JOURNAL OF STATISTICAL MODELING & ANALYTICS (JOSMA)
(ISSN: 2180-3102)

UNIVERSITI MALAYA

Multicollinearity in Binomial Regression: A Comparison between Conditional expectations and residuals (CERES) and Partial residual (PR) Plots for Detection

Nasir Saleem[1*], Atif Akbar[1], A. H. M. Rahmatullah Imon[2] & Javaria Ahmad khan [1]

[1]*Department of Statistics, Bahauddin Zakariya University, Multan, Pakistan*
[2] *Department of mathematical sciences, Ball State University, USA.*

* *Corresponding Authors: nasirsaleem160@gmail.com*

RESEARCH ARTICLE

Received 20 November 2023
Accepted 5 April 2024
Published 30 May 2024

Abstract

Conditional expectations and residuals (CERES) and partial residual (PR) plots have been used in linear regression model for the identification of multicollinearity. But not much work has been done on how they perform in generalized linear models (GLM). Binomial regression model (BRM) is a very important type of GLM which has wide applications in dealing with heart disease and many other types of data. In this paper we have offered a comparison between CERES and PR plots in BRM to detect the multicollinearity problem. At first, we have developed a comparison tool and then apply them to real-world and simulated data. We observe the performance of these plots on the detection of a possible multicollinearity separately. We observe that both these plots perform well in order to diagnose this problem for a real data. However, the overall performance of the CERES plot is found better as compared to the PR plots.

Keywords: Binomial regression model, CERES, Diagnostics, GLM, Multicollinearity and PR plots.

1. Introduction

Probably, regression analysis is the most popular statistical method where the connection between a dependent variable and independent variables is described. No exact relationship exists due to the existence of some factors which cannot be explained by the relationship, and known as errors. It is assumed in classical regression that those errors are normally distributed and so that the response variable also follows a normal distribution. But usually in real life problems, ideal condition does not discover and we have to adopt alternative method, named as generalized linear model (GLM) (Nelder and Wedderburn, 1972). GLM is a lithe generalized approach of a linear regression modal (LRM) that tolerates the other than normal distribution of the response variable via a link function. In this paper, we considered a response follows a binomial distribution which has wide applications where survival from disease, from accident etc. are under study. VIF is a method to check multicollinearity in regressors.

The parameter estimation of GLM depends on some standard assumptions. There are number of different diagnostic tests which are designed to find problems with the assumptions of any statistical procedure. In the multiple regression model, there is one basic assumption is that there is no perfect multicollinearity. This issue arises when two or combinations of variables are correlated. It has several consequences. Multicollinearity often causes a wrong sign problem (Mullet, 1976). It may inflate the variances of the estimators and consequently the significant estimators may find insignificant.

In this paper, we are going to use PR and CERES plots. PR plots was presented by Ezekiel (1924), to observe the direction of a regressor variable graphically. Larsen and McCleary (1972) should get credit for the name PR plots. PR plots also called component plus residual plots. Many authors extended this idea and made augmented PR plots (Mallows, 1986**)** and CERES plots (Cook, 1993). Further, the properties of PR plots were explored by Cook (1993) and Cook and Croos-Dabrera (1998). Berk and Booth (1995) compared PR plots with several other diagnostic plots. Fowlkes (1987) suggested an adaption of PR plots for logistic regression. Landwehr (1983) suggested the use of these plots for logistic regression. Fowlkes (1987) and Landwehr *et al.* (1984) claimed that PR plots are helpful in detecting nonlinearity in binary logistic regression. Landwehr and Pregibon (1993) studied our considered plots for GLM using canonical links. Kahng and Lee (2004) discussed the usefulness of CERES plots in GLM. Parkand Hastie (2007) discussed the technique of algorithm for regularized the GLM. Imran and Imran and Akbar (2020) discussed the construction of PR plots using response residuals for the inverse Gaussian regression model (IGRM). Saleem et al. (2022) used and compared the CERES and PR plots in detecting the heteroscedasticity problem using Liver cancer and simulated data. Hussain and Akbar (2022) also discussed the importance of partial residual plots by using chemical species data. Saleem et al. (2022) used and compared the CERES and PR plots in detecting the outlier's problem using Liver cancer and simulated data.

Now in this research, CERES and PR plots are created for binomial regression model (BRM). This will provide simple, powerful and wide applicable technique to the researchers for computational ease. These plots provide suitable diagnostics for model specification. This research discovers such idea while offering the importance of CERES and PR plots in regression diagnostics without examine the predictable tests. Finally, we will compare CERES and PR plots, and also identify which plot performs better in the detection of multicollinearity.

The rest of the paper is organized as, in section 2, construction of the comparison tools of CERES and PR plots for BRM. Section 3, discussed real data example heart disease data. Section 4, monte Carlo simulation and section 5 results and discussion.

2. Construction of the Comparison Tool

In this section, performance of CERES and PR plots will be compared in binomial regression for the detection of multicollinearity. The model is,

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

where $X = (X_1, X_2, \dots, X_p)'$ is the matrix of $n \times p$ explanatory variables, ε is $n \times 1$ vector and $\mathrm{Y} = (y_1, y_2, \dots, y_p)'$ is an $n \times 1$ vector of response and follows a binomial distribution with pdf

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

The mean and variance of y are $n\mu$ and $n\mu(1 - \mu)$, respectively. In logistic regression, which is main example in present research, $Y^*|X$ is a binomial (n, μ) random variable, its p may depend on X. n is independent and vary from observation to observation. The response $Y = Y^*/n$ is then the observed fraction of successes from a typical binomial trial and its link function can be written as (McCullagh and Nelder,1983; Cook and Cross-Debrrera,1998)

$$\eta = \theta = h(\mu) = log(\tfrac{\mu}{1-\mu}) \tag{3}$$

$\mu(\eta)$ = log (1 + exp (η)), and $v(\psi)$ =1/n.

Cook (1993) investigated the performance of PR plots and found that those are strongly depends on the conditional expectation $E(X_1|X_2)$, also he showed that if X_2 is linear in the $E(X_1|X_2)$, then performance of plots is outclass. Berk and Booth (1995) presented their research where they compared the methods for identifying the $g(X_2)$, with several plots, i.e. CERES plots (Cook 1993), PR plots, standard residual plots and a numerical method. This method is based on algorithm which was proposed by Breiman and Friedman (1985). They also described that PR plots are mostly used in GLM to examine the predictor transformation. But the results shows that effectiveness of PR plots can be limited in different aspects and they may not perform well in GLM. Most of the observation in CERES plot are very close to each other that is conjunction between the points, this is a case of severe multicollinearity. But on the other hand, in PR plots most of observation are not close to each other there is lack of multicollinearity.

On basis of this literature, we focus on PR plots of GLM and present a general definition of PRs Conclusions regarding CERES plots can be obtained in a straightforward way from the developments for PR plots. A (p_2 + 1) dimensional Cartesian coordinate plot of the scalar α versus the p_2-dimensionless vector b will occasionally be denoted by $\{a,\, b\}$, with the understanding that the first argument is assigned to the vertical axis and the coordinates of b are assigned to the "horizontal" axes. The data is summarized by fitting,

$$\eta_f(x|b) = h(\mu_f) = b_0 + {b_1}'X_1 + {b_2}'\imath(X_2) \tag{4}$$

where $\imath(X_2)$ is a user-defined function of X_2 and adaptation of will be discussed later. Estimated coefficients $\hat{b}_j$, j= 0, 1, 2, based on (4) are assumed to be obtained by minimizing a convex objective function. (for details see, Cook and Cross-Debrrera,1998)

A PR plot for X_2 is obtained by first setting $\imath(X_2) = X_2$ then constructing the (p_2 + 1)-dimensional plot $\{\widehat{pr}_2\,, X_2\}$, where

$$\widehat{pr}_2 = (y - \hat{u}_f)h'\,(\hat{u}_f) + \hat{b}_2'X_2 \tag{5}$$

is the PR for X_2, $h'(.)$ is the first derivative of h(.)

The CERES plot for X_2 is then the (p_2+ 1)-dimensional plot $\{\widehat{cr}_2, X_2\}$,

$$\widehat{cr}_2 = (y - \hat{\mu}_f)h'\,(\hat{\mu}_f) + \hat{b}_2'\tilde{E}(X_1|X_2) \tag{6}$$

where $\imath(X_2) = \tilde{E}(X_1|X_2)$. 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) described the construction of $\tilde{E}(X_1|X_2)$. So, in case if responses are binary the CERES and PR plots for BRM can be constructed by using equations (5) and (6).

The first derivative of the binomial regression link function given in equation (3) is

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

Hence, the fitted model by using log link for BRM can be expressed as

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

where $\hat{\beta}_0$, $\hat{\beta}_1'$, $\hat{\beta}_2'$ are the estimators; $\hat{\mu}_f$ denotes the fitted model; and x_i are the predictors. Similarly, for the model with p explanatory variables, the CERES and PR plots can be expressed as

$$\widehat{pr}_i = (y - \hat{u}_f)h'\,(\hat{u}_f) + \hat{b}_i'X_i \qquad i = 1, 2, \dots, p. \tag{7}$$

$$\widehat{cr}_i = (y - \hat{\mu}_f)h'\,(\hat{\mu}_f) + \hat{b}_i'\tilde{E}(X_i|X_i) \;\; i = 1, 2, \dots, p. \tag{8}$$

and model for px_is is

$$\hat{\mu}_f = \frac{e^{\hat{\beta}_0 + \hat{\beta}'_1 x_1 + \hat{\beta}'_2 x_2 + \ldots \hat{\beta}'_p x_i}}{1 + e^{\hat{\beta}_0 + \hat{\beta}'_1 x_1 + \hat{\beta}'_2 x_2 + \ldots \hat{\beta}'_p x_i}} \tag{9}$$

3. Real Data Example: A Heart Disease Data

Here, we consider a heart disease dataset to observe the performance of CERES and PR plots in the detection of multicollinearity for BRM. The methodology developed in the previous section is implemented here on the heart disease data previously used by by Ozkake *et al*. (2018). In this data, coronary heart disease is regarded as response variable (Y) with two explanatory variables, Age (X_1), and Age-group referred (X_2). The data contains 100 observations. The purpose of this study was to edify the major factors influencing heart disease. The Y (output variable) follows a binomial distribution and therefore we use a BRM here. The CERES and PR plots of BRM generated by real data are presented in Figures 1 and 2. As it is above mentioned, model has two predictors, so there are two possible CERES and PR plots that can be attained.

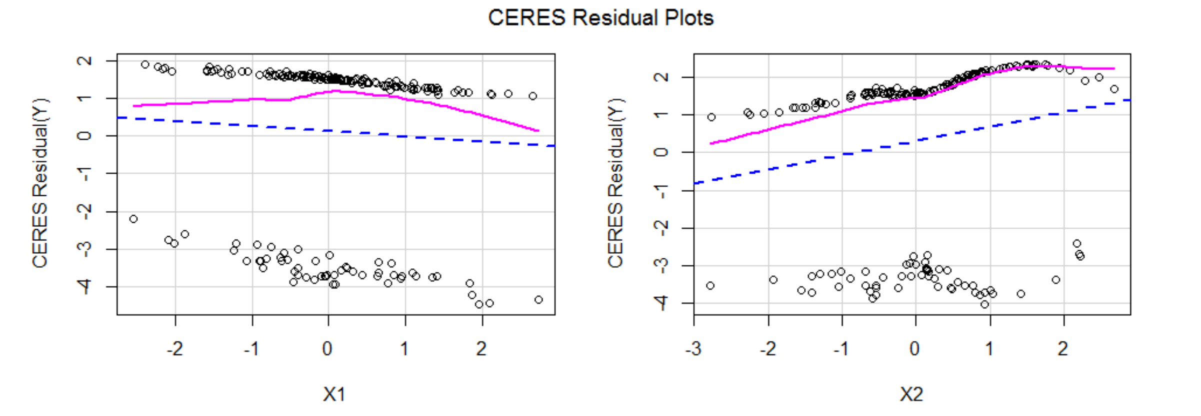
From Table 1, the inference related to BRM for Heart disease data and also the multicollinearity can be observed on the basis of variance inflation factor (VIF). The VIF is a method to check multicollinearity in regressor. If the value of VIF is less than five there is no multicollinearity in your data set. If the value of VIF lies between 5 and 10 there is high multicollinearity. If the values of VIF is greater than 10 there is severe multicollinearity. According to our results there is a severe multicollinearity in data set.

Table 1. Binomial Regression Analysis for Heart Disease Data

Variables	Coefficient	SE	T	Pr (>|t|)	VIF
Constant	-0.4555	0.5233	-0.87	0.386	
X_1	0.1780	0.02425	0.73	0.465	43.109
X_2	0.0213	0.1269	0.17	0.867	43.109

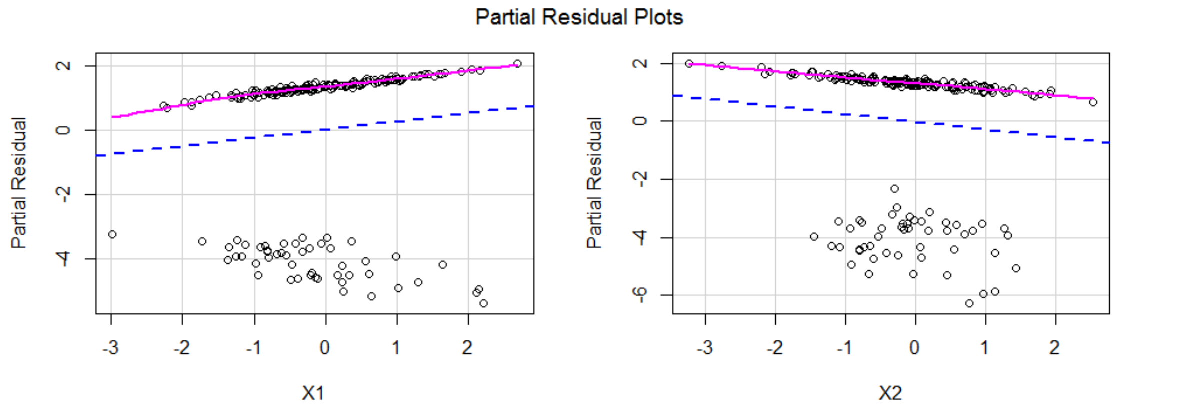
Note. S=0.430992, $R^2 = 26.5\%\%$, $R^2(adj) = 21.0\%$,Pearson correlation of X_1 and X_2 =0.988, p-value = 0.00

$$\hat{Y} = [-0.4555 + 0.1780X_1 + 0.0213X_2]$$



(a). CERES Plot (X_1 = Age) for multicollinearity **(b). CERES plot ($X_2 =$ Age − Group) for multicollinearity**

Figure 3.1: CERES plots for BRM for Heart disease data



(c). PR Plot (X_1 = Age) for multicollinearity **(d). PR plot ($X_2 =$ Age − Group) for multicollinearity**

Figure 3.2: PR plots for BRM for Heart disease data

In this heart disease data set, multicollinearity, were clearly observed in Fig 3.1, (a and b), are the CERES plots while, Figure 3.2, so (c and d) are PR plots, respectively. In these figures, CERES residuals and PR plots were plotted against each regressor i.e. X_1 and X_2 respectively. It is also observed that CERES and PR plots show the multicollinearity. It was found that several observations are very adjacent to each other that is conjunction between the points, shows the multicollinearity among the points. we also observed that CERES plots and PR plots for both of regressors ($X_1$and X_2) can clearly detect the multicollinearity, respectively.

4. Monte Carlo Simulation

In this study, multicollinearity is introduced in the data by following Amin *et al*. (2019). The simulation is conducted using the R software. The monte Carlo scheme and the relevant model for this simulation is given as

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

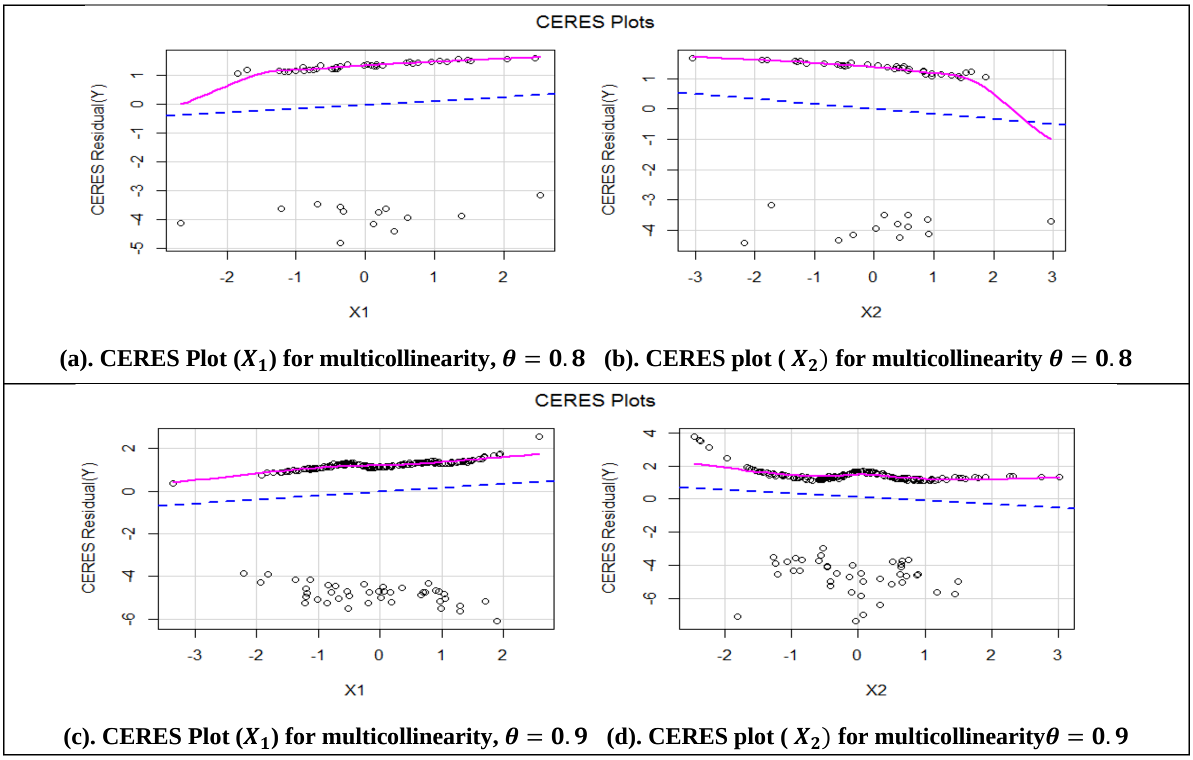
where $Z_{ij} \sim N(0,1)$ and θ is the level of multicollinearity set as 0.8, 0.9, 0.95, and 0.99 in the above simulation equation. These values are the multicollinearity level to checked multicollinearity effect. It is interesting note that when sample size is n=25 and level of collinearity (θ) are 0.8, 0.9, 0.95, and 0.99 increased as VIF and correlation increase. (see, Amin *et al*. 2019)

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

The output 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 have selected four different sample sizes, i.e., *n* is selected as 25, 50, 100, and 200 with 10,000 replications. The numerical results are presented in Table 2 and the Figures 3 to 10 represents their graphical representation.

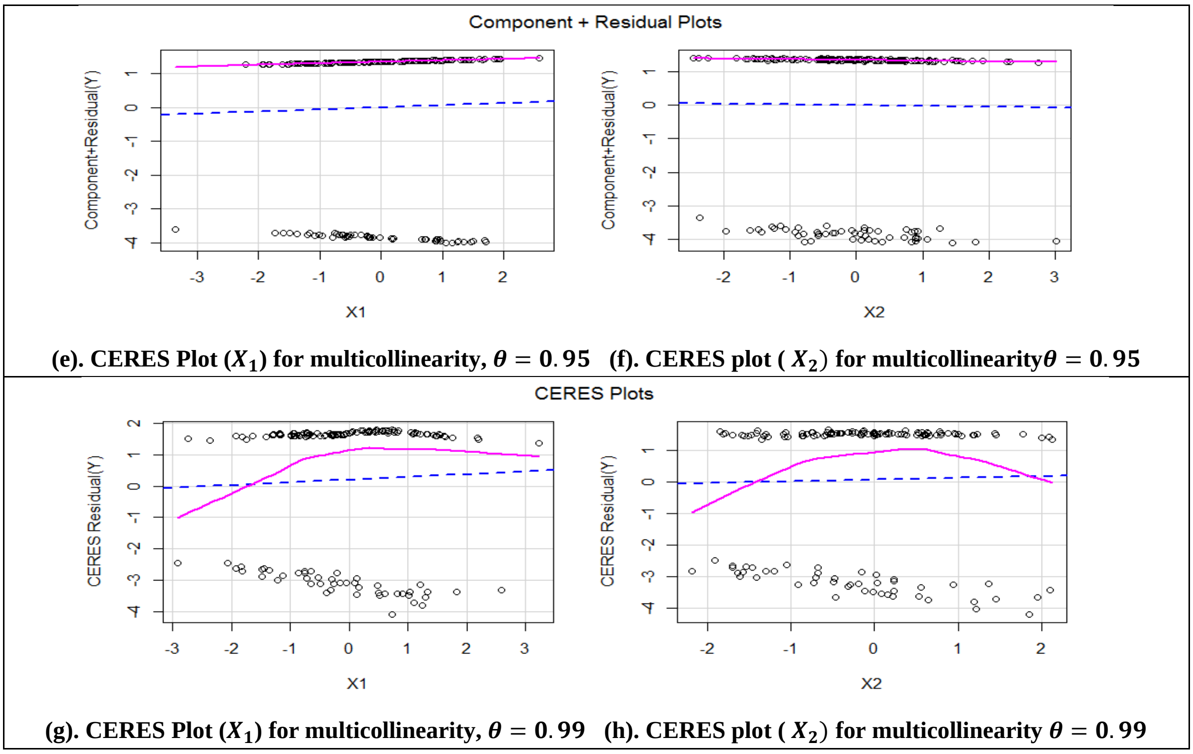
Table 2. Binomial Regression Analysis for Simulated Data

Sample size (n)	Level of collinearity (θ)	VIF	Correlation
n = 25	0.80	5.255	0.899
	0.90	5.772	0.909
	0.95	8.557	0.939
	0.99	8.943	0.942
n = 50	0.80	9.626	0.890
	0.90	10.070	0.942
	0.95	14.470	0.954
	0.99	18.019	0.967
n = 100	0.80	8.650	0.937
	0.90	14.528	0.962
	0.95	15.435	0.972
	0.99	15.886	0.988
n = 200	0.80	17.828	0.976
	0.90	24.422	0.981
	0.95	30.214	0.983
	0.99	37.633	0.986



(a). CERES Plot (X_1) for multicollinearity, $\theta = 0.8$ **(b). CERES plot (X_2) for multicollinearity $\theta = 0.8$**

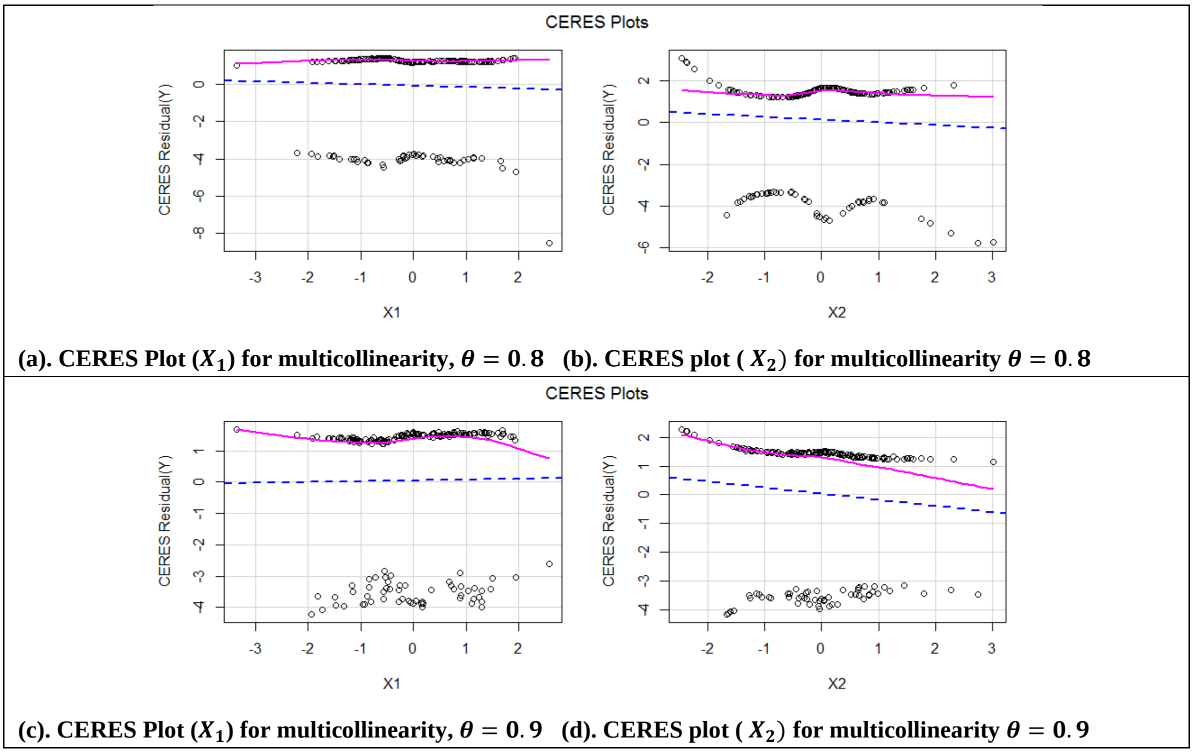
(c). CERES Plot (X_1) for multicollinearity, $\theta = 0.9$ **(d). CERES plot (X_2) for multicollinearity$\theta = 0.9$**



(e). CERES Plot (X_1) for multicollinearity, $\theta = 0.95$ **(f). CERES plot (X_2) for multicollinearity$\theta = 0.95$**

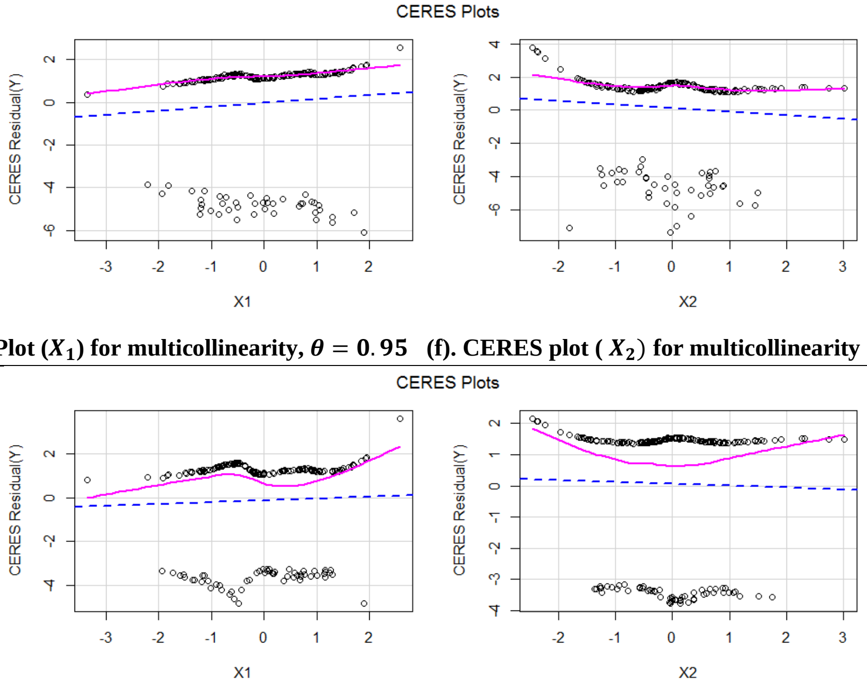
(g). CERES Plot (X_1) for multicollinearity, $\theta = 0.99$ **(h). CERES plot (X_2) for multicollinearity $\theta = 0.99$**

Figure 4.1: CERES plots when $n = 25$



(a). CERES Plot (X_1) for multicollinearity, $\theta = 0.8$ **(b). CERES plot (X_2) for multicollinearity $\theta = 0.8$**

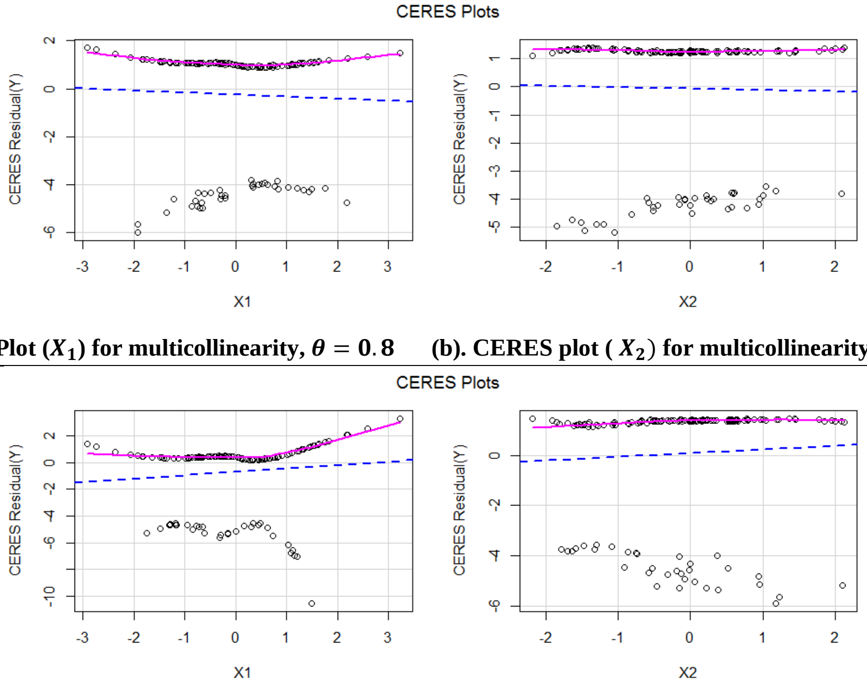
(c). CERES Plot (X_1) for multicollinearity, $\theta = 0.9$ **(d). CERES plot (X_2) for multicollinearity $\theta = 0.9$**



(e). CERES Plot (X_1) for multicollinearity, $\theta = 0.95$ **(f). CERES plot (X_2) for multicollinearity $\theta = 0.95$**

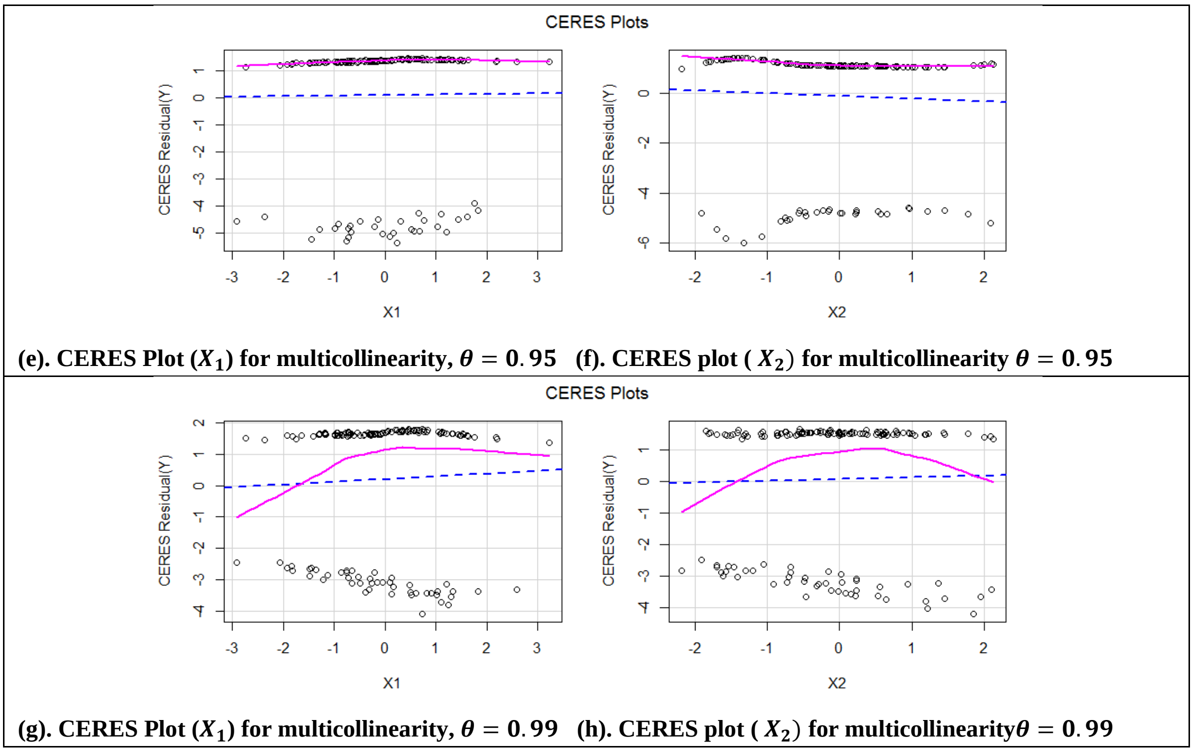
(g). CERES Plot (X_1) for multicollinearity, $\theta = 0.99$ **(h). CERES plot (X_2) for multicollinearity $\theta = 0.99$**

Figure 4.2: CERES plots when $n = 50$



(a). CERES Plot (X_1) for multicollinearity, $\theta = 0.8$ **(b). CERES plot (X_2) for multicollinearity $\theta = 0.8$**

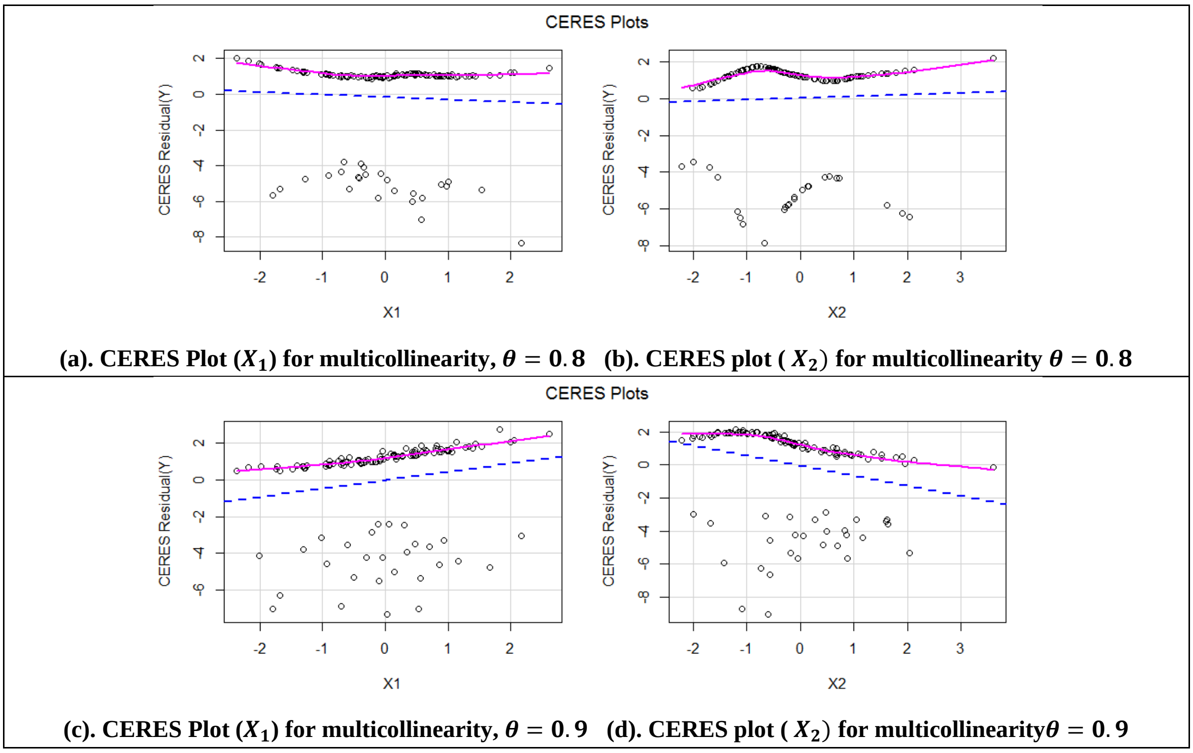
(c). CERES Plot (X_1) for multicollinearity, $\theta = 0.9$ **(d). CERES plot (X_2) for multicollinearity $\theta = 0.9$**



(e). CERES Plot (X_1) for multicollinearity, $\theta = 0.95$ **(f). CERES plot (X_2) for multicollinearity $\theta = 0.95$**

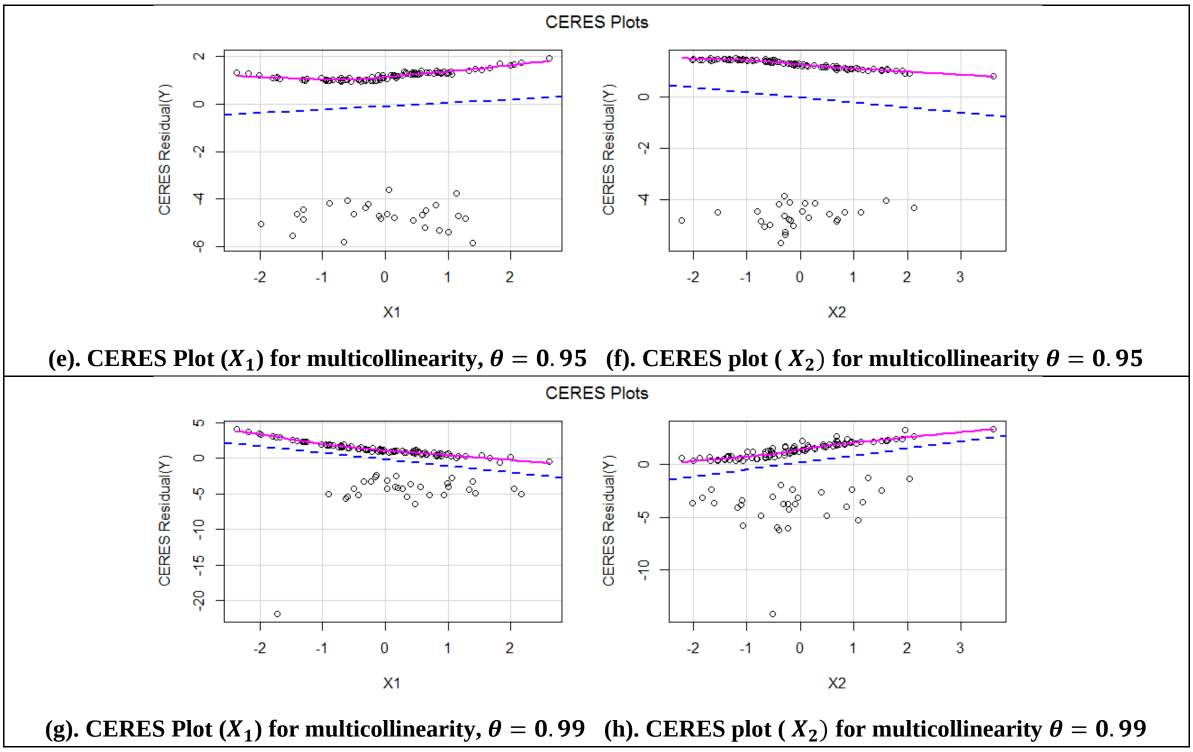
(g). CERES Plot (X_1) for multicollinearity, $\theta = 0.99$ **(h). CERES plot (X_2) for multicollinearity$\theta = 0.99$**

Figure 4.3: CERES plots when $n = 100$



(a). CERES Plot (X_1) for multicollinearity, $\theta = 0.8$ **(b). CERES plot (X_2) for multicollinearity $\theta = 0.8$**

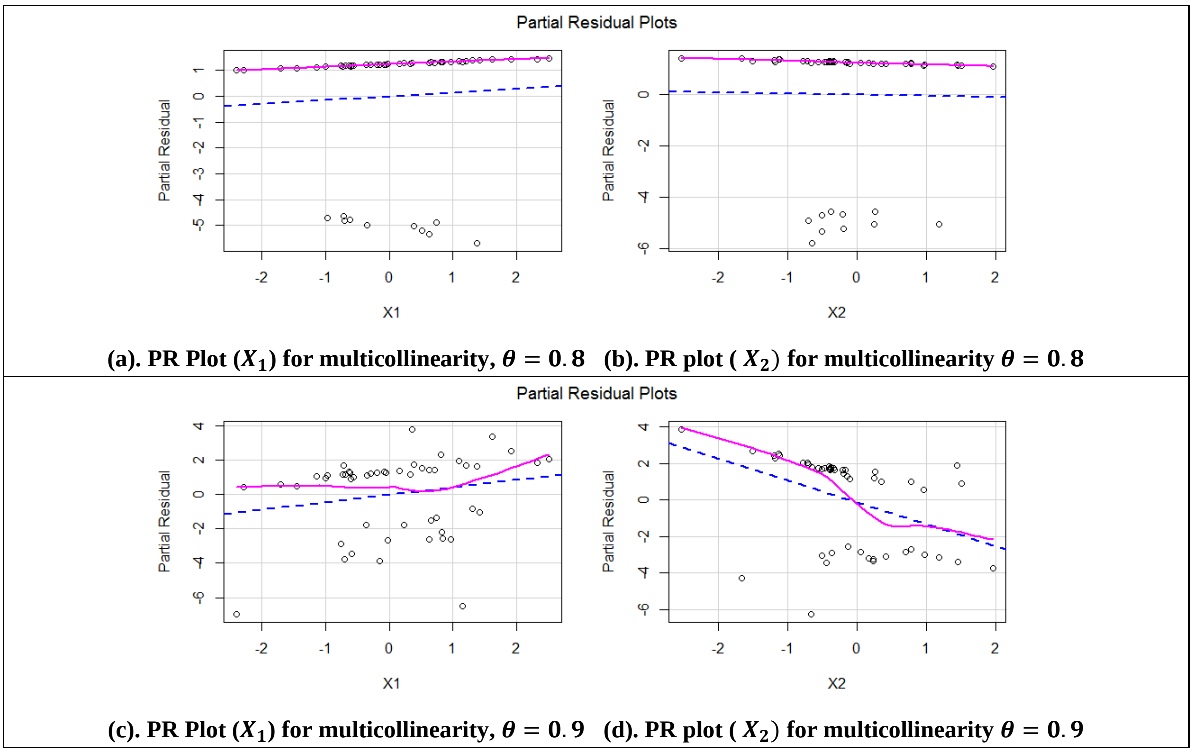
(c). CERES Plot (X_1) for multicollinearity, $\theta = 0.9$ **(d). CERES plot (X_2) for multicollinearity$\theta = 0.9$**



(e). CERES Plot (X_1) for multicollinearity, $\theta = 0.95$ (f). CERES plot (X_2) for multicollinearity $\theta = 0.95$

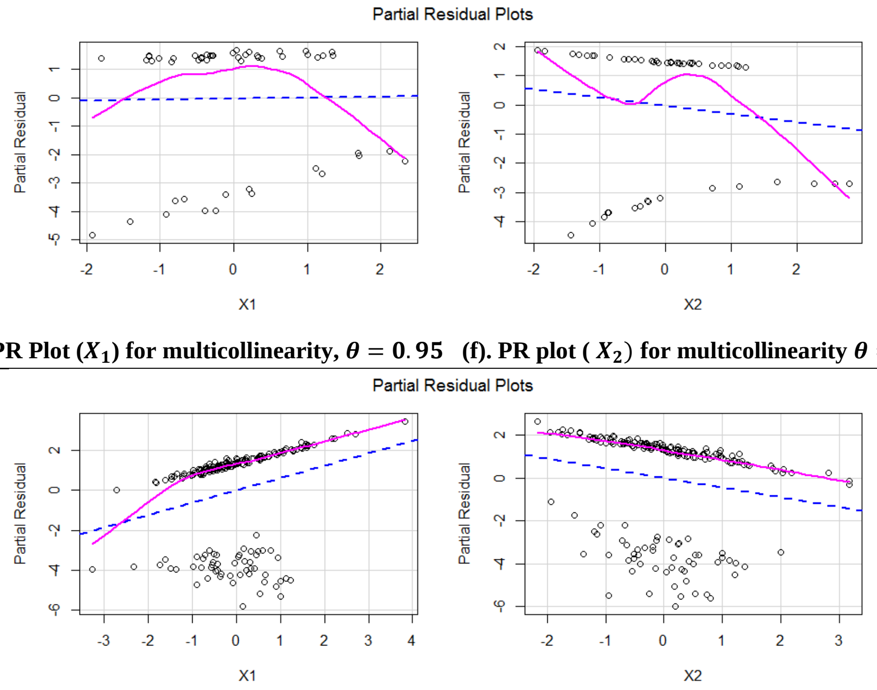
(g). CERES Plot (X_1) for multicollinearity, $\theta = 0.99$ (h). CERES plot (X_2) for multicollinearity $\theta = 0.99$

Figure 4.4: CERES plots when *n*=200



(a). PR Plot (X_1) for multicollinearity, $\theta = 0.8$ (b). PR plot (X_2) for multicollinearity $\theta = 0.8$

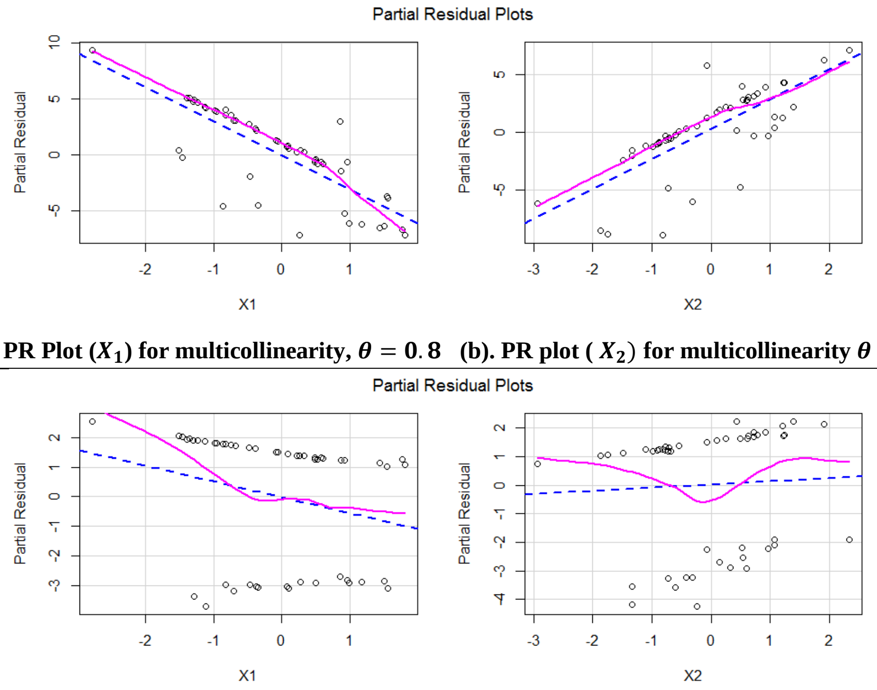
(c). PR Plot (X_1) for multicollinearity, $\theta = 0.9$ (d). PR plot (X_2) for multicollinearity $\theta = 0.9$



(e). PR Plot (X_1) for multicollinearity, $\theta = 0.95$ **(f). PR plot (X_2) for multicollinearity $\theta = 0.95$**

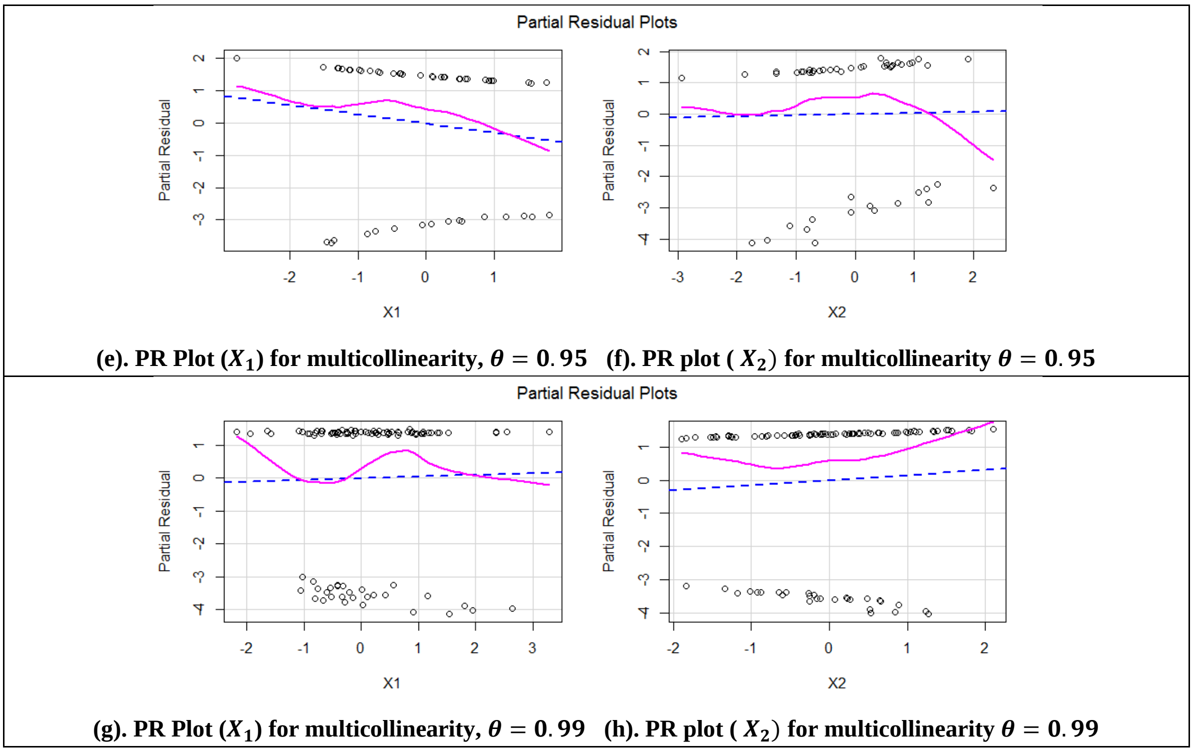
(g). PR Plot (X_1) for multicollinearity, $\theta = 0.99$ **(h). PR plot (X_2) for multicollinearity $\theta = 0.99$**

Figure 4.5: PR plots when $n = 25$



(a). PR Plot (X_1) for multicollinearity, $\theta = 0.8$ **(b). PR plot (X_2) for multicollinearity $\theta = 0.8$**

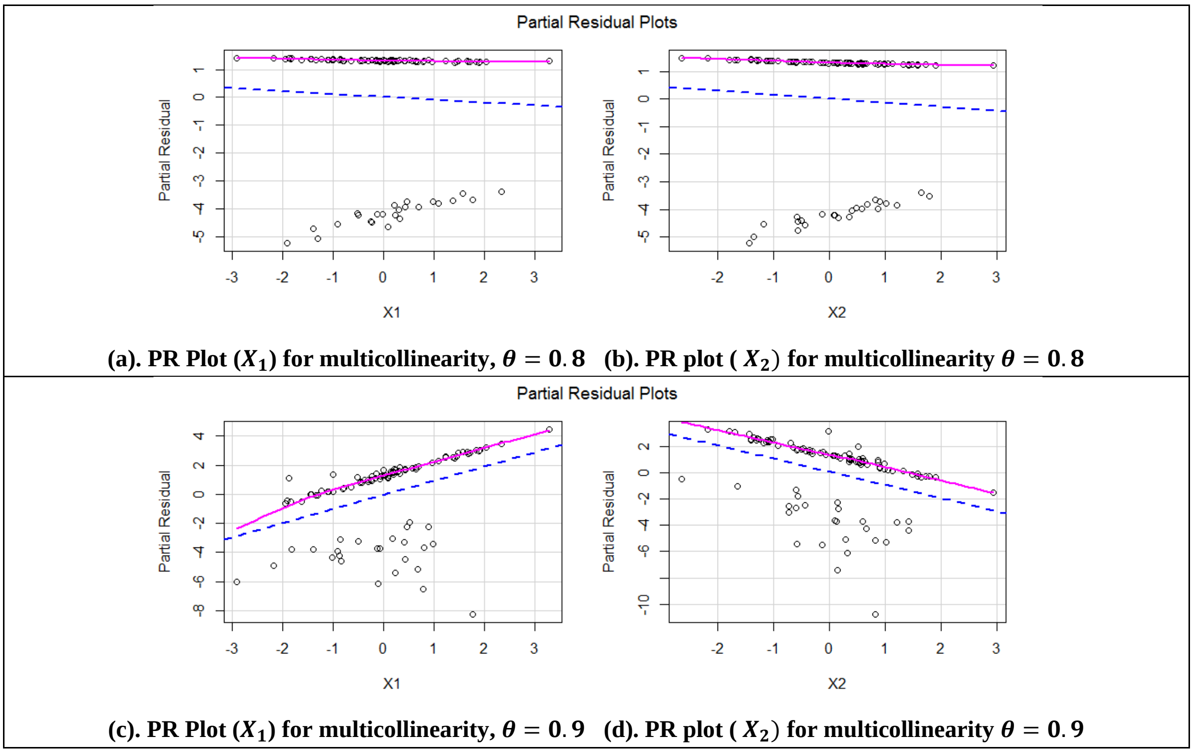
(c). PR Plot (X_1) for multicollinearity, $\theta = 0.9$ **(d). PR plot (X_2) for multicollinearity$\theta = 0.9$**



(e). PR Plot (X_1) for multicollinearity, $\theta = 0.95$ **(f). PR plot (X_2) for multicollinearity $\theta = 0.95$**

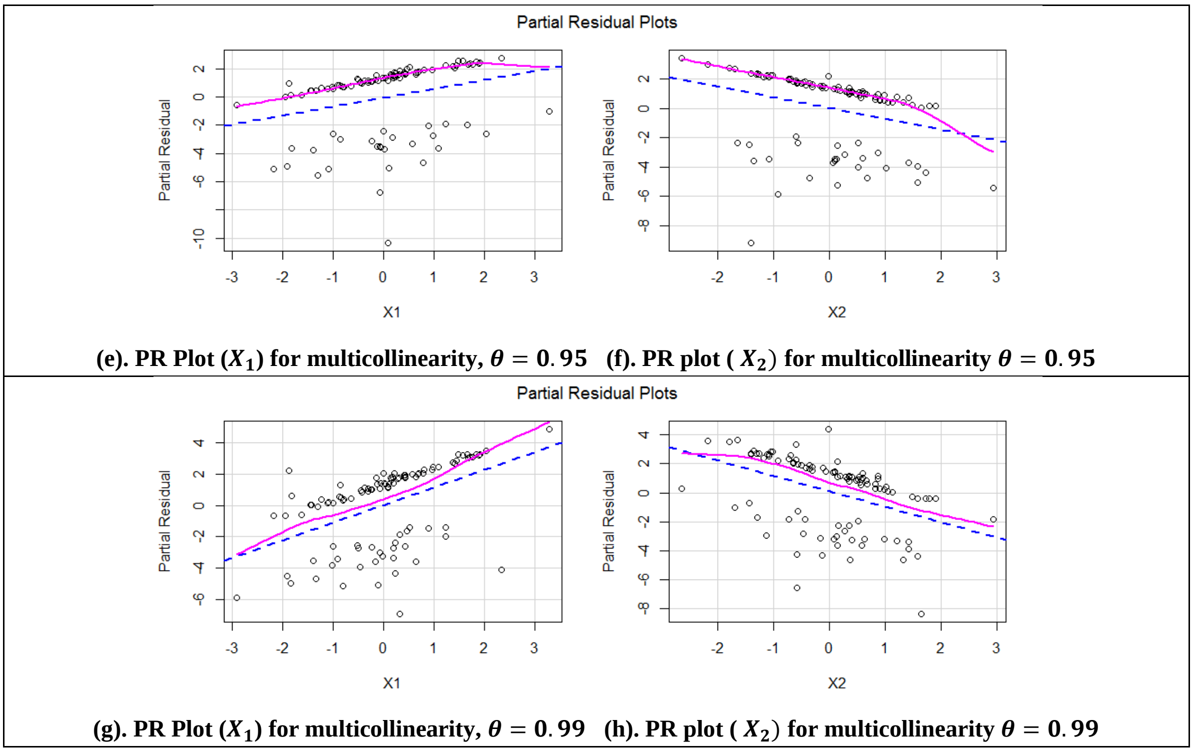
(g). PR Plot (X_1) for multicollinearity, $\theta = 0.99$ **(h). PR plot (X_2) for multicollinearity $\theta = 0.99$**

Figure 4.6: PR plots when $n = 50$



(a). PR Plot (X_1) for multicollinearity, $\theta = 0.8$ **(b). PR plot (X_2) for multicollinearity $\theta = 0.8$**

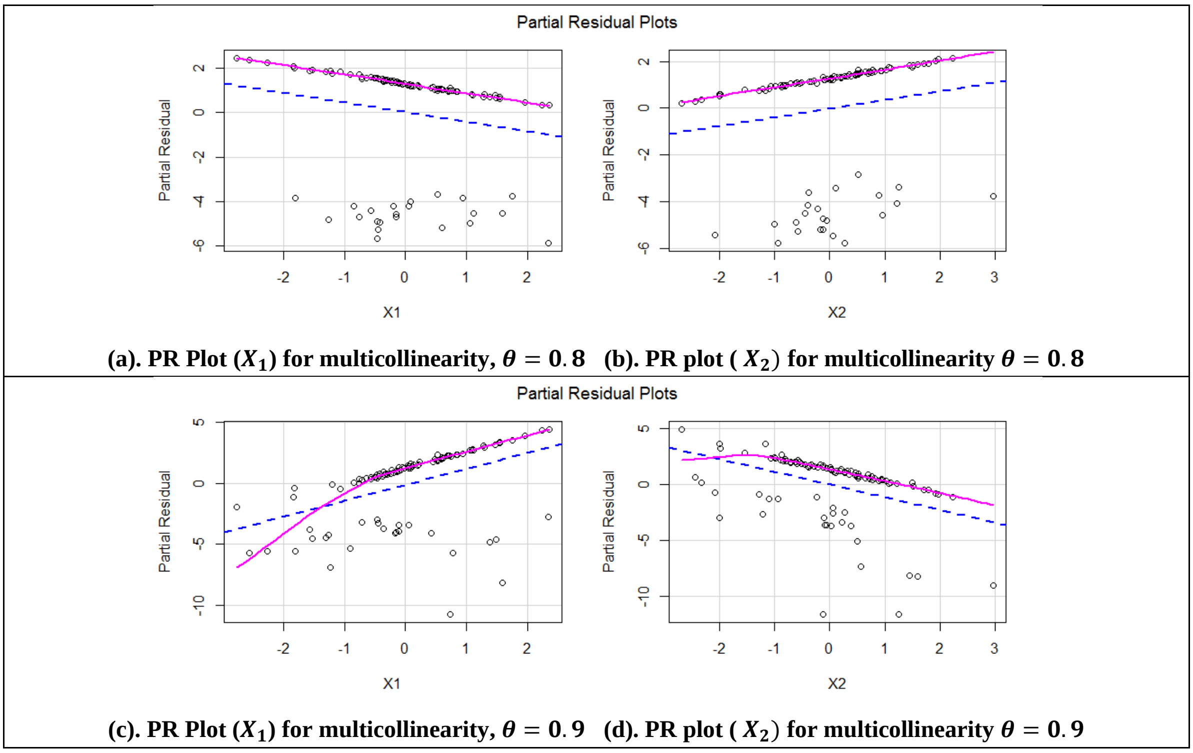
(c). PR Plot (X_1) for multicollinearity, $\theta = 0.9$ **(d). PR plot (X_2) for multicollinearity $\theta = 0.9$**



(e). PR Plot (X_1) for multicollinearity, $\theta = 0.95$ (f). PR plot (X_2) for multicollinearity $\theta = 0.95$

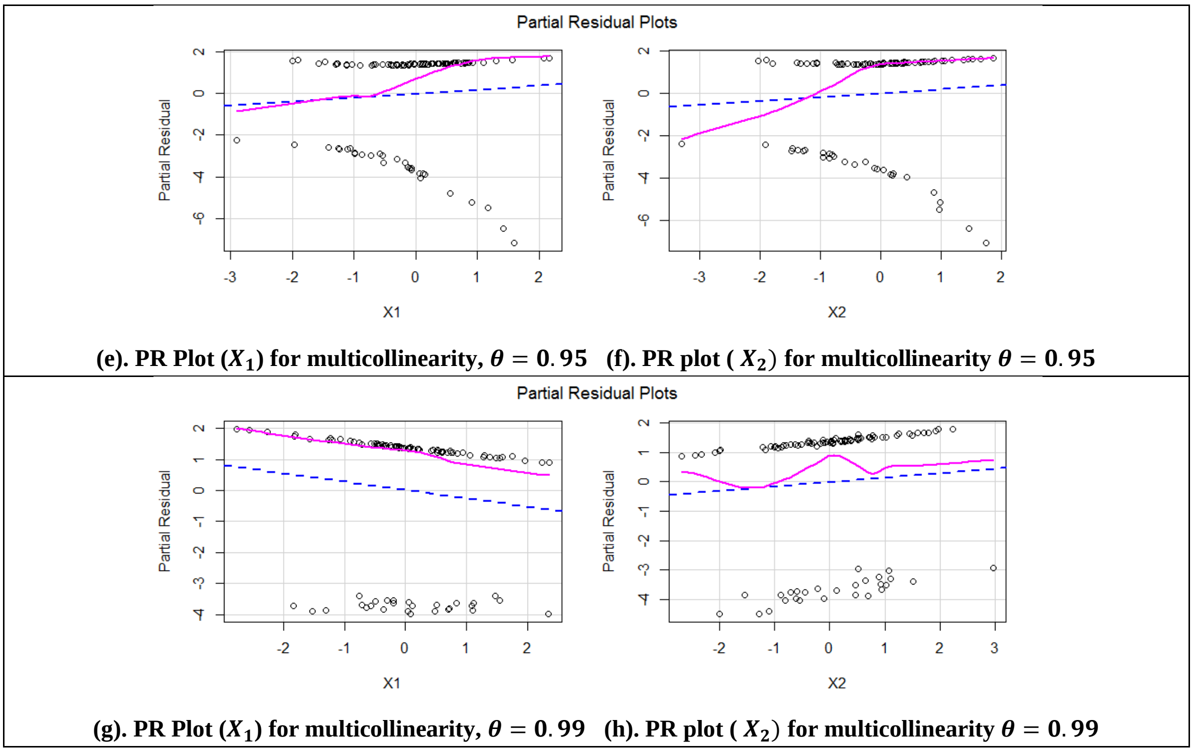
(g). PR Plot (X_1) for multicollinearity, $\theta = 0.99$ (h). PR plot (X_2) for multicollinearity $\theta = 0.99$

Figure 4.7: PR plots when $n = 100$



(a). PR Plot (X_1) for multicollinearity, $\theta = 0.8$ (b). PR plot (X_2) for multicollinearity $\theta = 0.8$

(c). PR Plot (X_1) for multicollinearity, $\theta = 0.9$ (d). PR plot (X_2) for multicollinearity $\theta = 0.9$



(e). PR Plot (X_1) for multicollinearity, $\theta = 0.95$ **(f). PR plot (X_2) for multicollinearity $\theta = 0.95$**

(g). PR Plot (X_1) for multicollinearity, $\theta = 0.99$ **(h). PR plot (X_2) for multicollinearity $\theta = 0.99$**

Figure 4.8: PR plots when $n = 200$

From Table 2, we notice the existence of multicollinearity through the higher values of the VIF and the correlation coefficients. It is intriguing to note that the VIF values increases with the increase in sample size and level of correlation. Figures 4.1 to 4.4 present the CERES plots while Figures 4.5 to 4.8 present the PR plots. It is observed that both the CERES and the PR plots detect multicollinearity successfully. Because, both type of plots show that various observations are very close to each other that is conjunction between the points, which exhibits the problem of multicollinearity among the points. The multicollinearity in the CERES plots is more evident as compared to the PR plots. According to the results of CERES plots most of the observation are very close to each other’s that is conjunction between the points show a severe multicollinearity. But on the other hand, in PR Plots most of observation are not close to each other.

5. Conclusion

This paper discussed the comparison of CERES and PR plots for the detection of multicollinearity in a BRM. The CERES and PR plots are graphical methods to detect multicollinearity. First, we developed a methodology of CERES and PR plots for BRM. Then apply a real data set (heart disease data). Find a model coefficient summary and VIF. The value of VIF is greater than 10, show a multicollinearity in data set. In BRM GLM, discussed the situations in which the CERES and PR plots provide the useful detection. This article addresses the theoretical development and implementation of CERES and PR plots for BRM GLM, and also illustrations are made on its advantages. Using real and simulated data, we have discussed and reviewed the detection of violations of assumptions in BRM. The CERES and PR plots are more useful to handle such situation. Results exhibits that both methods can successfully detect the multicollinearity problem in BRM. But the CERES plot performs better than the PR plots in the detection of multicollinearity.

6. Acknowledgements

Authors would like to thanks the referees, an associate editor, 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.

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.

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

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

Ezekiel, M. (1924). A method of handling curvilinear correlation for any number of variable. *Journal of the American Statistical Association*, 19(148), 431-453.

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

HosmerJr, D. W., Lemeshow, S., &Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

Özkale, M. R., Lemeshow, S., & Sturdivant, R. (2018). Logistic regression diagnostics in ridge regression. *Computational Statistics*, *33*, 563-593.

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

Kahng, M. W., & Lee, E. J. (2004). CERES plot in generalized linear models. *Communications for Statistical Applications and Methods*, 11(3), 575-582.

Landwehr, J. (1983). Using PR plots to detect nonlinearity in multiple regression. *Unpublished manuscript. Bell Laboratories, Murray Hill, New Jersey*.

Landwehr, J. M., and Pregibon, D. (1993), Comments on 'Improved added variable and PR 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.

Larsen, W. A., &McCleary, S. J. (1972). The use of PR plots in regression analysis. *Technometrics*, 14(3), 781-790.

Mallows, C. L. (1986). Augmented PRs. *Technometrics*, 28(4), 313-319.

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

Mullet, G.M. (1976), Why regression coefficients have the wrong sign, *Journal of Quality Technology*. 8, 121-126.

Nelder. J. A &Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society* A 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.

Saleem, N., Akbar, A., Shareef, S., Mamun, A., Imon, R., & Ahmad, S. (2022). Performance comparison between CERES and PR plots in detecting the heteroscedasticity problem using Liver cancer and simulated data. Global journal of engineering and technology, 3(1), 2583-3359.

Hussain, Z., & Akbar, A. (2022). Diagnostics through Residual Plots in Binomial Regression Addressing Chemical Species Data. *Mathematical Problems in Engineering*, *2022*.

Saleem, N., Akbar, A., Imon, A. R., & Al Mamun, A. S. M. (2022). Detection of Outliers in Binomial Regression Using CERES and Partial Residual Plots. *Journal of Statistical Modeling & Analytics (JOSMA)*, *4*(2).