

Research Article

Influence Diagnostic Methods in the Poisson Regression Model with the Liu Estimator

Aamna Khan ¹, Muhammad Amanullah,¹ Muhammad Amin ², Randa Alharbi ³,
Abdisalam Hassan Muse ⁴ and M. S. Mohamed⁵

¹Department of Statistics, Bahauddin Zakariya University, Multan, Pakistan

²Department of Statistics, University of Sargodha, Sargodha, Pakistan

³Department of Statistics, Faculty of Science, University of Tabuk, Tabuk, Saudi Arabia

⁴Department of Mathematics (Statistics Option) Programme, Pan African University,
Institute of Basic Science, Technology and Innovation (PAUSTI), Nairobi 6200-00200, Kenya

⁵Department of Mathematics, College of Science, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia

Correspondence should be addressed to Abdisalam Hassan Muse; abdisalam.hassan@amoud.edu.so

Received 16 August 2021; Accepted 30 August 2021; Published 9 September 2021

Academic Editor: Ahmed Mostafa Khalil

Copyright © 2021 Aamna Khan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

There is a long history of interest in modeling Poisson regression in different fields of study. The focus of this work is on handling the issues that occur after modeling the count data. For the prediction and analysis of count data, it is valuable to study the factors that influence the performance of the model and the decision based on the analysis of that model. In regression analysis, multicollinearity and influential observations separately and jointly affect the model estimation and inferences. In this article, we focused on multicollinearity and influential observations simultaneously. To evaluate the reliability and quality of regression estimates and to overcome the problems in model fitting, we proposed new diagnostic methods based on Sherman–Morrison Woodbury (SMW) theorem to detect the influential observations using approximate deletion formulas for the Poisson regression model with the Liu estimator. A Monte Carlo method is done for the assessment of the proposed diagnostic methods. Real data are also considered for the evaluation of the proposed methods. Results show the superiority of the proposed diagnostic methods in detecting unusual observations in the presence of multicollinearity compared to the traditional maximum likelihood estimation method.

1. Introduction

Nowadays, there are several distributions available in the literature that can be used to remove noise and then predict data. Similarly, there is a persistent record of concern in modeling count data which has several applications in biosciences and other disciplines [1–4]. The focus of this effort is on dealing with the issues that occur after modeling the count data. For the prediction and analysis of count data, it is valuable to study the factors that influence the performance of the model and the decision based on the analysis of that model. Considering the suitable statistical modeling, when the dependent variable is count data, one of the most used statistical models is the Poisson regression

model (PRM). For accurate statistical inferences, the standard ordinary least square (OLS) regression sets some important assumptions related to the model's errors [5]. Generally, numerous problems may arise when a count variable model is estimated using the OLS method, because of the level of noise. For the analysis of count data, PRM provides the most relevant results. According to McCullagh and Nelder [6], the PRM belongs to the family of GLM. The maximum likelihood ML estimation method is used to estimate the PRM estimates instead of the OLS method.

In the PRM, when the explanatory variables are linearly correlated, then the ML method is very sensitive [7]. Some biased estimators were introduced in the literature to handle the multicollinearity, i.e., Stein, ridge, Lasso, regularization,

and Liu estimators; see [1, 3] and [8–10] for more details. The most popular one is the ridge estimator, but it has some limitations, i.e., selecting the ridge parameter, where the ridge rule is based on two normal distributions. It is a shrinkage rule because it depends on the slope. In contrast, Lasso is based on the slope and the intercept. The best choice is to adopt a Liu estimator to avoid the hindrances of the ridge estimator. The Liu estimator is an ace in this regard as it avoids the disadvantages of the ridge estimator [10], where the main advantage of the ridge is easy to use, and it can be written in the explicate and the objective formulas. In the literature, various studies are available for the PRM to overcome the presence of collinearity [7, 11–16].

To evaluate the reliability and quality of regression estimates and to overcome the problems in model fitting, diagnostic techniques have been developed. Although regression diagnostics have been developed methodologically and theoretically for linear regression models together with multicollinearity (see [17–24]), some studies about the influence diagnostics in the GLM with uncorrelated explanatory variables are available in the literature. Pregibon [25] proposed the influence diagnostics for logistic regression using the one-step methods. For further discussion on influential diagnostics about GLM, see [26–32].

Influence diagnostics in the GLM with correlated explanatory variables is very limited. Özkale et al. [33] proposed the first study on influence diagnostics for logistic ridge regression. Amin et al. [34] worked on the influence diagnostics for the gamma ridge regression model. Khan et al. [35] assessed the performance of influence diagnostics in the PRM with a ridge estimator. Recently, Khan et al. [36] examined the superiority of influence diagnostics in the PRM with two-parameter estimator and, further, Amin et al. [37] discussed the influence diagnostics for the inverse Gaussian ridge regression model.

The available literature showed that no study in the GLM is available for influence diagnostics with the Liu estimator. Though, the Poisson Liu regression (PLR) diagnostics have got no thoughtful attention up till now. Thus, our present work is an effort to fill this gap. So, in the present work, we proposed diagnostic methods for the PRM under the Liu estimator, which prove to be the competed method.

The remaining of the study is organized as follows: we focused on the formulation of influence diagnostic measures for the PRM under the Liu estimator (LE). Next, in Sections 4 and 5, we conducted a Monte Carlo study using two, four, and six independent variables to examine the level of detection percentage of newly developed diagnostic measures and, finally, we proved the efficacy of proposed measures with the help of real world application.

1.1. Model Specification and Estimator. Suppose the model can be written as

$$y = X\beta + \varepsilon, \quad (1)$$

where $y = \{y_i; i = 1, 2, \dots, n\}$ are the observation, $X = \{x_{ij}; i = 1, 2, \dots, n, j = 1, 2, \dots, p\}$ is a matrix, $\beta = \{\beta_i; i = 1, 2, \dots, p\}$ are the unknown parameters, and

$\varepsilon \sim N_n(0, I\sigma^2)$ with ε_i and $\varepsilon_j (i \neq j)$ being independent. We assume the observation is the result of the integration form $(X\beta)$ and try to solve this problem by finding differentiation matrix. The PRM is applicable for real data, especially when the response variable y_i often comes in the form of count data that are known. Let y_i follow a Poisson distribution with μ_i , as its parameter. The probability mass function for PRM is used to describe the relationship when y_i , the response variable occurs as count data.

$$f(y_i; \mu_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} \mu_i > 0, \quad y = 0, 1, 2, \dots \quad (2)$$

The PRM belongs to the GLM with log link function as

$$\ln(\hat{\mu}) = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p, \quad (3)$$

where b_0 is the intercept and b_1, b_2, \dots, b_p are the set of coefficients. The estimated mean function is defined by $\hat{\mu}_i = \exp(x_i' \beta)$.

Here x_i is the i^{th} row of independent variable $X_{n \times p}$ and $\beta_{p \times 1}$ of coefficients, where p represents the number of explanatory variables.

Assume that all y_i are independent; then, the joint log-likelihood is defined as

$$l(\mu; y_i) = \sum_{i=1}^n \left(y_i \ln(\exp(x_i' \beta)) - \exp(x_i' \beta) - \ln \left(\prod_{i=1}^n y_i! \right) \right), \quad (4)$$

For finding the best value of β , we have to solve the following relation:

$$\frac{\partial l(\mu; y_i)}{\partial \beta} = 0. \quad (5)$$

Since the systems of equations are nonlinear, so the MLE with iterative reweighted least-squared (IRLS) algorithm is used to estimate the regression coefficients as explicit formulas:

$$\hat{\beta}_{\text{ML}} = (X' \hat{W} X)^{-1} X' \hat{W} \hat{z}, \quad (6)$$

where $\hat{W} = \text{diag}(\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_n)$ and $\hat{z}_i = \log(\hat{\mu}_i) + (y_i - \hat{\mu}_i / \hat{\mu}_i)$.

In the presence of multicollinearity, the $X' \hat{W} X$ matrix becomes ill conditioned, and because of this problem, it gets complicated to draw effective inferences. To overcome these effects of multicollinearity, we use the generalization of Liu [6] to define PLRE.

$$\hat{\beta}_d = (X' \hat{W} X + I)^{-1} (X' \hat{W} X + dI) \hat{\beta}_{\text{ML}}, \quad (7)$$

where $0 \leq d \leq 1$. Here, the important step is selecting shrinkage parameter d as the optimal value of d which affects the performance of PLRE. Furthermore, if $d = 1$, then $\hat{\beta}_{\text{ML}} = \hat{\beta}_d$. Recently, Qasim et al. [38] recommended the optimum Liu parameter for the Liu estimator in the PRM as

$$\begin{aligned} \hat{d}_1 &= \max(0, \min(m_j)), \quad j = 1, 2, \dots, p, \\ \hat{d}_2 &= \max(0, \min(h_j)), \end{aligned} \quad (8)$$

where $m_j = (\hat{\alpha}_j^2 - 1/\max(1/\hat{\lambda}_j) + \hat{\alpha}_j^2)$ and $h_j = (\hat{\alpha}_j^2 - 1/\max(1/\hat{\lambda}_j) + \hat{\alpha}_{\max}^2)$ and where $\hat{\alpha}_j^2$ is the j^{th} element of $\gamma\beta$ and the columns of orthogonal matrix γ represent the eigenvectors of $X^t\widehat{W}X$ matrix, such that $X^t\widehat{W}X = \gamma^t\Lambda\gamma$, where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$.

2. The PRM Diagnostics

2.1. Hat Matrix, Leverage, and Residuals for the PRM. Hat matrix H is a common measure used to compute leverages. According to Davison and Tsai [39], the hat matrix H in the PRM is

$$H = \widehat{W}^{1/2} X (X^t \widehat{W} X)^{-1} X^t \widehat{W}^{1/2}. \quad (9)$$

The diagonal elements of H are interpreted as leverages, i.e., $h_{ii} = \text{diag}(H)$. For the computation of regression

diagnostic measures, residuals play the most important role (Belsley et al. [18]). Let χ_i symbolize the Pearson residual, so for the case of PRM, we defined it as

$$\chi_i = \frac{y_i - \exp(x_i^t \widehat{\beta}_{\text{ML}})}{\sqrt{\exp(x_i^t \widehat{\beta}_{\text{ML}})}}, \quad (10)$$

Similarly, we find the standardized Pearson residual as

$$\chi'_i = \frac{\chi_i}{\sqrt{(1 - h_{ii})}}. \quad (11)$$

Another useful residual that proves to be of great help for detecting unusual observations is termed as the deviance residual. The i^{th} deviance residual for the PRM is defined by

$$d_i = \text{sign}(y_i - \exp(x_i^t \widehat{\beta}_{\text{ML}})) \sqrt{2 \left[y_i \log \left(\frac{y_i}{\exp(x_i^t \widehat{\beta}_{\text{ML}})} \right) - (y_i - \exp(x_i^t \widehat{\beta}_{\text{ML}})) \right]}, \quad (12)$$

where the sign is the sign function [31].

2.2. Influence Diagnostic Methods. Pregibon [25] was the first to work on logistic regression diagnostics tool and proposed the influence diagnostic measures using one-step approximations. The proposed influence diagnostics take account of Cook's distance, change in deviance, and change in Pearson χ^2 . For PRM Cook's distance, C_i is suggested as

$$C_i = \frac{(\widehat{\beta}_{\text{ML}} - \widehat{\beta}_{\text{ML}(i)})' X^t \widehat{W} X (\widehat{\beta}_{\text{ML}} - \widehat{\beta}_{\text{ML}(i)})}{(p+1)}. \quad (13)$$

The i^{th} C_i measures the overall change in the fitted model when the i^{th} observation is deleted from the model. The one-step approximation for the expression $(\widehat{\beta}_{\text{ML}} - \widehat{\beta}_{\text{ML}(i)})$ is defined as

$$\Delta \widehat{\beta}_{\text{ML}(i)} = (\widehat{\beta}_{\text{ML}} - \widehat{\beta}_{\text{ML}(i)}) = \frac{(X^t \widehat{W} X)^{-1} x_i \widehat{W}_{ii}^{1/2} \chi_i}{(1 - h_{ii})}, \quad (14)$$

where \widehat{W}_{ii} are the i^{th} diagonal elements of weight matrix after the removal of i^{th} observation. Furthermore, (13) can also be approximated as

$$C_i = \frac{\chi_i^2}{(p+1)} \frac{h_{ii}}{(1 - h_{ii})}. \quad (15)$$

Hardin and Hilbe [40] suggested the cut point for detecting the unusual observations in GLM as $(4/(n-1))$; this process is used to specify the window in GLM.

Pregibon [25] suggested change in Pearson χ^2 as another influence diagnostic measure to detect the influential observations. Applying one-step approximation, we defined $\Delta \chi_i^2$ as

$$\Delta \chi_i^2 = \chi_{(i)}^2 - \chi_0^2 = \frac{\chi_i^2}{1 - h_{ii}}, \quad (16)$$

where χ_0^2 is used to represent the squared Pearson residuals of the complete data set and $\chi_{(i)}^2$ signifies the squared Pearson residuals of the data set without the i^{th} observation, respectively. This statistic is employed to study the effect of i^{th} observation on the goodness of fit of the model and the estimates. On similar grounds, Pregibon [25] suggested that another statistic for measuring the impact of i^{th} observation on the goodness of fit of a model is the change in deviance statistic. The one-step linear approximation for change in deviance statistic is defined as

$$\Delta d_i^2 = d_{(i)}^2 - d_0^2 = d_i^2 + \frac{\chi_i^2 h_{ii}}{1 - h_{ii}}, \quad (17)$$

where for complete data set d_0 is used to represent the squared deviance residuals and the squared deviance residual $d_{(i)}$ are found after the removal of i^{th} observation, respectively. We suggested a simplified form of equation (17) by replacing χ_i^2 by d_i^2 as

$$\Delta d_i^2 = \frac{d_i^2}{1 - h_{ii}}. \quad (18)$$

The cut-off value for change in deviance statistic is 3.84 to detect the unusual observations [25].

The difference of fits (DFFITS) suggested by Belsley et al. [18] is another common influence measure. After deleting the i^{th} observation, DFFITS assesses the change in fit of model. For GLM, it is given as

$$\text{DFFITS}_i = \frac{\widehat{\mu}_i - \widehat{\mu}_{i(i)}}{\sqrt{h_{ii}}}, \quad (19)$$

where $\hat{\mu}_i$ is used to represent the predicted regressand of complete data set and $\hat{\mu}_{i(i)}$ represents the predicted regressand after deleting the i^{th} case. Furthermore, it can also be written as

$$\text{DFFITS}_i = \frac{\sqrt{W_{ii}}x_i'(\hat{\beta}_{\text{ML}} - \hat{\beta}_{\text{ML}(i)})}{\sqrt{h_{ii}}}. \quad (20)$$

By using the SMW theorem, (20) is retransformed as

$$\text{DFFITS}_i = |t_i| \sqrt{\frac{h_{ii}}{1 - h_{ii}}}, \quad (21)$$

where $t_i = \chi_i' \sqrt{(n - p - 1/n - p - \chi_i'^2)}$ is termed as the jackknife Pearson residual and $\text{DFFITS}_i > 2\sqrt{(p + 1)/n}$ shows that the i^{th} observation as influential. The second matrix will be introduced in the next section.

3. Influence Measures in Poisson Liu Regression Model (PLRM)

3.1. Hat Matrix, Leverages, and Residual in PLRM. Hat matrix H_d for the PLRM is defined as

$$H_d = X\hat{W}^{1/2}(X'\hat{W}X + I)^{-1}(X'\hat{W}X + dI)(X'\hat{W}X)^{-1}W^{1/2}X'. \quad (22)$$

The leverages are the Liu hat diagonals that proved helpful in detecting influential cases with some modifications. As for $d > 0$, $h_{di} < h_i$ for $i = 1, 2, \dots, n$ and as d increases, h_{di} decreases monotonically.

Using the Liu estimator, the Pearson residuals for PLRM are defined as

$$\chi_{di} = \frac{y_i - \exp(x_i'\hat{\beta}_d)}{\sqrt{\exp(x_i'\hat{\beta}_d)}}. \quad (23)$$

The standardized form of Pearson residuals with multicollinear independent variables is given as

$$\chi_{di}' = \frac{\chi_{di}}{\sqrt{(1 - h_{dii})}}. \quad (24)$$

3.2. Influence Diagnostics for PLRM. The approximate case deletion formulas using SMW theorem [41] are found for the identification of influential observations.

Theorem 1. After the deletion of i^{th} row from $\hat{\beta}_d$, we write $\hat{\beta}_{d(i)}$ as

$$\hat{\beta}_{d(i)} = [X_{(i)}'\hat{W}X_{(i)} + I]^{-1}[X_{(i)}'\hat{W}X_{(i)} + dI]\hat{\beta}_{\text{ML}(i)}, \quad (25)$$

where $X_{(i)}$ represents the X matrix without the i^{th} row. Using the SMW theorem, we approximated $\hat{\beta}_{d(i)}$.

Proof. Letting $K = \sqrt{\hat{W}}X$, $v = \sqrt{\hat{W}}z$, and $s = y - \hat{\mu}$, then $\hat{\beta}_{\text{ML}}$ and $\hat{\beta}_d$ stated by (6) and (7) become

$$\begin{aligned} \hat{\beta}_{\text{ML}} &= (K'K)^{-1}K'v, \\ \hat{\beta}_d &= (K'K + I)^{-1}(K'K + d\hat{\beta}_{\text{ML}}). \end{aligned} \quad (26)$$

Let $\hat{\beta}_{\text{ML}(i)}$ and $\hat{\beta}_{d(i)}$ represent the ML and PLRE of β after deleting the i^{th} observation, respectively. Thus, we have

$$\begin{aligned} \hat{\beta}_{\text{ML}(i)} &= (K_{(i)}'K_{(i)})^{-1}X_{(i)}'s_{(i)}, \\ \hat{\beta}_{d(i)} &= (K_{(i)}'K_{(i)} + I)^{-1}(K_{(i)}'K_{(i)} + d\hat{\beta}_{\text{ML}(i)}). \end{aligned} \quad (27)$$

With the help of SMV theorem, $\hat{\beta}_{\text{ML}(i)}$ can be improved as

$$\hat{\beta}_{\text{ML}(i)} = \left[(K'K)^{-1} + \frac{(K'K)^{-1}k_i'k_i(K'K)^{-1}}{1 - m_i} \right] (X's - x_i's_i), \quad (28)$$

where $k_i' = \sqrt{\hat{W}_{ii}}x_i'$ is the i^{th} row vector of the K matrix and $m_i = k_i'(K'K)^{-1}k_i$ solves the first part of R.H.S of (27)

$$\begin{aligned} (K_{(i)}'K_{(i)} + I)^{-1} &= [(K'K + I) - k_i'k_i]^{-1} = (K'K + I)^{-1} \\ &\quad + \frac{(K'K + I)^{-1}k_i'k_i(K'K + I)^{-1}}{1 - m_{di}}, \end{aligned} \quad (29)$$

where $m_{di} = k_i'(K'K + I)^{-1}k_i'$. We also have

$$\begin{aligned} (K_{(i)}'K_{(i)} + I)^{-1}(K_{(i)}'K_{(i)} + d\hat{\beta}_{\text{ML}}) &= [(K'K + I) - k_i'k_i]^{-1}(K'K + d\hat{\beta}_{\text{ML}} - k_i'k_i) \\ &= (K'K + I)^{-1}(K'K + d\hat{\beta}_{\text{ML}}) + \frac{(K'K + I)^{-1}k_i'k_i(K'K + I)^{-1}(K'K + d\hat{\beta}_{\text{ML}})}{1 - m_{di}} - \frac{(K'K + I)^{-1}k_i'k_i}{1 - m_{di}}. \end{aligned} \quad (30)$$

Now,

$$\begin{aligned}
\widehat{\beta}_{d(i)} &= (K'K + I)^{-1}(K'K + d\widehat{\beta}_{ML}) - \frac{(K'K + I)^{-1}k'_i}{1 - m_{di}} \left(\frac{s_i}{\sqrt{\widehat{W}_{ii}}} - k_i(K'K + I)^{-1}(K'K + d\widehat{\beta}_{ML}) \right) \\
&= \widehat{\beta}_d - \frac{(K'K + I)^{-1}k'_i}{1 - m_{di}} \left(\frac{s_i}{\sqrt{\widehat{W}_{ii}}} - k_i\widehat{\beta}_d \right) \\
&= \widehat{\beta}_d - \frac{(X'\widehat{W}X + I)^{-1}x'_i\sqrt{\widehat{W}_{ii}}\chi_{di}}{1 - m_{di}} \\
\widehat{\beta}_d - \widehat{\beta}_{d(i)} &= \frac{(X'\widehat{W}X + I)^{-1}x'_i\sqrt{\widehat{W}_{ii}}\chi_{di}}{1 - m_{di}}.
\end{aligned} \tag{31}$$

Hence, the theorem is completed.

Following [42] for the PLRE, the Cook's distance is redefined as

$$C_{di} = \frac{(\widehat{\beta}_d - \widehat{\beta}_{d(i)})'(X'\widehat{W}X)(\widehat{\beta}_d - \widehat{\beta}_{d(i)})}{(p+1)}. \tag{32}$$

The i^{th} observation is considered as influential if the distance between $\widehat{\beta}_{d(i)}$ and $\widehat{\beta}_d$ is larger. Another version can be expressed as

$$\begin{aligned}
C_{di} &= \frac{1}{(p+1)} [\widehat{\beta}_d - \widehat{\beta}_{d(i)}]' (X'\widehat{W}X + I) (X'\widehat{W}X + dI)^{-1} \\
&\quad \times (X'\widehat{W}X) (X'\widehat{W}X + dI)^{-1} (X'\widehat{W}X + I) [\widehat{\beta}_d - \widehat{\beta}_{d(i)}].
\end{aligned} \tag{33}$$

Using the Liu estimator, we defined the change in Pearson chi-square as

$$\Delta\chi_{di}^2 = \Delta\chi_{d(i)}^2 = \Delta\chi_{d0}^2 = \frac{1}{1 - h_{dii}} \left[\frac{(y_i - \widehat{\mu}_{d(i)})^2}{\widehat{\mu}_{d(i)}} \right], \tag{34}$$

where the squared Liu Pearson residuals χ_{d0}^2 are used to represent the complete data set and $\chi_{d(i)}^2$ computed without i^{th} observation. Correspondingly, with Liu estimator, we formulated the change in deviance statistic as

$$\Delta d_d^2 = d_{d(i)}^2 - d_{d0}^2 = d_{di}^2 + \frac{\chi_{di}^2 h_{dii}}{1 - h_{dii}}, \tag{35}$$

where d_{d0}^2 and $d_{d(i)}^2$ represent the squared Liu deviance residuals with complete data and the squared Liu deviance residuals computed without i^{th} observation, and

$$d_{di} = s_i \sqrt{2 \left(y_i \log \left(\frac{y_i}{\exp(x'_i \widehat{\beta}_d)} \right) - (y_i - \exp(x'_i \widehat{\beta}_d)) \right)}, \tag{36}$$

where s_i is the sign function of $(y_i - \exp(x'_i \widehat{\beta}_d))$.

Following (19), we give the DFFITS for PLRM as

$$\text{DFFITS}_{di} = \frac{\widehat{\mu}_{di} - \widehat{\mu}_{d(i)}}{\sqrt{h_{dii}}} = \frac{\sqrt{\widehat{W}_{ii}} x'_i (\widehat{\beta}_d - \widehat{\beta}_{d(i)})}{\sqrt{h_{dii}}}, \tag{37}$$

where $\widehat{\mu}_{d0}$ and $\widehat{\mu}_{d(i)}$ represent the predicted regressand of the complete data set and predicted regressand after deleting the i^{th} case.

Using the SMW theorem, we simplified (37) as

$$\text{DFFITS}_{di} = |t_{di}| \sqrt{\frac{h_{dii}}{1 - h_{dii}}}, \tag{38}$$

where $t_{di} = \chi_{di} \sqrt{(n - p' / n - p - \chi_{di}^2)}$ is the i^{th} Pearson Jackknife residual with Liu estimator. \square

4. Simulation Study

In this section, we summarize the results of the PRM and the PLRM influence diagnostics using the Monte Carlo simulation scheme. We follow the same simulation scheme used by many researchers to see [43, 44]. The response variable y is generated from the Poisson distribution with mean function μ_i as defined by

$$\mu_i = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}), \quad i = 1, 2, \dots, n. \tag{39}$$

We set simulation with $p = 2, 4, 6$ explanatory variables with various sample sizes plus the mild to severe levels of collinearity. We assumed sample sizes $n = 25, 50, 100, 150, 200$. Moreover, we generated the regressors using the following formula:

$$x_{ij} = \sqrt{(1 - \rho^2)} z_{ij} + \rho z_{ip} + 1, \quad j = 1, 2, \dots, p + 1, i = 1, 2, \dots, n. \tag{40}$$

We consider the different collinearity levels as $\rho^2 = 0.75, 0.85, 0.95, 0.99$, and we assume the arbitrary values of regression coefficients in such a way that $\sum_{j=1}^p \beta_j^2 = 1$. Now we generated few influential observations in the regressors by using the expression $X_{ij} = X_{ij} + \alpha_0$, $i = 15$ and $j = 1, 2, \dots, p$, where $\alpha_0 = \bar{X}_j + 6$. All the analyses are performed using the *R* software with 1000 replications.

TABLE 1: Influence diagnostics detections (%) with d_1 and $p = 2$.

n	ρ^2	C_i	$\Delta\chi_i^2$	Δd_i^2	DFFITS_i	C_{di}	$\Delta\chi_{di}^2$	Δd_{di}^2	DFFITS_{di}
25	0.75	79.2	47.5	47.5	96.1	91.8	77	76.9	91.8
	0.85	77.6	42.4	42.4	95.1	91.1	75.7	75.8	90.9
	0.95	79.5	44.9	44.9	95.1	90.2	77.1	77.1	90.2
	0.99	75.6	40	40	94.5	90.3	72.2	72.2	90.3
50	0.75	80.9	37.1	37.1	97	91.9	71	71.1	91.9
	0.85	81	35.9	35.9	97.6	91	70.7	70.5	91
	0.95	79.3	32.6	32.6	97.3	90.8	68.8	68.7	90.8
	0.99	78.5	33.2	33.2	96.1	91.2	68.2	68.1	91.2
100	0.75	89.5	42.2	42.2	98.6	95.9	73.9	73.7	95.9
	0.85	87.5	37.9	37.9	99	94.4	72.8	72.7	94.4
	0.95	86.3	35.7	35.7	97.9	94.8	70.1	70.1	94.8
	0.99	86.9	39.1	39.1	98.4	95.4	74.2	74.4	95.4
150	0.75	91	45	45	99.2	96.1	76.8	76.9	96.1
	0.85	89.1	39.3	39.3	98.7	95.3	72.6	72.6	95.3
	0.95	90.8	38.3	38.3	99.2	96	73	72.9	96
	0.99	88.2	39.7	39.7	99	94.9	71.8	71.8	94.9
200	0.75	91.6	47.4	47.4	99.3	96.5	75.4	75.5	96.5
	0.85	91.4	42.7	42.7	99.1	96.7	73.5	73.5	96.7
	0.95	90.7	41.1	41.1	99.9	95.3	72.5	72.5	95.3
	0.99	93.8	41.3	41.3	99.3	96.8	75.6	75.5	96.8

4.1. Results and Discussion. The study results of the calculations of the identification of the unusual observations with LE in the presence of mild to severe multicollinearity are provided in Tables 1–6 with $p = 2, 4$, and 6 with defined optimum d_1 and d_2 . From Tables 1–3 with $p = 2$, it is clear that performed C_{di} is good as compared to the C_i method for different sample sizes with multicollinearity. The influence detection of $\Delta\chi_{di}^2$ and Δd_{di}^2 methods is identical and performs significantly better than $\Delta\chi_i^2$ and Δd_i^2 , respectively. However, it is observed that their performance does not occur in a better way than that of the C_i method for all the combinations of n, p and ρ . Comparable effects are observed on DFFITS_i method, and it is found that the detection percentage of influential observations by DFFITS_i method is better than C_i , although the performance of DFFITS_{di} related to C_{di} is equally better. Furthermore, as we increased the sample sizes, the percentage of detecting the influential observation of the developed measures equally increases. Moreover, from Tables 4–6, we observed that the newly developed diagnostic measures performed more efficiently with d_2 , but d_1 give better detection percentage as compared to d_2 . Furthermore, varying the regressors size affects the functioning of C_{di} method and DFFITS_{di} method, respectively. $\Delta\chi_{di}^2$ and Δd_{di}^2 values are larger than $\Delta\chi_i^2$ and Δd_i^2 , respectively. Further the changing pattern of sample size and multicollinearity together with their effects are demonstrated explicitly to study the performance of newly developed measures through graphs; see Figures 1, 2, and 3 for defined d_1 and d_2 . Considering Figures 1–3, with defined combinations of sample sizes, regressors, and collinearity levels, we clearly observe a positive increase in the performance of newly developed measures.

5. Application: English League Football Data

For the illustration of the proposed diagnostic methods, we analyze the football English League data set which is also available in Table 7. The said data comprise $n = 20$ observations with one response variable, i.e., the number of won matches (y) and $p = 5$ explanatory variables, i.e., the number of yellow cards (X_1), the number of red cards (X_2), goals won (X_3), goals conceded (X_4), and the number of points earned (X_5). Algamal and Alanaz [11] also used this data set. After checking the χ^2 , the goodness of fit test found that it is well fitted to the Poisson distribution. The said data are multicollinear as the condition index $\text{CI} = 31.274$.

From Table 8, it is found that all methods commonly identify the 1st observation as the influential observation. Change in chi-square statistic and change in deviance statistic with ML estimator do not detect any of the observation as influential. Furthermore, observation 19th was detected as influential by DFFITS without Liu estimator and by all of the proposed diagnostics.

The effect of deleting the highlighted observations on the estimates of PRM and PLRM is presented in Table 9. We found a maximum change in PRM and PLRM estimates after the removal of the 1st observation that was detected by all selected and proposed measures. The second observation that was identified by just DFFITS_i and all proposed measures is the 19th. After the deletion of detected observations, we found the maximum change on $\hat{\beta}_2$ and $\hat{\beta}_3$. After examining these results, it was noted that in the presence of multicollinearity, PLRM diagnostic measures efficiently detect the influential observations. Furthermore, we incorporated index plots to summarize the efficacy of the proposed measures in Figure 4.

TABLE 2: Influence diagnostics detections (%) with d_1 and $p = 4$.

n	ρ^2	C_i	$\Delta\chi_i^2$	Δd_i^2	DFFITs $_i$	C_{di}	$\Delta\chi_{di}^2$	Δd_{di}^2	DFFITs $_{di}$
25	0.75	75.9	57	57	94.8	91	82.6	82.6	90.9
	0.85	69.7	48.9	48.9	93.7	87.8	76.4	76.4	87.7
	0.95	67.3	45.6	45.6	93.5	85.4	74.3	74.3	85.3
	0.99	65.7	44.4	44.4	93.9	85.3	74.3	74.2	85.2
50	0.75	78.6	47.8	47.8	97.4	91.5	77.9	77.9	91.4
	0.85	77.8	44.9	44.9	97.7	90.6	77.5	77.4	90.4
	0.95	74.1	38.6	38.6	95.1	88.5	72.6	72.6	88.5
	0.99	71.7	35.8	35.8	95.9	86.7	68.8	68.7	86.7
100	0.75	86.1	50.4	50.4	98.8	94.4	79.4	79.3	94.4
	0.85	84.8	44.6	44.6	97.9	93.8	77.7	77.7	93.8
	0.95	81.5	39.7	39.7	97.6	91.3	72.5	72.3	91.3
	0.99	82.2	40	40	98.5	93.7	75	75	93.7
150	0.75	86.6	49.6	49.6	98.8	94.5	78.2	78.2	94.5
	0.85	87.5	46.7	46.7	98.8	95.3	78.4	78.5	95.3
	0.95	85.8	42.5	42.5	98.6	94.7	74.3	74.2	94.7
	0.99	86.8	41.4	41.4	99.2	94.5	76.3	76.4	94.5
200	0.75	90.1	50.5	50.5	99	96.6	80.1	80.1	96.6
	0.85	89.8	48	48	99.3	96.3	79.3	79.3	96.3
	0.95	87.7	46.9	46.9	99.2	94.8	75.9	75.9	94.8
	0.99	88.1	39.9	39.9	98.8	95.2	73.7	73.7	95.2

TABLE 3: Influence diagnostics detections (%) with d_1 and $p = 6$.

n	ρ^2	C_i	$\Delta\chi_i^2$	Δd_i^2	DFFITs $_i$	C_{di}	$\Delta\chi_{di}^2$	Δd_{di}^2	DFFITs $_{di}$
25	0.75	73.1	54.5	54.5	93.7	88.6	78.1	78.1	88.4
	0.85	72.3	52.6	52.6	93.6	87.4	78.8	78.8	87.4
	0.95	68.4	47	47	93.9	85.9	76.5	76.5	85.7
	0.99	67.4	45.1	45.1	93	84.5	72.4	72.3	84.5
50	0.75	77.3	46.2	46.2	97.2	91.2	76.9	76.9	91.2
	0.85	74	43.2	43.2	96.5	89.8	74.7	74.7	89.8
	0.95	72.4	38.4	38.4	96.7	89.6	71.6	71.5	89.5
	0.99	73.1	35.2	35.2	95.9	87.5	68.5	68.7	87.5
100	0.75	83.8	47	47	98.7	95.1	77.2	77.1	95.1
	0.85	84.1	46.6	46.6	98.3	94.2	78.7	78.7	94.2
	0.95	82	41.9	41.9	97.8	94	75.4	75.4	94
	0.99	81.4	38.4	38.4	98.5	91.6	71.4	71.3	91.6
150	0.75	88.9	51.5	51.5	98.8	95.4	80.3	80.3	95.4
	0.85	86	47.7	47.7	99.1	94.4	78.1	78.2	94.4
	0.95	86.5	42.7	42.7	98.6	94.2	75.2	75.4	94.2
	0.99	83.9	40.3	40.3	98.7	93.6	71.4	71.4	93.6
200	0.75	90.7	50.1	50.1	99.1	95.3	81.2	81.2	95.3
	0.85	88.8	45.8	45.8	99.2	96.9	76.3	76.3	96.9
	0.95	88	44.8	44.8	99.2	95.5	74.7	74.8	95.5
	0.99	87.7	41.9	41.9	99.3	94.4	75.7	75.7	94.4

TABLE 4: Influence diagnostics detections (%) with d_2 and $p = 2$.

n	ρ^2	C_i	$\Delta\chi_i^2$	Δd_i^2	DFFITs $_i$	C_{di}	$\Delta\chi_{di}^2$	Δd_{di}^2	DFFITs $_{di}$
25	0.75	77.7	44.6	44.6	95.4	91.3	75	75	91.3
	0.85	74.6	38.9	38.9	95.1	89.9	73	73	89.9
	0.95	76.8	42.4	42.4	94.5	89.9	74.4	74.6	89.9
	0.99	75.4	40.6	40.6	95.2	90.3	71.3	71.3	90.3
50	0.75	78.2	35.9	35.9	97.6	91.6	68	68	91.5
	0.85	81	34.5	34.5	98	92.6	69.5	69.6	92.6
	0.95	79.1	35.1	35.1	97	90.8	68.9	68.7	90.7
	0.99	80.2	34.3	34.3	97.2	91.5	67.5	67.5	91.5

TABLE 4: Continued.

n	ρ^2	C_i	$\Delta\chi_i^2$	Δd_i^2	DFFITs _{i}	C_{di}	$\Delta\chi_{di}^2$	Δd_{di}^2	DFFITs _{di}
100	0.75	87.9	40.1	40.1	98.5	94.3	73.1	73.2	94.3
	0.85	86.6	42.3	42.3	98.2	94	72	72	94
	0.95	88.4	37.1	37.1	98.6	94.8	74.2	74.3	94.8
	0.99	88.8	42.3	42.3	98.6	94.1	73.5	73.4	94.1
150	0.75	91.1	46.1	46.1	98.9	96.4	76.1	76.1	96.4
	0.85	90.3	39.2	39.2	99.8	95.6	72.4	72.4	95.6
	0.95	90.2	38.8	38.8	99	95.9	72	72.1	95.9
	0.99	89.8	39.1	39.1	99.2	95.5	74.7	74.7	95.5
200	0.75	93	44.9	44.9	99.3	96.4	75.9	76	96.4
	0.85	93.2	43.6	43.6	99.6	96.6	77.6	77.7	96.6
	0.95	91.9	41.9	41.9	99.2	96.9	73.4	73.4	96.9
	0.99	92.6	40.3	40.3	99.5	96.3	74.3	74.2	96.3

TABLE 5: Influence diagnostics detections (%) with d_2 and $p = 4$.

n	ρ^2	C_i	$\Delta\chi_i^2$	Δd_i^2	DFFITs _{i}	C_{di}	$\Delta\chi_{di}^2$	Δd_{di}^2	DFFITs _{di}
25	0.75	72.7	47.5	47.5	94.6	89.1	76.1	76.2	89.1
	0.85	73.5	46.1	46.1	94.7	88.7	77.4	77.4	88.7
	0.95	69.9	41.2	41.2	91.9	86.5	73.7	73.8	86.5
	0.99	70.3	42.5	42.5	94.2	87.6	73.6	73.6	87.6
50	0.75	79.6	40.6	40.6	96.4	91.2	75.1	75.1	91.2
	0.85	76.5	37.1	37.1	97.4	91.2	73.6	73.4	91.2
	0.95	78.9	38.8	38.8	96.8	89.6	74	74	89.6
	0.99	74.4	36.1	36.1	97.7	90.5	71	70.9	90.5
100	0.75	87	45.5	45.5	99.1	95.7	76.2	76.1	95.7
	0.85	84.1	41.2	41.2	97.8	93.7	73.5	73.5	93.7
	0.95	84.3	38.4	38.4	98.4	94.5	73.5	73.5	94.5
	0.99	86	38.3	38.3	98.9	94.5	74.4	74.3	94.5
150	0.75	89.5	46.3	46.3	99.1	95.9	77.3	77.3	95.9
	0.85	89.5	44.4	44.4	99.2	96	77.7	77.6	96
	0.95	89.6	41.9	41.9	99.1	95.8	76.3	76.4	95.8
	0.99	86.8	38.8	38.8	98.4	94.3	71.5	71.5	94.3
200	0.75	90.9	47.3	47.3	98.8	96.4	79.9	79.8	96.4
	0.85	89.3	43.8	43.8	99.2	95.5	76	75.9	95.5
	0.95	89.6	41.5	41.5	99.6	95.5	75.8	75.8	95.5
	0.99	89.6	41.9	41.9	99.1	95.9	72.6	72.7	95.9

TABLE 6: Influence diagnostics detections (%) with d_2 and $p = 6$.

n	ρ^2	C_i	$\Delta\chi_i^2$	Δd_i^2	DFFITs _{i}	C_{di}	$\Delta\chi_{di}^2$	Δd_{di}^2	DFFITs _{di}
25	0.75	75.9	57	57	94.8	91	82.6	82.6	90.9
	0.85	69.7	48.9	48.9	93.7	87.8	76.4	76.4	87.7
	0.95	67.3	45.6	45.6	93.5	85.4	74.3	74.3	85.3
	0.99	65.7	44.4	44.4	93.9	85.3	74.3	74.2	85.2
50	0.75	78.6	47.8	47.8	97.4	91.5	77.9	77.9	91.4
	0.85	77.8	44.9	44.9	97.7	90.6	77.5	77.4	90.4
	0.95	74.1	38.6	38.6	95.1	88.5	72.6	72.6	88.5
	0.99	71.7	35.8	35.8	95.9	86.7	68.8	68.7	86.7
100	0.75	86.1	50.4	50.4	98.8	94.4	79.4	79.3	94.4
	0.85	84.8	44.6	44.6	97.9	93.8	77.7	77.7	93.8
	0.95	81.5	39.7	39.7	97.6	91.3	72.5	72.3	91.3
	0.99	82.2	40	40	98.5	93.7	75	75	93.7
150	0.75	86.6	49.6	49.6	98.8	94.5	78.2	78.2	94.5
	0.85	87.5	46.7	46.7	98.8	95.3	78.4	78.5	95.3
	0.95	85.8	42.5	42.5	98.6	94.7	74.3	74.2	94.7
	0.99	86.8	41.4	41.4	99.2	94.5	76.3	76.4	94.5
200	0.75	90.1	50.5	50.5	99	96.6	80.1	80.1	96.6
	0.85	89.8	48	48	99.3	96.3	79.3	79.3	96.3
	0.95	87.7	46.9	46.9	99.2	94.8	75.9	75.9	94.8
	0.99	88.1	39.9	39.9	98.8	95.2	73.7	73.7	95.2

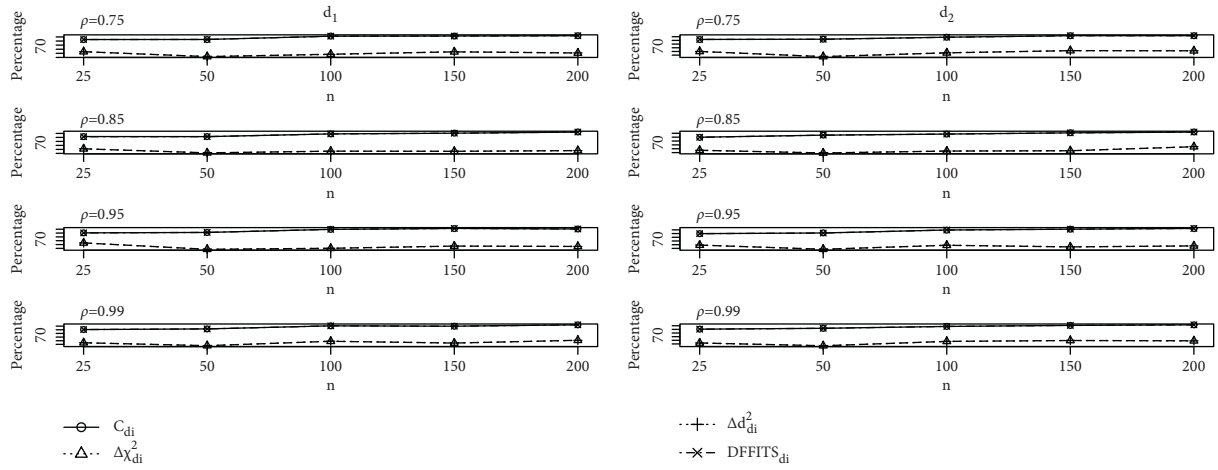


FIGURE 1: Graph of influential observation detection percentages of proposed measures for $p = 2$.

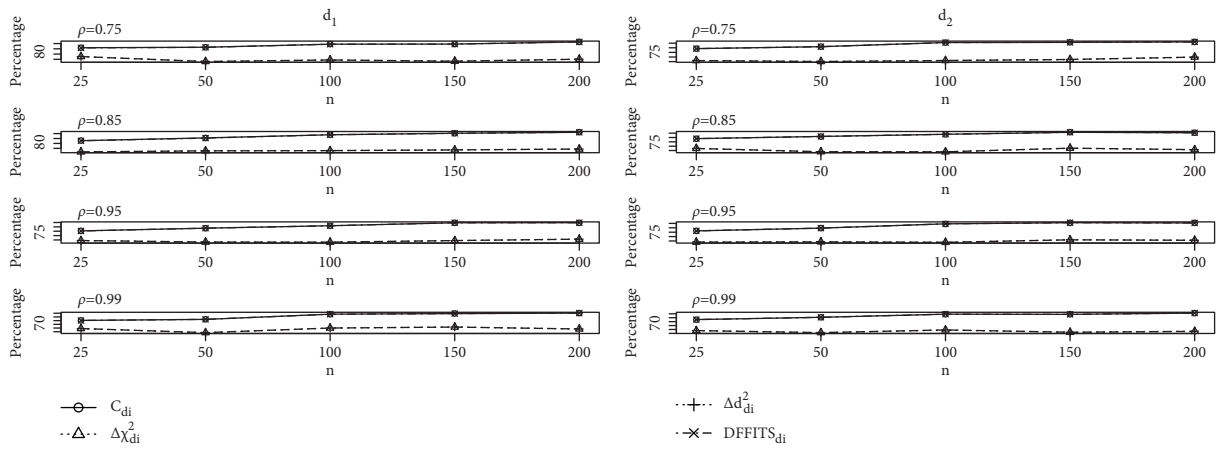


FIGURE 2: Graph of influential observation detection percentages of proposed measures for $p = 4$.

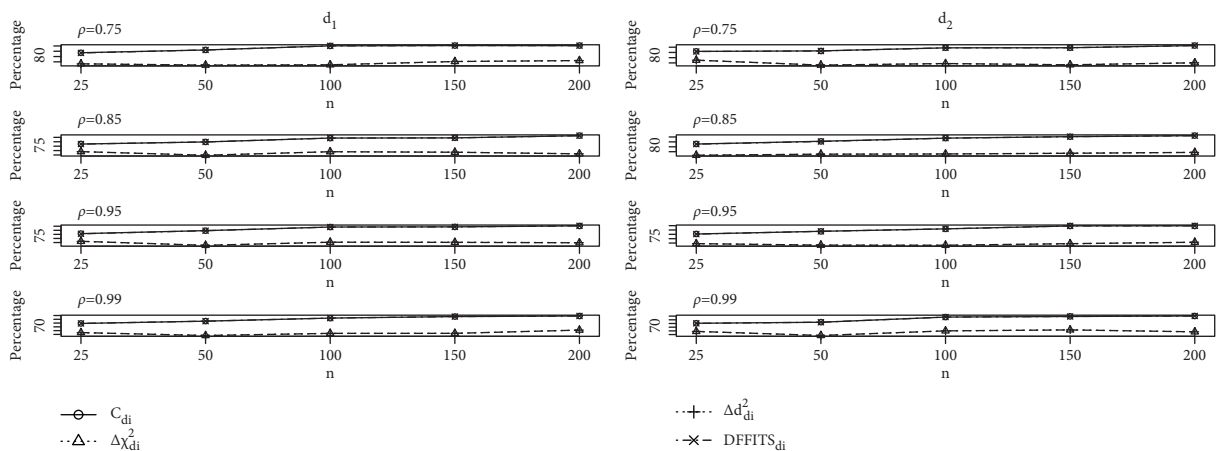


FIGURE 3: Graph of influential observation detection percentages of proposed measures for $p = 6$.

TABLE 7: English League football data.

Y (no. of won matches)	X1 (no. of yellow cards)	X2 (no. of red cards)	X3 (goals own)	X4 (goals conceded)	X5 (no. of points earned)
30	72	0	33	52	93
26	62	0	26	60	86
23	71	4	39	41	78
22	54	0	42	36	76
23	68	3	44	33	75
18	78	2	29	25	69
17	72	2	44	18	61
12	59	2	48	-7	46
12	52	3	67	-12	46
12	80	0	51	-8	45
12	78	5	64	-17	45
12	72	1	63	-15	44
11	70	2	56	-15	44
12	77	0	63	-13	41
12	56	0	70	-25	41
11	65	2	55	-16	40
11	84	5	68	-28	40
9	67	5	80	-43	34
5	77	1	53	-26	28
6	78	4	69	-40	24

TABLE 8: Influential observations detected using the PRM and the PLRM.

Procedures	PRM	Procedures	PLRM \hat{d}_1
C_i	1	C_{di}	1,19
$\Delta\chi_i^2$	—	$\Delta\chi_{di}^2$	1,19
Δd_i^2	—	Δd_{di}^2	1,19
DFFITs _i	1,6,14,19	DFFITs _{di}	1,6,7,8,10,14,15,16,18,19,20

TABLE 9: Absolute change (%) in the estimates after deletion of influential observations.

Estimates	Influential observations					
	1 st		19 th		1 st and 19 th	
	PRM	PLRM	PRM	PLRM	PRM	PLRM
$\hat{\beta}_0$	69.1	81.0	27.6	23.0	11.9	52.0
$\hat{\beta}_1$	3944.7	17.8	1448.5	47.2	3341.4	56.8
$\hat{\beta}_2$	510.0	160.4	139.8	54.4	392.3	189.6
$\hat{\beta}_3$	355.4	626.9	77.7	75.7	119.2	485.2
$\hat{\beta}_4$	144.7	50.9	70.5	19.5	13.9	28.4
$\hat{\beta}_5$	58.1	46.2	20.9	10.5	12.2	31.5

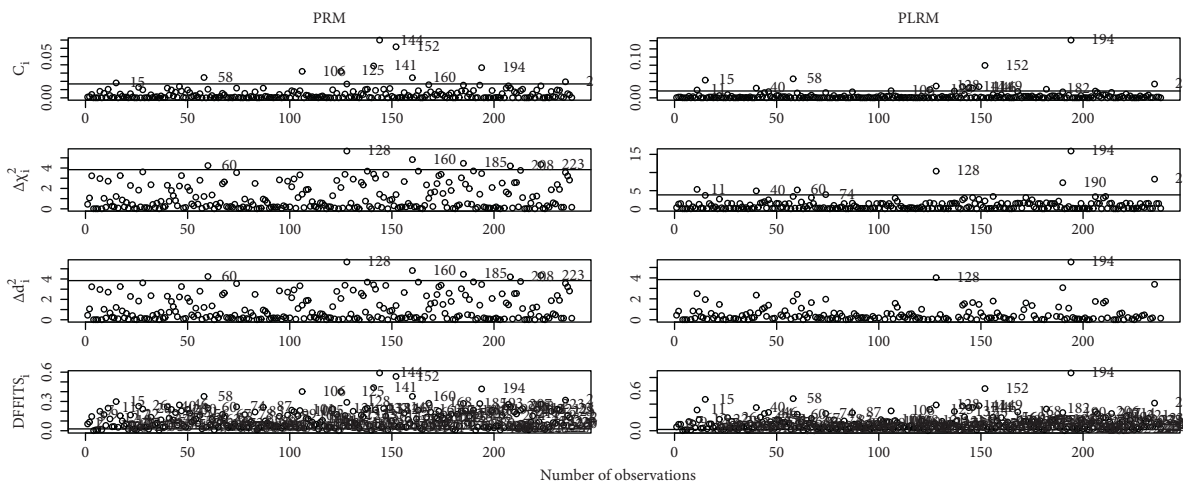


FIGURE 4: Graphical representation of influential observations in English League Football data.

6. Conclusion

This study introduces diagnostic measures for Poisson Liu regression using biased estimator to handle influential observations and multicollinearity simultaneously in the PRM. As discussed earlier, multicollinearity affects the performance of traditional ML estimator in PRM. Therefore, we adopted the Liu estimator due to its efficient statistical properties to solve multicollinearity and influential observations in PRM. The simulation results support the performance of new diagnostic measures as the detection percentage of ML estimators and the existing measures turn out to be the worst with increasing the sample size, number of regressors, and the level of multicollinearity. The results proved that the suggested measures proved more beneficial for the identification of influential observations together with multicollinearity. Hence, it is suggested that these proposed measures guide the user to handle the issue of multicollinearity with robust estimator support efficiently.

Data Availability

All data are included in the paper with their links.

Conflicts of Interest

The authors declare there are no conflicts of interest.

Acknowledgments

The authors are thankful for the Taif University researchers supporting project number TURSP-2020/160, Taif University, Taif, Saudi Arabia.

References

- [1] Z. Y. Algamal, "Biased estimators in poisson regression model in the presence of multicollinearity: a subject review," *Al-Qadisiyah Journal for Administrative and Economic Sciences*, vol. 20, pp. 37–43, 2018.
- [2] A. Majid, M. Amin, and M. N. Akram, "On the Liu estimation of bell regression model in the presence of multicollinearity," *Journal of Statistical Computation and Simulation*, 2021.
- [3] M. Amin, M. N. Akram, and B. M. G. Kibria, "A new adjusted Liu estimator for the poisson regression model," *Concurrency and Computation: Practice and Experience*, Wiley, Hoboken, NJ, USA, 2021.
- [4] F. Sami, M. Amin, and M. M. Butt, "On the ridge estimation of the conway-maxwell poisson regression model with multicollinearity: methods and applications," *Concurrency and Computation: Practice and Experience*, Wiley, Hoboken, NJ, USA, 2021.
- [5] W. B. Cohen, T. K. Maierperger, S. T. Gower, and D. P. Turner, "An improved strategy for regression of biophysical variables and landsat ETM+ data," *Remote Sensing of Environment*, vol. 84, no. 4, pp. 561–571, 2003.
- [6] P. McCullagh and J. A. Nelder, *Generalized Linear Models, CRC Monographs on Statistics & Applied Probability*, Springer Verlag, Berlin, Germany, 1989.
- [7] K. Månsson and G. Shukur, "A poisson ridge regression estimator," *Economic Modelling*, vol. 28, no. 4, pp. 1475–1481, 2011.
- [8] C. Stein, *Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution*, Stanford University, Stanford, CA, USA, 1956.
- [9] A. E. Hoerl and R. W. Kennard, "Ridge regression: biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.
- [10] K. Liu, "A new class of biased estimate in linear regression," *Communication in Statistics—Theory and Methods*, vol. 22, pp. 393–402, 1993.
- [11] Z. Y. Algamal and M. M. Alanaz, "Proposed methods in estimating the ridge regression parameter in poisson regression model," *Electronic Journal of Applied Statistical Analysis*, vol. 11, pp. 506–515, 2018.
- [12] N. K. Rashad and Z. Y. Algamal, "A new ridge estimator for the poisson regression model," *Iranian Journal of Science and Technology, Transactions A: Science*, vol. 43, no. 6, pp. 2921–2928, 2019.
- [13] A. Alkhateeb and Z. Algamal, "Jackknifed Liu-type estimator in poisson regression model," *Journal of the Iranian Statistical Society*, vol. 19, no. 1, pp. 21–37, 2020.
- [14] M. Qasim, K. Månsson, M. Amin, B. M. Golam Kibria, and P. Sjölander, "Biased adjusted poisson ridge estimators-method and application," *Iranian Journal of Science and Technology, Transactions A: Science*, vol. 44, no. 6, pp. 1775–1789, 2020.
- [15] M. Amin, M. N. Akram, and M. Amanullah, "On the james-stein estimator for the poisson regression model," *Communications in Statistics—Simulation and Computation*, pp. 1–13, 2020.
- [16] A. F. Lukman, I. Dawoud, B. M. G. Kibria, Z. Y. Algamal, and B. Aladeitan, "A new estimator for the multicollinear poisson regression model: simulation and application," *Scientific Reports*, vol. 11, pp. 1–11, 2021.
- [17] R. D. Cook, "Detection of influential observation in linear regression," *Technometrics*, vol. 19, no. 1, pp. 15–18, 1977.
- [18] D. A. Belsley, E. Kuh, and R. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, Wiley, Hoboken, NJ, USA, 1980.
- [19] S. Chatterjee and A. S. Hadi, "Influential observations, high leverage points, and outliers in linear regression," *Statistical Science*, vol. 1, pp. 379–393, 1986.
- [20] E. Walker and J. B. Birch, "Influence measures in ridge regression," *Technometrics*, vol. 30, no. 2, pp. 221–227, 1988.
- [21] A. Jahufer and C. Jianbao, "Assessing global influential observations in modified ridge regression," *Statistics & Probability Letters*, vol. 79, no. 4, pp. 513–518, 2009.
- [22] M. A. Ullah, G. R. Pasha, and M. Aslam, "Local influence diagnostics in the modified ridge regression," *Communications in Statistics—Theory and Methods*, vol. 42, no. 10, pp. 1851–1869, 2013.
- [23] D. Peña, "A new statistic for influence in linear regression," *Technometrics*, vol. 47, no. 1, pp. 1–12, 2005.
- [24] M. Kashif, M. A. Ullah, and M. Aslam, "Influential diagnostics with pena's statistic for the modified ridge regression," *Communications in Statistics—Simulation and Computation*, 2019.
- [25] D. Pregibon, "Logistic regression diagnostics," *The Annals of Statistics*, vol. 9, pp. 705–724, 1981.
- [26] A. H. Lee, "Diagnostic displays for assessing leverage and influence in generalized linear models," *Australian Journal of Statistics*, vol. 29, no. 3, pp. 233–243, 1987.
- [27] D. A. Williams, "Generalized linear model diagnostics using the deviance and single case deletions," *Journal of Royal Statistical Society Series C*, vol. 36, pp. 741–749, 1987.

- [28] R. J. O. H. Hines and E. M. Carter, "Improved added variable and partial residual plots for the detection of influential observations in generalized linear models," *Journal of Royal Statistical Society: Series C (Applied Statistics)*, vol. 42, no. 1, pp. 3–16, 1993.
- [29] W. Thomas and R. D. Cook, "Assessing influence on regression coefficients in generalized linear models," *Biometrika*, vol. 76, no. 4, pp. 741–749, 1989.
- [30] M. Amin, M. Amanullah, and M. Aslam, "Empirical evaluation of the inverse gaussian regression residuals for the assessment of influential points," *Journal of Chemometrics*, vol. 30, no. 7, pp. 394–404, 2016.
- [31] M. Amin, M. Amanullah, and G. M. Cordeiro, "Influence diagnostics in gamma regression model with adjusted deviance residuals," *Communications in Statistics—Simulation and Computation*, vol. 46, no. 9, pp. 6959–6973, 2017.
- [32] Z. Y. Algamal, "Diagnostic in poisson regression models," *Electronic Journal of Applied Statistical Analysis*, vol. 5, pp. 178–186, 2012.
- [33] M. R. Özkale, S. Lemeshow, and R. Sturdivant, "Logistic regression diagnostics in ridge regression," *Computational Statistics*, vol. 33, no. 2, pp. 563–593, 2018.
- [34] M. Amin, M. A. Ullah, and M. Aslam, "Influence diagnostics in gamma ridge regression model," *Journal of Statistical Computation and Simulation*, vol. 89, pp. 536–556, 2020.
- [35] A. Khan, M. Amanullah, and M. Amin, "Poisson regression diagnostics with ridge estimation," *Communications in Statistics Simulation and Computation*, 2021.
- [36] A. Khan, M. Amanullah, H. M. Aljohani, and S. A. M. Mubarak, "Influence diagnostics for the poisson regression model using two-parameter estimator," *Alexandria Engineering Journal*, vol. 60, no. 5, pp. 4745–4759, 2021.
- [37] M. Amin, M. Faisal, and M. N. Akram, "Influence diagnostics in the inverse gaussian ridge regression model: applications in chemometrics," *Journal of Chemometrics*, vol. 35, pp. 1–20, 2021.
- [38] M. Qasim, B. M. G. Kibria, K. Månsson, and P. Sjölander, "A new poisson Liu regression estimator: method and application," *Journal of Applied Statistics*, vol. 47, pp. 1–14, 2019.
- [39] A. C. Davison and C. L. Tsai, "Regression model diagnostics," *International Statistical Review/Revue Internationale de Statistique*, vol. 60, no. 3, pp. 337–353, 1992.
- [40] J. W. Hardin and J. M. Hilbe, *Generalized Estimating Equations*, Chapman and Hall/CRC, Boca Raton, FL, USA, 2012.
- [41] C. R. Rao, *Linear Statistical Inference and its Applications*, John Wiley & Sons, Hoboken, NJ, USA, 1973.
- [42] A. Jahufer, "Detecting global influential observations in Liu regression model," *Open Journal of Statistics*, vol. 03, no. 01, pp. 5–11, 2013.
- [43] M. Mittlböck and H. Heinzl, "Measures of explained variation in gamma regression models," *Communications in Statistics—Simulation and Computation*, vol. 31, no. 1, pp. 61–73, 2002.
- [44] A. E. Clark and C. G. Troskie, "Ridge regression—a simulation study," *Communications in Statistics—Simulation and Computation*, vol. 35, no. 3, pp. 605–619, 2006.