^T (X^T X - kI_p) \hat{\beta}_{OLS} \quad (10)$$

where $k = \frac{\hat{\sigma}^2}{2\hat{\alpha}_i^2 + (\hat{\sigma}^2 / \lambda_i)}$ and $\hat{\beta}_{OLS}$ is the OLS estimator

The robust M-version of (10) as defined by Majeed *et al.* (2022) and Adejumo *et al.* (2024), is expressed as:

$$\hat{\beta}_M^{KL} = (X^T X + kI_p)^{-1} (X^T X - kI_p) \hat{\beta}_M. \quad (11)$$

where $k_M = \frac{\hat{\sigma}_{iM}}{2\alpha_{iM}^2 + \frac{\hat{\sigma}_M^2}{\lambda_i}}$ which is the estimated parameter for the robust M- KL estimator,

λ_i is the i th Eigen value of $X^T X$, $\hat{\alpha}_{iM} = Q^T \hat{\beta}_M$, $\sigma_M^2 = \frac{(e_{iM})^T (e_{iM})}{n-p}$ which is the MSE of each of the robust estimator with full data.

Such that $e_{iM} = Y_i - \hat{Y}_i^M$ and $\hat{\beta}_M^{KL}$ is the KL-M estimator

The fitted value for the robust M-KL estimator can be derived as follows.

$$\hat{Y}_i^{KL-M} = X \hat{\beta}_M^{KL}. \quad (12)$$

$$\hat{Y}_i^{KL-M} = X (X^T X + k_M I_p)^{-1} (X^T X - k_M I_p) \hat{\beta}_M. \quad (13)$$

$$\hat{Y}_i^{KL-M} = X (X^T X + k_M I_p)^{-1} (X^T X - k_M I_p) (X^T X)^{-1} X^T Y. \quad (14)$$

$$\hat{Y}_i^{KL-M} = H^{KL-M} Y. \quad (15)$$

where $H^{KL-M} = X (X^T X + k_M I_p)^{-1} (X^T X - k_M I_p) (X^T X)^{-1} X^T$,

This is equivalent to the hat matrix of the OLSE when $k_M = 0$. Since H^{KL-M} is not an idempotent, therefore it is not a projection matrix and the i th diagonal element of H^{KL-M} can be written as:

$$h_{ii}^{KL-M} = x_i (X^T X + k_M I_p)^{-1} (X^T X - k_M I_p) (X^T X)^{-1} x_i^T. \quad (16)$$

where x_i denotes the i th row of the matrix X

Similarly, the i th fitted value written in terms of element H^{KL-M} is as follows:

$$\hat{Y}_i^{KL-M} = \sum_{i=1}^n h_{ii}^{KL-M} Y_i. \quad (17)$$

Therefore, taking the partial derivative of $\hat{Y}_i^{KL-MM}$ with respect to Y_i ;

$$\frac{\partial \hat{Y}_i^{KL-M}}{\partial Y_i} = h_{ii}^{KL-M} \quad (18)$$

While, the i th KL_M residual is given as:

$$e_i^{KL-M} = Y_i - \hat{Y}_i^{KL-M} \quad (19)$$

3.1 Cook's D in Robust M-KL Estimator

The Cook's D version for the robust M-KL estimator is derived by substituting the Ordinary Least Squares (OLS) estimator $\hat{\beta}_{OLS}$ and the diagonal elements of the hat matrix used in the conventional Cook's D (Cook, 1977) with the corresponding robust M-KL estimator and its associated hat matrix H^{M-KL} diagonal elements. Following the procedures of the traditional Cook's D formulation and applying the Sherman–Morrisson–Woodbury theorem (Rao, 1973), the Cook's D measure for the robust M-KL estimator, denoted as D_i^{KL-M} , is defined as follows:

$$D_i^{KL-M} = \frac{(r_{iKL-M})^2}{(p+1)} \left(\frac{h_{ii}^{KL-M}}{1-h_{ii}^{KL-M}} \right) \quad (20)$$

$$\text{where } r_{iKL-M} = \frac{e_i^{KL-M}}{\sqrt{MSE_M(1-h_{ii}^{KL-M})}}$$

such that, $MSE_M = \frac{(e_i^M)^T e_i^M}{n-p}$, $e_i^{KL-M} = \sum_{i=1}^n (Y_i - \hat{Y}_i^{KL-M})^2$ and $e_i^M = Y_i - \hat{Y}_i^M$. h_{ii}^{KL-M} is the diagonal element of the hat matrix H^{KL-M} for the KL_M estimator when data is complete. The cut-off point is $D_i^{KL-M} > 1$ this implies that the observation that its $D_i^{KL-M} > 1$ is influential.

3.2 DFFITs in robust M-KL Estimator

The DFFITs version of robust M- KL estimator is hereby proposed following the conventional influential measure for DFFITs developed based on OLS estimator and it can be expressed as:

$$DFFITs_i^{KL-M} = \frac{|\hat{Y}_i^{KL-M} - \hat{Y}_{(-i)}^{KL-M}|}{S_{(-i)KL-M} \sqrt{h_{ii}^{KL-M}}} \quad (21)$$

where $\hat{Y}_i^{KL-M}$ is the fitted model of the KL_M estimator when the data is complete and $\hat{Y}_{(-i)}^{KL-M}$ is the fitted model for the KL_M estimator when the i th observation is deleted, $S_{(-i)KL-M}$ is the KL_M standard error estimated when the i th observation is deleted and h_{ii}^{KL-M} is the diagonal elements of the hat matrix $X(X^T X + k_M I_p)^{-1}(X^T X - k_M I_p)(X^T X)^{-1} X^T$. (22)

The observation for which $DFFITs_i^{KL-M} / > 2 \sqrt{\frac{p}{n}}$ for large data, such observation can be considered to be influential.

3.2.1 Approximate Case Deletion formulas for KL_M estimator

If i th row is deleted from $\hat{\beta}_{KL_M}$, then $\hat{\beta}_{(i)KL_M}$ can be written as:

$$\hat{\beta}_{(i)KL_M} = (G_{(i)} + kI_p)^{-1} (G_{(i)} - kI_p) \hat{\beta}_{(i)M}, \quad (23)$$

where $G_{(i)} = X_{(i)}^T X_{(i)}$ and $X_{(i)}$ is the Matrix X without the i th row. $X_{(i)}$ is scaled so that $G_{(i)}$ is in correlation form.

Hence, $\hat{\beta}_{(i)KL_M}$ is approximately obtained by applying the Sherman-Morrison-Woodbury (SMW) theorem by Rao, (1973). It is therefore obtained as follows:

$$\hat{\beta}_{(i)KL_M} = (G - x_i^T x_i + kI_p)^{-1} (G - x_i^T x_i - kI_p)^{-1} \hat{\beta}_{(i)M}. \quad (24)$$

$$\hat{\beta}_{(i)KL_M} = (F_k - x_i^T x_i)^{-1} (G - x_i^T x_i - kI_p) \hat{\beta}_{(i)M}. \quad (25)$$

where, $F_k = G + kI_p$,

Therefore,

$$\hat{\beta}_{(i)KL_M} = \left(F_k^{-1} + \frac{F_k^{-1} x_i^T x_i F_k^{-1}}{1 - x_i F_k^{-1} x_i^T} \right) (G - kI_p - x_i^T x_i) \hat{\beta}_{(i)M}. \quad (26)$$

$$\text{where, } (F_k - x_i^T x_i)^{-1} = \left(F_k^{-1} + \frac{F_k^{-1} x_i^T x_i F_k^{-1}}{1 - x_i F_k^{-1} x_i^T} \right), \text{ such that, } L_{ii} = x_i F_k^{-1} x_i^T. \quad (27)$$

By expanding (28), it leads to (29)

$$\hat{\beta}_{(i)KL_M} = F_k^{-1} (F - kI_p) \hat{\beta}_{(i)M} - \left(\frac{F_k^{-1} x_i^T x_i F_k^{-1}}{1 - L_{ii}} \right) \hat{\beta}_{(i)M} + \left(\frac{F_k^{-1} x_i^T x_i F_k^{-1}}{1 - L_{ii}} \right) (G - kI_p) \hat{\beta}_{(i)M} \quad (28)$$

$$\hat{\beta}_{(i)KL_M} = \hat{\beta}_{KL_M} + \frac{F_k^{-1} x_i^T}{1 - L_{ii}} [x_i F_k^{-1} (G - kI_p) \hat{\beta}_{(i)M} - x_i^T F_k^{-1} x_i \hat{\beta}_{(i)M} - (1 - L_{ii}) \hat{\beta}_{(i)M}] \quad (29)$$

$$\hat{\beta}_{(i)KL_M} = \hat{\beta}_{KL_M} + \frac{F_k^{-1} x_i^T}{1 - L_{ii}} [\hat{y}_{iKL_M} - L_{ii} y_i - (1 - L_{ii}) y_i]. \quad (30)$$

where, $\hat{y}_{iKL_M} = x_i F_k^{-1} (G - kI_p) \hat{\beta}_{(i)M} = x_i F_k^{-1} (G - kI_p) G^{-1} X^T y_i$ and

$$\hat{\beta}_{KL_M} = F_k^{-1} (G - kI_p) \hat{\beta}_{(i)M}, \quad (31)$$

$$\hat{\beta}_{(i)KL_M} = \hat{\beta}_{KL_M} + \frac{F_k^{-1} x_i^T}{1 - L_{ii}} (\hat{y}_{iKL_M} - y_i) \quad (32)$$

$$\hat{\beta}_{(i)KL_M} = \hat{\beta}_{KL_M} - \frac{F_k^{-1} x_i^T}{1 - L_{ii}} (y_i - \hat{y}_{iKL_M}) \quad (33)$$

$$\hat{\beta}_{(i)KL_M} = \hat{\beta}_{KL_M} - \frac{F_k^{-1} x_i^T e_i^{KL_M}}{1 - L_{ii}}. \quad (34)$$

where, $e_i^{KL_M} = (y_i - \hat{y}_{iKL_M}) = (1 - h_{ii}^{KL_M}) y_i$

$$\text{Therefore, } \hat{\beta}_{KL_M} - \hat{\beta}_{(i)KL_M} \cong \frac{e_i^{KL_M} F_k^{-1} x_i^T}{1 - L_{ii}}. \quad (35)$$

3.2.2 The approximate case deletion formula of Cook's D version in the KL-M estimator

The approximate version of D_{iKL_MM} can be written as:

$$D_{iKL_M} = \frac{1}{ps^2} \left[\frac{e_i^{KL_M} F_k^{-1} x_i^T}{1 - L_{ii}} \right]^T G \left[\frac{e_i^{KL_M} F_k^{-1} x_i^T}{1 - L_{ii}} \right]. \quad (36)$$

$$= \frac{1}{ps^2} \frac{x_i^T}{(1 - L_{ii})^2} F_k^{-1} (e_i^{KL_SM})^2 G F_k^{-1} x_i^T,$$

where, $F_k^{-1} = (G + kI_p)^{-1}$,

Therefore

$$\begin{aligned} D_{iKL_M} &= D_{iRKL}^* = \frac{1}{ps^2} \left[\frac{e_i^{KL_M}}{1 - L_{ii}} \right]^2 x_i^T (G + kI_p)^{-1} (G + kI_p)^{-1} x_i^T \\ &= \frac{1}{ps^2} \left[\frac{e_i^{KL_M}}{1 - L_{ii}} \right]^2 \sum_{j=1}^n L_{ij}^2 = \frac{1}{ps^2} \left[\frac{e_i^{KL_M}}{1 - L_{ii}} \right]^2 \sum_{j=1}^n L_{ij}^2 \end{aligned} \quad (37)$$

Also the approximate version of D_{iKL_M} can be derived as:

$$D_{iKL_M}^{a*} = \frac{1}{ps^2} \left[\frac{e_i^{KL_M}}{1 - L_{ii}} \right]^2 x_i (G + kI_p)^{-1} (G - kI_p)^{-1} G x_i^T (G - kI_p) (G + kI_p)^{-1} x_i^T \quad (38)$$

$$D_i^{KL-MM} = \frac{1}{pS^2} \left[\frac{e_i^{KL-M}}{1-L_{ii}} \right]^2 \sum_{j=1}^n (h_{ij}^{KL-M})^2. \quad (39)$$

3.2.3 The approximate case deletion formula of the DFFITs version in the KL_M estimator

The approximate case deletion formula for DFFITs in the KL_M estimator can be expressed as:

$$DFFIT S_i^{KL-M} = \left[\frac{L_{ii}}{1-L_{ii}} \right] \frac{e_i^{KL-M}}{S(x_i \hat{\beta}_{KL-M})} \quad (40)$$

$$DFFIT S_i^{KL-M} = \left[\frac{L_{ii}}{1-L_{ii}} \right] \frac{e_i^{KL-M}}{S_{(i)} \sqrt{\sum_{j=1}^n (h_{ij}^{KL-M})^2}} \quad (41)$$

where $S_{(i)} = \sqrt{\frac{(n-p)S^2 - \frac{e_i^2}{1-h_{ii}}}{n-p}}$ such that $L_{ii} = x_i F_K^{-1} x_i^T$ and $F_K^{-1} = (G + kI_p)^{-1}$.

4. Application to Real-life data set

Two datasets were used in this study to evaluate the performance of the proposed influential diagnostics: the Manufacturing dataset and the Longley dataset. Both datasets have been widely used by several researchers in previous studies.

4.1 Application to Manufacturing data

The manufacturing dataset used in this study was originally adopted by Hussein and Abdalla (2012). The regression model for the dataset is expressed as the linear model below:

$$y = \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3 \quad (42)$$

where y represents the product value in the manufacturing sector, X_1 denotes the value of imported intermediate goods, X_2 represents imported capital commodities, and X_3 indicates the value of imported raw materials. Hussein and Abdalla (2012) reported that the dataset suffers from multicollinearity, as reflected in the high Variance Inflation Factor (VIF) values of 128.29, 103.43, and 70.87. Similarly, Lukman *et al.* (2014) confirmed this finding and further identified the presence of outliers in the dataset. The value of the ridge parameter k used in this study follows that of Yasin and Murat (2016) and Kashif *et al.* (2019), which was computed as 0.0012.

Table 1: Comparison of the Five Most Influential Points Identified by Existing and Proposed Diagnostic Measures Using the Manufacturing Dataset

Author(s)	Year	Existing Method	
		Influential points in order	Method
Cook	1977	15, 16, 31, 30 and 20	Cooks D in OLS
Ayinde <i>et al.</i>	2015	30, 31, 16, 21 and 14	Cooks D in M-estimator
Ayinde <i>et al.</i>	2015	15, 30, 31, 14 and 12	Cooks D in MM-estimator
Ayinde <i>et al.</i>	2015	15, 16, 31, 30 and 20	DIFFTs based on OLS
Ayinde <i>et al.</i>	2015	31, 15, 30, 14 and 12	DIFFTs in MM-estimator
Adejumo <i>et al.</i>	2025	15, 16, 31, 30 and 20	Cooks in KL estimator
Adejumo <i>et al.</i>	2025	4, 3, 1, 5 and 2	DFFITs in KL estimator
Proposed Method			
Cook_KL_M	Proposed	15, 16, 31, 30 and 20	Cooks D in Robust-M KL estimator
DFFITs_KL_M	Proposed	4, 3, 1, 5 and 2	DFFITs in Robust-M KL estimator

From Table 1, it can be observed that the newly proposed measure, Cooks_KL_M, performed comparably to those of Ayinde *et al.* (2015), Adejumo *et al.* (2025), and Cook (1977), as all identified the same influential cases 15, 16, 31,

30, and 20 in identical order. Meanwhile, the new influential measure DFFITs_KL_M detected the same cases as reported by Adejumo *et al.* (2025), namely 4, 3, 1, 5, and 2. This suggests that the DFFITs_KL and DFFITs_Robust-M_KL influential diagnostics are effective in identifying additional influential observations, even when issues of outliers or multicollinearity have been mitigated. This is evident since the cases they identified are distinct from those detected by some of the existing measures.

4.2 Application to Longley data set

The Longley dataset is an economic dataset that has been extensively used by several researchers, including Jahufer and Jianbao (2009), Jahufer (2013), Yasin and Murat (2016), Ullah *et al.* (2013), Kashif *et al.* (2019), Lukman and Ayinde (2018), and Longley (1967). The regression model for the dataset is expressed as follows:

$$y = \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3 + \gamma_4 x_4 + \gamma_5 x_5 + \gamma_6 x_6 \quad (43)$$

where y represents the total derived employment, x_1 is the gross national product implicit price deflator, x_2 is the gross national product, x_3 denotes unemployment, x_4 represents the size of the armed forces, x_5 is the non-institutional population aged 14 years and over, and x_6 denotes time. Walker and Birch (1988) reported that the scaled Condition Number (CN) of the dataset is 43.275, indicating potential multicollinearity. Similarly, Liu (1993) found that the dataset suffers from multicollinearity, as reflected in the high Variance Inflation Factor (VIF) values of 128.29, 103.43, and 70.87.

Table 2: Comparison of the Five Most Influential Observations Identified by Existing and Proposed Diagnostic Measures Using the Longley Dataset

Author(s)	Year	Existing Method	
		Influential points in order	Method
Cook	1977	5, 16, 4, 10, 15	Cooks D based on OLS
Jahufer	2013	15, 4, 1, 6, 16 16, 10, 4, 6, 1	Cooks D and DFFITs in Liu regression
Ullah <i>et al.</i>	2013	16, 10, 4, 6, 5, and 16, 5, 4, 10, 15	Influential points in Liu regression for $d = 0.1, 0.5$ and 0.9
Yasin and Murat	2016	16, 10, 6, 1 and 4	Influential points in Two-parameter Ridge
Lukman and Ayinde	2018	16, 10, 4, 5 and 15 3, 10, 11, 4 and 5 10, 3, 4, 15 and 16, 10, 4, 15, 5 and 16	DFFITs in Two-parameter Ridge_Liu
Kashif <i>et al.</i>	2018	15, 10, 5, 4 and 16	Pena's Statistic for the Liu regression for $d = 0.1, 0.5, 0.9$ and 1.0
Kashif <i>et al.</i>	2019	16, 15, 5, 4 and 10	Modified Pena's Statistic for the Modified Ridge
Adejumo <i>et al.</i>	2024	10, 9, 11, 1 and 8	Regression for $K = 0.0002$
Adejumo <i>et al.</i>	2024	16, 10, 6, 4 and 13	DFFITs in New Two-Parameter
Adejumo <i>et al.</i>	2025	16, 6, 8, 13 and 7	Cooks in New Two-Parameter
Adejumo <i>et al.</i>	2025	10, 11, 5, 14 and 6	Cooks in KL estimator
Dawoud and Hussein	2025	5, 15, 6, 13 and 1	DFFITs in KL estimator
Dawoud and Hussein	2025	16, 6, 8, 13 and 7	DFFITs in Dawoud-Kibria
Proposed Method			
Cook_KL_M	Proposed	16, 6, 8, 13 and 7	Cooks in Robust-M KL estimator
DFFITs_KL_M	Proposed	2, 16, 12, 1 and 9	DFFITs in Robust-M KL estimator

When applied to the economic dataset, as shown in Table 2, the proposed influential measure Cook's in Robust-M KL estimator detected all the cases identified by Cook's in KL estimator proposed by Adejumo *et al.* (2025). However, the DFFITs in Robust-M KL estimator detected cases 16, 1, and 9 similar to some of the existing

measures, though not in the same order. Meanwhile, cases 2, 12, and 9 were detected in a different order, and notably, cases 2 and 12 identified by DFFITs_KL_M were not captured by any of the existing influential measures. This clearly suggests that the newly proposed diagnostic tool, DFFITs_KL_M, is capable of identifying more influential points than the pre-existing diagnostic methods.

4.3 Discussion

The introduction of the Robust-M KL estimator–based influential diagnostics represents a significant advancement in regression modeling where multicollinearity and outliers coexist. The Cook's_KL_M measure proved to be consistent with established approaches, confirming its reliability in identifying major influential observations without losing interpretability. This consistency aligns with the findings of previous authors such as Ayinde *et al.* (2015) and Adejumo *et al.* (2025), thereby reinforcing its methodological soundness. The DFFITs_KL_M diagnostic, however, demonstrated even stronger performance by identifying unique influential observations that were not detected by existing diagnostic tools. This improvement indicates that integrating robustness through M-estimation with the bias-correction mechanism of the Kibria–Lukman framework enhances the sensitivity of the estimator to subtle influential effects that classical diagnostics often overlook. The proposed measures also showed consistent performance across both real-life datasets analyzed in this study. This uniformity highlights the adaptability and strength of the KL-M estimator in delivering accurate and resilient diagnostic assessments, particularly in applied economic and industrial datasets where anomalies and high interdependence among variables are common. The empirical analyses further emphasize the superiority of the proposed diagnostic measures Cook's_KL_M and DFFITs_KL_M over existing classical and KL-based statistics. Their enhanced performance is rooted in two key structural properties of the robust M-KL estimator:

- (i) the M-estimation component, which effectively down weights outliers in both the response and predictor spaces, and
- (ii) the two-parameter shrinkage mechanism of the KL framework, which stabilizes parameter estimates in the presence of multicollinearity.
- (iii) Together, these properties enable the proposed diagnostics to remain sensitive to influential points while avoiding the masking and swamping effects commonly associated with OLS-based measures.

Across both datasets, classical diagnostics such as Cook's Distance (OLS), DFFITs (OLS), and traditional leverage measures tended to underestimate the influence of several observations. This behavior is expected since OLS-based diagnostics rely on residuals and the hat matrix that become unstable under multicollinearity. In contrast, Cook's_KL_M produced more stable and consistent assessments of global influence due to its reliance on the bias-reduced KL-M hat matrix. As a result, it successfully identified influential points even in highly correlated predictor structures. The superior performance of DFFITs_KL_M was even more remarkable. Unlike the OLS-based DFFITs which often fails to detect cases with moderate leverage but strong residual impact the robust KL-M version identified additional influential points that were confirmed to have substantial effects on fitted values and parameter stability. This improvement arises from the robustness inherent in M-estimation, which down weights extreme residuals and produces case-deletion fitted values that more accurately reflect each observation's true contribution.

Moreover, the approximate deletion formulas developed in this study proved to be both effective and computationally efficient. They produced influence values nearly identical to those obtained from exact case-deletion computations, thereby enabling fast and reliable diagnostics even for medium- and large-scale datasets. This computational advantage makes the proposed measures particularly valuable for practitioners working with high-dimensional or contamination-prone data.

Practical Guidance for Applied Researchers

The findings of this study offer several practical recommendations:

- i. Use Cook's_KL_M when the goal is to assess overall influence of each observation on the regression model.
It is especially useful when multicollinearity is present, as it avoids the inflated influence estimates that result from unstable OLS leverages.
- ii. Use DFFITs_KL_M when the interest lies in measuring the impact of individual observations on fitted values.
This measure is more sensitive to subtle influential patterns and is recommended as the primary diagnostic tool when both outliers and multicollinearity coexist.
- iii. Prefer the KL-M diagnostics over OLS-based diagnostics when data contamination is suspected or when the correlation among predictors is high.
The KL-M framework provides a more reliable assessment because it is not easily distorted by extreme values.
- iv. Cross-check influential cases identified by multiple methods.
Observations flagged by Cook's_KL_M, DFFITs_KL_M, and at least one classical measure should be examined closely, as they likely exert substantial influence on the model.
- v. Interpret influence diagnostics alongside residual, leverage, and fitted value plots.
Although the KL-M diagnostics are robust, visual inspection enhances understanding of the nature and source of influence.

Hence, the proposed diagnostics provide researchers with a more stable and accurate means of detecting influential points in regression models affected by multicollinearity and outliers. Their superior performance in real-data applications confirms their practical value and highlights their potential for broader use in applied statistical modelling.

4.4 Summary of Findings

The study focused on developing and evaluating new influential diagnostic measures within the Robust-M KL estimator framework that can effectively address the simultaneous challenges of outliers and multicollinearity in regression analysis. The key findings are as follows: The Cooks_KL_M diagnostic consistently identified the same influential cases as those detected by earlier studies using traditional Cook's Distance and KL-based diagnostics, thereby confirming its reliability and consistency. The DFFITs_KL_M diagnostic exhibited greater sensitivity, successfully detecting additional influential observations that were overlooked by existing methods. Both proposed diagnostics demonstrated stability across multiple datasets, indicating their robustness under varying data conditions. Moreover, analyses based on the Manufacturing and Longley datasets revealed that these measures effectively identified influential cases even when the effects of multicollinearity and outliers were minimized. Overall, the study concludes that the KL-M framework offers a more comprehensive and dependable approach for detecting influential cases in complex regression settings.

5 Conclusion

This study introduced two new influential diagnostic measures Cook's_KL_M and DFFITs_KL_M within the framework of the Robust-M Kibria–Lukman estimator to address the dual challenges of multicollinearity and outliers in regression analysis. By integrating the robustness properties of M-estimation with the stabilizing two-parameter shrinkage mechanism of the KL estimator, the proposed diagnostics offer a more reliable and resilient approach to detecting influential observations in contaminated and highly correlated datasets. The empirical results from the two real-life datasets clearly demonstrate the superiority of the proposed diagnostics over classical OLS-based and KL-based influential measures. Cook's_KL_M provided stable and consistent assessments of global influence even in the presence of strong interdependencies among predictors, while DFFITs_KL_M exhibited enhanced sensitivity by identifying influential cases overlooked by existing diagnostics. These improvements underscore the value of combining robustness with bias reduction to prevent masking and swamping effects that commonly undermine traditional influence measures. Furthermore, the approximate deletion formulas for Cook's_KL_M and DFFITs_KL_M proved computationally efficient, producing results nearly identical to exact case-deletion methods. This efficiency makes the proposed diagnostics suitable for routine use in medium- to large-scale data settings.

Overall, the findings of this study highlight the practical importance of adopting robust and bias-corrected diagnostic tools in regression modeling, especially in applied fields such as economics, industry, and environmental sciences, where data irregularities are common. Future research may extend these diagnostics to generalized linear models, high-dimensional settings, variable selection frameworks, and Bayesian shrinkage estimators. Such advancements will further broaden the applicability of robust KL-based diagnostic methods and strengthen their role in modern regression analysis.

Conflict of interest: Authors hereby declare that there is no conflict of interest

References

- Abidoye, A. O., Ajayi, I.M. Lukman, A. F. & Ogunjobi, J. O. (2022). Unbiased Modified Two-Parameter Estimator for the Linear Regression Model. *Journal of Scientific Research*. Vol. 14(3), pp 785- 795.
- Adejumo, T. J., Ayinde, K., Okegbade, A. I., Akomolafe, A. A., Oshuporu, O. A. and Koleoso, S.O. (2024). Robust-M Kibria-Lukman Estimator for Linear Regression Model ith Outliers In the x-direction: Simulations and Applications. *Science World Journal*. Vol. 19(2). Doi.org/10.4314/swj.v19i2.21.
- Adejumo. T. J., Amin, M., Ayinde, K. and Adewuyi, E. T. (2025). Influential diagnostics in the Linear model with Kibria-Lukman Estimator. *Communications in Statistics-Simulation and Computation*. Doi.org/10.1080/03610918.20252484628.
- Alkasadi, N. A., Ibrahim, S., Abuzaid, A. H. M., Yusoff, M. I., Hamid, H., Waozhe, L. and Abdrasak, A. (2019). Outlier Detection in Multiple circular Degression Model using DFFITs statistics. *Sains Malaysiana* 46 (7) (2019). 1557 – 1563. <http://dx.doi.org/10.17576/ism-2019-4807> – 25.
- Arumairajan, S. & Kayathiri, S. (2002). A New Stochastic Restricted Two-Parameter Estimator in Multiple Linear Regression Model. *Vavuniya Journal of Science* 1,38-47
- Asar, Y. and Erisoglu, M. (2016). Influence diagnostics in Two-parameter Ridge Regression. *Journal of Data Science* 14, (2016). 33-52. Doi: 10.6339/JDS. 201601_14(1).0003.
- Ayinde, K., Lukman, A. F. and Arowolo, O. T. (2015). Robust regression diagnostics of influential observations in linear regression model. *Open Journal of Statistics*. 5, 1- 11.
- Bagheri, A., and Midi, H. (2009). Robust Estimations as a Remedy for multicollinearity caused by multiple High leverage points. *Journal of Mathematics and Statistics* 5 (4); 311 – 321.
- Bagheri, A., Midi, H. and Imon, A. H. M. R. (2010). The effect of collinearity – influential observations on collinear Data set. A montecarlo Simulation study. *Journal od Applied Sciences*. 10 (18). 2086 -2093.
- Belsley, D. A., Kuh, E. and Welsch, R. E. (1980). Regression Diagnostics: identifying influence Data and sources of collinearity. Wiley and sons, New York. <http://dx.doi.org/10.1002/0471725153>.

- Birkes, D. and Dodge, Y. D. (1993). *Alternative methods of regression*, Wiley, New York.
- Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics*. 19; 15 – 18
- Cook, R. D. and Weisberg, S. (1982). *Residual and Influence in Regression*. Chapman and Hall, New York.
- Dawoud, I. and Hussein, E. (2025). Detection of Influential Observations for the regression model in the presence of multicollinearity: theory and methods. *Communications in Statistics - Theory and Methods*. Doi: [10.1080/03610926.2024.244910](https://doi.org/10.1080/03610926.2024.244910).
- Emami, H. and Emami, M. (2016). New influence diagnostics in ridge regression, *Journal of Applied Statistics*. 43(3). 476-489. Doi: [10.1080/02664763.1070804](https://doi.org/10.1080/02664763.1070804).
- Emami, H. and Emami, M. (2017). Influence diagnostics in Modified Liu –type Estimator. *Calcutta Statistical Association Bulletin* 68(1&2) 82 – 91. DOI: [10.1177/0008068316668426](https://doi.org/10.1177/0008068316668426).
- Gujarati, D. N. (2003). *Basic Econometrics*, New Delhi; Tata McGrawHill. New York.
- Hussein, E. (2021). Leverage and Influential observations on the Liu type estimator in the linear regression model with the severe collinearity. *Heliyon*. Doi.org/10.1016/j.heliyon.2021.e07792.
- Hussein, Y. A. and Abdalla, A. A. (2012). Generalized Two stages Ridge Regression Estimator for Multicollinearity and Autocorrelated errors. *Canadian Journal on Science and Engineering Mathematics*, 3(3), 79 - 85.
- Jahufer, A. (2013). Detecting Global influential observations in Liu Regression Model. *Open journal of statistics*. 3, 5 – 11. <http://dx.doi.org/10.4236/ojs.2013.31002>.
- Jahufer, A. and Jianbao, C. (2008). Assessing global influential observations in modified ridge regression. *Statistics and probability letters*. 79 (2008). 513 -518. <http://dx.doi.org/10.1080/00401766.1970.10488634>.
- Jegede, S.L., Lukman, A. F., Ayinde, K. and Odeniyi, K. A. (2022). Jackknife Kibria-Lukman M- Estimator: Simulation and Application. *Journal of the Nigeria Society of Physical Science*. Vol. 4, pp 251 – 264. Doi: [10.46481/jnsps.2022.664](https://doi.org/10.46481/jnsps.2022.664).
- Kashif, M., Amanullah, M. and Aslam, M. (2018). Pena’s statistic for the Liu regression. *Journal of Statistic for the Liu regression*. *Journal of Statistical Computation and Simulation* 88 (13): 2473 – 2488.
- Kashif, M., Ullah, M. A. and Aslam, M. (2019). Influential diagnostic with Pena’s statistic for the Modified ridge regression. *Communication in Statistics-Simulation and Computation*. Doi: [10.1080/03610918.2019.1634204](https://doi.org/10.1080/03610918.2019.1634204).
- Kibria, B. M. and Lukman, A. F. (2020). A new Ridge-Type Estimator for the linear Regression model. *Simulations and Applications, Hindawi scientifica Vol. 2020*. <https://doi.org/10.1155/20209758378>.

- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2004). *Applied Linear Statistical Models* (5th ed.). New York, NY: McGraw-Hill/Irwin.
- Leiva, V., Barros, M. Paula, G. and Galea, M. (2006). Influence diagnostics in log – Birnbaum banders regression models with Censored data. *Journal of Computational statistics and data analysis*. 51, (2007), 5694 – 5707.
- Liu, K. (1993). A new class of biased estimate in linear regression. *Journal of Communications in statistics. Theory and Methods*. 22:2, 393 – 402. Doi:10.1080/03610929308831027.
- Longley, J. W. (1967). An appraisal of least squares programs for electronic computer from the point of view of the use. *Journal of American Statistical Association* 62: 819 – 841.
- Lukman, A. F., Arowolo, O. and Ayinde, K. (2014). Some Robust Ridge Regression for Handling Multicollinearity and Outliers. *International Journal of Sciences. Basic and Applied Research (IJSBAR)*, 16(2), 192 -202.
- Lukman, A. F. and Ayinde, K. (2016). Detecting observations in Two-Parameter Liu-Ridge Estimator. *Journal of Data Science*. 207218, Doi:10.6339/JDS.201804_16(2).0001.
- Lukman, A. F. and Ayinde, K. (2020). Detecting influential observations in Two-parameter Liu-Ridge Estimator. *Journal of Data Science*, 16(2).0001, 201 -218. Doi:10.6339/JDS.201804.
- Lukman, A. F., Ayinde, K., Okunola, A. O., Akanbi, O. B. and Onate, C. A. (2018). Classification-Based Ridge Estimation Techniques of Alkhamisi Methods. *Journal of probability and Statistical Sciences*. 16(2), 2018. 165 – 181.
- Majeed, A, Ahmad, S., Aslam, M. & Kashif M. (2022). A robust Kibria-Lukman estimator for linear regression model to combat multicollinearity and outliers. *Concurrency and Computation: Practice and Experience*. Doi.org/10.1002/cpe.7533
- Meloun, M. and Militky, J. (2001). Detection of Single influential points in OLS regression model building. *Analytical Chimica Acta*. Doi: 1.1016/S0003.267(01)01040-6. 169 -191.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis* (6th ed.). John Wiley & Sons
- Peng, L. Y., Midi, H., Rana, S. and Fitrianto, A. (2016). Identification of multiple outliers in a Generalized linear model with continuous variables. *Journal of Mathematical problems in Engineering* . Vol. 2016. <http://dx.doi.org/10.1155/2016/5840523>.
- Rao, C. R. (1973). *Linear Statistical Inference and its applications*, 22: John Wiley and Sons.
- Seber, G. A. F., & Lee, A. J. (2012). *Linear regression analysis* (2nd ed.). John Wiley & Sons.
- Ullah, M. I., Ahmed, N., and Ahmed, S. (2013). Influential observations in Liu regression models. *Pakistan Journal*

of Statistics, 29(3), 343–356.

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

Welsch, R. E. (1982). Influence function and regression diagnostics. Modern Data Analysis. New york. Academic Press.

Welsch, R. E. and Kuh, E. (1977). Linear regression diagnostics. Technical Report 923-77. Sloan school of management, Massachusetts Institute of Technology

Yasin, A. and Murat, E. (2016). Influence Diagnostics in Two Parameter Ridge Regression. *Journal of Data Science*, 14, 33 – 52.