Behavior of Residuals in Cook's Distance for Beta Ridge Regression Model (BRRM)

JAVARIA AHMAD KHAN[a], ATIF AKBAR[a], B. M. GOLAM KIBRIA[b]
[a]Department of Statistics,
Bahuddin Zakariya University, Multan,
PAKISTAN
[b]Department of Mathematics and Statistics,
Florida International University, Miami, FL 33199,
USA

Abstract: Beta ridge regression is used to tackle the sensitivity of maximum likelihood estimation when regressors are linearly correlated in Beta generalized linear model. Cook's distance is one of the renowned and classic tools for detection of outliers. In this article, we propose to use Cook's distance with different residuals in the Beta ridge regression model. Simulated and real data are provided for illustration purposes. It has been observed that a class of weighted residuals performs better in outliers' detection but there is no impact of small or large shrinkage parameter on detection.

Key-Words: Beta regression; Cook's distance; Influence diagnostics; Multicollinearity; Outliers; Ridge regression; Residuals.

Received: October 13, 2022. Revised: September 12, 2023. Accepted: October 15, 2023. Published: November 14, 2023.

1 Introduction

The generalized linear model (GLM) is a continuation of the linear regression model for modeling a non-normal response variable through a canonical link function. GLM unified different statistical models such as the linear regression model (LRM), the Beta regression model (BRM), the Poisson regression model (PRM), Gamma regression model and others [1].

Multicollinearity is one the severe problems in multiple regression and occurred when the explanatory variables of model are highly correlated. Consequently, it makes maximum likelihood estimates unstable and inefficient [2]. Similarly, outliers are those data points which has lack of neighboring values and significantly different from other data points. Basically, outliers represent uncommon values of data. In many situations, outliers can influence the results, e.g. in term of bias. There are some statistical methods used to detect outliers, such as index plot, Cook's distance and potential residual plot, see Hadi [3]. Usually, researchers suggest removing such values but this cause loss of information [4].

Ferrari and Cribari-Neto [5] proposed to use beta regression models (BRM) for percentages, proportions, rates and fractions, to investigate the influence of a continuous variable that assumes values on the open interval (0, 1). They also proposed MLE estimators to estimate the model parameters. Espinheira et al. [6, 7] proposed some residuals and likelihood distance method for influential diagnostics. Rocha and Simas [8] generalized the Espinheira et al. [7] results and constructed some residuals, and a Portmanteau test for serial correlation. Ferrari and Pinheiro [9] and Simas et al. [10] tried to modify the MLE for the beta regression models. Anholeto et al. [11] studied the adjusted Pearson residuals and for the beta regression. Espinheira et al. [12] proposed a model selection criterion which is directly related to the leverage, residuals and influence of the observations. Pereira [13] proposed quantile residuals for BRM.

So, in literature there is a plenty of work which deals with the problem of outliers. Recently, the multicollinearity issue has been considered for beta regression models and for that many estimators have been introduced such as the ridge estimator [14, 15], modified ridge-type estimator [16], Liu estimator [17], Liu-type estimator [18], two-parameters estimator [19] and Dawoud-Kibria estimator [2]. Seifollahi and Bevrani [20] proposed James-Stein type estimators.

The available literature reveals that the impact of residual is yet not studied specifically with

Cook's distance when multicollinearity is present in the data. This article addresses that gap and will discuss how residuals effect the performance of Cook's distance in the presence of multicollinearity for BRM and its related inferences.

Such combination is also considered by other researchers, but they focus on other regression models. Like; for LRM: Lukman et al. [21], Pati et al. [22], Ibrahim and Yahya [23], Majid et al. [24, 25], Arum et al. [26] and Lukman et al. [27] considered this combination. For GLM: Arum et al. [28] considered this combination of problem and presented the robust modified jackknife ridge estimator for the Poisson regression.

The organization of the paper is as follows: The beta ridge regression model and cook's distance are discussed in section 2. A simulation study has been conducted in section 3. For illustration purposes, a real-life data is analyzed in section 4. Some concluding remarks are outlined in section 5.

2 Beta Regression Model

In this section, we will summarize the beta ridge regression model, Cook's distance and the associated residuals.

2.1 The Beta Ridge Regression Model

Let $y_i = (y_1,\ y_2, \dots, y_n)'$, $i = 1, 2, \dots, n;$ be the vector of the response variable, which independently comes from Beta (α, ϕ) distribution; with shape parameter α and scale parameter ϕ and the probability density function of the Bata distribution is given as;

$$f(y; \alpha, \phi) = \frac{y^{\alpha-1}(1-y)^{\phi-1}}{B(\alpha,\phi)}, \quad x \in (0,1), \alpha > 0, \phi > 0 \qquad (1)$$

where $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\phi)}{\Gamma(\alpha+\phi)}$. The mean and variance of beta distribution are $\frac{\alpha}{\alpha+\phi}$ and $\frac{\alpha\phi}{(\alpha+\phi)^2(\alpha+\phi+1)}$, respectively with link function $g(\mu_i) = \eta_i = x_i^T\beta = logit(\mu_i)$. $x_i^T = (1, x_{i1}, \dots, x_{ip})'$ is the matrix of $(p+1)$ explanatory variables and $\beta = (\beta_0, \beta_1, \dots, \beta_p)'$ is a vector of regression coefficients [29].

With the use of the iterative reweighted least-squares (IRLS) algorithm with initial values of β and φ as in Ferrari and Cribari-Neto [5] and Espinheira et al. [12], the Beta maximum likelihood (BML) estimator of the parameter β is provided as

$$\hat{\beta}_{BML} = (X'\widehat{W}X)^{-1}X'\widehat{W}z,$$

where $z = \hat{\eta} + \widehat{W}^{-1}\hat{T}(y^* - \mu^*)$ and $\widehat{W} = diag(\widehat{W}_1, \dots, \widehat{W}_n)$.

It is a well-known assumption in the multiple regression that the independent variables are not linearly correlated. In practical situations, explanatory variables may be linearly correlated, which cause the problem of multicollinearity [30]. In the presence of multicollinearity, confidence interval become wider, the variance of the MLE turn very large and the inference based on this estimator may not be reliable [31]. However, there are number of tools to combat multicollinearity in linear regression, such as James-Stein estimator [20], principal component estimator [26], ridge regression estimator [14], improved ridge estimators [33], modified ridge regression estimator [27], Liu-type estimator [34], restricted and unrestricted two-parameter estimator [19], mixed ridge estimator [35] and etc.

To reduce the effects of multicollinearity in the BR model, Abonazel and Taha [15] and Qasim et al. [14] introduced the BRR estimator as an alternative to the BML estimator and is given as:

$$\hat{\beta}_{BRR} = (X'\widehat{W}X + kI_p)^{-1}X'\widehat{W}z, \qquad k > 0$$

2.2 Cook's Distance and Residuals

Cook's distance (CD) was first proposed by Cook [36] for the LRM and Pregibon [37] used this technique for GLM, to identify the outlier. CD measures the overall change in the fitted model when the i^th observation is deleted from the model. The CD statistic is modified for the BRM as,

$$CD_i = \frac{(\hat{\beta}_{ML} - \hat{\beta}_{ML(i)})' X'\widehat{W}X(\hat{\beta}_{ML} - \hat{\beta}_{ML(i)})}{(k+1)\hat{\phi}}, \qquad (3)$$

where $\hat{\beta}_{ML}$ is the estimated BRM coefficients vector and $\hat{\beta}_{ML(i)}$ is the estimated BRM coefficients vector after deleting the i^{th}

observation. McCullagh and Nelder [38], simplify the equation (3) as

$$CD_i = \frac{{\pi_i}^2}{(k+1)} \frac{h_{ii}}{1-h_{ii}}, \quad (4)$$

where π_i is the i[th] residual, which is explained in Table 1. The largest value of CD indicates that the i[th] observation is the outlier. Cutoff point for the detection of outlier (s) using CD statistics in the BRRM is 2*mean (Cook's distance) [39]. We are going to consider following residuals in CD; which are summarized in Table 1.

Table 1. Summary of Residuals

Sr. #	Residual	Notation	Reference
1	Pearson Residual (P Res)	$r_t = \frac{y_t - \hat{f}(x_t)}{\sqrt{\hat{f}(x_t)}}$	Davison and Snell [40]
2	Adjusted Pearson Residuals (AP Res)	$r_t^a = \frac{r_i - \hat{E}(r_i)}{\sqrt{\widehat{Var}(r_i)}}$	Anholeto et al. [11]
3	Deviance Residual (D Res)	$r_t = sign(y_t - \mu_t)\sqrt{d_t}$	Davison and Snell [40]
4	Working Residual (W Res)	$r_t = \frac{y_t - \hat{f}(x_t)}{\hat{f}(x_t)}$	Davison and Snell [40]
5	Response Residual (R Res)	$r_t = y_t - \hat{f}(x_t)$	Davison and Snell [40]
6	Weighted Residual (W Res)	$r_t^* = \frac{y_t^* - \hat{\mu}_t^*}{\sqrt{\phi v_t}},$	Espinheira et al. [6]
7	Standardized Weighted Residual (SW Res)	$r_t^\omega = \frac{y_t^* - \hat{\mu}_t^*}{\sqrt{v_t}}$	Espinheira et al. [6]
8	Adjusted Standardized Weighted Residual (ASW Res)	$r_t^{\omega a} = \frac{r_i - \hat{E}(r_t^\omega)}{\sqrt{\widehat{Var}(r_t^\omega)}}$	Anholeto et al. [11]
9	Standardized Weighted 2 Residual (SW2 Res)	$r_t^{\omega\omega} = \frac{r_t^\omega}{\sqrt{(1 - h_{tt})}}$	Espinheira et al. [6]

3 Simulation

The following scheme is considered for the generation of simulated datasets:

I. The dependent variable of the BRM is generated from Beta distribution as $y_i = B(\alpha, \phi)$ for $i = 1, 2, \ldots, n$, where $\mu_i = E(y_i) = 3$ is arbitrary mean and $\phi = 5$ is dispersion parameter.

II. Two explanatory variables x_1 and x_2 are kept fixed through the whole simulation study. To introduce multicollinearity, we followed Saleem et al. [41] and Hussain and Akbar [42]. They used $x_{ij} = \sqrt{1-\rho^2} Z_{ij} + \rho Z_{ij}(p+1),\ i = 1, 2, \ldots, n; j = 1, 2$ where $Z_{i1}, Z_{i2}, Z_{i3}, Z_{i4}$ are independent standard normal pseudo random numbers, and ρ is defined as degree of multicollinearity, $\rho = 0.8, 0.9, 0.95$.

III. Then, we introduce the outliers in X_{ij} at 5th, 10th, 15th, 20th, and 25th point as $x_{ij} = a_0 + x_{ij}$, for $i = 5, 10, 15, 20, 25; j = 1, 2$, where $a_0 = \bar{x}_j + 100$.

IV. Sample size is considered $n = 25, 50, 100, \text{and } 200$ with 1000 replications.

V. We used detection rate in percentage.

The simulated outlier detection rate in percentage of the BRM Cook's distance under considering different factors such as sample size, dispersion parameter, levels of multicollinearity and different type of residuals are presented in Tables 2-4 for k=0.2, 0.5 and 0.9 respectively. For a better picture, we obtained the average detection rate along with standard deviation for all methods over all n and ρ and presented them in the last two rows of Tables 2- 4.

Table 2: Estimated outlier detection rate (%) of the BRM influence diagnostics with different residuals when k = 0.2

n	ρ	P Res	AP Res	D Res	R Res	W Res	SW Res	ASW Res	SW2 Res	W Res
25	0.8	19.8	15.3	23.7	20.2	22.5	22.5	22.2	22.4	18.0
	0.9	24.6	19.1	26.2	24.3	25.2	25.2	26.6	25.4	21.5
	0.95	23.6	18.7	27.4	22.9	24.5	24.5	25.3	23.9	21.1
50	0.8	67.7	64.5	75.5	69.9	80.7	80.7	78.6	71.7	62.4
	0.9	81.9	76.0	85.7	83.1	87.6	87.6	86.3	83.5	74.0
	0.95	86.1	81.8	89.4	87.3	91.4	91.4	90.4	89.5	81.7
100	0.8	98.3	97.9	98.7	98.7	99.3	99.3	99.5	99.0	96.0
	0.9	98.4	98.5	99.2	99.3	99.3	99.3	99.5	99.0	97.6
	0.95	98.0	97.7	99.2	98.4	99.5	99.5	99.8	98.9	96.3
200	0.8	100.0	100.0	99.8	100.0	99.9	99.9	99.9	99.8	99.5
	0.9	99.9	99.9	100.0	100.0	100.0	100.0	100.0	99.9	98.9
	0.95	99.9	99.8	100.0	99.9	100.0	100.0	100.0	100.0	99.3
Mean	74.9	72.4	77.1	75.3	77.5	77.5	77.3	76.1	72.2	74.9
SD	28.0	29.9	27.1	28.3	27.9	27.9	27.5	27.9	28.8	28.0

Table 3: Estimated outlier detection rate (%) of the BRM influence diagnostics with different residuals when k = 0.5

n	ρ	P Res	AP Res	D Res	R Res	W Res	SW Res	ASW Res	SW2 Res	W Res
25	0.8	22.8	12.8	24.9	22.9	25.5	25.5	20.1	24.5	16.6
	0.9	15.5	8.4	22.1	16.9	23.2	23.2	14.8	20.1	10.9
	0.95	17.6	9.2	21.8	18.0	24.3	24.3	21.9	20.5	13.0
50	0.8	87.5	72.5	89.0	90.0	89.7	89.7	85.4	88.5	80.5
	0.9	80.6	68.1	85.6	82.6	89.4	89.4	84.6	85.4	71.4
	0.95	86.1	70.4	88.3	88.0	91.1	91.1	86.8	88.4	77.4
100	0.8	99.4	96.1	99.7	99.7	99.6	99.6	99.0	99.5	97.6
	0.9	98.5	94.4	99.5	99.0	99.6	99.6	98.9	99.6	94.3
	0.95	98.9	95.2	99.7	99.4	99.9	99.9	99.4	99.7	96.7
200	0.8	100.0	99.5	100.0	100.0	100	100	100.0	100.0	99.6
	0.9	99.9	99.8	100.0	99.9	100	100	99.9	100.0	99.7
	0.95	99.9	99.7	99.9	99.9	100	100	100.0	100.0	99.2
Mean	75.6	68.8	77.5	76.4	78.5	78.5	75.9	77.2	71.4	75.6
SD	35.0	37.3	33.4	34.9	33.0	33.0	34.9	33.9	36.2	35.0

Table 4: Estimated outlier detection rate (%) of the BRM influence diagnostics with different residuals when k = 0.9

n	ρ	P Res	AP Res	D Res	R Res	W Res	SW Res	ASW Res	SW2 Res	W Res
25	0.8	21.5	12.1	22.8	20.5	23.4	23.4	23.5	22.4	19.9
	0.9	22.0	10.8	23.4	21.7	26.8	26.8	22.1	24.2	16.2
	0.95	20.3	8.5	23.4	20.8	26.4	26.4	23.5	27.2	17.4
50	0.8	93.0	65.4	93.9	94.7	93.5	93.5	85.6	92.7	84.3
	0.9	89.7	60.0	91.2	92.4	93.7	93.7	81.7	91.8	78.4
	0.95	81.9	54.5	88.6	84.7	91.9	91.9	76.4	89.3	70.4
100	0.8	98.3	88.9	99.2	98.7	99.7	99.7	97.8	99.4	95.1

	0.9	99.7	90.7	99.7	99.9	100.0	100.0	98.4	99.8	96.7
	0.95	99.0	89.7	99.9	99.5	99.9	99.9	98.9	99.7	95.7
200	0.8	100.0	99.1	100	100	100	100	100	100	99.5
	0.9	100.0	99.2	100	100	100	100	100	100	99.5
	0.95	100.0	99.4	100	100	100	100	100	100	99.7
Mean	77.1	64.9	78.5	77.7	79.6	79.6	75.7	78.9	72.7	77.1
SD	34.1	36.2	33.6	34.5	32.8	32.8	32.7	33.0	34.4	34.1

Tables 2-4 shown the performance of Cook's distance in BRRM. Here above mentioned three different shrinkage parameters are considered, which showed very low impact on detection percentage for all methods except Pearson residual. It is interesting to note that rate of identification of outliers is highly influenced by sample size; percentage increases when sample size increased and detected all outliers when sample size is very large. This behavior is almost same for all the residuals.

By examining the results, according to ρ (levels of multicollinearity) it can be observed that it does not affect the detection rate. It may be noticed that for large sample size all residuals performed almost equally good but for small sample size weighted, standardized weighted and adjusted standardized weighted perform better.

In our study, SWeighted2 residual is one of the best options which performs good with Cook's distance in detection of outliers, also Espinheira et al. [6] showed that SWeighted2 residual is best choice to be used in likelihood displacement (LD).

4 Application: Crude Oil Conversion Data

This empirical application is based on a data set from Prater [51]. It has four explanatory variables, first is the crude oil gravity (x_1), which is measured using the index suggested by the American Petroleum Institute and this variable measure the density of a liquid. Second is the vapour pressure of the crude oil (x_2) and this variable is measured using the Reid vapour pressure defined as the pressure needed to keep a liquid from vaporizing at 100 degrees Fahrenheit. Third is the temperature (degrees Fahrenheit) at which 10 percent of crude oil has vaporized (x_3) and the temperature (degrees Fahrenheit) at which all the gasoline is vaporized (x_4). The proportion of crude oil converted to gasoline after distillation and fractionation is a dependent variable (y).

Atkinson [43] used LRM to analyzed this data set and examined that the error term is not symmetrical and transformed the dependent variable. Then, Lemonte et al. [44] used this data set and considered that dependent variable follows a beta distribution. Ferrari and Cribari-Neto [5] used the data for detection of outliers and found observation 4 as an influential. Then Qasim et al. [14] used the same data set and showed that there exist some multicollinearity issues, especially between variables x_2 and x_3.

Now, we considered both problems (outliers and multicollinearity) and examine the impact of different residuals in cook's distance. Our considered residuals mentioned in Table 1 are determined by using Beta ridge regression model.

Table 5: Detection of outliers using different residuals varying shrinkage parameter.

Residuals \ *k*	**0.1**	**0.5**	**0.9**
P Res	1, 2, 14, 31	1, 2, 14, 31	1, 2, 14, 31
AP Res	1, 2, 14, 31	1, 2, 14, 31	1, 2, 14, 31
D Res	1, 2, 14, 31	1, 2, 14, 31	1, 2, 14, 31
W Res	1, 2, 11, 21, 25, 31	1, 2, 11, 21, 25, 31	1, 2, 11, 21, 25, 31
R Res	2, 3, 7, 14, 31	2, 3, 7, 14, 31	2, 3, 7, 14, 31
W Res	1, 2, 14, 21, 31	1, 2, 14, 21, 31	1, 2, 14, 21, 31
SW Res	1, 2, 14, 21, 31	1, 2, 14, 21, 31	1, 2, 14, 21, 31
ASW Res	1, 2, 14, 21, 31	1, 2, 14, 21, 31	1, 2, 14, 21, 31
SW2 Res	1, 2, 14, 21, 31	1, 2, 14, 21, 31	1, 2, 14, 21, 31

Table 5, presents the outliers by using different residuals and shrinkage parameter (k). It is evident that results are not influenced by k. Detected outliers are same by using any k. All weighted residuals detect same outliers but working residual detect maximum numbers of outliers (6). When we use response residual in Cook's distance, then quite different outliers are detected.

5 Some Concluding Remarks

This paper considers the Cook's distance for the BRRM with different residuals. Comparisons of residuals with Cook's distance are assessed through a simulation study and by a real data set which yielded important conclusions. Firstly, Cook's distance for BRRM can be helpful to determine the choice of residual related to motive of the study. If the goal of the researcher is to detect the outlier (s) then class of weighted residual is a best choice. Chances of detection are high for large sample sizes. It is also observed that detection of outliers is not affected by shrinkage parameter, which is evident from both simulation study and a real data analysis. In future, impact of these outliers on coefficients can be examined to observe their influence.

References:

[1]. Nelder J, Wedderburn R, Generalized Linear Models, *Journal of the Royal Statistical Society*, *Series A* (General*)*, Blackwell Publish, Vol.135, No.3, 1972, pp. 370-384.

[2]. Abonazel MR., Dawoud I, Awwad FA, Lukman AF, Dawoud–Kibria Estimator for Beta Regression Model: Simulation and Application, Frontiers in *Applied Mathematics* and *Statistics*, Vol.8, 2022, pp.775068.

[3]. Hadi AS, A New Measure of Overall Potential Influence in Linear Regression, *Computational Statistics and Data Analysis*, Vol.14, No.1, 1992, pp. 1–27.

[4]. Khan JA, Akbar A, Empirical Performance of Nonparametric Regression over LRM and IGRM Addressing Influential Observations, *Journal of Chemometrics*, Vol.33, No.7, 2019, pp. e3143.

[5]. Ferrari S, Cribari-Neto F, Beta Regression for Modelling Rates and Proportions, *Journal* of *Applied Statistics*, Vol.31, No.7, 2004, pp. 799–815.

[6]. Espinheira PL, Ferrari SL, Cribari-Neto F, On Beta Regression Residuals. *Journal of Applied Statistics*, Vol.35, No.4, 2008a, pp. 407-419.

[7]. Espinheira PL, Ferrari SL, Cribari-Neto F, Influence diagnostics in beta regression. *Computational Statistics & Data Analysis*, Vol.52, No.9, 2008b, pp. 4417-4431.

[8]. Rocha AV, Simas AB, Influence Diagnostics in A General Class of Beta Regression Models, *Test*, Vol.20, 2011, pp. 95-119.

[9]. Ferrari SL, Pinheiro EC, Improved Likelihood Inference in Beta Regression, *Journal of Statistical Computation and Simulation*, Vol.81, No.4, 2011, pp. 431-43.

[10]. Simas AB, Barreto-Souza W, Rocha AV, Improved Estimators for A General Class of Beta Regression Models, *Computational Statistics & Data Analysis*, Vol.54, No.2, 2010, pp. 48-66.

[11]. Anholeto T, Sandoval MC, Botter DA, Adjusted Pearson Residuals in Beta Regression Models, *Journal of Statistical Computation and Simulation*, Vol.84, No. 5, 2014, pp. 999-1014.

[12]. Espinheira PL, da Silva LCM, Silva ADO, Ospina R. Model Selection Criterion Beta Regression for Machine Learning, *Machine Learning* and *Knowledge Extraction,* Vol.1, 2019, pp. 427–49.

[13]. Pereira GHA, On Quantile Residuals in Beta Regression, *Communications in Statistics-Simulation and Computation,* Vol.48, No.1, 2019, pp. 302-316.

[14]. Qasim M, Månsson K, Kibria BMG, On Some Beta Ridge Regression Estimators: Method, Simulation and Application, *Journal of Statistical Computation and Simulation*, Vol.91, 2021, pp. 1699–712.

[15]. Abonazel MR, Taha IM., Beta Ridge Regression Estimators: Simulation and Application, Communications *in Statistics - Simulation* and *Com putation*, Vol.52, No.9, 2023, pp. 4280-4292.

[16]. Swindel BF, Good Ridge Estimators Based on Prior Information, *Communications* in Statistics *-Theory and Method*, Vol.5, No.11, 1976, pp. 1065–1075.

[17]. Kejian L, A New Class of Blased Estimate in Linear Regression, *Communications* in *Statistics -Theory and Method,* Vol.22, No.2, 1993, pp. 393–402.

[18]. Kejian L, Using Liu-type Estimator to Combat Collinearity, *Communications* in *Statistics -Theory and Method*, Vol.32, No.5, 2003, pp. 1009–1020.

[19]. Özkale MR, Kaciranlar S, The Restricted and Unrestricted Two-Parameter Estimators, *Communications* in Statistics *-Theory and Method,* Vol. 36, No. 15, 2007, pp. 2707–2725.

[20]. Seifollahi S, Bevrani H, James-Stein Type Estimators in Beta Regression Model: Simulation and Application, *Hacettepe Journal of Mathematics and Statistics*, 2023, pp. 1-20.

[21]. Lukman A, Arowolo O, Ayinde K. Some robust ridge regression for handling multicollinearity and outlier,

International Journal of Sciences: Basic and Applied Research, Vol.16, No.2, 2014, pp. 192-202.

[22]. Pati KD, Using standard error to find the best robust regression in presence of multicollinearity and outliers. In *2020 International Conference on Computer Science and Software Engineering (CSASE)*, *IEEE*, 2020, pp. 266-271

[23]. Ibrahim SA, Yahya WB, Effects of Outliers and Multicollinearity on Some Estimators of Linear Regression Model, In *Edited Proceedings of 1st International Conference*, Vol.1, 2017, pp. 204-209.

[24]. Majid A, Aslam M, Ahmad S, Altaf S, Afzal S, Robust estimation of the distributed lag model with multicollinearity and outliers. *Communications in Statistics-Simulation and Computation*, 2022, pp. 1-15.

[25]. Majid A, Ahmad S, Aslam M, Kashif M, A robust Kibria–Lukman estimator for linear regression model to combat multicollinearity and outliers. *Concurrency and Computation: Practice and Experience*, Vol.35, No.4, 2023, pp. e7533.

[26]. Arum KC, Ugwuowo FI, Oranye HE, Alakija TO, Ugah TE, Asogwa OC, Combating Outliers and Multicollinearity in Linear Regression Model using Robust Kibria-Lukman Mixed with Principal Component Estimator, Simulation and Computation, *Scientific African*, 2023, pp. e01566.

[27]. Lukman AF, Ayinde K, Binuomote S, Clement OA, Modified Ridge-Type Estimator to Combat Multicollinearity: Application to Chemical Data, *Journal of Chemometrics*, Vol.33, No.5, 2019, pp. e3125.

[28]. Arum KC, Ugwuowo FI, Oranye HE, Robust Modified Jackknife Ridge Estimator for the Poisson Regression Model with Multicollinearity and outliers, *Scientific African*, 2022, pp. e01386.

[29]. Forbes C, Evans M, Hastings N, Peacock B, Statistical Distributions, John Wiley & Sons, 2011.

[30]. Frisch R, *Statistical Confluence Analysis by Means of Complete Regression Systems* (Vol. 5), Universitetets Økonomiske Instituut, 1934.

[31]. Qasim M, Kibria BMG, Månsson K, Sjölander P, A New Poisson Liu Regression Estimator: Method and Application, *Journal* of *Applied Statistics*, Vol.47, No.12, 2020, pp. 2258–2271.

[32]. Amin M, Akram MN, Amanullah M, On the James-Stein Estimator for the Poisson Regression Model, *Communications in Statistics-Simulation and Computation*, Vol.51, No.10, 2020, pp. 5596-608.

[33]. Singh B, Chaubey YP, On Some Improved Ridge Estimators, *Statistische Hefte*, Vol.28, No.1, 1987, pp. 53–67.

[34]. Alheety MI, Kibria BMG, Modified Liu-type Estimator Based on (r− k) Class Estimator, *Communications in Statistics - Theory and Methods*, Vol.42, No.2, 2013, pp. 304–319.

[35]. Li Y, Yang H, A New Stochastic Mixed Ridge Estimator in Linear Regression Model, S*tatistical Papers*, Vol.51, No.2, 2010, pp. 315–323.

[36]. Cook RD, Detection of Influential Observation in Linear Regression. *Technometrics*, Vol.19, No.1, 1977, pp. 15-18.

[37]. Pregibon D, Logistic Regression Diagnostics. *The Annals of Statistics*, Vol. 9, No. 4, 1981, pp. 705-724.

[38]. McCullagh P, Nelder JA, Generalized Linear Models, Chapman and Hall, London, 1989.

[39]. Hardin JW, Hilbe JM, *Generalized Linear Models and Extensions*, Stata press, 2007.

[40]. Davison AC, Snell EJ (ed.) *Residuals and Diagnostics. In: Statistical Theory and Modelling*, Chapman & Hall, 1991.

[41]. Saleem N, Akbar A, Imon AR, Al Mamun ASM, Detection of Outliers in Binomial Regression Using CERES and Partial Residual Plots, *Journal of Statistical Modeling & Analytics (JOSMA)*, Vol.4, No.2, 2022.

[42]. Hussain Z, Akbar A. Diagnostics through Residual Plots in Binomial Regression Addressing Chemical Species Data, *Mathematical Problems in Engineering*, 2022.

[43]. Atkinson AC, *Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*, New York: Oxford University Press, 1985.

[44]. Lemonte AJ, Ferrari SL, Cribari-Neto F, Improved Likelihood Inference in Birnbaum–Saunders Regressions, *Computational Statistics* and *Data Analysis*, Vol.54, No.5, 2010, pp. 1307–1316.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

Javaria Ahmad Khan: conceived the idea, programming, writing, performed the computations and final approval of the version to be published. Atif Akbar: editing, proof reading and final approval for publication. B M Golam Kibria: proofreading, formatting and critical review.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.