Detection of Outliers in Binomial Regression Using CERES and Partial Residual Plots

Nasir Saleem[1], Atif Akbar[1], A. H. M. Rahmatullah Imon[2] & Abu Sayed Md. Al Mamun[3*]

[1]*Department of Statistics, Bahauddin Zakariya University, Multan, Pakistan*

[2] *Department of mathematical sciences, Ball State University, USA.*

[3]*Department of Statistics, University of Rajshahi, Rajshahi-6205, Bangladesh*

** Corresponding author:: mithun_stat@yahoo.com*

Abstract

Conditional expectations and residuals (CERES) and partial residual (PR) plots have been used in linear regression model for the identification of outliers. But not much work has been done on how they perform in generalized linear models (GLM). Binomial regression model is a very important type of GLM which have wide applications in dealing with Liver cancer and many other types of data. In this paper, CERES and PR plots is used in binomial regression to detect the outliers. Through real data set, the performance of these plots on the detection of possible outliers is observed separately. The CERES plot performs well in order to diagnose this problem. However, the performance of the CERES and PR plot is found very similar to detect outlier by using simulated data but visualization of CERES plot is better as compared to the PR plots.

Keywords: Binomial regression, CERES and PR plots, Diagnostics, Generalized linear model, Outliers.

1. Introduction

In classical regression it is assumed that the error and consequently the response follow a normal distribution. But in reality, this assumption does not hold and we used the generalized linear model (GLM), which was developed by Nelder and Wedderburn (1972), is a flexible generalization of a linear regression model that allows the distribution of the response other than normal. The GLM has wide applications in modeling Liver cancer and many other types of data. The estimation GLM parameters and their optimality heavily depend on some standard assumptions and violations of this assumption misinterpret the estimated parameter and misleading the entire statistical inference. Outlier is one of the important reasons for violating the standard assumption of a regression model. Therefore, diagnostics are required to check the standard assumption before estimating parameter.

Many formal diagnostic tests are available, and these tests are based on certain regularity conditions and are therefore more computationally intensive. Under various conditions in the GLM class, conditional expectations and residuals (CERES) and partial residual (PR) plots may provide useful information. PR plots, according to Fowlkes (1987) and Landwehret al. (1984), can be used to determine nonlinearity in binary logistic regression. The technique of algorithm for regularized GLMs was explored by Park and Hastie (2007). These plots for GLMs are also studied by Landwehr and Pregibon (1993). Imran and Akbar (2020) addressed the construction of PR using response residuals for the inverse Gaussian regression model in order to investigate structure and utility for visualizing outliers, Multicollinearity, heteroscedasticity, and curvature as a function of selected predictors. Cook (1998) also discussed about the curvature of CERES and PR plots. However, identification of outliers using CERES and PR plots was not established.

In this study, CERES and PR plots for binomial regression models (BRM) are generated and the diagnostics provided by these plots are examined for model specification. Applied sciences often require techniques that are simple, efficient, and widely applicable, as well as ones that are computationally simple. Learning and applying computationally intensive statistical methods is difficult for practitioners. This article investigates this concept while emphasizing the significance of CERES and PR plots in regression diagnostics without the use of traditional measures like a statistical test. The main objective of the study was to identify the outliers by using CERES and PR plots. The comparison of CERES and PR plots for the identification of outliers using real and simulated data was also made.

The paper is organized as follows: in Section 2, we present the material and methods which include introduction of GLM, BRM and the construction of CERES and PR in BRM. The numerical example is given in Section 3 and simulation study is presented in the next Section. Finally, conclusion of the study is presented in Section 5.

2. Material and Methods

In this section, first generalized linear model and binomial regression model is introduced. Then construction procedure of CERES and PR plots in binomial regression model is described.

2.1 Generalized linear model (GLM)

In this section, we describe CERES and PR plots in detecting outliers in binomial regression. Let us consider the model

$$Y = f(X) + \varepsilon, \tag{1}$$

where $Y = (y_1, y_2, \dots, y_p)'$ is an $n \times 1$ vector of response; $X = (X_1, X_2, \dots, X_p)'$ is a $n \times 1$ covariate matrix; and ε is $n \times 1$ random vector. The conditional distribution of Yon X for a GLM for a set of n observations due to McCullagh and Nelder (1983) is,

$$d_{y|x}(y|\theta,\psi) = \exp\left\{\frac{\theta y - \mu(\theta)}{\upsilon(\psi)} + w(y,\psi)\right\}, \quad (2)$$

where $\mu(.), \upsilon(.)$, w(.,.) are well-known smooth functions; θ is an unknown scalar-valued parameter that is dependent on X; and is ψ an unknown dispersion parameter.

$$E(Y|X) = \frac{\partial \mu}{\partial \theta} = \mu(x) \text{ and } V(Y|X) = \left\{\frac{\partial^2 u}{\partial \theta^2}\right\} \upsilon(\psi).$$

There is no consideration of the dispersion parameter ψ; when calculating $\mu(x)$, as a result, $\upsilon(\psi)$ is presumed to be established. This function's log-likelihood function β is,

$$\iota(\beta) = lnL(\beta) = \exp\left\{\frac{\theta y - \mu(\theta)}{\upsilon(\psi)} + w(y,\psi)\right\}.$$

The predictors are portioned as $X' = (X_1', X_2')$, where X_j is $p_j \times 1$, j = 1, 2. The regression function can be modeled according to Cook and Croos-Debrera (1998) is as follows,

$$\eta(x) = h\big(\mu(x)\big) = \alpha_0 + \alpha_1' X_1 + g(X_2) \quad (3)$$

and $g(X_2)$ is an unknown function of X_2 and is assumed as $(X_2) = \alpha_2' X_2$. Assume that the regression function has a parametric form and that it is given by

$$\eta(x) = h\big(\mu(x)\big) = \alpha_0 + \alpha_1' X_1 + \alpha_2' X_2,$$

In Eq. (3) the term '$h\big(\mu(x)\big)$' refers to a relation function centred on a monotonic and differentiable probability distribution and $(\alpha_0 + \alpha_1' + \alpha_2')'$ is consisting of a vector of unknown parameters $(p_1 + 1)\times$ 1 vector. The regression function, $\mu(x)=h^{-1}(\eta(x))$, is a function of x or function of η depending on interest and concerns.

2.2 Binomial regression model (BRM)

The binomial response variable's probability density function is given by

$$f(y;\, n, \mu) = \binom{n}{y} \mu^y (1-\mu)^{n-y} y = 0,1,2, \dots, n.$$

It can be written as $y \sim$ binomial (y; n, μ). The mean and variance of y are, $E(y) = n\mu$ and var(y) = $n\mu(1 - \mu)$ respectively. In logistic regression, which serves as a running example in this article, we begin with a binomial (n, μ) random variable $Y^*|X$, where the unknown probability of "success" p, may depend on X. The known index n may vary from observation to observation but is assumed to be independent of X.

The observed fraction of successes from a standard binomial trial is then Y = Y*/n. In terms of Eq. (2) & (3)

$$\eta = \theta = h(\mu) = \log(\frac{\mu}{1-\mu}) \tag{4}$$

$\mu(\eta) = \log(1 + \exp(\eta))$, $v(\psi) = 1/n$ and g is an unknown scalar-valued function. Cook (1993) looked at how well PR plots could depict g in the special case of additive-error models where the relation is the identity function, $\eta = \mu$, and the conditional distribution of Y/X can be defined as

$$Y/X = \alpha_0 + \alpha_1' X_1 + g(X_2) + \varepsilon, \tag{5}$$

where ε is unaffected by X and has a mean of 0. Cook's research revealed that the output of PR plots is highly influenced by the conditional expectation $E(X_1/X_2)$, with the best results obtained when the $E(X_1/X_2)$, is linear in the value of X_2. Consider summarizing the data by fitting

$$\eta_f(x/b) = h(\mu_f) = b_0 + b_1'X_1 + b_2'\iota(X_2) \tag{6}$$

where $b' = (b_0, b_1', b_2')$and $\iota(X_2)$ is a user-defined X_2 function. The equipped model is indicated by the subscript f on η_f and μ_f. Based on Eq. (6) it is assumed that Estimated coefficients $\hat{b}_j$, j= 0, 1, 2, are obtained by minimizing a convex objective function.

$$\hat{b}' = (\hat{b}_0, \hat{b}_1', \hat{b}_2') = \arg\min_b L_N(b), \tag{7}$$

$$L_N(b) = \frac{1}{N}\sum_{i=1}^{N} L(\eta_f(x_i|b), y_i) = \frac{1}{N}\sum_{i=1}^{N} L(b_0 + b_1'X_{i1} + b_2'\iota(X_{i2}), y_i)$$

$L(.\ ,.)$ is a convex objective function with respect to its first argument that is chosen by the consumer. Since it contains ordinary least squares, maximum probability, and some robust estimates, this class is not very restrictive. For logistic regression with the relation provided in Eq. (4) for example, the objective function corresponding to maximum likelihood is

$$L(\eta_f(x|b), y) = n\{\log(1 + \exp(\eta_f)) - y\eta_f\}. \tag{8}$$

The maximum likelihood estimates are obtained from Eq. (2), (3) & (6). While maximum likelihood estimation is commonly used, it is not needed for the purposes of this article. The class of convex objective functions is a generalization of the class of objective functions corresponding to Eq. (7)

$$L(\eta_f, y) = L(y - \eta_f)$$

used by Cook (1993) for additive-error models (5).

2.3 Construction of PR and CERES plots in BMR

A PR plot for X_2 is obtained by first setting $\iota(X_2) = X_2$ and fitting Eq. (6) & (7) then constructing the (p_2 + 1)-dimensional plot $\{\widehat{pr}_2, X_2\}$, where

$$\widehat{pr}_2=(y-\hat{u}_f)h'(\hat{u}_f) + \hat{b}_2'X_2 \quad (9)$$

is the partial residual for X_2, $h'(.)$ is the first derivative of $h(.)$ with respect to u, $\hat{b}$ obtained from Eq. (7) and $\hat{u}_f(x) = h^{-1}(\eta_f(x|\hat{b}))$ is the regression function u_f evaluated at $\hat{b}$. The subscript "2" in $\widehat{pr}_2$ is intended to remind that the partial residuals are for X_2.

To form a CERES plot for X_2let us set $\iota(X_2)$ equal to a function $E(X_1|X_2)$ that captures the behavior of $\tilde{E}(X_1|X_2)$. This function may be $E(X_1|X_2)$ if known, an estimate $E(X_1|X_2)$ based on smoothing, or a parameterized class of functions that includes $E(X|X_2)$ as a special case. Once $\iota(X_2) = \tilde{E}(X_1|X_2)$ is specified, we fit Eq. (6) & (7). The CERES plot for X_2 is then the (p_2+ 1)-dimensional plot $\{\widehat{cr}_2, X_2\}$, where

$$\widehat{cr}_2 =(y-\hat{\mu}_f)h'(\hat{\mu}_f) + \hat{b}_2'\tilde{E}(X_1|X_2) \quad (10)$$

is the CERES residual for X_2 constructed from the quantities defined in (8) but based on $\iota(X_2) = \tilde{E}(X_1|X_2)$. A CERES plot reduces to a PR plot when $\hat{b}_2'\tilde{E}(X_1|X_2)$ is a linear function of X_2. Cook (1993) provided further discussion on the construction of$\tilde{E}(X_1|X_2)$

Partial residuals as defined in Eq. (9) reduce to the usual definition of partial residuals in additive-error models Eq. (5) because then the link is the identity function and $h' = 1$. For logistic regression Eq. (4)

$$(y-\hat{\mu}_f)h'(\hat{\mu}_f) = \frac{y-\hat{\mu}_f}{\hat{\mu}_f(1-\hat{\mu}_f)}$$

and the partial residuals Eq. (9) reduce to those defined by Landwehr *et al.* (1984) when the response is binary. Recall that in our formulation, $y = y^* | n$. Generally, the first term on the right of Eq. (9) can be interpreted in terms of η as the score scaled by the expected information per observation, all evaluated at $\hat{\mu}_f$, that

$$(y-\mu)\,h'(\mu) = \frac{\partial\, log\, d_{y|x}/\partial\eta}{-E\{\partial^2 log d_{y|x}/\partial\eta^2\}}$$

because $$E\left(\partial\, log\, d_{y|x}/\partial\eta\right) = 0$$

and $$-E\{\partial^2 log d_{y|x}/\partial\eta^2\} = E\left(\partial\, log\, d_{y|x}/\partial\eta\right)^2$$

$(y - \mu)\, h'(\mu)$ can also be interpreted as the standardized score weighted by the inverse of its standard deviation. If we let $\hat{\eta}_f(X) = \hat{\eta}_f\,(X|\hat{b})$, the quantity $\hat{\eta}_f + (y - \hat{u}_f)h'\,(\hat{u}_f)$ The adjusted dependent variable, which is used in iterative estimation techniques such as the Newton-Raphson process, is often referred to as the adjusted dependent variable (McCullagh and Nelder, 1983). Expression Eq. (8) coincides with all partial residual definitions that we are aware of, including those of Collett (1991) and McCullagh and Nelder (1983). However, maximum likelihood estimation is not needed, and 'g' may be a function of multiple predictors, necessitating the use of three-dimensional plots when $p_2 = 2$. Fitting a regression curve to the PR plot $\{\widehat{pr}_2, X_2\}$ should yield a useful approximation of 'g' up to a linear transformation if the correlation between $g(X_2)$ and the regression function $E(\widehat{pr}_2\,|X_2)$ is sufficiently high. Because obtaining a closed-form for $E(\widehat{pr}_2\,|X_2)$ is difficult.

An approximation is used to study the relationship between $E(\widehat{pr}_2\,|X_2)$ and $g(X_2)$ and to use $g(X_2) = bX_2$. So the CERES and PR plots for BRM can be constructed by using Eq. (9) & (10). The first derivative of the binomial regression link function given in equation (4) is

$$h'\,(\hat{\mu}_f) = \frac{1}{\mu(1-\mu)}$$

Hence the fitted model by using log link for binomial regression can be expressed as

$$\hat{\mu}_f = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1' x_1 + \hat{\beta}_2' x_2}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1' x_1 + \hat{\beta}_2' x_2}}$$

where the regression estimators are $\hat{\beta}_0, \hat{\beta}_1', \hat{\beta}_2'$ the fitted model is $\hat{\mu}_f$ and the predictors are x_i. Similarly, the CERES and partial residual for a model with p explanatory variables can be expressed as

$$\widehat{pr}_i = (y - \hat{u}_f)h'\,(\hat{u}_f) + \hat{b}_i' X_i \qquad i = 1, 2, \ldots, p. \tag{11}$$

$$\widehat{cr}_i = (y - \hat{\mu}_f)h'\,(\hat{\mu}_f) + \hat{b}_i' \tilde{E}(X_i|X_i) \qquad i = 1, 2, \ldots, p. \tag{12}$$

In addition, for p explanatory variables, the fitted model is

$$\hat{\mu}_f = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1' x_1 + \hat{\beta}_2' x_2 + \ldots \hat{\beta}_p' x_i}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1' x_1 + \hat{\beta}_2' x_2 + \cdots \hat{\beta}_p' x_i}} \tag{13}$$

3. Numerical Example

In this section we considered a real data to check the performance of CERES and PR plots in the detection of outliers in binomial regression. This techniques discussed in the previous section is enforce here on the Liver cancer data used first by Zelterman (1999) and also later by Atkinson and Riani (2001). Zelterman (1999) quoted data on the incidence of liver cancer in mice, which we reproduce in (Appendix A_1), given the number of mice developing cancer and the total number tasted, which forms the binomial

denominator. There are eight doses, units unspecified, and observation are taken at nine unequally spaced times, making 72 observations in all. In this data, Liver cancer is regarded as regressand variable (Y) with two regressors, i.e. dose of a patient (X_1), and months of study (X_2). The data contains 72 observations. The response variable follows a binomial distribution and therefore a binomial regression model is applied here. The CERES and PR plots of BRM for the Liver cancer data are shown in Figures 3.1 and 3.2. Because we have two predictors in the model, so there are two possible CERES and PR plots can be obtained. The summary of binomial regression model for Liver cancer data is presented in Table 3.1. Based on the result, it shows that both of independents variables are significant (p-value < 0.05).

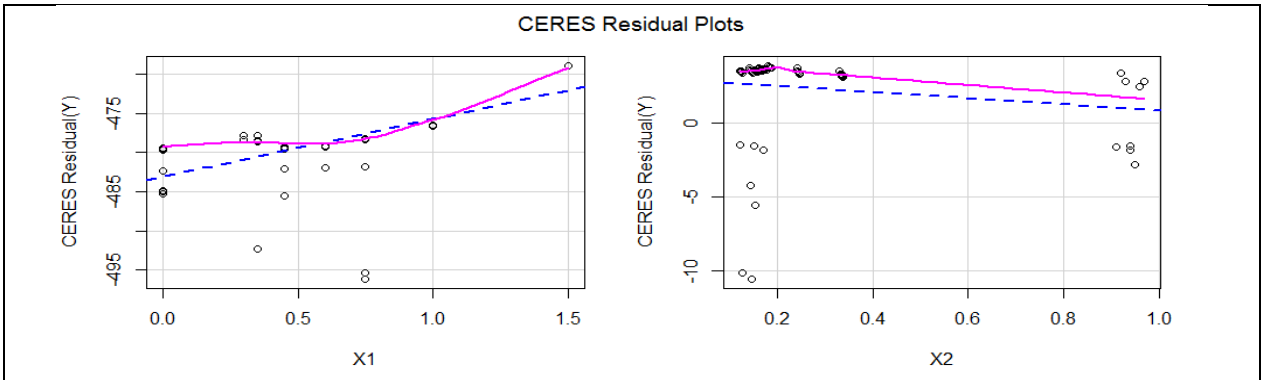
Table 3.1 Binomial Regression Analysis for Liver Cancer Data

Predictors	Coefficients	Standard error	t-test	p-value
Constant	0.411	0.124	3.32	0.001
X_1	0.1972	0.0905	2.18	0.033
X_2	0.01788	0.00588	3.04	0.003

$R^2 = 16.88\%,\ R^2(adj) = 14.47\%$

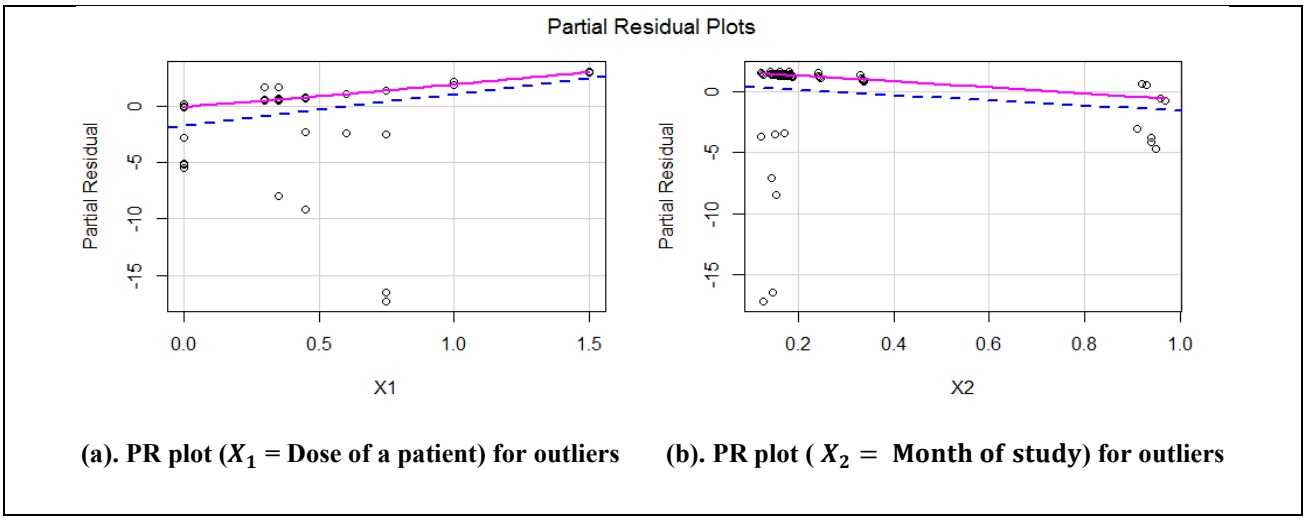
$$\hat{Y} = 0.411 + 0.1972\,\text{X1} + 0.01788\,\text{X2}$$

First we applied Grubbs test to check whether or not there is any outlier in Liver cancer data. Here the test statistics value of Grubbs test is 3.107 and p-value is 0.00047, which shows that outlier exist in the Liver cancer data.



(a). CERES Plot (X_1 = Dose of a patient) for outliers (b). CERES plot (X_2 = Month of study) for outliers

Figure 3.1 CERES plots for binomial regression model for Liver Cancer Data



(a). PR plot (X_1 = Dose of a patient) for outliers **(b). PR plot (X_2 = Month of study) for outliers**

Figure 3.2 PR plots for binomial regression model for Liver Cancer Data

By using the real data set, outliers are clearly visible in Figure 3.1 (a) & (b) and present the CERES plots of X_1 = Dose of a patient and X_2 = Month of study , respectively. Figure 3.2 (a) & (b) present the PR plots, of X_1 = Dose of a patient and X_2 = Month of study , respectively. In Figures 3.1 & 3.2 CERES and PR are plotted against each regressor respectively. The Liver cancer data is used first by Zelterman (1999) and also later by Atkinson and Riani (2001). The Zelterman used liver cancer data in his study, and they identified outlier observations are 11, 12, 20, 42, 48 and 67. In our study, it is found that outliers observations in CERES plots are 11th, 12th, 19th, 44, 51th and 67th, on the other hand outliers observations in PR plots are 5th, 11th, 12th, 27th, 46th and 62th. Three identified observations (11th, 12th, and 67th) by CERES plots and two identified observaions (11th and 12th) by PR plots are coincide with the identified outlier observation by Zelterman (1999). From both the plots, it is found that CERES plot identify outlier more accurately and gives better visual diagnostics for outliers as compared to PR residual plots. In the next section, the detection of outliers is made by using simulated datasets.

4. Monte Carlo Simulation

We use the Monte Carlo simulation used by Amin et al. (2019). In this study, the simulation's computational scheme and related model are as follows:

$$X_{ij} = \sqrt{(1-\theta^2)}Z_{ij} + \theta Z_{i(j+1)} \qquad i = 1,2,\ldots,n;\ j = 1,2,\ldots,p$$

where Z_{ij} is provided by the standard normal distribution, i.e. Z_{ij} N(0,1), and θ is the degree of multicollinearity in the above simulation equation, which is set to 0.8, 0.9, 0.95, and 0.99. We replaced the 19th, 21th, 23th and 25th observations in the entire data set with outlying observations in the independent variables X's.

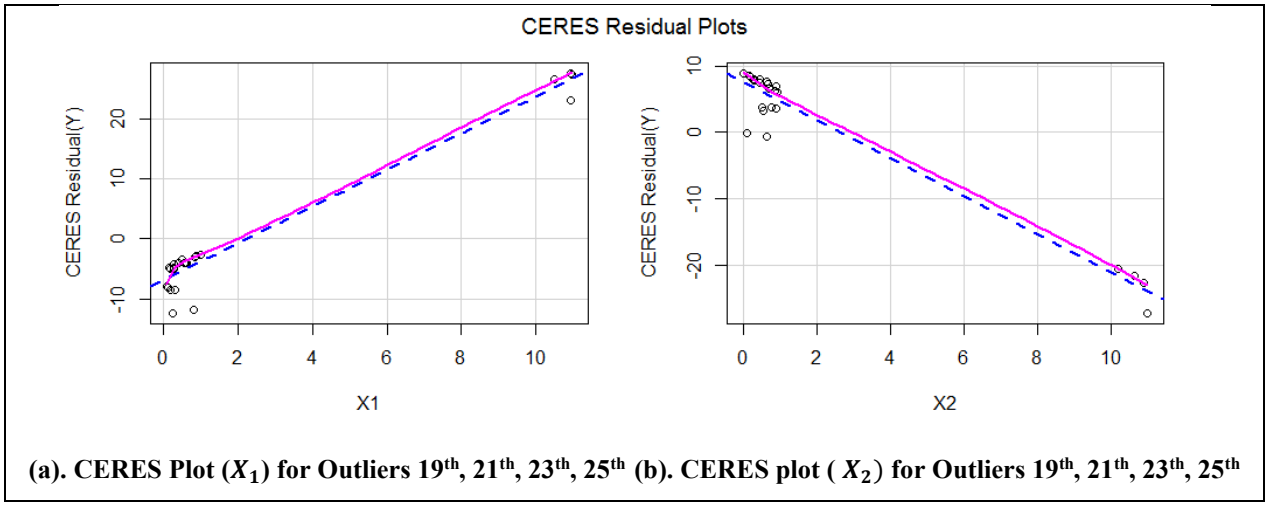
$$X_{ij} = X_{ij} + \alpha_o \qquad \text{i =19, 21, 23, 25.} \quad \text{j = 1, 2, … , p,}$$

where,

$$\alpha_o = \bar{X}_j + 10$$

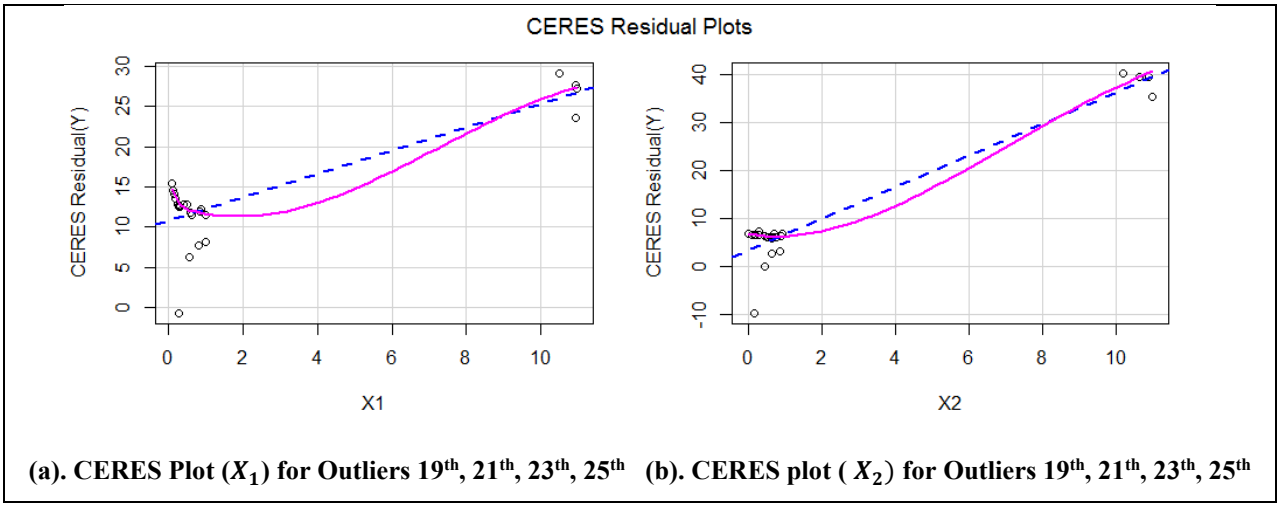
$$\hat{\mu}_i = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1' x_1 + \hat{\beta}_2' x_2}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1' x_1 + \hat{\beta}_2' x_2}}$$

The response variable is generated randomly as $y \sim B(1, \hat{\mu}_i)$. The regression coefficients are considered to be fixed as $\beta_0 = \beta_1 = \beta_2 = 1$.We chose four different sample sizes, with n being 25, 50, 100, and 200, respectively. Each of the result is based on 10,000 simulations using the R software. The performance of the CERES and PR plots in BRM are assessed. The graphical displays of the CERES and the PR plots are presented in Figures 4.1 to 4.8.



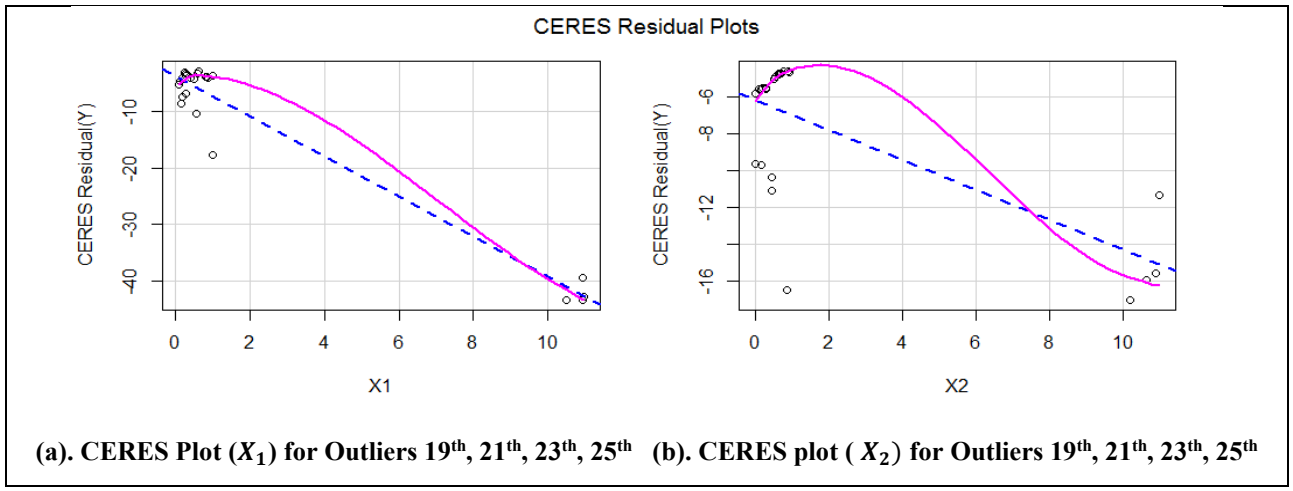
(a). CERES Plot (X_1) for Outliers 19th, 21th, 23th, 25th (b). CERES plot (X_2) for Outliers 19th, 21th, 23th, 25th

Figure 4.1 CERES plots for binomial regression model for simulated data, $n = 25$



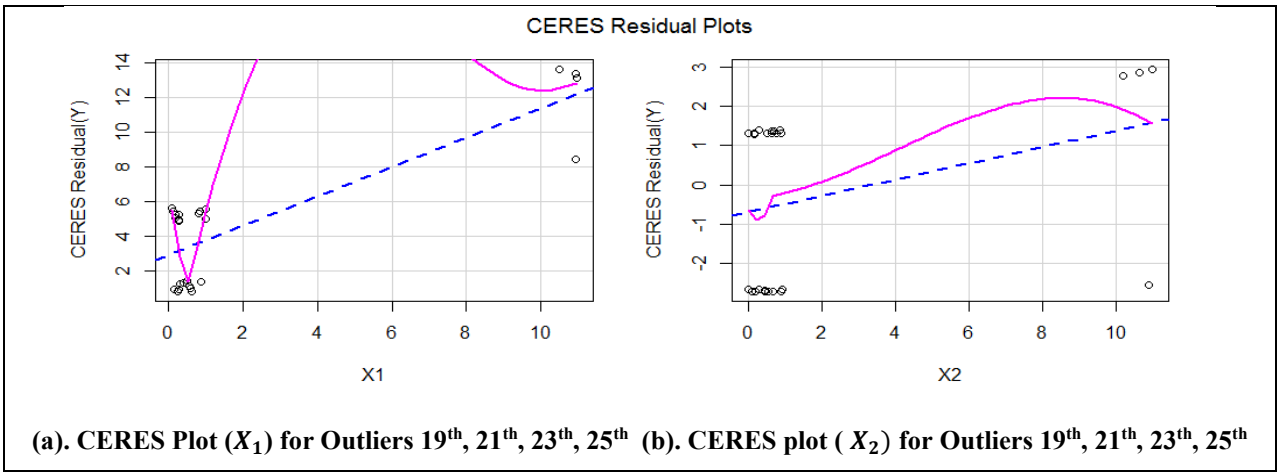
(a). CERES Plot (X_1) for Outliers 19th, 21th, 23th, 25th (b). CERES plot (X_2) for Outliers 19th, 21th, 23th, 25th

Figure 4.2 CERES plots for binomial regression model for simulated data, $n = 50$



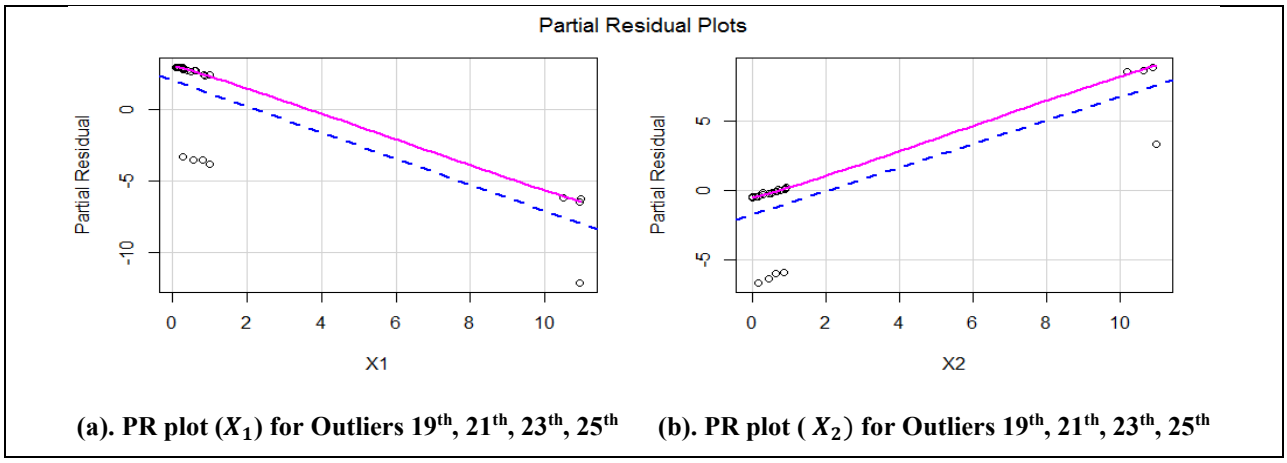
(a). CERES Plot (X_1) for Outliers 19th, 21th, 23th, 25th **(b). CERES plot (X_2) for Outliers 19th, 21th, 23th, 25th**

Figure 4.3 CERES plots for binomial regression model for simulated data, $n = 100$



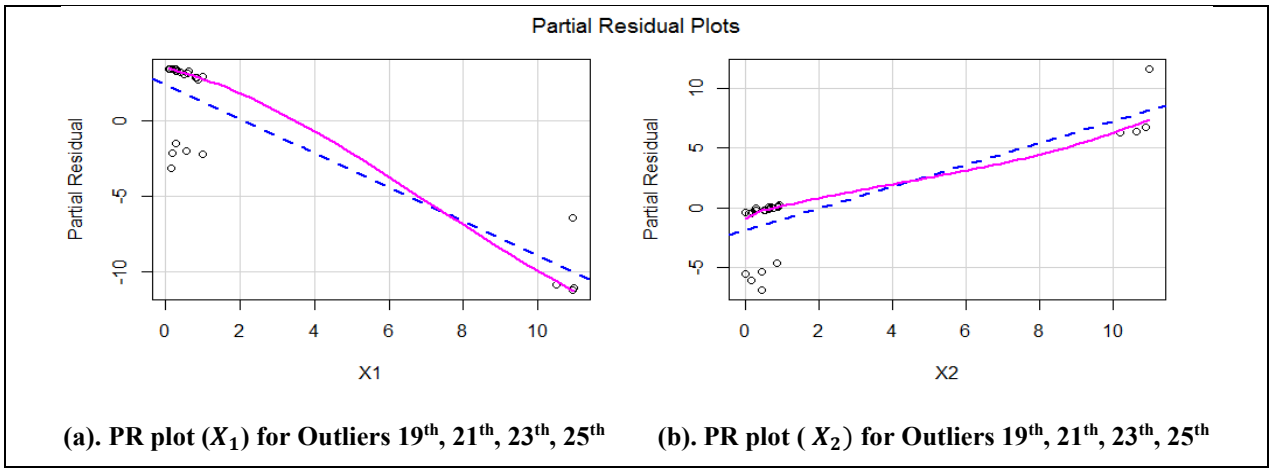
(a). CERES Plot (X_1) for Outliers 19th, 21th, 23th, 25th **(b). CERES plot (X_2) for Outliers 19th, 21th, 23th, 25th**

Figure 4.4 CERES plots for binomial regression model for simulated data, $n = 200$



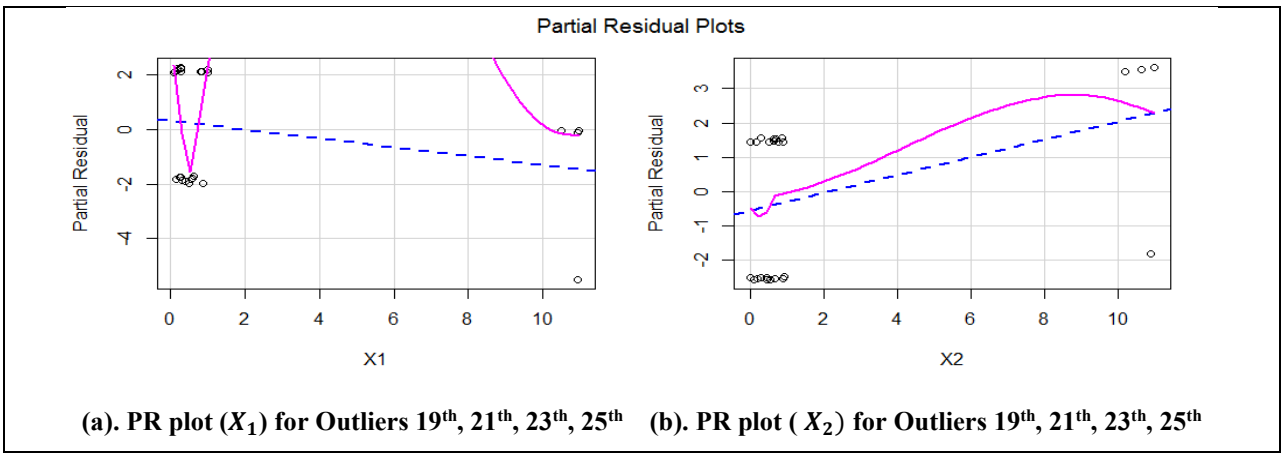
(a). PR plot (X_1) for Outliers 19th, 21th, 23th, 25th **(b). PR plot (X_2) for Outliers 19th, 21th, 23th, 25th**

Figure 4.5 PR plots for binomial regression model for simulated data, $n = 25$



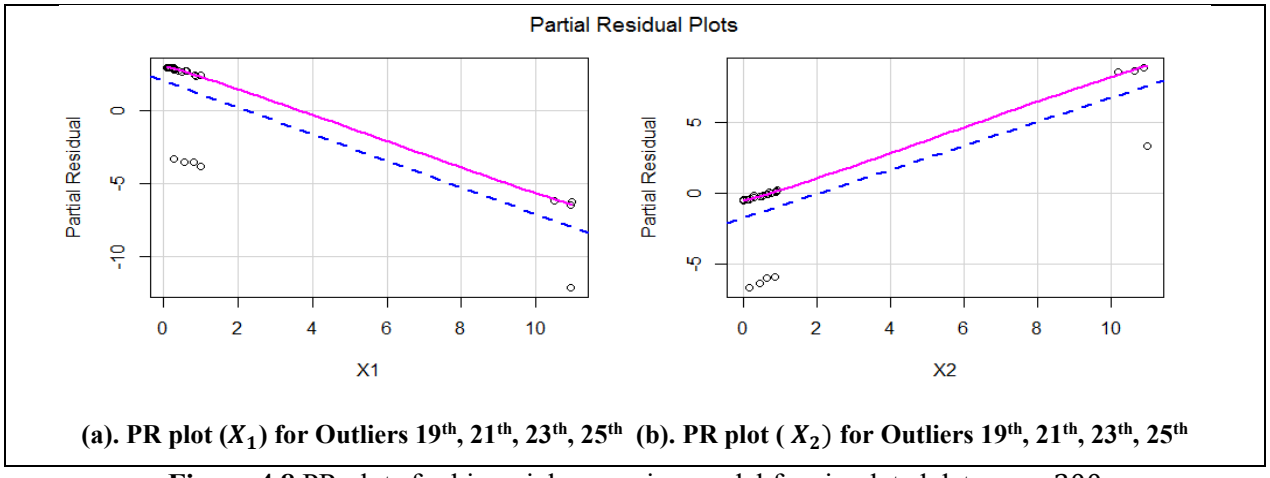
(a). PR plot (X_1) for Outliers 19th, 21th, 23th, 25th **(b). PR plot (X_2) for Outliers 19th, 21th, 23th, 25th**

Figure 4.6 PR plots for binomial regression model for simulated data, $n = 50$



(a). PR plot (X_1) for Outliers 19th, 21th, 23th, 25th **(b). PR plot (X_2) for Outliers 19th, 21th, 23th, 25th**

Figure 4.7 PR plots for binomial regression model for simulated data, $n = 100$



(a). PR plot (X_1) for Outliers 19th, 21th, 23th, 25th **(b). PR plot (X_2) for Outliers 19th, 21th, 23th, 25th**

Figure 4.8 PR plots for binomial regression model for simulated data, $n = 200$

By using the simulated data set, outliers is clearly detected in Figures 4.1 to 4.4, by using CERES plots, while Figures 4.5 to 4.8, i.e PR plots also detected outliers but not clearly deected. For different sample sizes, outliers is detected by both CERES and PR plots and found observations 19^{th}, 21^{th}, 23^{th} and 25^{th} are outliers i.e both the CERES and the PR plots correctly identify outliers.The CERES and PR plots is plotted against each regressors *i.e* (x_1, x_2). All the figures were made that clearly detects the outlier's issue. In comparison to the PR plots, the outliers in the CERES plots are far away from the CERES residuals trend line. In comparison to CERES and PR plots, the CERES plot provides a clearer visual diagnostic for outliers as compare to PR plots.

5. Conclusions

This article addresses the implementation of CERES and PR plots for the identification of outliers in a binomial regression model and then compare these two plots based on real life and simulated data. The real life data shows that CERES performs better in detecting outlier in binomial regression model than PR plots. However, in case of simulated data, both of these plots can successfully detect the all outliers in binomial regression model but the visualization to detect is better in CERES plots than PR Plots. Therefore, CERES plots can be used as a diagnostic tool to detect outlier in binomial regression model. Furthermore, this research can be extended by increasing the contamination level and calculating the swamping and masking rate in binomial regression model.

6. Acknowledgements

Authors would like to thanks the referees and the editor for their very careful readings and invaluable comments that led to improvement in the presentation of this article.

7. References

Amin, M., Amanullah, M., Aslam, M., & Qasim, M. (2019). Influence diagnostics in gamma ridge regression model. *Journal of Statistical Computation and Simulation*, 89(3): 536-556.

Atkinson, A. C., & Riani, M. (2001). Regression diagnostics for binomial data from the forward search. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 50(1): 63-78.

Berk, K. N., & Booth, D. E. (1995). Seeing a curve in multiple regression. *Technometrics*, 37(4): 385 398.

Breiman, L. & Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlations (with discussion). *Journal of the American Statistical Association*. 80 (391): 580–619.

Collett, D. (1991). *Modeling Binary data*, London: Chapman and Hall.

Cook, R. D. (1993). Exploring partial residual plots. *Technometrics*, 35(4): 351-362.

Cook, R. D., Croos-Dabrera, R., (1998). Partial residual plots in generalized linear models. *Journal of the American Statistical Association*, 93(442): 730–739.

Fowlkes, E. B. (1987), Some diagnostics for binary logistic regression via smoothing. *Biometrika*, 74: 503-515.

Imran, M., & Akbar, A. (2020). Diagnostics via partial residual plots in inverse Gaussian regression. *Journal of Chemometrics*, 34(1):e3203.

Landwehr, J. M., and Pregibon, D. (1993), Comments on 'Improved added variable and partial residual plots for the detection of influential observations in generalized linear model' by R. J. O' Hara Hines and E. M. Carter, *Applied Statistics*, 42: 16-19.

Landwehr, J. M., Pregibon, D., and Shoemaker, A, C. (1984), Graphical methods for assessing logistic regression models, *Journal of the American Statistical Association*, 79: 61-83.

McCullagh, P. & Nelder, J. A. (1983). *Generalized Linear Models. London*: Chapman and Hall.

Nelder. J. A. & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society*. Series A (General), 135: 370-84.

Park, M. Y., & Hastie, T. (2007). L1-regularization path algorithm for generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4): 659-677.

Zelterman, D. (1999) *Models for Discrete Data*. Oxford university Press.