Sigma J Eng Nat Sci, Vol. 43, No. 3, pp. 760–776, June, 2025

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2025.00070

Research Article

# Influential observations detection in the gamma-pareto regression model under different link functions with standardized and adjusted deviance residuals: simulation and application

**Nasir SALEEM[1,*], Atif AKBAR[1], Saima LAEEQ[2], Shakeel AHMAD[1], Herlina HANUM[3]**

[1]*Department of Statistics, Bahauddin Zakariya University, Multan, 60800, Pakistan*
[2]*Department of Statistics, Government Graduate College for Women, Multan, 60800, Pakistan*
[3]*Department of Mathematics, Universitas Sriwijaya, 36557, Indonesia*

## ARTICLE INFO

*Article history*
Received: 04 April 2024
Revised: 12 June 2024
Accepted: 02 July 2024

**Keywords:**
Adjusted Deviance Residuals; Cook's Distance; Deviance Residuals, DFFITS, Gamma-Pareto Regression; Generalized linear Model; Influential observations; Link functions; Standardized Deviance Residuals

## ABSTRACT

This study compares the performance of link functions for diagnostic methods to diagnose influential observations in the Gamma-Pareto regression model (G-PRM). Three link functions, i.e. inverse, identity, and log are considered to identify which link function gives the best results. For our investigation, we employed standardized deviance residuals (SDR) and adjusted deviance residuals (ADR). We used Cook's distance (CD) and Difference of fit (DIFFITS) as diagnostic methods. We compare the performance of influence diagnostics with the link functions using the simulation study and a real-life application. Results show that the CD with the log link function is a good method for small dispersion. For large dispersion and small sample sizes, the performance of the DIFFITS with inverse and identity link functions is better than the CD method. Similarly, for large dispersion and sample sizes, the CD (with identity and log link functions) and DFFITS with inverse link function give the same performance.

**Cite this article as:** Saleem N, Akbar A, Laeeq S, Ahmad S, Hanum H. Influential observations detection in the gamma-pareto regression model under different link functions with standardized and adjusted deviance residuals: simulation and application. Sigma J Eng Nat Sci 2025;43(3):760–776.

## INTRODUCTION

In practical life, we often deal with the data set in which the response variable is continuous and follow Gamma-Pareto distribution (GP-D). In such circumstances, the G-PRM GLM is preferred. GP-D has widespread applications in the area of health care economics, medical science, meteorology, automobile insurance claims and reaction rate etc. [1] used GP-D for the purpose of forecasting using rainfall data. [2] discussed the G-PD and its area of applications used floyd river data for annual flood discharge rates. Herlina hanum is the inventor of G-PRM. The motive of the current study is the non-availability of work in this area. Although the work on different link functions is available but the comparison is made here for influence diagnostic under different link function in the G-PRM.

***Corresponding author.**
*E-mail address: nasirsaleem160@gmail.com

*This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic*

The gamma-Pareto distribution (G-PD) is a continuous distribution. A phenomenon (the response variable) is explained by the regression model using other phenomena (explanatory variables). The development of a classical regression model is predicated on the normality of the response variables. This assumption applies to the model's parameters as well as the tests validity. The response variable is not always normally distributed in real data. An extended generalized linear model (GLM) is developed for data with an exponential family distribution. The mean of the response variable is connected to the linear form of the explanatory variables using the GLM link functions. According to [3] the link functions is a monotone differentiable function. The form of the link function depends on the response variables probability distribution, which is the basis for the development of GLM.

Regression analysis results can be greatly impacted by a single observation. It may result in a misleading covariance matrix and inaccurate coefficient estimates. For the regression models to produce accurate estimates, these observations must be located and eliminated. In order to diagnostic a model and evaluate how well it fits, residuals are important. Only raw residuals are used by the linear model (LM) to evaluate the model diagnostics. In contrast, the GLM provides a variety of residual structures, including the working, Pearson, deviance, Anscombe, and likelihood residuals. In order to affect GLM influence diagnostics, the Pearson and deviance and likelihood residuals are the most often utilized residuals. There are different in probability distributions for these residuals.

**Objective of the Study**

We found from the literature that the majority of researchers used an identity link function with deviance residuals to focus on G-PRM diagnostics. But not focused on other link functions and deviance residual form like SDR and ADR. There are various link functions such as identity, inverse and log, and diagnostic methods are CD and DFFITS which can be applied to evaluate the model's performance more effectively. Therefore, the purpose of this study is to compare the effectiveness and performance of various link functions for identifying influential observations as well as the diagnostic processes methods for identifying influential observations using deviance SDR and ADR.

**Organized of the Paper**

This paper is organized as follows: In section wise, advantages and disadvantages of the G-PRM, literature review. In next Section discussed methodology, next Section of the G-PRM and its estimation methods, next Section presents the G-PRM residuals with derivation of standardized and adjusted deviance residuals, next Section describe the influence diagnostics methods in G-PRM. In next Section defined a Monte Carlo simulation, next Section a Simulation design and next Section present a simulation result. In next Section present an application of data. Finally, Section gives away conclusion of the research work.

**Advantages and Dis-Advantages of the Gamma-Pareto Regression Model**

Here is the advantages and disadvantages of the Gamma-Pareto Regression Model (G-PRM). Advantages, the G-PRM is a flexible regression model. The G-PRM can model a wide range of skewed and heavy-tailed data, making it a flexible choice for modeling continuous outcomes. The Robust to outliers, G-PRM is more robust to outliers compared to traditional linear regression, as it uses a robust link function. Interpretable coefficients, the coefficients in a G-PRM model have a similar interpretation to those in linear regression, making it easy to understand the relationships between variables. Extension to other models, the G-PRM can be extended to other models, such as Gamma-Pareto ridge regression model. Disadvantages, computational complexity, G-PRM can be computationally intensive, especially for large datasets, due to the need to estimate the shape parameter. Sensitive to starting values, the estimation algorithm can be sensitive to the starting values, which can affect the convergence of the model. The G-PRM requires a sufficient sample size to estimate the shape parameter accurately, making it less suitable for samples. Not appropriate for all types of data, G-PRM assumes a continuous outcome variable, making it less appropriate for categorical or count data. In summary, G-PRM offers flexibility and robustness but may require careful consideration of computational complexity, starting values, and sample size. we apply some distribution of fitting test to response variable. After applying a distribution fitting test the response variable follow a GP-D. According to literature we know that, the G-PRM is used when the response variable is continuous and follow a GP-D. In this study we used a rate data set is taken from [4], the dependent variable reaction rate ($y$) is continuous and follow a GP-D that's why we select a G-PRM.

**Literature Review**

Alzaatreh [2] invent a G-PD and discussed the mathematical relationship between G-PD and GD. [5] employed G-PD to model and forecast extreme monthly rainfall, so this makes sense given that the G-PD evolved from the GD. The G-PD based regression model. Regression models for non-normal response variables usually take the form of GLM. [1] convert a G-PD in the exponential family distribution of member. And derived an inverse, identity and log link function for G-PRM. Consequently, GLM could be used to develop the regression modeling for the G-PRM. After that they derived a parameter estimation method for G-PRM. GLM G-PD is analytically developed by [6]. The gamma distribution (GD) is the basis for GLM gamma, which is applied frequently. When GLM gamma is used for analysis, the right skew data are

frequently fit. [1,6] examined the relationship between the explanatory variable and the distributed response variable is continuous in a simulated G-PD using GLM gamma. The application of modeling gamma-Pareto distributed data with GLM gamma in monthly rainfall estimation with TRMM data was covered by [6]. In order to map the safety continuum and estimate crashes, [7] discussed the Shifted Gamma-Generalized Pareto Distribution model. The new Log-Gamma-Pareto Distribution is created by [8]. A new Gamma-Pareto (IV) distribution and its uses were presented by [9]. The gamma generalized pareto distribution and its applications in survival analysis were covered by [10]. Exponentiated gamma-Pareto distribution was applied to bladder cancer susceptibility by [11]. The weighted gamma-pareto distribution and its use were covered by [12]. The introduction of generalized linear models (GLMs) allows for the investigation of dependent variable dependence on two independent variables. Another variation of the GLM was discussed by [13]. According to [14], the GLM in fact goes against the non-influential observation assumption. Influence diagnostics were first introduced for linear models (LMs) by [15]. These impact diagnostics were covered by [16] in a number of dimensions. According to [17], [18], and [19], influence diagnostics in the GLM continue to be the main topic of debate. When evaluating influential observations in influence diagnostics, the Pearson residuals are frequently used. Additionally, [19] demonstrated the use of deviance residuals in influence diagnostics. The two primary theories of adjusted residuals still in use are the adjusted deviance residuals provided by [20] and the adjusted Pearson residuals suggested by [21] based on [22]. The aim of these theories is to attain normality. [23] found that an examination of the adjusted Pearson residuals (APR) in the exponential family of nonlinear models yields comparable outcomes. Several methods have been put forth in the literature to diagnose significant observations or points for the LM, including [24], [25], [26], and [27]. Conversely, [28] provided a method for evaluating partial influence in the GLM. One approach to evaluating the impact on the GLM regression coefficients was suggested by [29].

[30] described the importance of Beta regression residuals-based control charts with different link functions. For this purposed used an application to the thermal power plants data. Further, the three criteria are used for performance checking: the average of the run length, the standard deviation of the run length, and the median of the run length. And also evaluate the performance of the proposed control chart in two ways: by monitoring the intercepts and by monitoring the slope coefficients. [31] investigate the influential observation detection in the logistic regression under different link functions and used pearson residuals. And used a real-life application to urine calcium oxalate crystals data. [32] used Deviance and Pearson residuals-based control charts with different link functions for monitoring logistic regression profiles, an application to COVID-19 data. [34] discussed a comparison of link functions for the estimation of logistic ridge regression with real life an application to urine data. A Monte Carlo simulation study and a real dataset are considered and using scalar mean squared error as performance evaluation criteria. [34] described the performance of link functions in the beta ridge regression model. For the suitable link functions the evaluation criteria is minimum MSE. [35] discussed a comparison of some link functions for binomial regression models with application to school drop-out rates in East Java. For the suitable link function, the evaluation criteria are Akaike information criterion (AIC), Bayesian information criterion (BIC), log likelihood (LL) and R-squared. [36] used generalized Weibull linear models with different link functions in survival analysis. For the goodness of fit models under different link functions, fit measures such as deviance, AIC and BIC. Now we discussed the methodology, link function, residuals and diagnostics methods of the G-PRM is given below.

## MATERIALS AND METHODS

### Gamma-Pareto Regression Model and Estimation Methods

The probability density function of the gamma-pareto response variable y is given by [2],

$$f(y;\alpha,\beta,\gamma)=\frac{\gamma^{-1}}{\beta^{\alpha}\Gamma(\alpha)}\left(log\left(\frac{y}{\gamma}\right)\right)^{\alpha-1}\left(\frac{y}{\gamma}\right)^{-\left(\frac{1}{\beta}+1\right)} \quad (1)$$

with α, $\beta$,$\gamma$> 0 and y > $\gamma$.

The mean and variance of y are, $E(a(y))=\alpha\beta, V(a(y))=\alpha\beta^2$ respectively.

According to [1,6], Eq. (1) can be modified with parameters $\alpha=\frac{1}{\phi}$ and $\beta=\mu\phi$. Under these parameters, the gamma-pareto density for y is given by

$$f(y;\mu,\phi)=\frac{\gamma^{-1}}{(\mu\phi)^{\frac{1}{\phi}}\Gamma\left(\frac{1}{\phi}\right)}\left(log\left(\frac{y}{\gamma}\right)\right)^{\frac{1}{\phi}-1}\left(\frac{y}{\gamma}\right)^{-\left(\frac{1}{\mu\phi}+1\right)} \quad (2)$$

with $y\geq 0, \mu>0\ and\ \phi>0$.

It may also be noted that the mean and variance of y are given by

$$E(y)=\mu \text{ and } V(y)=\phi V(\mu)=\phi\mu^2.$$

For the ith observation, let $x_{i1}, x_{i2}, \ldots, x_{ip}$ represent the p non-stochastic regressors. According to [1,6], link function of the G-PRM for the mean of the given response variable y is given by

$$g(\mu_i)=\eta_i=X_i^T\beta,\ i=1,2,\ldots,n.$$

where $x_i^T = (1, x_{i1}, x_{i2}, \dots, x_{ip})$, $\beta^T = (\beta_0, \beta_1, \dots, \beta_p)$ is a vector regression coefficient including intercept. And $x_{i1}$, $x_{i2}, \dots, x_{ip}$ represent the p non-stochastic regressors.

For the G-PRM, this link function is either identity link function $g(\mu_i) = \eta_i = X_i^T\beta$, inverse link function $g(\mu_i) = \eta_i = \frac{1}{X_i^T\beta}$, and log link function $g(\mu_i) = \eta_i = \log(X_i^T\beta)$.

Finding the likelihood function's derivative with respect to $\beta_j$ is the first step in estimating the parameter $\beta_j$ using maximum likelihood. By Eq. (2)

$$\frac{\partial l}{\partial \beta_j} = U_j = \sum_{i=1}^{N}\left[\frac{\partial l_i}{\partial \beta_j}\right] = \sum_{i=1}^{N}\left[\frac{\partial l_i}{\partial \tau_i}\frac{\partial \tau_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \beta_j}\right] \quad (3)$$

Now

$$\frac{\partial l_i}{\partial \tau_i} = a(y)b'(\tau) + c'(\tau) = \beta^{-2}\left(log\left(\frac{y_i}{\gamma}\right) - \mu_i\right)$$

$$\frac{\partial \tau_i}{\partial \mu_i} = \frac{1}{\frac{\partial \mu_i}{\partial \tau_i}} = \frac{1}{\frac{\partial \alpha\beta}{\partial \beta}} = \frac{1}{\alpha}$$

$$\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i}x_{ij}$$

Where $\frac{\partial \mu_i}{\partial \eta_i}$ based on the GLM link function. So, the score for $\beta_j$ in GLM Gamma-Pareto is

$$\frac{\partial l}{\partial \beta_j} = U_j = \sum_{i=1}^{N}\alpha^{-1}\beta^{-2}\left(log\left(\frac{y_i}{\gamma}\right) - \mu_i\right)\frac{\partial \mu_i}{\partial \eta_i}x_{ij} \quad (4)$$

Finally, jth score is presented.

$$U_j = \sum_{i=1}^{N}\left[var\left(log\left(\frac{y_i}{\gamma}\right) - \mu_i\right)\right]^{-1}\left(log\left(\frac{y_i}{\gamma}\right) - \mu_i\right)\frac{\partial \mu_i}{\partial \eta_i}x_{ij}$$

The variance $U_j$ is

$$var(U_j) = \zeta_{jk} = \sum_{i=1}^{N}\frac{x_{ij}x_{ik}}{\left[var\left(log\left(\frac{y_i}{\gamma}\right)\right)\right]}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 = X^TWX$$

Where,

$$W = \frac{1}{\left[var\left(log\left(\frac{y_i}{\gamma}\right)\right)\right]}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2$$

Since the estimators of $\beta_j$ is not in close form.

Iterative weighted least squares (IWLS) were proposed by [3] as a method for estimating $\beta_j$.

It's the IWLS.

$$X^TWX\, b^{(m)} = X^TWz$$
$$b^{(m)} = (X^TWX)^{-1}(X^TWz) \quad (5)$$

And now, Using W and $var(U_j)$ for G-P and obtained the iteration for $\beta_j$ as

$$X^TWX\, b^{(m)} = \sum_{k=1}^{p}\sum_{i=1}^{N}\frac{x_{ij}x_{ik}}{\left[var\left(log\left(\frac{y_i}{\gamma}\right)\right)\right]}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 b_k^{(m-1)} + \frac{\left(log\left(\frac{y_i}{\gamma}\right) - \mu_i\right)x_{ij}}{\left[var\left(log\left(\frac{y_i}{\gamma}\right)\right)\right]}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)$$

$$z_i = \sum_{i=1}^{N}x_{ij}b_k^{(m-1)} + \left(log\left(\frac{y_i}{\gamma}\right) - \mu_i\right)\frac{\partial \mu_i}{\partial \eta_i}$$

**Gamma-Pareto Regression Model with Standardized and Adjusted Deviance Residuals**

Many types are available of GLM residuals in literature [14]. But we used only deviance residual and its types standardized deviance residual and adjusted deviance residual form.

The deviance residuals for the G-PRM are given by

$$R_{dr} = sign(y_i - \hat{y}_i)\sqrt{|d_i|} \quad (6)$$

where $d_i = -2\left\{\ln\left(\frac{y_i}{\hat{y}_i}\right) - \left(\frac{y_i - \hat{y}_i}{\hat{y}_i}\right)\right\}$ and sign $(y_i - \hat{y}_i)$ is signum function, which is defined as

$$sign(y_i - \hat{y}_i) = \begin{cases} + \; if \; y_i > \hat{y}_i \\ 0 \; if \; y_i = \hat{y}_i \\ - \; if \; y_i < \hat{y}_i \end{cases}$$

For the G-PRM, this link function is either identity link function $g(\mu_i) = \eta_i = X_i^T\beta$, inverse link function $g(\mu_i) = \eta_i = \frac{1}{X_i^T\beta}$, and link log function $g(\mu_i) = \eta_i = \log(X_i^T\beta)$ are fitted model $\hat{y} = \eta_i$.

Eq. (6) is used to present the standardized deviance residuals.

$$R_{sdr} = \frac{R_{dr}}{\sqrt{\phi(1-h_{ii})}} \quad (7)$$

Since $h_{ii}$ is the ith diagonal element of the hat matrix $H = W^{\frac{1}{2}}X(X^TWX)^{-1}X^TW^{\frac{1}{2}}$

Adjusted residuals were first introduced by [22]. According to [21] and [20], the adjusted deviance residuals for both methods. The adjusted deviance residuals are defined by using Eq. (6)

$$R_{adr} = \frac{R_{dr} - E(R_{dr})}{\sqrt{V(R_{dr})}} \quad (8)$$

The adjusted deviance residuals are normally distributed by [20].

**Influence Diagnostics, Gamma-Pareto Regression Model**

A bad value in the LM has an impact on the model estimates and inferences, as noted by [37]. These poor values could be influential that have an impact or be outliers. An outlier is produced by an extreme value in the response

variable, whereas an influential observation is produced by an extreme value in the explanatory variable. A portion of these is covered here for the G-PRM influence diagnostics since the GLM employing deviance residuals (standardized and adjusted) has not yet any attention. The reason for this is that the GLM influence diagnostics under various GLM residuals have received little consideration. [18] was the first to study residuals in the GLM. Different GLM residuals are used to compute the GLM influence assessment tools.

A diagnostic measure known as influence that has received a lot of attention in the literature, DFFITS is defined as the scaled difference between the fitted value of the complete data set and the fitted value following the deletion of the ith observation.

$$DFFITS_i = \frac{y_i - \hat{y}_i}{\sqrt{\hat{\phi}_i\, h_{ii}}} \tag{09}$$

Eq. (09) can also be written as

$$DFFITS_i = \frac{\hat{W}_{ii}^{\frac{1}{2}} x_i^T (y_i - \hat{y}_i)}{\sqrt{\hat{\phi}_i\, h_{ii}}} \tag{10}$$

$$DFFITS_i = |t_i| \sqrt{\frac{h_{ii}}{1-h_{ii}}} \tag{11}$$

Where,

$$t_i = spr_{pi} \sqrt{\frac{n-p-1}{n-p-(sdr_{pi})^2}} \tag{11.1}$$

The DFFITS for standardized deviance residuals used Eq. (7)

$$DFFITS_i = |t_i| \sqrt{\frac{h_{ii}}{1-h_{ii}}} \tag{12}$$

$$t_i = R_{sdr} \sqrt{\frac{n-p-1}{n-p-(R_{sdr})^2}} \tag{12.1}$$

The DFFITS for adjusted deviance residuals used Eq. (8)

$$DFFITS_i = |t_i| \sqrt{\frac{h_{ii}}{1-h_{ii}}} \tag{13}$$

$$t_i = R_{apr} \sqrt{\frac{n-p-1}{n-p-(R_{adr})^2}} \tag{13.1}$$

where $hii = diag(H)$ is the $i$th hat matrix H diagonal element for the G-PRM [13], $H = \widehat{W}^{\frac{1}{2}} X \left(X^T \widehat{W} X\right)^{-1} X^T \widehat{W}^{\frac{1}{2}}$. These diagonal elements are utilized for influence diagnostics and are also referred to as leverages. In order to influence additional diagnostic measures, the leverages serve as an indicator. If the data is small, then an observation is considered influential if the DFFITS value is greater than one [27]. In the case of large data sets, an observation is considered influential when the ith value of DFFITS exceeds $2\sqrt{\frac{p+1}{n}}$,[16]. The impact of the ith influential observation on the fitted and estimated values is measured using DFFITS. Similarly, we can substitute other forms of standardized and adjusted G-PRM residuals for the purpose of detection influential observations. We apply the same cut-off point for the DFFITS computation with standardized and adjusted G-PRM residuals in order to compare the outcomes with the conventional use of standardized and adjusted residuals.

Here is second diagnostic measure, the most widely used measures, such as Cook's distance (CD), are included. [15] first proposed the $CD_i$ statistic for the LM to quantify the impact of the influential observation on the LM estimates. When the ith observation is removed from the model, $CD_i$ calculates the overall change in the fitted model. For the G-PRM case, $CD_i$ is given

$$CD_i = \frac{(\hat{\beta} - \hat{\beta}_i)^T X^T \widehat{W} X (\hat{\beta} - \hat{\beta}_i)}{(P+1)\hat{\phi}} \tag{14}$$

After simplification, Eq. (14) becomes

The cook`s distance for standardized deviance residuals used eq. (7)

$$CD_{(i)R_{sdr}} = \frac{(R_{sdr})^2}{(P+1)} \frac{h_{ii}}{(1-h_{ii})} \tag{15}$$

The cook`s distance for adjusted deviance residuals used eq. (8)

$$CD_{(i)R_{adr}} = \frac{(R_{adr})^2}{(P+1)} \frac{h_{ii}}{(1-h_{ii})} \tag{16}$$

According to [38], this diagnostic is used to assess the impact of an influential observation solely on $\hat{\beta}$. When $CD_i$ is large, it means that the ith observation has influential. [15] proposed that the use of a cut point is another method for detecting the influential observation. i.e., $CD_i \geq F\alpha, (p + 1, n - p - 1)$. Influential observations are not detected by this cut point in certain GLM cases. An additional cut-off points in the GLM for identifying influential observations is $\frac{4}{n-1}$, as discussed by [14]. We employ an identical cut-off point for $CD_i$ for all forms of the GPRM residuals in our comparison of standardized and adjusted GPRM residuals for the identification of influential observations.

**Monte Carlo Simulation**

This section will compare, using a Monte Carlo simulation study, the performance influence diagnostics under various link functions, as well as the SDR and ADR residuals. In our study, we compared the effectiveness

of Gamma-Pareto regression diagnostics by taking into account different sample sizes with different dispersion parameters.

**Simulation Design**

The purpose of this section is to demonstrate the efficacy of the G-PRM standardized and adjusted deviance residuals for influence diagnostics through simulation. The independent variables comprise four influential points. To compare the performance of identity, inverse and log link functions of the G-PRM residuals with diagnostic methods Cook`s distance and DFFITS, we take into consideration the following Monte Carlo scheme. We used algorithm of [1,6] to generate response variable which follows a gamma- Preto regression model and data generation is as follows: $y_i \sim G - P(\alpha, \beta, \gamma)$, where $\hat{y}_i = E(y_i) = (\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3})$ identity, $\hat{y}_i = E(y_i) = (\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3})^{-1}$ inverse and $\hat{y}_i = E(y_i) = \log(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3})$ log link function, $i = 1, 2, \dots, n$ is mean function and $\phi$ is dispersion parameter $\phi = 0.04, 0.11, 0.17, 0.33, 0.67, 2, 5, 10$ which is thought to have arbitrary values. For the true parameters, we choose the following arbitrary values as $\beta_0 = 0.05$, $\beta_1 = 0.0025$, $\beta_2 = 0.005$ *and* $\beta_3 = 0.0001$ [39,40] and $\gamma$ is minimum value of response variables. In this case, the design matrix X has no influential points of sample sizes n = 25, 50, 100 and 200 generated as $X_i \sim N(-1,1)$, $i = 1,2, \dots, n$; *and* $j = 1,2,3$, and then we make $10^{th}$, $15^{th}$, $20^{th}$, $25^{th}$, points in the X as $X_{ij} = \alpha_0 + X_{ij}$, $i = 10, 15, 20$ *and* $25$, *and* $j = 1,2,3$, where $\alpha_0 = \bar{X}_j + 100$. For the estimation of G-PRM, the link functions used is inverse, identity and log link functions. These simulation results are perform using the R software. The simulation is run 10000 times to test the influential observation detection percentages for each of the G-PRM under different link functions.

## RESULTS AND DISCUSSION

The simulation results of the Gamma-Pareto regression influence diagnostics under different link functions are given in Tables 1-8 The summary of the simulation results is as follows.

- In table 1, for dispersion level is $\phi = 0.04$, the performance of the Cook's Distance and DFFITS procedures with inverse, identity and log link functions. The Cook's Distance and DFFITS approach with the log link function gives the larger influential observations detections percentages of the generated influential observations as compared to the Cook's Distance and DFFITS with inverse and identity link functions. The graphical results of the table 1, is presented with index plot in Figure 1.
- In table 2, for dispersion level is $\phi = 0.11$, the performance of the Cook's Distance and DFFITS procedures with inverse, identity and log link functions. The Cook's Distance and DFFITS approach with the log link function gives the larger influential observations detections percentages of the generated influential observations as compared to the Cook's Distance and DFFITS with inverse and identity link functions. The dispersion level is $\phi = 0.04$ and $\phi = 0.11$, in table 1 and 2, a log link function with diagnostic measures Cook`s Distance and DFFITS detect a large percentage influential observation. The graphical results of the table 2, is presented with index plot in figure 2.
- In table 3, for dispersion level is $\phi = 0.17$, the performance of the Cook's Distance and DFFITS diagnostics with inverse, identity and log link functions. The Cook's Distance method with the inverse link function gives the larger influential observations detections percentages of the generated influential observations as compared to the DFFITS. For all sample sizes the performance of the Cook's Distance is better than the DFFITS method. While the DFFITS method with the log link function gives the larger influential observations detections percentages of the generated influential observations as compared to the Cook`s Distance. For all sample sizes the performance of the DFFITS is better than the Cook's Distance method. The graphical results of table 3 is presented in figure 3.
- When the dispersion level is further increase $\phi = 0.33$ the results is almost same as when dispersion is $\phi = 0.11$ in favor of inverse link function in table 4. These results are also verified and prominent with index plot in figure 4.
- When the dispersion level is $\phi = 0.67$ the Cook's distance and DFFITS diagnostic methods performance with inverse, identity and log link functions for both SDR and ADR are almost same diagnostics influential observations detection percentages are true for all sample sizes in table 5. The visualized results are show index plot in figure 5.
- For dispersion level is $\phi = 2$ the Cook's distance and DFFITS diagnostic methods performance with identity link function for both SDR and ADR are larger diagnostics influential observation detection percentages as compare to the inverse and log link functions. But on the other hand, DFFITS with inverse link function better diagnose as compare to the identity and log link function with all sample sizes in table 6. These results are also verified and prominent with index plot in figure 6.
- For large dispersion level are $\phi = 5, 10$ the Cook's distance and DFFITS diagnostic methods performance with inverse link function for both SDR and ADR are larger diagnostics influential observations detection percentages as compare to the identity link and log link functions for all sample sizes in table 7 and 8. These results are also verified and prominent with index plot in figure 7 and 8 respectively.

**Table 1.** Performance of different link functions with standardized and adjusted deviance residuals for the detection of influential observations when $\phi$ =0.04

| Sample size n | Influential Observations | Cook's Distance | | | | | | DFFITS | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\phi$ =0.04 | | | | | | $\phi$ = 0.04 | | | | | |
| | | Inverse | | Identity | | Log | | Inverse | | Identity | | Log | |
| | | SDR | ADR | SDR | ADR | SDR | ADR | SDR | ADR | SDR | ADR | SDR | ADR |
| 25 | 10 | 84.7 | 80.4 | 82.9 | 77.5 | 83.6 | 79.1 | 84.6 | 79.8 | 84.8 | 80.5 | 97.1 | 84.7 |
| | 15 | 75.8 | 68.8 | 74 | 66.5 | 72.5 | 66.5 | 75.9 | 70.6 | 74.3 | 67.5 | 95.3 | 74 |
| | 20 | 64.3 | 57.8 | 65.7 | 58.3 | 64 | 56 | 64.2 | 56.5 | 62.6 | 55.1 | 90.1 | 62.9 |
| | 25 | 56.2 | 48.2 | 56 | 48.6 | 52.1 | 45.8 | 54 | 46.9 | 54.6 | 47.4 | 81.9 | 52.7 |
| 50 | 10 | 87.2 | 82.7 | 87.4 | 83.5 | 88.9 | 83.9 | 88.3 | 84.6 | 88.8 | 85.4 | 98.8 | 89.4 |
| | 15 | 82.4 | 78.5 | 84.3 | 78.9 | 83.6 | 79.3 | 84.2 | 80.7 | 82.4 | 79.4 | 97.8 | 84.5 |
| | 20 | 78 | 73 | 79.8 | 74.9 | 79.6 | 75.2 | 79.1 | 74.1 | 78.3 | 74.7 | 95.6 | 76.9 |
| | 25 | 71.4 | 66.6 | 72.4 | 67.7 | 75.5 | 71.4 | 73.6 | 68.5 | 72.8 | 67.9 | 93.4 | 75 |
| 100 | 10 | 91.7 | 88.9 | 91.7 | 88.3 | 91.2 | 88 | 92.8 | 89.3 | 90.9 | 88.6 | 98.8 | 91.4 |
| | 15 | 87.7 | 84.7 | 88.1 | 85.3 | 87.3 | 83.9 | 88.8 | 87.2 | 88.6 | 85.7 | 98.2 | 88 |
| | 20 | 84.1 | 81.3 | 85.6 | 82.2 | 84.9 | 81.3 | 85.5 | 82.9 | 83.6 | 80.4 | 97.1 | 86.4 |
| | 25 | 79.8 | 76.7 | 82.3 | 79.3 | 83 | 80.3 | 83.1 | 80.9 | 85.3 | 82.5 | 94.3 | 80.3 |
| 200 | 10 | 92.7 | 90.4 | 95 | 93.5 | 93.3 | 91.3 | 92.7 | 90.6 | 93.3 | 90.9 | 99 | 93.9 |
| | 15 | 90.1 | 87.8 | 93.1 | 91.7 | 91 | 89.6 | 91.2 | 89.5 | 92.1 | 89.9 | 97.8 | 91.5 |
| | 20 | 88.6 | 87.1 | 90.7 | 89 | 90.4 | 89.1 | 89.4 | 87.7 | 91.1 | 89.8 | 97.9 | 90.1 |
| | 25 | 88.1 | 86.8 | 88.2 | 86.4 | 89.6 | 87.9 | 88.6 | 87.2 | 88 | 86.4 | 97.6 | 88.8 |

**Table 2.** Performance of different link functions with standardized and adjusted deviance residuals for the detection of influential observations when $\phi$ = 0.11

| Sample size n | Influential Observations | Cook's Distance | | | | | | DFFITS | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\phi$ = 0.11 | | | | | | $\phi$ = 0.11 | | | | | |
| | | Inverse | | Identity | | Log | | Inverse | | Identity | | Log | |
| | | SDR | ADR | SDR | ADR | SDR | ADR | SDR | ADR | SDR | ADR | SDR | ADR |
| 25 | 10 | 84 | 80 | 86.7 | 81.7 | 83.8 | 78 | 82 | 75.7 | 85.3 | 79.8 | 97.8 | 85.2 |
| | 15 | 75 | 69.9 | 75 | 66.9 | 75 | 69.3 | 75.5 | 68.2 | 76.6 | 70.8 | 94 | 74.4 |
| | 20 | 60.8 | 56 | 65.4 | 58.1 | 62.8 | 56.7 | 65.5 | 59.3 | 61.6 | 54.1 | 89.4 | 63 |
| | 25 | 47.1 | 45.3 | 54.4 | 47.7 | 53.5 | 46.1 | 51 | 43.6 | 55.2 | 48.5 | 83.1 | 55 |
| 50 | 10 | 88.3 | 83.9 | 88.9 | 84.7 | 89.3 | 85.6 | 88.2 | 83.7 | 90.1 | 85.5 | 98.4 | 88.3 |
| | 15 | 84.2 | 79.2 | 82.2 | 77.6 | 82.9 | 78.1 | 84.6 | 80.6 | 80.7 | 77.1 | 96.5 | 82.1 |
| | 20 | 78 | 74.4 | 76.8 | 73.2 | 77.4 | 72.2 | 79.3 | 76.3 | 79.8 | 75 | 95.1 | 78.6 |
| | 25 | 69.8 | 66.7 | 72.8 | 68.1 | 70.5 | 66.7 | 71.9 | 67.7 | 75.3 | 70 | 92.7 | 72.1 |
| 100 | 10 | 90.9 | 87.2 | 92.5 | 88.9 | 91.7 | 88.9 | 92 | 89.2 | 90.4 | 87 | 98.3 | 90.7 |
| | 15 | 88.4 | 85.7 | 88.4 | 85.7 | 88.1 | 83.9 | 89.1 | 85.6 | 87.7 | 85.6 | 97.7 | 87.6 |
| | 20 | 87.9 | 85.2 | 85 | 82.9 | 86.3 | 83.8 | 84.4 | 81.5 | 84.3 | 81.7 | 96.8 | 84 |
| | 25 | 81.1 | 78.5 | 83.9 | 80.4 | 82.5 | 80.3 | 82.6 | 80.3 | 82.5 | 79.1 | 96 | 83.4 |
| 200 | 10 | 93.2 | 90.8 | 92.8 | 91.1 | 94.1 | 92.2 | 93 | 91.4 | 92.4 | 90.4 | 98.3 | 91.5 |
| | 15 | 92 | 90.1 | 92.6 | 91.3 | 93 | 90.3 | 89.8 | 88.5 | 91.2 | 89.4 | 98.2 | 91.4 |
| | 20 | 90.5 | 88.8 | 88.4 | 87.2 | 89.7 | 87.9 | 90.5 | 89.1 | 91 | 89.8 | 97.8 | 90.3 |
| | 25 | 87.1 | 85.8 | 85.3 | 83.7 | 87.3 | 84.9 | 87.5 | 86.1 | 88 | 85.9 | 96.4 | 87.5 |

**Table 3.** Performance of different link functions with standardized and adjusted deviance residuals for the detection of influential observations when $\phi$ =0.17

| Sample size n | Influential Observations | Cook's Distance | | | | | | DFFITS | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\phi$ = 0.17 | | | | | | $\phi$ = 0.17 | | | | | |
| | | Inverse | | Identity | | Log | | Inverse | | Identity | | Log | |
| | | SDR | ADR | SDR | ADR | SDR | ADR | SDR | ADR | SDR | ADR | SDR | ADR |
| **25** | 10 | 87.4 | 80.9 | 84.2 | 79.8 | 84.4 | 80.3 | 84.7 | 79.2 | 82.4 | 76.5 | 98 | 86.9 |
| | 15 | 77.5 | 68.6 | 73.2 | 66.4 | 74.1 | 67.6 | 75.3 | 69 | 77.7 | 70.1 | 94.3 | 75.8 |
| | 20 | 62.9 | 57 | 64.2 | 57.8 | 64.7 | 58.1 | 63.4 | 57.1 | 64.6 | 58.3 | 90.4 | 64.3 |
| | 25 | 49.0 | 42.2 | 56.4 | 48.4 | 53.1 | 47 | 54.8 | 47 | 55.6 | 47.2 | 85.3 | 53.2 |
| **50** | 10 | 86.6 | 83 | 87.2 | 83.4 | 88.1 | 84.2 | 88.4 | 84.8 | 87.3 | 82.4 | 98.6 | 89.3 |
| | 15 | 83.9 | 81.3 | 84.5 | 80.5 | 82.2 | 78.3 | 83.6 | 78.8 | 80.9 | 76.8 | 97.5 | 85 |
| | 20 | 79.3 | 76.8 | 75.9 | 71.4 | 76.3 | 72.4 | 75.2 | 70.2 | 77.4 | 72.1 | 94.6 | 76.4 |
| | 25 | 70 | 67 | 73.1 | 68.5 | 72.3 | 68.2 | 73.4 | 69 | 75.3 | 71.2 | 92.6 | 72 |
| **100** | 10 | 91.4 | 88.2 | 92.5 | 88.7 | 90.1 | 88 | 91 | 88.2 | 90.9 | 88.2 | 98.6 | 90.4 |
| | 15 | 86.5 | 84.7 | 89.2 | 86.7 | 88 | 84.1 | 88.8 | 85.1 | 88.3 | 85.4 | 98.5 | 89.5 |
| | 20 | 83.6 | 82.4 | 86.2 | 83.9 | 84.4 | 80.6 | 82.9 | 80 | 83.7 | 80.4 | 96.9 | 84.4 |
| | 25 | 80.9 | 78.9 | 83.4 | 80.6 | 83 | 81 | 81.8 | 79.4 | 82.9 | 79.6 | 94.3 | 80.9 |
| **200** | 10 | 92.4 | 91 | 93.2 | 91.6 | 90.7 | 88.8 | 92.1 | 90 | 93.4 | 91.6 | 99.1 | 93.8 |
| | 15 | 91.8 | 90.5 | 93.5 | 92.2 | 91.6 | 90.7 | 90.8 | 89.4 | 92 | 89.7 | 99.1 | 92.8 |
| | 20 | 88.8 | 87.2 | 90.8 | 89.5 | 88.6 | 86.9 | 88.9 | 87 | 90.9 | 88.7 | 97.8 | 88.1 |
| | 25 | 87.2 | 86 | 87.8 | 86.4 | 88.1 | 86.2 | 89.4 | 86.9 | 90.1 | 88.6 | 97 | 88.7 |

**Table 4.** Performance of different link functions with standardized and adjusted deviance residuals for the detection of influential observations when $\phi$ =0.33

| Sample size n | Influential Observations | Cook's Distance | | | | | | DFFITS | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\phi$ = 0.33 | | | | | | $\phi$ = 0.33 | | | | | |
| | | Inverse | | Identity | | Log | | Inverse | | Identity | | Log | |
| | | SDR | ADR | SDR | ADR | SDR | ADR | SDR | ADR | SDR | ADR | SDR | ADR |
| **25** | **10** | 88.8 | 79.1 | 83.4 | 78.2 | 85.2 | 79.3 | 85.1 | 79.5 | 84.9 | 80 | 97.6 | 84.7 |
| | **15** | 74.8 | 70 | 73.4 | 68.3 | 74.3 | 67.3 | 74.2 | 68 | 72.9 | 67.8 | 93.1 | 71.4 |
| | **20** | 56.5 | 53.6 | 61.5 | 55.4 | 67.4 | 58.3 | 64 | 58.2 | 63.9 | 56.9 | 90.5 | 65.9 |
| | **25** | 47 | 48.5 | 54.3 | 48 | 53.2 | 45.6 | 54.3 | 46.5 | 54.7 | 46.8 | 83.3 | 53.6 |
| **50** | **10** | 86.6 | 82.6 | 88.2 | 84.2 | 89.7 | 86.2 | 89.4 | 85.6 | 89.9 | 86.3 | 98.7 | 90.4 |
| | **15** | 83.6 | 79.8 | 82.6 | 79.8 | 83.7 | 79.7 | 85.3 | 81.3 | 82.7 | 79.4 | 96.2 | 82.1 |
| | **20** | 77.4 | 72.9 | 79.4 | 74.9 | 80.1 | 75.8 | 77.4 | 72.4 | 79.4 | 75 | 94.5 | 78.2 |
| | **25** | 70.7 | 67.3 | 72.9 | 67.8 | 70.9 | 65.1 | 71.9 | 67.1 | 71.7 | 66.1 | 92.8 | 75.1 |
| **100** | **10** | 92.3 | 89.8 | 91.2 | 88 | 92.6 | 89.4 | 91.4 | 88.5 | 91.3 | 88.5 | 98.9 | 92.3 |
| | **15** | 88.1 | 85 | 89.5 | 87 | 87.7 | 84.6 | 89.5 | 86.3 | 88 | 84.7 | 96.7 | 86.5 |
| | **20** | 84.4 | 82.5 | 83.6 | 80.4 | 84 | 81.4 | 84.9 | 82.4 | 85.8 | 83 | 97.7 | 86 |
| | **25** | 80.7 | 77.5 | 82.8 | 80 | 82.3 | 80.1 | 82.3 | 80.2 | 83 | 79.7 | 94.8 | 81.5 |
| **200** | **10** | 93.9 | 91.2 | 94 | 91.7 | 92.6 | 89.5 | 93.1 | 90.6 | 92.8 | 91.6 | 98.5 | 93.2 |
| | **15** | 90.9 | 88.8 | 90.8 | 89.1 | 90.8 | 89.3 | 92 | 90.5 | 90.3 | 88.7 | 98.8 | 91.1 |
| | **20** | 91.7 | 89.3 | 90.6 | 88.3 | 90 | 88.6 | 89.8 | 88.2 | 88.7 | 86.5 | 98.1 | 91 |
| | **25** | 87 | 84 | 88.3 | 87.2 | 89.4 | 87 | 87 | 84.6 | 89.9 | 88.3 | 97.9 | 89 |

**Table 5.** Performance of different link functions with standardized and adjusted deviance residuals for the detection of influential observations when $\phi$ =0.67

| Sample size n | Influential Observations | Cook's Distance | | | | | | DFFITS | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\phi$ = 0.67 | | | | | | $\phi$ = 0.67 | | | | | |
| | | Inverse | | Identity | | Log | | Inverse | | Identity | | Log | |
| | | SDR | ADR | SDR | ADR | SDR | ADR | SDR | ADR | SDR | ADR | SDR | ADR |
| **25** | **10** | 83.8 | 78.1 | 84.1 | 79.3 | 85.9 | 81.6 | 85.2 | 80.1 | 83.4 | 79.2 | 97.2 | 83.8 |
| | **15** | 74.3 | 68.6 | 74.7 | 68.6 | 74.1 | 68.6 | 75.4 | 68.6 | 74.1 | 68.6 | 92.8 | 74.6 |
| | **20** | 62.8 | 56.7 | 64.3 | 57.6 | 62.2 | 57 | 62.5 | 56.7 | 65.4 | 58.3 | 89 | 63 |
| | **25** | 56.7 | 50.6 | 53 | 45.1 | 57.3 | 48.7 | 52.7 | 45.9 | 54 | 46.6 | 84.8 | 53 |
| **50** | **10** | 89 | 85.2 | 88.3 | 85.2 | 87.1 | 83.4 | 89.8 | 85.6 | 86.8 | 82.9 | 97.9 | 88.6 |
| | **15** | 81.3 | 76.5 | 84.4 | 80.3 | 83.5 | 79.1 | 81 | 77.3 | 82.7 | 78.7 | 96.8 | 82.6 |
| | **20** | 77.4 | 72.3 | 77.5 | 74.2 | 76.1 | 71 | 78.6 | 74.5 | 79.4 | 74.4 | 95.3 | 79.4 |
| | **25** | 72.9 | 69.7 | 72.7 | 67 | 72.8 | 68.4 | 70.3 | 66.1 | 73.5 | 68.5 | 92 | 69.3 |
| **100** | **10** | 90.7 | 87.1 | 92.5 | 89.9 | 90.1 | 88.3 | 93.9 | 90.7 | 91.9 | 88.9 | 98.5 | 91 |
| | **15** | 88.7 | 86.6 | 87.8 | 85.2 | 89 | 86.1 | 89.5 | 86.8 | 88.2 | 85.7 | 97.9 | 91.1 |
| | **20** | 84.7 | 81.9 | 85.9 | 82.7 | 85.3 | 82.7 | 84.2 | 81.7 | 84.9 | 82.1 | 97.1 | 87.8 |
| | **25** | 84.7 | 81.5 | 81.9 | 78.6 | 83.2 | 80.9 | 81.3 | 78.8 | 82.3 | 79.5 | 96 | 81.7 |
| **200** | **10** | 93.2 | 91.4 | 93.8 | 91.3 | 94.2 | 92.1 | 93.2 | 91.6 | 93.4 | 91.9 | 98.9 | 93 |
| | **15** | 91 | 88.5 | 90.8 | 88.8 | 91.2 | 89.7 | 92.6 | 89.9 | 91.3 | 89.4 | 97.3 | 90.2 |
| | **20** | 90.6 | 89 | 89.9 | 89.1 | 90 | 88.4 | 89 | 86.7 | 90.4 | 88.8 | 98.6 | 89.3 |
| | **25** | 87.8 | 85.5 | 88.8 | 88 | 89.7 | 88.2 | 88.2 | 86.1 | 87.9 | 86.1 | 96.6 | 88.6 |

**Table 6.** Performance of different link functions with standardized and adjusted deviance residuals for the detection of influential observations when $\phi$ =2

| Sample size n | Influential Observations | Cook's Distance | | | | | | DFFITS | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\phi$ = 2 | | | | | | $\phi$ = 2 | | | | | |
| | | Inverse | | Identity | | Log | | Inverse | | Identity | | Log | |
| | | SDR | ADR | SDR | ADR | SDR | ADR | SDR | ADR | SDR | ADR | SDR | ADR |
| **25** | **10** | 85.4 | 80.7 | 84.5 | 83.4 | 83.4 | 78.5 | 84 | 79.1 | 82.2 | 79 | 97.4 | 85.3 |
| | **15** | 75.5 | 70.1 | 74.1 | 76.6 | 76.6 | 70.9 | 76.3 | 71 | 72.2 | 67.2 | 94.7 | 75.5 |
| | **20** | 62.5 | 56.6 | 62 | 62.1 | 62.1 | 56.3 | 63.5 | 56 | 63.3 | 57.2 | 89.6 | 63.4 |
| | **25** | 53.6 | 43.4 | 53.3 | 53.6 | 53.6 | 45.6 | 54.9 | 47.8 | 57.6 | 50.2 | 84.1 | 53.6 |
| **50** | **10** | 89.1 | 85 | 89.3 | 86.5 | 86.5 | 82.3 | 87.4 | 83.6 | 88.1 | 83.9 | 97.7 | 89.5 |
| | **15** | 82.3 | 78.4 | 83.5 | 85.7 | 85.7 | 80.8 | 83.3 | 79.6 | 84.5 | 80.8 | 97.2 | 83.6 |
| | **20** | 77.6 | 74.5 | 77.7 | 80.5 | 80.5 | 75.7 | 80.1 | 74.7 | 77.7 | 74.1 | 94.7 | 78.9 |
| | **25** | 73.3 | 70.1 | 73.6 | 73.3 | 73.3 | 68.9 | 73.5 | 69.3 | 74.6 | 69.7 | 93.8 | 75.2 |
| **100** | **10** | 92.8 | 90.5 | 91.8 | 90.5 | 90.5 | 87.2 | 89.9 | 88.1 | 90 | 86.8 | 98.1 | 90.4 |
| | **15** | 88.3 | 85.5 | 90 | 87.7 | 87.7 | 85.4 | 88.4 | 85.8 | 86.9 | 83.9 | 97.4 | 89.5 |
| | **20** | 85 | 83.2 | 84.8 | 86.9 | 86.9 | 84.2 | 85.6 | 82.5 | 84.5 | 81.4 | 97.2 | 83.8 |
| | **25** | 82.5 | 80.3 | 81.1 | 81.4 | 81.4 | 78.6 | 81.2 | 79.6 | 82.9 | 80 | 95.4 | 82.5 |
| **200** | **10** | 92.9 | 90.6 | 94.6 | 93.2 | 94.8 | 90.7 | 95.1 | 92.5 | 93 | 90.1 | 98.9 | 93.2 |
| | **15** | 91.6 | 89.8 | 92.8 | 91 | 93.2 | 89.5 | 92.4 | 90.7 | 90.8 | 89.4 | 98.2 | 90.7 |
| | **20** | 90.3 | 89.6 | 89.8 | 90.9 | 91 | 88.5 | 89.4 | 88.3 | 89.7 | 88.1 | 97.6 | 89.9 |
| | **25** | 86.9 | 85.1 | 88.4 | 87.9 | 90.9 | 86.1 | 89.9 | 88.6 | 88.2 | 87.4 | 96.8 | 86.3 |

**Table 7.** Performance of different link functions with standardized and adjusted deviance residuals for the detection of influential observations when $\phi$ =5

| Sample size n | Influential Observations | Cook's Distance | | | | | | DFFITS | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\phi$ = 5 | | | | | | $\phi$ = 5 | | | | | |
| | | Inverse | | Identity | | Log | | Inverse | | Identity | | Log | |
| | | SDR | ADR | SDR | ADR | SDR | ADR | SDR | ADR | SDR | ADR | SDR | ADR |
| **25** | **10** | 82 | 76.3 | 85.1 | 79.2 | 84 | 78.7 | 84.7 | 79.1 | 85.2 | 79.8 | 96.9 | 82.1 |
| | **15** | 76.8 | 69.1 | 76.9 | 70 | 73.8 | 68.2 | 75.6 | 69.7 | 74.8 | 68.7 | 92.2 | 74 |
| | **20** | 65.4 | 58.3 | 63.7 | 56.1 | 66.2 | 60.1 | 64.2 | 56.7 | 64.9 | 57.6 | 89.4 | 63.7 |
| | **25** | 56.1 | 48.8 | 56.6 | 48.2 | 53.8 | 46.2 | 52.9 | 47.2 | 54.3 | 46.1 | 84.6 | 54.5 |
| **50** | **10** | 89.2 | 85.3 | 88.3 | 84.8 | 88.3 | 84.7 | 90.3 | 85.4 | 88 | 83.5 | 98.4 | 91.2 |
| | **15** | 83.7 | 79.9 | 81.5 | 77.6 | 84.7 | 82 | 84.9 | 80.4 | 82.6 | 77.6 | 96.2 | 83 |
| | **20** | 77.7 | 74.5 | 75.9 | 71.7 | 75 | 71.5 | 77.9 | 74.1 | 78.5 | 74.3 | 94.5 | 77.4 |
| | **25** | 73.1 | 68.7 | 71.6 | 67.5 | 74 | 69.1 | 75.2 | 69 | 74 | 69.1 | 92.2 | 74.6 |
| **100** | **10** | 92.2 | 89.3 | 89.1 | 86.7 | 91.6 | 88.5 | 90 | 86.2 | 90.4 | 87.7 | 98.5 | 90.3 |
| | **15** | 89.1 | 87.3 | 87.1 | 84.5 | 89.2 | 86.2 | 87.7 | 85.2 | 88.3 | 85.3 | 97 | 88 |
| | **20** | 85.6 | 82.7 | 83.8 | 81.3 | 84.7 | 81.8 | 84.6 | 82.2 | 86 | 83.1 | 97.1 | 84.9 |
| | **25** | 83.5 | 80 | 83.2 | 80.3 | 83.1 | 80.1 | 83.9 | 81.4 | 82.9 | 80.3 | 95.9 | 81.7 |
| **200** | **10** | 94.3 | 91.3 | 93 | 90.5 | 94 | 91.5 | 91 | 90.1 | 93.7 | 91.3 | 98.6 | 92.4 |
| | **15** | 91.7 | 89.8 | 89.9 | 87.6 | 92.8 | 91.1 | 91.1 | 89.7 | 90 | 87.3 | 98.9 | 92.5 |
| | **20** | 89.2 | 87.5 | 88.7 | 87.4 | 90.4 | 88.6 | 91.4 | 89.6 | 89.3 | 87.6 | 98.3 | 90.6 |
| | **25** | 88.4 | 87.2 | 88.2 | 86.7 | 86.7 | 84.8 | 89.9 | 88 | 89.6 | 88.2 | 97.4 | 88.7 |

**Table 8.** Performance of different link functions with standardized and adjusted deviance residuals for the detection of influential observations when $\phi$ =10

| Sample size n | Influential Observations | Cook's Distance | | | | | | DFFITS | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $\phi$ =10 | | | | | | $\phi$ =10 | | | | | |
| | | Inverse | | Identity | | Log | | Inverse | | Identity | | Log | |
| | | SDR | ADR | SDR | ADR | SDR | ADR | SDR | ADR | SDR | ADR | SDR | ADR |
| **25** | **10** | 93.8 | 91.9 | 82.4 | 76.2 | 84.9 | 79.8 | 85 | 82.2 | 83.5 | 78.9 | 97.1 | 82.9 |
| | **15** | 91.7 | 89.4 | 76.8 | 70.4 | 76.2 | 70 | 73.3 | 66.9 | 73.2 | 68 | 94.8 | 75.2 |
| | **20** | 87.9 | 85.4 | 63.9 | 56.7 | 63.6 | 56.7 | 64.4 | 58.7 | 64.5 | 58.1 | 89.1 | 61.6 |
| | **25** | 90 | 88.3 | 53.2 | 47.8 | 54.7 | 48.4 | 53.1 | 46.1 | 53.8 | 47.2 | 84 | 54.8 |
| **50** | **10** | 95 | 94 | 88 | 84.4 | 88 | 85 | 88.7 | 84.5 | 87.8 | 84.6 | 98.5 | 88.8 |
| | **15** | 91.5 | 89.6 | 82.3 | 78.2 | 82.7 | 78.3 | 83.4 | 78.5 | 82 | 78.7 | 96.7 | 81.9 |
| | **20** | 91.2 | 89.3 | 76.4 | 72.4 | 78.8 | 74.4 | 77.8 | 73.1 | 79.5 | 75 | 94.5 | 76.7 |
| | **25** | 90 | 88.5 | 74.6 | 68.5 | 71.7 | 66.3 | 70.5 | 64.9 | 74.9 | 68.9 | 92 | 70.4 |
| **100** | **10** | 93.6 | 91.1 | 90.6 | 87.7 | 91.2 | 88.4 | 92 | 88.1 | 90.1 | 87.2 | 98.5 | 91.6 |
| | **15** | 92.4 | 91 | 88.3 | 85.7 | 88 | 85.8 | 89.4 | 88.1 | 87.2 | 84.4 | 97.2 | 89.2 |
| | **20** | 91.8 | 90.6 | 85.2 | 81.8 | 84.8 | 81.6 | 85.6 | 83.2 | 88.2 | 84.9 | 96.5 | 85.6 |
| | **25** | 90.1 | 88.8 | 82.2 | 79.8 | 79.6 | 76.8 | 80.6 | 78.1 | 81.7 | 79.5 | 95.2 | 83.3 |
| **200** | **10** | 93.2 | 91.8 | 93 | 90.1 | 92.6 | 90.3 | 94.6 | 92 | 93.1 | 90.6 | 98.2 | 94.8 |
| | **15** | 92.7 | 91.4 | 91.9 | 90.2 | 91.2 | 88.9 | 90.2 | 88.5 | 91.7 | 89.9 | 98.7 | 91.9 |
| | **20** | 92.1 | 90.7 | 90.2 | 89 | 88.7 | 87.2 | 90.5 | 88.8 | 89.7 | 87.8 | 98.7 | 90.8 |
| | **25** | 88.9 | 87.6 | 90.3 | 88.6 | 88.2 | 86.8 | 88.2 | 86.9 | 88 | 86.6 | 98 | 88 |

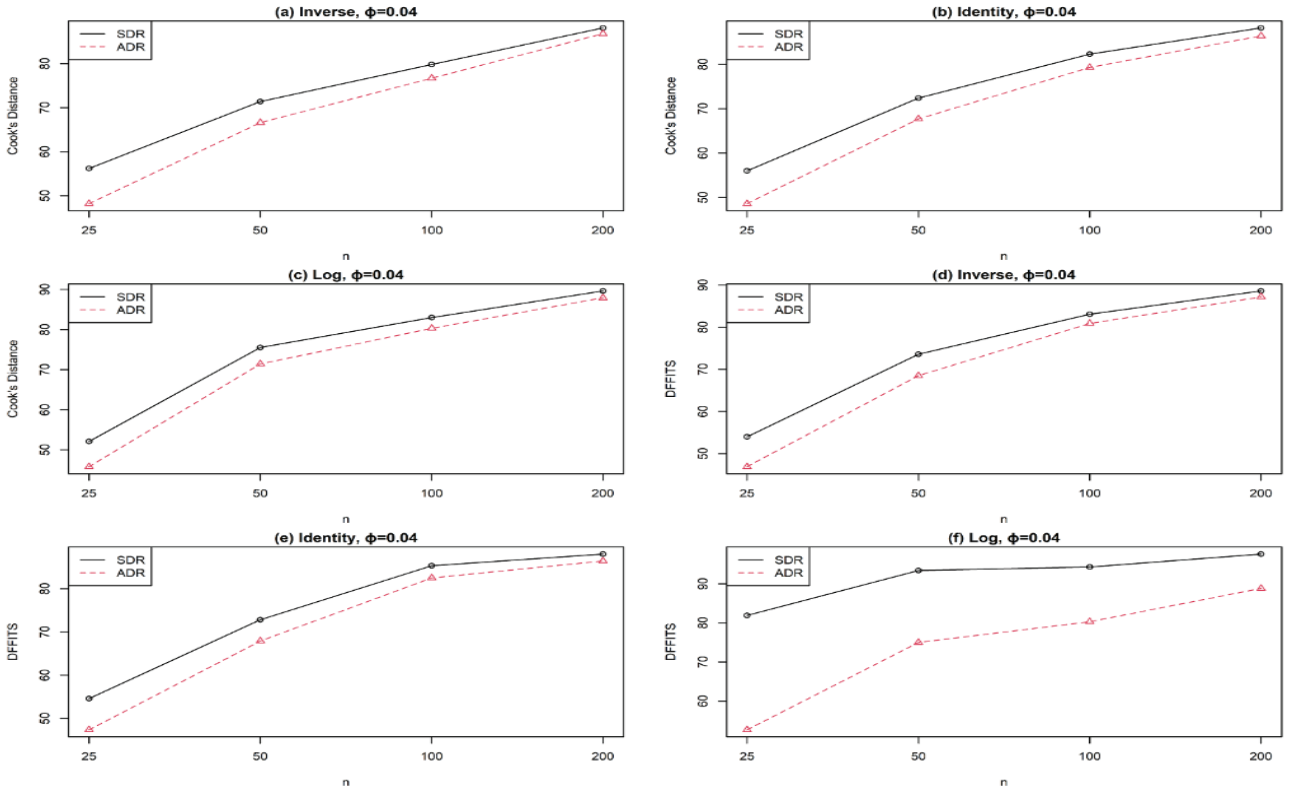


**Figure 1.** Index plots of CD and DFFITS under different link functions with $\phi$ =0.04

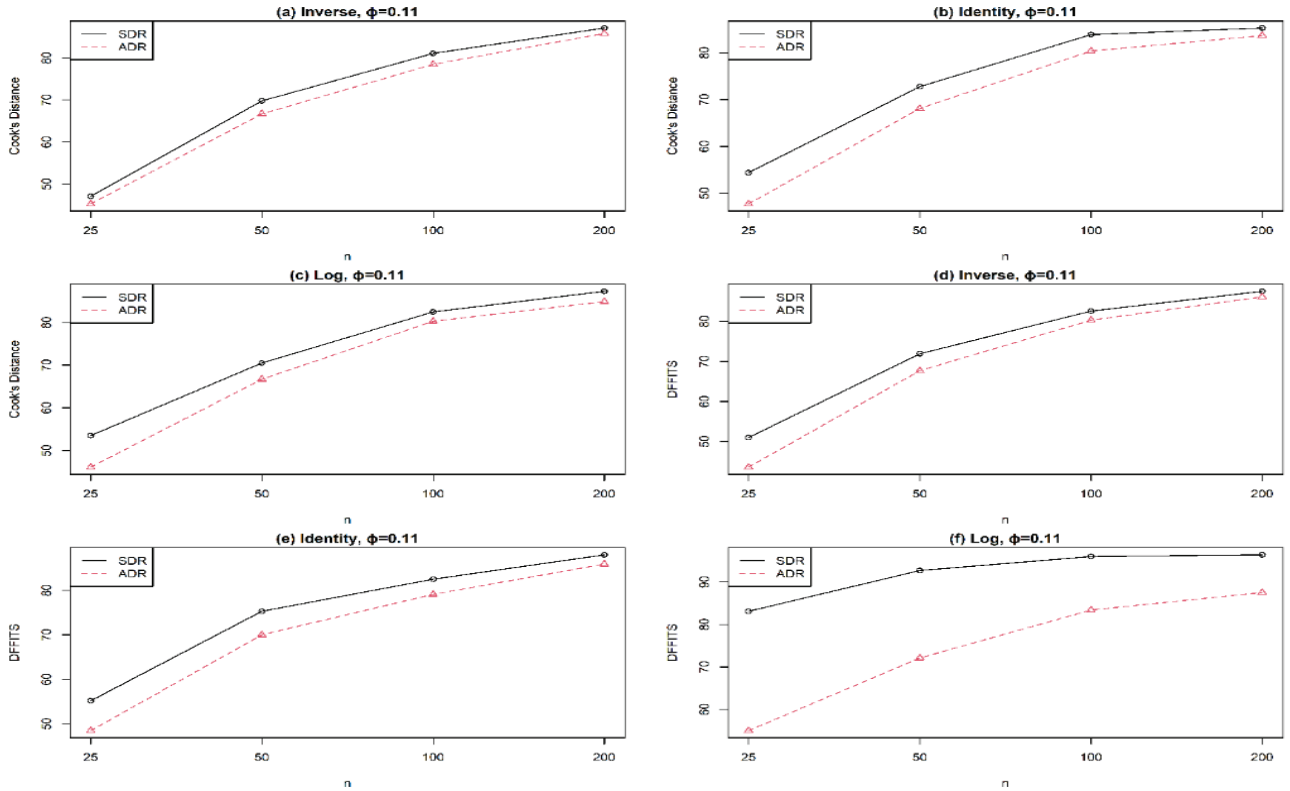


**Figure 2.** Index plots of CD and DFFITS under different link functions with $\phi$ =0.11

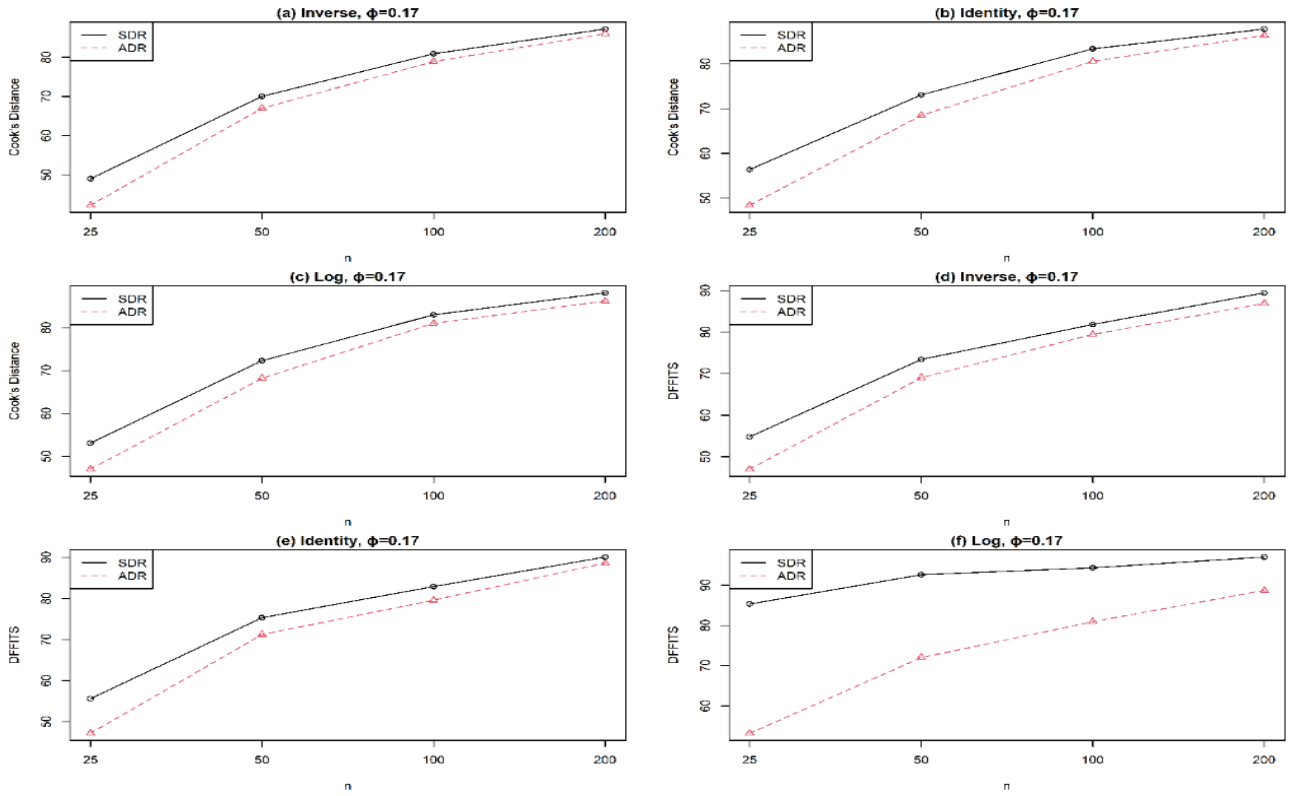


**Figure 3.** Index plots of CD and DFFITS under different link functions with $\phi$ =0.17

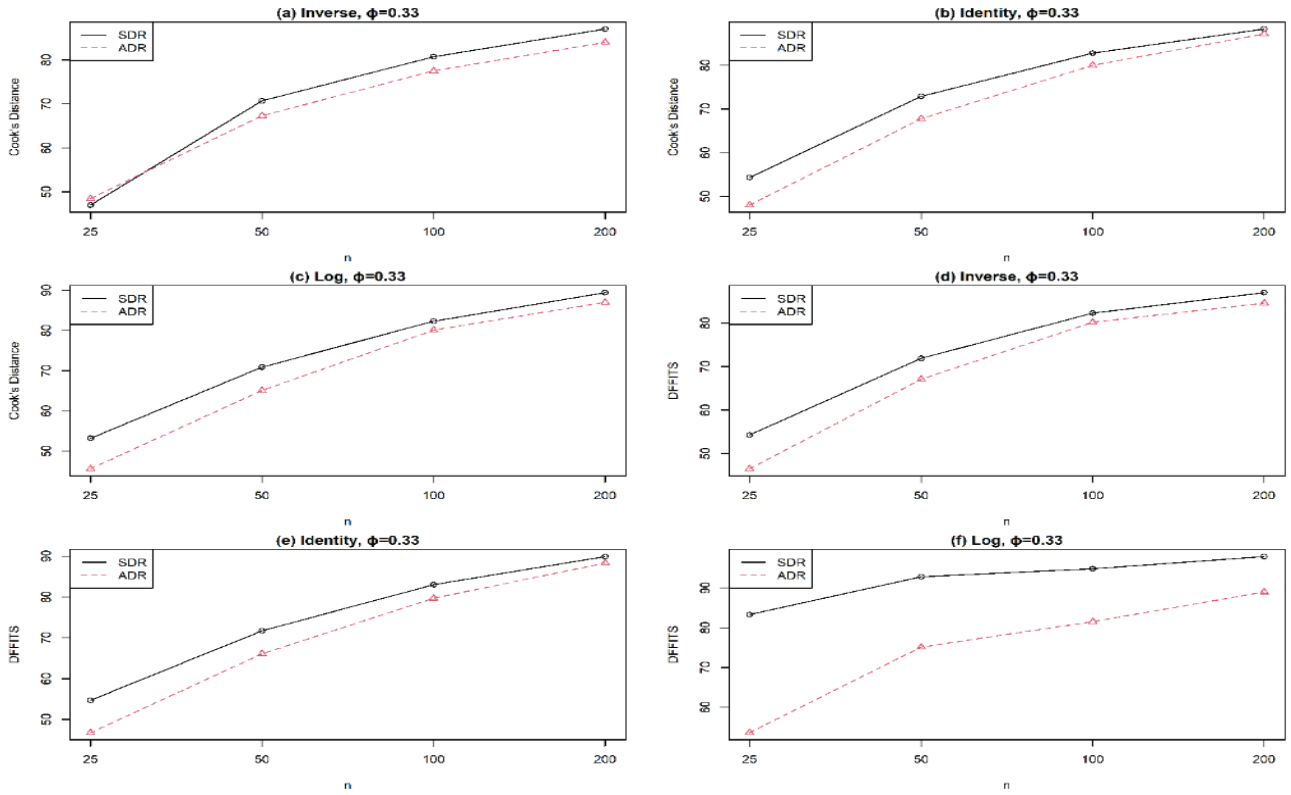


**Figure 4.** Index plots of CD and DFFITS under different link functions with $\phi$ =0.33

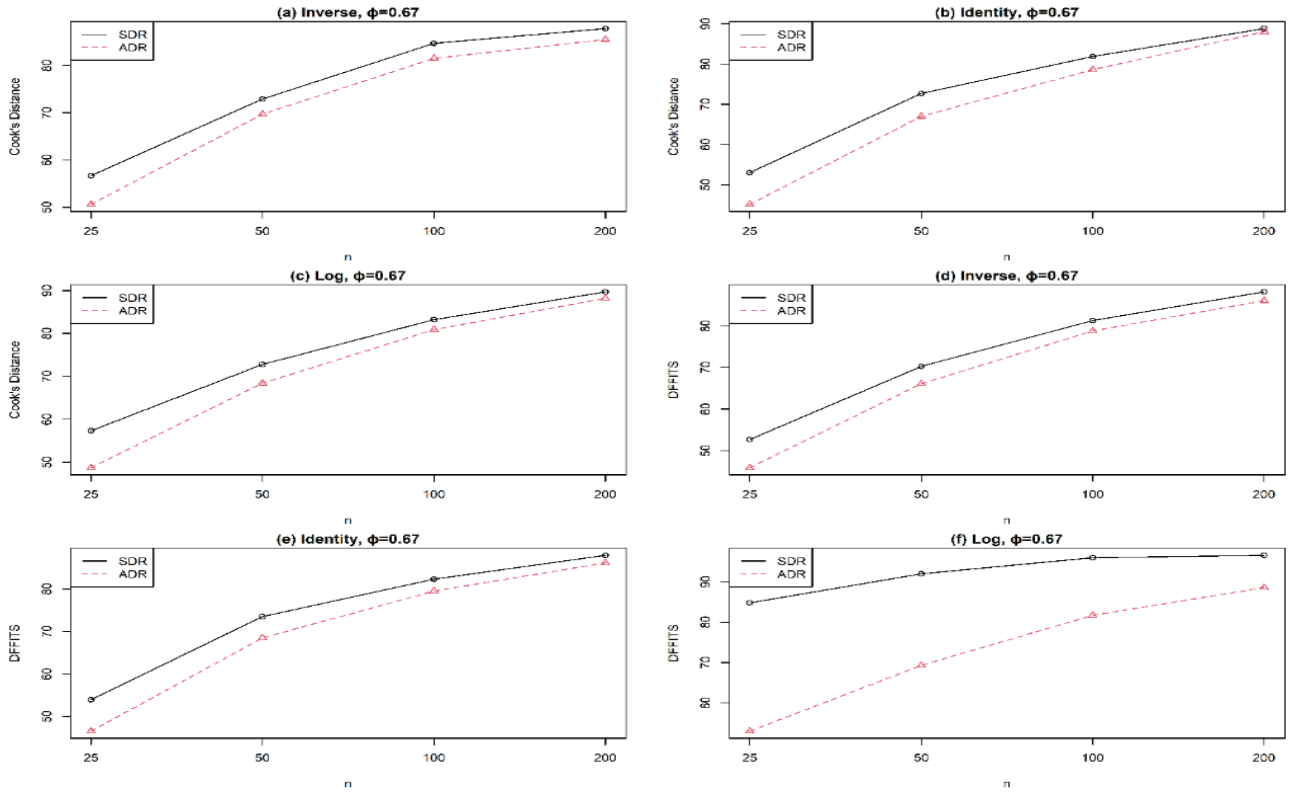


**Figure 5.** Index plots of CD and DFFITS under different link functions with $\phi$ =0.67

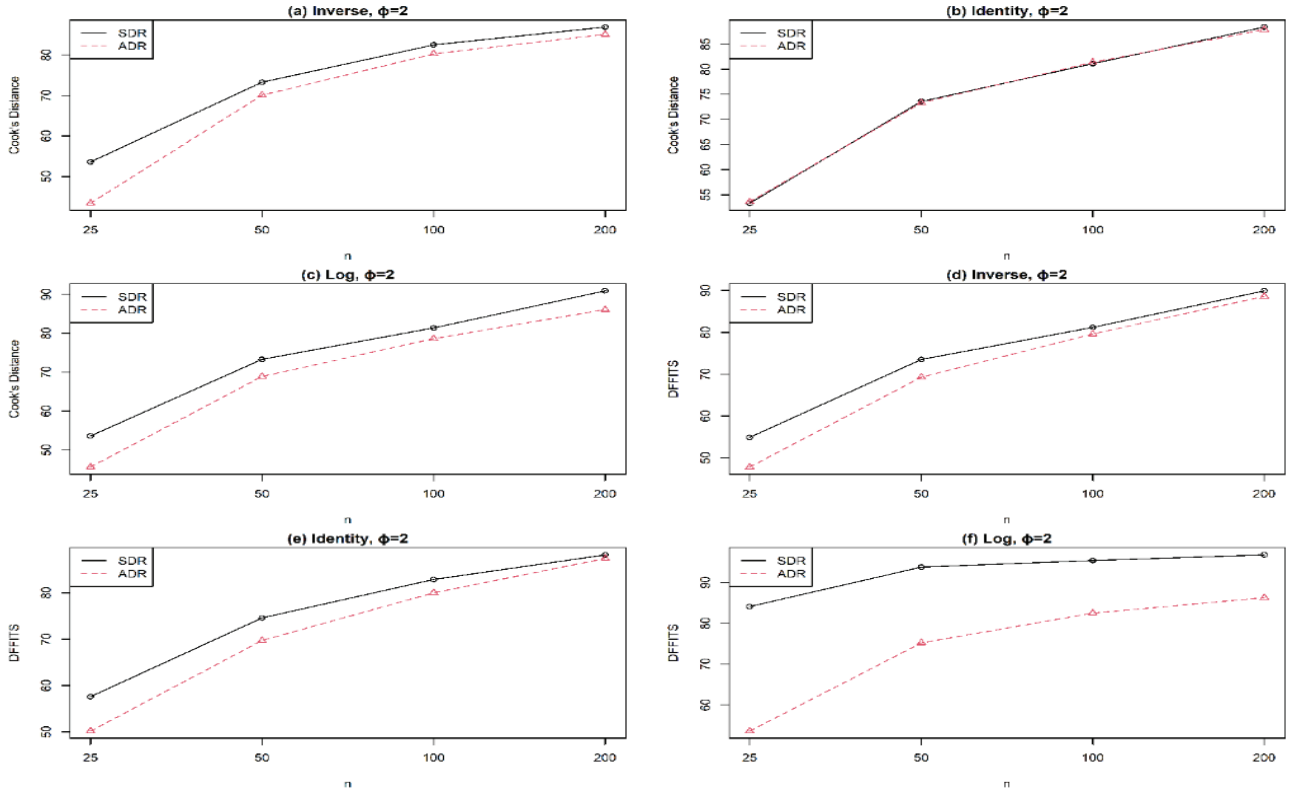


**Figure 6.** Index plots of CD and DFFITS under different link functions with $\phi$ = 2.

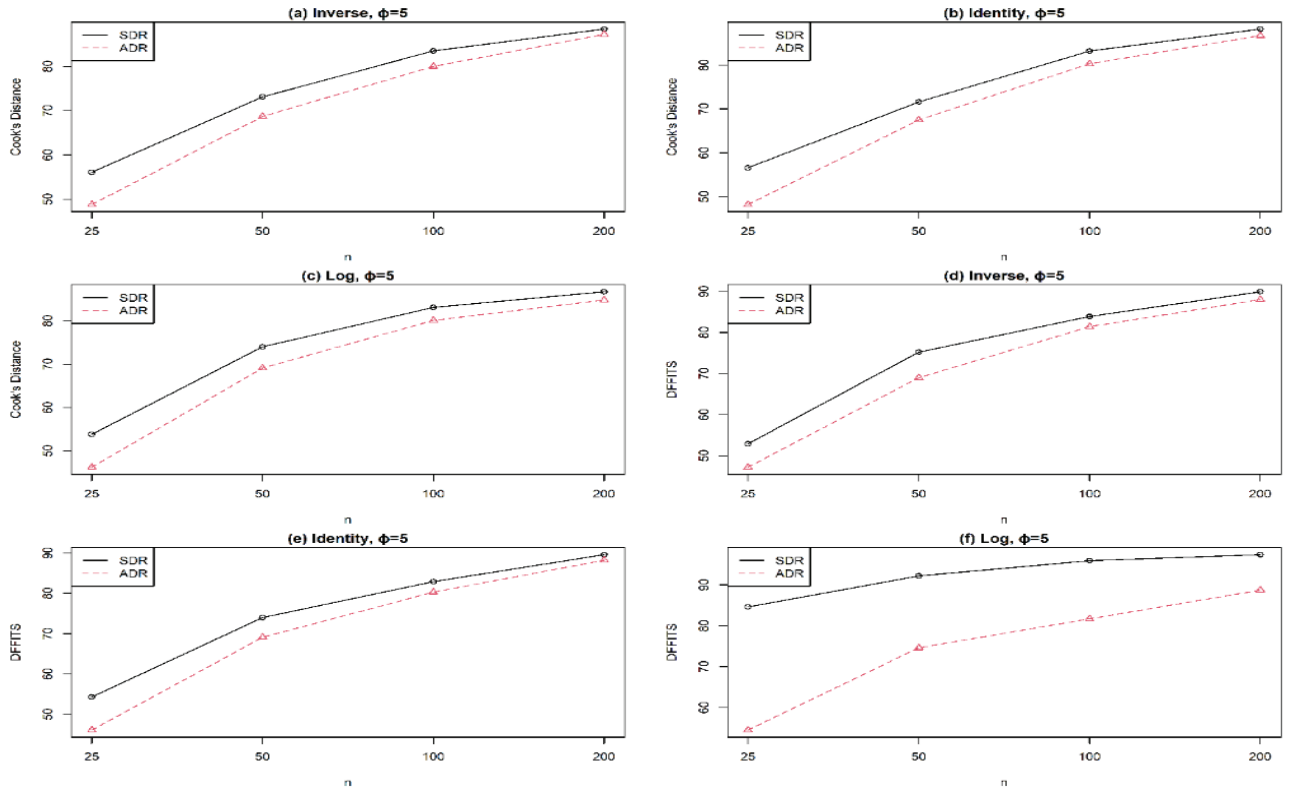


**Figure 7.** Index plots of CD and DFFITS under different link functions with $\phi = 5$.

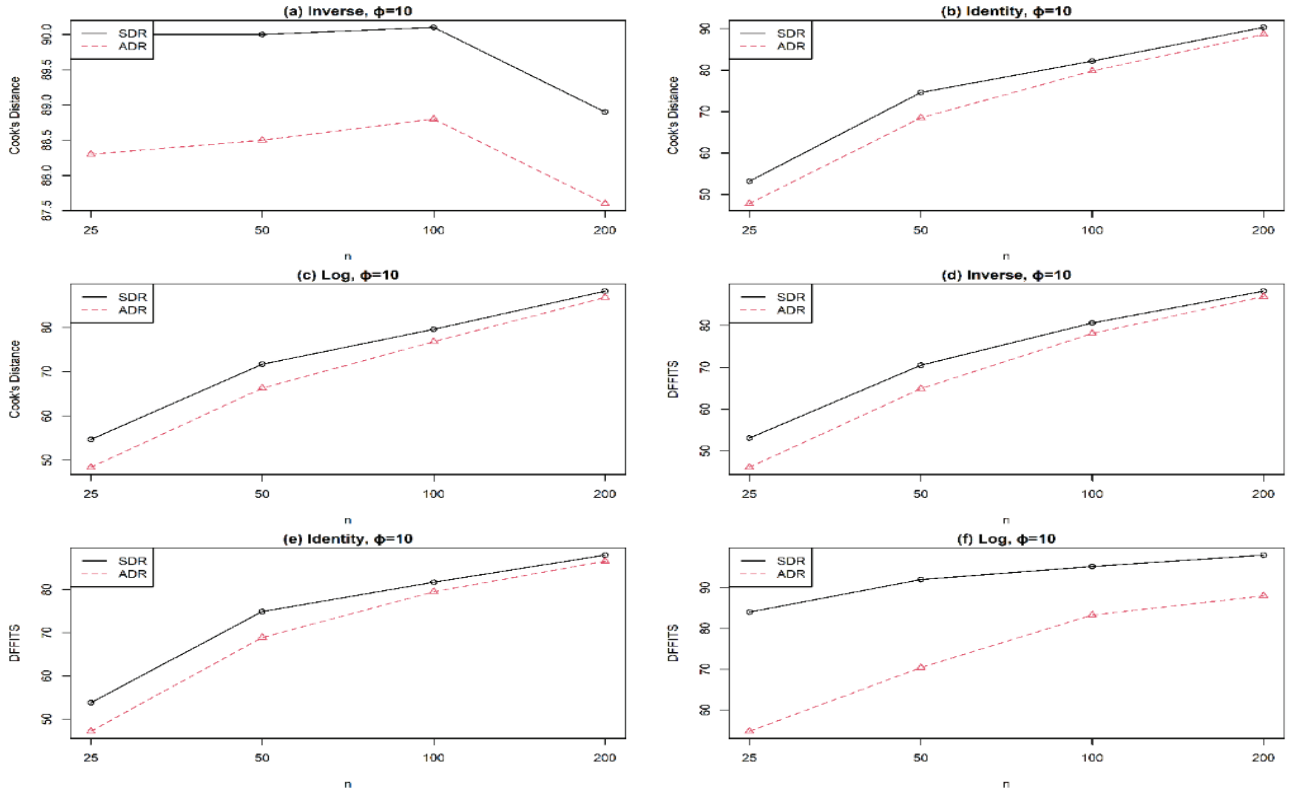


**Figure 8.** Index plots of CD and DFFITS under different link functions with $\phi = 10$.

**Application: Reaction Rate Dataset**

Now, we will illustrate the performance of the different link functions for G-PRM with the help of real-life application. The reaction rate data set is taken from [4]. The main objective of this data set is to determine the reaction rate of the catalytic isomerization of n-pentane to iso-pentane based on partial pressure of different independent variables (factors). These explanatory variables are used to speed up the reaction rate. This data set consists of $n = 24$ experimental data values with one dependent variable, i.e. reaction rate ($y$) and $p = 3$ explanatory variables, i.e. partial pressure of hydrogen ($x1$), partial pressure of n-pentane ($x2$) and partial pressure of iso-pentane ($x3$). Then Amin et al. [7,41,42,43] and [44] utilized this data set. As it is mentioned by [44], response variable follows a G-PD is required by following [1,6]. However, because of the positively skewed trend of the dependent variable, this data set is not well fitted to the normal distribution. This data set is well fitted to the Gamma distribution as tested by few tests of goodness of fit and the results are given in Table 9. In Table 11,12 and 13 are present a model coefficient summary as inverse, identity and log link function respectively and with and without influential observation.

The G-PRM is an appropriate regression model for this set of data. Influential observations have an impact on the G-PRM estimates just like they do on the other models. Therefore, identifying these important observations under various link functions is our primary concern. We have calculated the cook's distance and DFFITS and fitted the G-PRM under various link functions. In Table 10, present an influential observations summary. The diagnostic measures cook's distance and DFFITS under different link functions with SDR and ADR respectively. we observe that the Cook`s distance with SDR under inverse link function diagnosed 5,6,11 is influential observations. while, on the other hand the Cook`s distance with ADR under inverse link function diagnosed 5,6 is influential observations. Similarly, we observe that the Cook`s distance with SDR under identity link function diagnosed 6,22,24 is influential observation while, on the other hand the Cook`s distance with ADR under identity link function does not diagnosed any influential observations. It is interesting to note that the Cook`s distance with SDR and ADR under log link function diagnosed observations 5,6,7,8,11,12,16,19,22,23,24 and other 1,4-11,13,14,20-24 is influential observation respectively. Now we discussed second diagnostic measure is DFFITS under different link functions and deviance residual form such SDR and ADR. For DFFITS with SDR under inverse link function diagnosed 11,12 is influential observation. while, on the other hand DFFITS with ADR under inverse link function diagnosed only 6 is influential observations. Similarly, for DFFITS with SDR under identity link function diagnosed only 19 is influential observation while, on the other hand the DFFITS with ADR under identity link function does not diagnosed any influential observations. It is interesting to note that the DFFITS with SDR under log link function diagnosed observation 22,24 is influential observation. But DFFITS with ADR under log link function diagnosed 5,6,13,24 is influential observation. We now identify the observations that affect the G-PRM estimates and confirm the influence of the diagnostic process and the link function. To do this, we calculate the percentage change in the G-PRM estimates following the removal of any influential observations that we find. The results are shown in Table 14. Table 14 presents a comparison of the various diagnostic techniques under various link functions, allowing us to determine which technique correctly identifies the influential observations. We can see from Table 14 that the 7th observation is the most influential value. With the excluding of the identity link function, only the Cook's distance method was able to identify this observation under the inverse and log link functions with G-PRM estimates of $\beta_1$ and $\beta_2$ are impacted by this finding. The 22th observation is the second most important one. This observation was diagnosed by the cook's distance using only the inverse and log link functions, leaving out the identity link function. The G-PRM estimates of $\beta_0$ and $\beta_3$ are impacted by this finding. Likewise, under various link functions 11th, 24th influential observations, the cook's distance also affects the G-PRM estimate that is indicated in Table 14 with bold values. It has been observed that the 19th observation holds the most influence. The G-PRM estimate of $\beta_0$ is affected by the influential observation, which is only detected by the DFFITS method under the inverse link function and excludes the identity and log link functions.

So, the appropriate regression model to determine the reaction rate ($y$) based on these three explanatory variables such as $p = 3$ explanatory variables, i.e. partial pressure of hydrogen ($x_1$), partial pressure of n-pentane ($x_2$) and partial pressure of iso-pentane ($x_3$) is the G-PR model.

The fitted G-PRM for inverse link function using real data is given by.

$$\hat{y}_i = (-3.175[0.382,\ \mathrm{S}] + 0.059x_1[0.011,\ \mathrm{S}] - 0.067x_2[0.005, \mathrm{S}] + 0.004x_3[0.002, \mathrm{S}])^{-1}$$

The fitted G-PRM for identity link function using real data is given by.

$$\hat{y}_i = (0.178[0.093,\ \mathrm{N}] + 0.0004x_1[0.0002,\ \mathrm{N}] - 0.001x_2[0.0002, \mathrm{S}] + 0.003x_3[0.0007, \mathrm{S}])$$

The fitted G-PRM for log link function using real data is given by.

$$\hat{y}_i = log(0.946[0.343,\ \mathrm{S}] - 0.001x_1[0.0007,\ \mathrm{N}] + 0.009x_2[0.0002, \mathrm{N}] - 0.011x_3[0.001, \mathrm{S}])$$

where the square brackets contain the standard errors of the estimated parameters. The letter N represents the non-significance and S represents the significance of the regression coefficients.

## CONCLUSION

Like other models, influential observations also affect the G-PRM estimates. The G-PRM is estimated under

various link functions. So, in this study, we compare the performance of two influence diagnostic methods with different link functions for the identification of influential observations to find a suitable diagnostic method and a link function in the G-PRM. For these purposes, we consider inverse, identity and link functions. We use the two influence diagnostic techniques, i.e. Cook's distance and DIFFITS to diagnose the influential observations with considered link functions. To evaluate the performance of G-PRM diagnostic methods with different link functions, we use the Monte Carlo simulation and a real application. Simulation results show that for all sample sizes, the performance of the Cook`s Distance with log link functions is better than the DFFITS methods. While for large dispersion and small $n$, the performance of the cook's distance is better than the DIFFITS method. Similarly, for large dispersion and large $n$, the performance of the cook's distance and DIFFITS is the same in detecting the influential observations. The real-life application results also support the simulation results. We strongly recommended a Cook`s Distance for the G-PRM to the detection of influential observations under different link functions.

Future research recommendations, there are dimensions which still need to be explored. This study covers the influence diagnostics with different GLM residuals under different link functions in the G-PRM. These can be extended to GLM influence diagnostics with one biased (Ridge) estimation, modified ridge estimation, Liu estimation, modified Liu estimation and Stein estimation methods.

## AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

## DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

## CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## ETHICS

There are no ethical issues with the publication of this manuscript.

## REFERENCES

[1] Hanum H, Wigena AH, Djuraidah A, Mangku IW. Modeling gamma-Pareto distributed data using GLM gamma. Glob J Pure Appl Math 2016;12:3569–3575.

[2] Alzaatreh A, Famoye F, Lee C. Gamma-Pareto distribution and its applications. J Mod Appl Stat Methods 2012;11:7. [CrossRef]

[3] Dobson AJ. An introduction to generalized linear models. Boca Raton: Chapman and Hall/CRC; 2002. [CrossRef]

[4] Huet S, Bouvier A, Poursat MA, Jolivet E, Bouvier AM. Statistical tools for nonlinear regression: A practical guide with S-PLUS and R examples. New York: Springer; 2004. p. 233.

[5] Hanum HH. Modeling extreme rainfall with gamma. Appl Math Sci 2015;9:6029–6039. [CrossRef]

[6] Hanum H, Wigena AH, Djuraidah A, Mangku IW. The application of modeling gamma-Pareto distributed data using GLM gamma in estimation of monthly rainfall with TRMM data. Sriwijaya J Environ 2017;2:40–45. [CrossRef]

[7] Zheng L, Ismail K, Meng X. Shifted Gamma-Generalized Pareto distribution model to map the safety continuum and estimate crashes. Saf Sci 2014;64:155–162. [CrossRef]

[8] Ashour SK, Said DR, Fahim MA. The log-Gamma-Pareto distribution. Int J Sci Basic Appl Res 2014;16:357–374.

[9] Alzaatreh A, Ghosh I. A study of the Gamma-Pareto (IV) distribution and its applications. Commun Stat Theory Methods 2016;45:636–654. [CrossRef]

[10] De Andrade TA, Fernandez LMZ, Silva FG, Cordeiro GM. The gamma generalized Pareto distribution with applications in survival analysis. Int J Stat Probab 2017;6:141–156. [CrossRef]

[11] Alzaghal A. The exponentiated gamma-Pareto distribution with application to bladder cancer susceptibility. J Stat Appl Probab Lett 2020;7:39–56. [CrossRef]

[12] Dar AA, Ahmed A, Reshi JA. Weighted Gamma-Pareto distribution and its application. Pak J Stat 2020;36:287–304.

[13] McCullagh P. Generalized linear models. London: Routledge; 1989. [CrossRef]

[14] Hardin JW, Hilbe JM. Generalized linear models and extensions. College Station: Stata Press; 2007.

[15] Cook RD. Detection of influential observation in linear regression. Technometrics 1977;19:15–18. [CrossRef]

[16] Belsley DA, Kuh E, Welsch RE. Regression diagnostics: Identifying influential data and sources of collinearity. Hoboken: John Wiley & Sons; 2005.

[17] Preisser JS, Qaqish BF. Deletion diagnostics for generalised estimating equations. Biometrika 1996;83:551–562. [CrossRef]

[18] Pregibon D. Logistic regression diagnostics. Ann Stat 1981;9:705–724. [CrossRef]

[19] Williams D. Generalized linear model diagnostics using the deviance and single case deletions. J R Stat Soc Ser C Appl Stat 1987;36:181–191. [CrossRef]

[20] Pierce DA, Schafer DW. Residuals in generalized linear models. J Am Stat Assoc 1986;81:977–986. [CrossRef]
[21] Cordeiro GM. On Pearson's residuals in generalized linear models. Stat Probab Lett 2004;66:213–219. [CrossRef]
[22] Cox DR, Snell EJ. A general definition of residuals. J R Stat Soc Ser B Methodol 1968;30:248–265. [CrossRef]
[23] Simas AB, Cordeiro GM. Adjusted Pearson residuals in exponential family nonlinear models. J Stat Comput Simul 2009;79:411–425. [CrossRef]
[24] Cook RD, Weisberg S. Residuals and influence in regression. New York: Chapman and Hall; 1982.
[25] Atkinson AC. Plots, transformations, and regression: An introduction to graphical methods of diagnostic regression analysis. New York: Oxford University Press; 1985.
[26] Cook RD. Assessment of local influence. J R Stat Soc Ser B Stat Methodol 1986;48:133–155. [CrossRef]
[27] Chatterjee S, Hadi AS. Sensitivity analysis in linear regression. New York: John Wiley & Sons; 1988. [CrossRef]
[28] Lee AH. Assessing partial influence in generalized linear models. Biometrics 1988;44:71–77. [CrossRef]
[29] Thomas W, Cook RD. Assessing influence on regression coefficients in generalized linear models. Biometrika 1989;76:741–749. [CrossRef]
[30] Amin M, Noor A, Mahmood T. Beta regression residuals-based control charts with different link functions: An application to the thermal power plants data. Int J Data Sci Anal 2024;22:1–13. [CrossRef]
[31] Amin M, Fatima A, Akram MN, Kamal M. Influential observation detection in the logistic regression under different link functions: An application to urine calcium oxalate crystals data. J Stat Comput Simul 2024;94:346–359. [CrossRef]
[32] Cheema M, Amin M, Mahmood T, Faisal M, Brahim K, Elhassanein A. Deviance and Pearson residuals-based control charts with different link functions for monitoring logistic regression profiles: An application to COVID-19 data. Mathematics 2023;11:1113. [CrossRef]
[33] Hadia M, Amin M, Akram MN. Comparison of link functions for the estimation of logistic ridge regression: An application to urine data. Commun Stat Simul Comput 2024;53:4121–4137. [CrossRef]
[34] Mustafa S, Amin M, Akram MN, Afzal N. On the performance of link functions in the beta ridge regression model: Simulation and application. Concurr Comput Pract Exp 2022;34:e7005. [CrossRef]
[35] Prasetyo RB, Kuswanto H, Iriawan N, Ulama BSS. A comparison of some link functions for binomial regression models with application to school drop-out rates in East Java. AIP Conf Proc 2019;2194:012001. [CrossRef]
[36] Afsana-Al-Sharmin, Islam MA. Generalized Weibull linear models with different link functions. Adv Appl Stat 2017;50:367–384. [CrossRef]
[37] Atkinson AC. Two graphical displays for outlying and influential observations in regression. Biometrika 1981;68:13–20. [CrossRef]
[38] Ullah MA, Pasha GR. The origin and developments of influence measures in regression. Pak J Stat 2009;25:295–307.
[39] Amin M, Amanullah M, Aslam M. Empirical evaluation of the inverse Gaussian regression residuals for the assessment of influential points. J Chemometr 2016;30:394–404. [CrossRef]
[40] Amin M, Amanullah M, Cordeiro GM. Influence diagnostics in the gamma regression model with adjusted deviance residuals. Commun Stat Simul Comput 2017;46:6959–6973. [CrossRef]
[41] Amin M, Afzal S, Akram MN, Muse AH, Tolba AH, Abushal TA. Outlier detection in gamma regression using Pearson residuals: Simulation and an application. AIMS Math 2022;7:15331–15347. [CrossRef]
[42] Amin M, Amanullah M, Aslam M, Qasim M. Influence diagnostics in gamma ridge regression model. J Stat Comput Simul 2019;89:536–556. [CrossRef]
[43] Amin M, Qasim M, Amanullah M. Performance of Asar and Genç and Huang and Yang's two-parameter estimation methods for the gamma regression model. Iran J Sci Technol Trans A Sci 2019;43:2951–2963. [CrossRef]
[44] Yasin A, Amin M, Qasim M, Muse AH, Soliman AB. More on the ridge parameter estimators for the Gamma ridge regression model: Simulation and applications. Math Probl Eng 2022;2022:6769421. [CrossRef]