Dhaka Univ. J. Sci. 74(1): 95-110, 2026 (January)
Prof. M. Muhibur Rahman Memorial Issue
ISSN 1022 - 2502, E-ISSN 2408- 8528
DOI: https://doi.org/10.3329/dujs.v74i1.83556

Influential Observations Detection in the Gamma-Pareto Regression Model Under Different link Functions: an Application to Reaction Rate Data

Atif Akbar and Nasir Saleem*

Department of Statistics, Bahauddin Zakariya University (BZU), Multan, Pakistan,60800

(Received : 7 August 2025; Accepted : 1 January 2026)

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 pearson residuals (SPR) and adjusted pearson residuals (APR). 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.

Keywords: Cook's Distance; DFFITS; Link functions; Invers; Identity; Log; Gamma-Pareto, Regression Model; Standardized Pearson Residuals; Adjusted Pearson Residuals

I. Introduction

The gamma-Pareto distribution (G-PD) is invented by Ayman Alzaatreh and further extend in a form of Gamma-Pareto regression model (G-PRM) by Herlina Hanum. 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 test's 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 function. According to (Dobson and Barnett, 2002) 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.

Alzaatreh et al. (2012) developed the G-PD. The exponential family distribution is a member represented by the G-PD Hanum et al. (2016). Consequently, GLM could be used to develop the regression modeling for the Gamma-Pareto regression model (G-PRM). GLM G-PD is analytically developed by Hanum et al. (2016). 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. The mathematical relationship between G-PD and GD was mentioned by Alzaatreh et al. (2012). Hanum et al. (2015) 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. Hanum et al. (2016) examined the relationship between the explanatory variable and the distributed response variable 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 Hanum et al. (2017). In order to map the safety continuum and estimate crashes, Zheng et al. (2014) discussed the Shifted Gamma-Generalized Pareto Distribution model. The new Log-Gamma-Pareto Distribution is created by Ashour et al. (2014). A new Gamma-Pareto (IV) distribution and its uses were presented by Alzaatreh and Ghosh (2016). The gamma generalized pareto distribution and its applications in survival analysis were covered by De Andrade et al. (2017). Exponentiated gamma-Pareto distribution was applied to bladder cancer susceptibility by Alzaghal (2020). The weighted gamma-pareto distribution and its use were covered by Dar et al. (2020). The introduction of generalized linear models (GLMs) allows for the investigation of dependent variable dependence on two independent variables. Another

*Author for correspondence. e-mail: nasirsaleem160@gmail.com

variation of the GLM was discussed by McCullagh and Nelder (1989). According to Hardin and Hilbe (2012), the GLM in fact goes against the non-influential observation assumption. Influence diagnostics were first introduced for linear models (LMs) by Cook (1977). These impact diagnostics were covered by Belsley et al. (1980) in a number of dimensions. According to Preisser and Qaqish (1996), Pregibon (1981), and Williams (1987), 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, Williams (1987) 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 Pierce and Schafer (1986) and the adjusted Pearson residuals suggested by Cordeiro (2004) based on Cox and Snell (1968). The aim of these theories is to attain normality. Simas and Cordeiro (2009) 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 Cook and Weisberg (1982), Atkinson (1985), Cook (1986), and Chatterjee and Hadi (1988). Conversely, Lee (1986) provided a method for evaluating partial influence in the GLM. One approach to evaluating the impact on the GLM regression coefficients was suggested by Thomas and Cook (1989).

We found from the literature that the majority of researchers used an identity link function with Pearson residuals to focus on G-PRM diagnostics. But not focused on other link functions and pearson residual form like SPR and APR. There are various link functions such as identity, inverse and log, and diagnostic methods are Cook`s distance 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 SPR and APR.

This paper is organized as follows: In Section 2 discussed methodology, Section 2.1 the G-PRM and its estimation methods, Section 2.2 presents the G-PRM residuals with derivation of standardized and adjusted pearson residuals, Section 2.3 describe the influence diagnostics methods in G-PRM. In Section 3 defined a Monte Carlo simulation, Section 3.1 a Simulation design and Section 3.2 present a simulation result. In Section 4 present an application: Reaction rate dataset. Finally, Section 5 gives away conclusion of the research work.

II. Methodology

Gamma-pareto regression model and estimation methods

The probability density function of the gamma-pareto response variable y is given by Alzaatreh et al. (2012).

$$f(y;\alpha,\beta,\gamma)=\frac{\gamma^{-1}}{\beta^{\alpha}\Gamma(\alpha)}\left(log\left(\frac{y}{\gamma}\right)\right)^{\alpha-1}\left(\frac{y}{\gamma}\right)^{-\left(\frac{1}{\beta}+1\right)} \quad (2.1)$$

with α, β,γ> 0 and y > γ.

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

According to Hanum et al. (2016), Eq. (2.1) 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

$$f(y;\mu,\phi)=\frac{\gamma^{-1}}{(\mu\phi)^{\frac{1}{\phi}}\Gamma\left(\frac{1}{\phi}\right)}\left(log\left(\frac{y}{\gamma}\right)\right)^{\frac{1}{\phi}-1}\left(\frac{y}{\gamma}\right)^{-\left(\frac{1}{\mu\phi}+1\right)} \quad (2.2)$$

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 $x_{i1},x_{i2},\dots,x_{ip}$ represent the p non-stochastic regressors. According to Hanum et al. (2016), link function of the G-PRM for the mean of the given response variable y is given by

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

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 $g(\mu_i)=\eta_i=\log\left(X_i^T\beta\right)$.

Finding the likelihood function's derivative with respect to β_j is the first step in estimating the parameter β_j using maximum likelihood. By Eq. (2.2)

$$\frac{\partial l}{\partial\beta_j}=U_j=\sum_{i=1}^{N}\left[\frac{\partial l_i}{\partial\beta_j}\right]=\sum_{i=1}^{N}\left[\frac{\partial l_i}{\partial\tau_i}\frac{\partial\tau_i}{\partial\mu_i}\frac{\partial\mu_i}{\partial\beta_j}\right] \quad (2.3)$$

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 β_j in GLM Gamma-Pareto is

$$\frac{\partial l}{\partial\beta_j} = U_j = \sum_{i=1}^{N} \alpha^{-1}\beta^{-2}\left(log\left(\frac{y_i}{\gamma}\right) - \mu_i\right)\frac{\partial\mu_i}{\partial\eta_i}x_{ij} \qquad (2.4)$$

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 β_j is not in close form.

Iterative weighted least squares (IWLS) were proposed by (Dobson et al, 2002) as a method for estimating β_j.

It's the IWLS.

$$X^TWX\,b^{(m)} = X^TWz$$

$$b^{(m)} = (X^TWX)^{-1}(X^TWz) \qquad (2.5)$$

And now, Using W and $var(U_j)$ for G-P and obtained the iteration for β_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, residuals

Many types are available of GLM residuals in literature (Hardin and Hilbe, 2012). But we used only pearson residual and its types standardized pearson residual and adjusted pearson residual form.

The Pearson residuals in the G-PRM are given by

$$R_{pr} = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i}} = \frac{y_i - \hat{\mu}_i}{\hat{y}_i} \qquad (2.6)$$

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 function $g(\mu_i) = \eta_i = \log(X_i{}^T\beta)$ are fitted model $\hat{\mu}_i = \eta_i$.

The standardized Pearson residuals are present by using Eq. (2.6)

$$R_{spr} = \frac{R_{pr}}{\sqrt{\phi(1-h_{ii})}} \qquad (2.7)$$

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

The adjusted Pearson residuals is defined by using Eq. (2.6)

$$R_{apr} = \frac{R_{pr} - \hat{E}(R_{pr})}{\sqrt{\hat{V}(R_{pr})}} \qquad (2.8)$$

The adjusted Pearson residuals approximately follow a normal distribution (Cordeiro, 2004).

2.3. Influence diagnostics, Gamma-Pareto regression model

A bad value in the LM has an impact on the model estimates and inferences, as noted by (Atkinson, 1981). 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 pearson 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. (Pregibon, 1981) 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.

$$DFFITS_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\phi}_i\,h_{ii}}} \qquad (2.9)$$

Eq. (2.9) can also be written as

$$DFFITS_i = \frac{\hat{W}_{ii}^{\frac{1}{2}}x_i^T(y_i - \hat{\mu}_i)}{\sqrt{\hat{\phi}_i\,h_{ii}}} \qquad (2.10)$$

$$DFFITS_i = |t_i|\sqrt{\frac{h_{ii}}{1-h_{ii}}} \qquad (2.11)$$

The DFFITS for standardized pearson residuals used Eq. (2.7)

$$DFFITS_i = |t_i|\sqrt{\frac{h_{ii}}{1-h_{ii}}} \qquad (2.12)$$

$$t_i = R_{spr}\sqrt{\frac{n-p-1}{n-p-(R_{spr})^2}} \qquad (2.12.1)$$

The DFFITS for adjusted pearson residuals used Eq. (2.8)

$$DFFITS_i = |t_i|\sqrt{\frac{h_{ii}}{1-h_{ii}}} \qquad (2.13)$$

$$t_i = R_{apr}\sqrt{\frac{n-p-1}{n-p-(R_{apr})^2}} \quad (2.13.1)$$

where $h_{ii} = diag(H)$ is the *i*th hat matrix H diagonal element for the G-PRM McCullagh and Nelder (1989), $H = \widehat{W}^{\frac{1}{2}}X\left(X^T\widehat{W}X\right)^{-1}X^T\widehat{W}^{\frac{1}{2}}$. 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 (Chatterjee and Hadi, 1988). In the case of large data sets, an observation is considered influential when the ith value of DFFITS exceeds $2\sqrt{\frac{p+1}{n}}$ (Belsley et al, 1980). 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. Cook (1977) 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

$$CD_i = \frac{(\hat{\beta}-\hat{\beta}_i)^T X^T\widehat{W}X\,(\hat{\beta}-\hat{\beta}_i)}{(P+1)\hat{\phi}} \quad (2.14)$$

After simplification, Eq. (2.14) becomes

The cook`s distance for standardized pearson residuals used eq. (2.7)

$$CD_{(i)R_{spr}} = \frac{(R_{spr})^2}{(P+1)}\frac{h_{ii}}{(1-h_{ii})} \quad (2.15)$$

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

$$CD_{(i)R_{apr}} = \frac{(R_{apr})^2}{(P+1)}\frac{h_{ii}}{(1-h_{ii})} \quad (2.16)$$

According to Ullah and Pasha (2009), 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. Cook (1977) 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 Hardin and Hilbe (2012). 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.

III.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 SPR and APR 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 pearson 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 Hanum et al. (2016) 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 $\hat{y_i} = E(y_i) = (\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+\beta_3X_{i3})$ identity, $\hat{y_i} = E(y_i) = (\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+\beta_3X_{i3})^{-1}$ inverse and $\hat{y_i} = E(y_i) = \log(\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+\beta_3X_{i3})$ log link function, $i = 1,2,\ldots,n$ is mean function and ϕ is dispersion parameter ϕ = 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 $\beta_0 = 0.05, \beta_1 = 0.0025,\ \beta_2 = 0.005\ and\ \beta_3 = 0.0001$ (Amin et al, 2016 and 2017) and γ is minimum value of response variable. In this case, the design matrix X has no influential points of sample sizes n = 25, 50, 100 and 200 generated as $X_i \sim N(-1,1),\ \ i = 1,2,\ldots,n;\ and\ j = 1,2,3,$ and then we make 10th, 15th, 20th, 25th, points in the X as $X_{ij} = \alpha_0 + X_{ij},\ \ i = 10,15,20\ and\ 25, and\ j = 1,2,3,$ where $\alpha_0 = \overline{X}_j$ +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.

Simulation results

Tables 1–8 present the simulation results of the Gamma-Pareto regression influence diagnostics under various link functions. Here is a summary of the simulation's findings.

1. when dispersion level is ϕ=0.04 the Cook's distance and DFFITS diagnostic methods performance with log link function for both SPR and APR are larger diagnostics influential observations detection percentages as compare to the inverse and identity link functions. All sample sizes yield the same results as those mentioned above. These results are also verified and prominent with index plot in figure 1.

2. It is interesting to note that, when dispersion level is increase as ϕ=0.11 the Cook's distance and DFFITS diagnostic methods performance with inverse link function for both SPR and APR are larger diagnostics influential observations detection percentages as compare the identity and log link functions for all sample sizes. These results are also verified and prominent with index plot in figure 2.
3. While dispersion level is further increase as ϕ=0.17 and ϕ=0.33 the results is almost same as when dispersion is ϕ=0.11 in favor of inverse link function. These results are also verified and prominent with index plot in figure 3 and 4 respectively.
4. when dispersion level is ϕ=0.67 the Cook's distance and DFFITS diagnostic methods performance with inverse, identity and log link function for both SPR and APR are almost same diagnostics influential observations detection percentages are true for all sample sizes. These results are also verified and prominent with index plot in figure 5.
5. For dispersion level is ϕ=2 the Cook's distance and DFFITS diagnostic methods performance with identity link function for both SPR and APR 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. These results are also verified and prominent with index plot in figure 6.
6. For large dispersion level are ϕ=5, 10 the Cook's distance and DFFITS diagnostic method's performance with inverse link function for both SPR and APR are larger diagnostics influential observations detection percentages as compare to the identity link and log link functions for all sample sizes. 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 pearson residuals for the detection of influential observations when ϕ =0.04

Sample size *n*	Influential observation	Cook's Distance						DFFITS					
		ϕ =0.04						ϕ =0.04					
		Inverse		Identity		Log		Inverse		Identity		Log	
		SPR	APR	SPR	APR	SPR	APR	SPR	APR	SPR	APR	SPR	APR
25	**10**	84.9	80.7	82.8	77.5	83.5	78.8	84.1	79.9	84.7	80.6	83.9	79.3
	15	74.4	70.8	73.8	68.9	72.3	68	75.4	71.7	73.5	69.6	73.5	70.2
	20	63	56.9	61.4	57.4	60.5	57.5	61	56.7	60.5	57	60.2	56.3
	25	47.6	44.2	47.5	44.8	45.5	43	45.1	42.4	47.2	44.5	47.1	44.3
50	**10**	87.3	81.8	87.4	83.1	88.7	83.9	88.1	84.8	88.7	85.2	89.3	84.9
	15	82.8	79.4	84.1	80.2	83.8	80.9	84.3	81.3	83.1	80.3	84.1	81.4
	20	77.2	74.6	78.9	76.8	78.7	76.8	77.9	74.9	78.3	76.2	76.6	73.7
	25	70.6	67	71.3	69.1	73.6	71.6	72.2	69.6	71.5	69.2	74.4	71.5
100	**10**	91.7	89.3	91.7	88.1	91.2	87.9	92.8	89.7	90.8	87.9	91.7	87.8
	15	87.7	85.4	88	85.6	87.6	84.7	88.6	87	88.5	85.9	88.2	85.9
	20	83.9	81.9	85.9	83.5	84.3	82.3	85.4	83.6	83.7	81.4	86.3	84.4
	25	79.6	77.9	81.7	80.2	83.1	81.1	82.8	81.1	85	83.2	80.2	78.2
200	**10**	92.9	90.8	95	93.2	93.2	91.2	92.9	90.9	93.4	91.1	94.2	91.9
	15	89.9	88.3	93.2	92.2	91	89.6	91.1	89	92	90.7	91.1	89.2
	20	88.8	86.6	90.5	89.5	90.3	89.4	89.4	88.1	90.8	89.8	90.2	88.6
	25	88.3	86.5	88.2	87	89.4	87.7	88.6	88.2	87.9	87.1	88.9	87.2

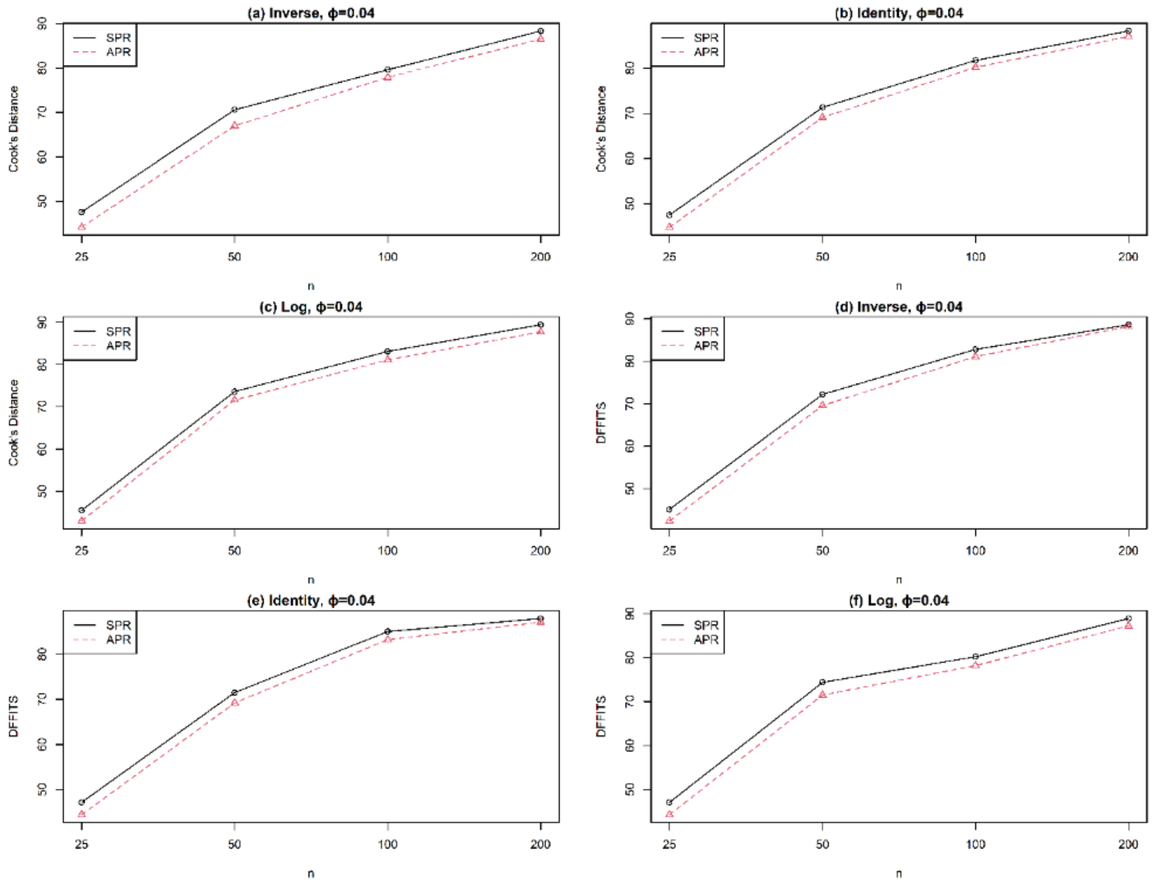

Fig.1. Index plots of Cook`s distance and DFFITS under different link functions with ϕ =0.04

Table 2. Performance of different link functions with standardized and adjusted pearson residuals for the detection of influential observations when ϕ =0.11

Sample size n	Influential observation	Cook's Distance						DFFITS					
		ϕ =0.11						ϕ =0.11					
		Inverse		Identity		Log		Inverse		Identity		Log	
		SPR	APR	SPR	APR	SPR	APR	SPR	APR	SPR	APR	SPR	APR
25	**10**	84	79.9	86.6	81.8	83.5	77.8	81.7	75.7	85.6	80.2	85.1	79.8
	15	75	70.4	73.3	69.8	74.2	70.9	74.2	70.1	75.7	72	74.1	69.8
	20	60.8	57.6	60.6	56.7	60.4	57	63.7	59.7	59	54.8	60.4	57
	25	47.1	44.7	47.6	43.5	44	42	44	41.3	46.4	43.8	48.7	45.7
50	**10**	88.3	84	88.9	84.9	89.4	85.7	88.1	84.1	90.1	85.8	88.7	84.6
	15	84.2	80.2	81.5	77.9	82.7	79.3	84.2	81.5	81	77.7	81.7	77.5
	20	78	75.7	76.3	73.8	76.2	73.9	78.8	76.2	78.5	76.5	78.5	75.7
	25	69.8	67.2	72.1	69.4	69.4	67.9	70.5	68.5	73.6	71.5	71.8	69.1
100	**10**	90.9	87.8	92.3	88.7	91.7	89	91.8	89.3	90.6	86.8	90.8	87.8
	15	88.4	85.7	88.7	86.2	87.8	84.6	88.8	85.3	87.9	85.7	87.8	84.9
	20	87.9	85.4	85.1	83.5	86.5	84	83.6	82	83.9	82	84.1	82.3
	25	81.1	79.7	83.6	81.7	81.9	80.8	82.8	80.4	82.1	80.3	83.6	81.7
200	**10**	93.2	90.7	92.7	91.3	94.2	92.1	93	90.2	92.4	90.5	91.6	89.4
	15	92	90.4	92.7	91.5	92.9	90.7	89.7	88.6	91.2	89.3	90.7	89.5
	20	90.5	89	88.5	87.1	89.7	88.2	90.5	89.2	91	89.9	90.7	89.3
	25	87.1	86.1	85.1	84.2	87.3	85.1	87.6	86.6	88	86.3	87.3	86.6

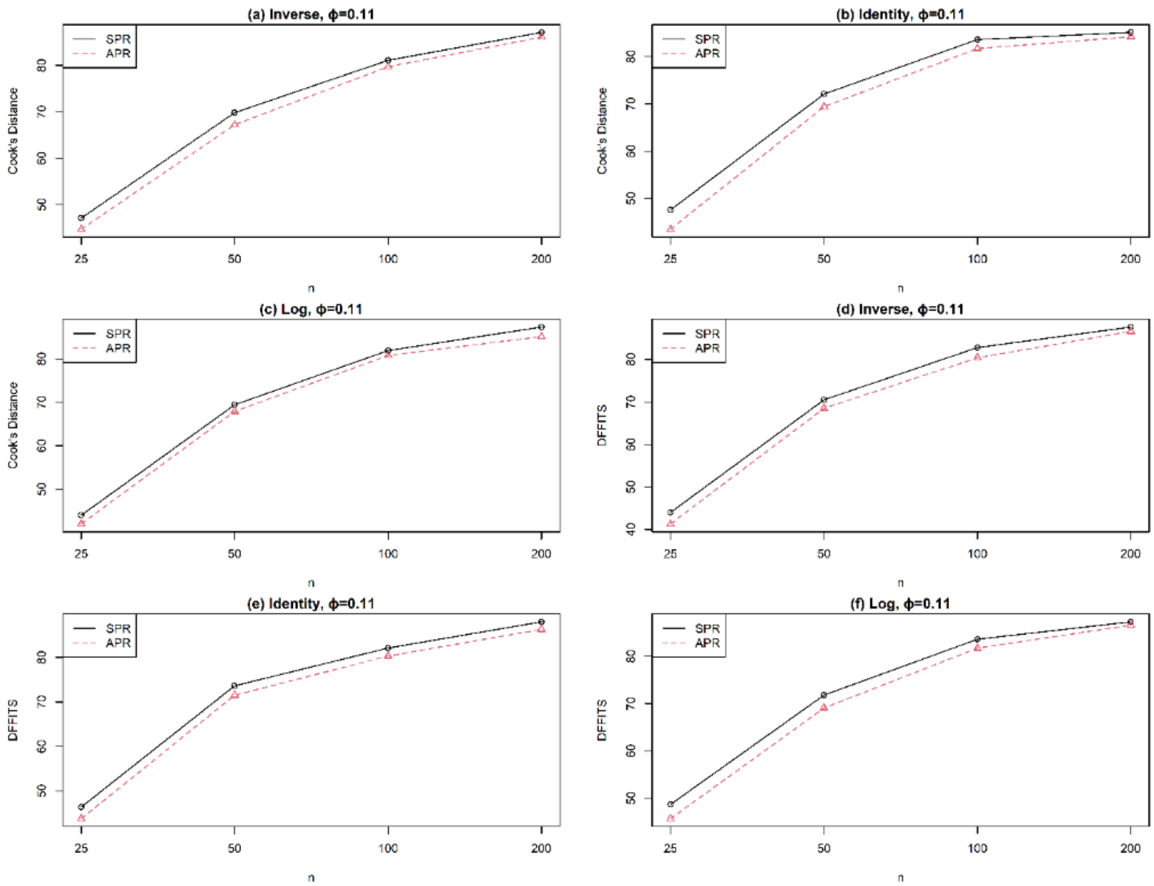

Fig. 2. Index plots of Cook`s distance and DFFITS under different link functions with ϕ =0.11

Table 3. Performance of different link functions with standardized and adjusted pearson residuals for the detection of influential observations when ϕ =0.17

Sample size n	Influential observation	Cook's Distance						DFFITS					
		ϕ =0.17						ϕ =0.17					
		Inverse		Identity		Log		Inverse		Identity		Log	
		SPR	**APR**	**SPR**	**APR**	**SPR**	**APR**	**SPR**	**APR**	**SPR**	**APR**	**SPR**	**APR**
25	**10**	87.4	80.9	84.1	79.9	84.6	80.1	84.6	79.4	82.5	75.8	86.8	81.7
	15	77.5	68.6	72	68.6	73	69.1	74.6	70.1	76.3	71.9	75.1	71.7
	20	62.9	57	61.5	57.9	62.6	59.3	60.9	58	63.1	59.4	62.1	59.5
	25	49.0	42.2	48.5	45.5	46.3	42.6	48	44.8	48.9	45.8	46.4	42.9
50	**10**	86.6	83	86.8	83.1	87.6	83.7	88.2	84.5	87.2	81.7	89.1	86
	15	83.9	81.3	84.5	81.5	81.7	78.9	83.7	80.3	80.4	77.8	85.2	82.1
	20	79.3	76.8	75.7	72.8	75.7	72.7	74.7	71.2	76.7	73.5	76.1	72.8
	25	70	67	72.7	70.1	72.3	69.9	71.8	69.5	74.8	72.9	71.5	68.7
100	**10**	91.4	88.2	92.2	88.9	90.2	87.7	91.2	88.3	90.9	88.3	90.7	87.5
	15	86.5	84.7	89.4	87.2	87.8	84.5	88.8	85.6	88.1	86.1	89.3	86.7
	20	83.6	82.4	86.4	84.6	83.7	81.3	82.8	80.3	83.2	81.3	84.1	82.3
	25	80.9	78.9	83.4	81.5	82.8	81.1	81	79.8	82.5	80.1	80.8	79.6
200	**10**	92.4	91	93.2	91.4	90.8	88.8	92.1	90.1	93.5	91.7	93.1	91.5
	15	91.8	90.5	93.5	91.9	91.7	90.8	90.9	88.9	92	89.6	92.8	90.8
	20	88.8	87.2	91	89.7	89.1	87.5	88.9	87.3	90.8	89.2	88.4	86.3
	25	87.2	86	87.9	86.6	88	86.8	89.1	87.5	89.9	88.8	88.3	87.6

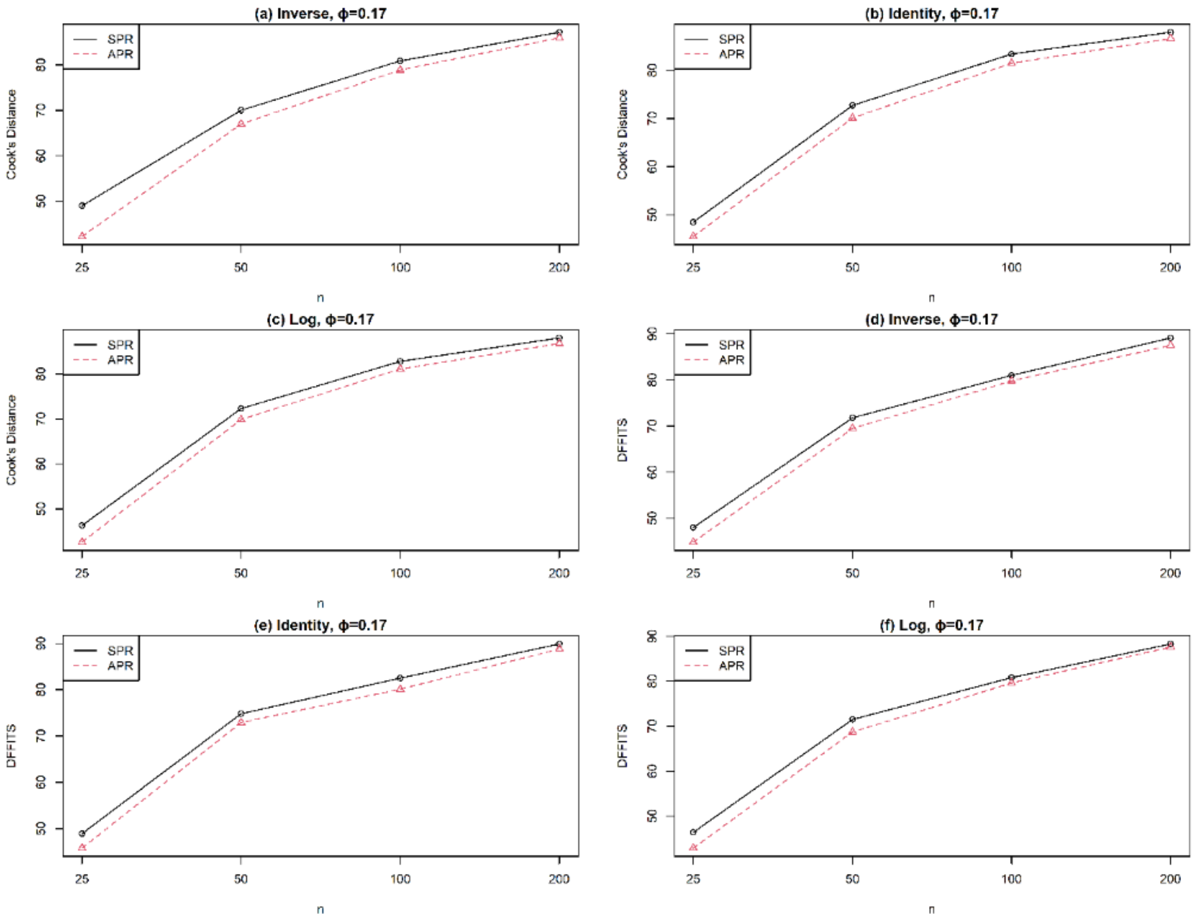

Fig. 3. Index plots of Cook`s distance and DFFITS under different link functions with ϕ =0.17

Table 4. Performance of different link functions with standardized and adjusted pearson residuals for the detection of influential observations when ϕ =0.33

Sample size n	Influential observation	Cook's Distance						DFFITS					
		ϕ =0.33						ϕ =0.33					
		Inverse		Identity		Log		Inverse		Identity		Log	
		SPR	APR	SPR	APR	SPR	APR	SPR	APR	SPR	APR	SPR	APR
25	**10**	88.8	79.2	84	78.5	84.5	79.2	84.9	78.8	84.3	79.8	84.6	80.2
	15	74.8	70.8	72.8	69.4	74.2	68.9	73.4	69.2	73	69.1	70.5	66.8
	20	56.5	52.8	59.9	55	62.8	59.1	59.6	56.4	60.4	57.1	63.6	61
	25	47	44.9	47.3	44.6	47.2	45	46.5	43.7	46.5	44.2	45.7	42.5
50	**10**	86.6	82.8	88.4	84.7	89.5	85.9	89.4	85.3	89.8	85.8	90.1	85.7
	15	83.6	80.9	82.7	80.2	83.6	80.7	84.8	82	82.4	80.1	81.9	78.9
	20	77.4	74.3	78.9	76.3	79.6	77.8	77.3	73.5	78.6	76.6	77.8	74.5
	25	70.7	68.9	72.4	69.4	69.5	66.5	71.2	68.7	70.2	67.6	74.1	72
100	**10**	92.3	89.4	91	87.5	92.4	88.8	91.4	88.6	91.3	88.8	91.8	88.7
	15	88.1	86.1	89.5	87.3	87.1	85.1	89.6	87.1	88.2	85.4	86.5	84.5
	20	84.4	82.5	83.6	81.4	83.6	81.9	84.8	83.5	85.9	83.8	86.1	84.1
	25	80.7	78.4	82.8	81.2	82.4	80.7	82	80.4	82.2	80.4	81.7	80.1
200	**10**	93.9	91.3	94	91.7	92.6	90.2	93.1	89.4	93	90.7	94	91.2
	15	90.9	89.3	91	89	90.7	88.8	91.8	90.3	90.5	88.8	91	89.4
	20	91.7	90	90.5	88.9	89.8	88.5	89.7	88.3	88.6	87	90.6	89.3
	25	87	85.3	88.1	87.2	89.3	87.9	86.6	85.2	89.9	88.3	88.6	87.6

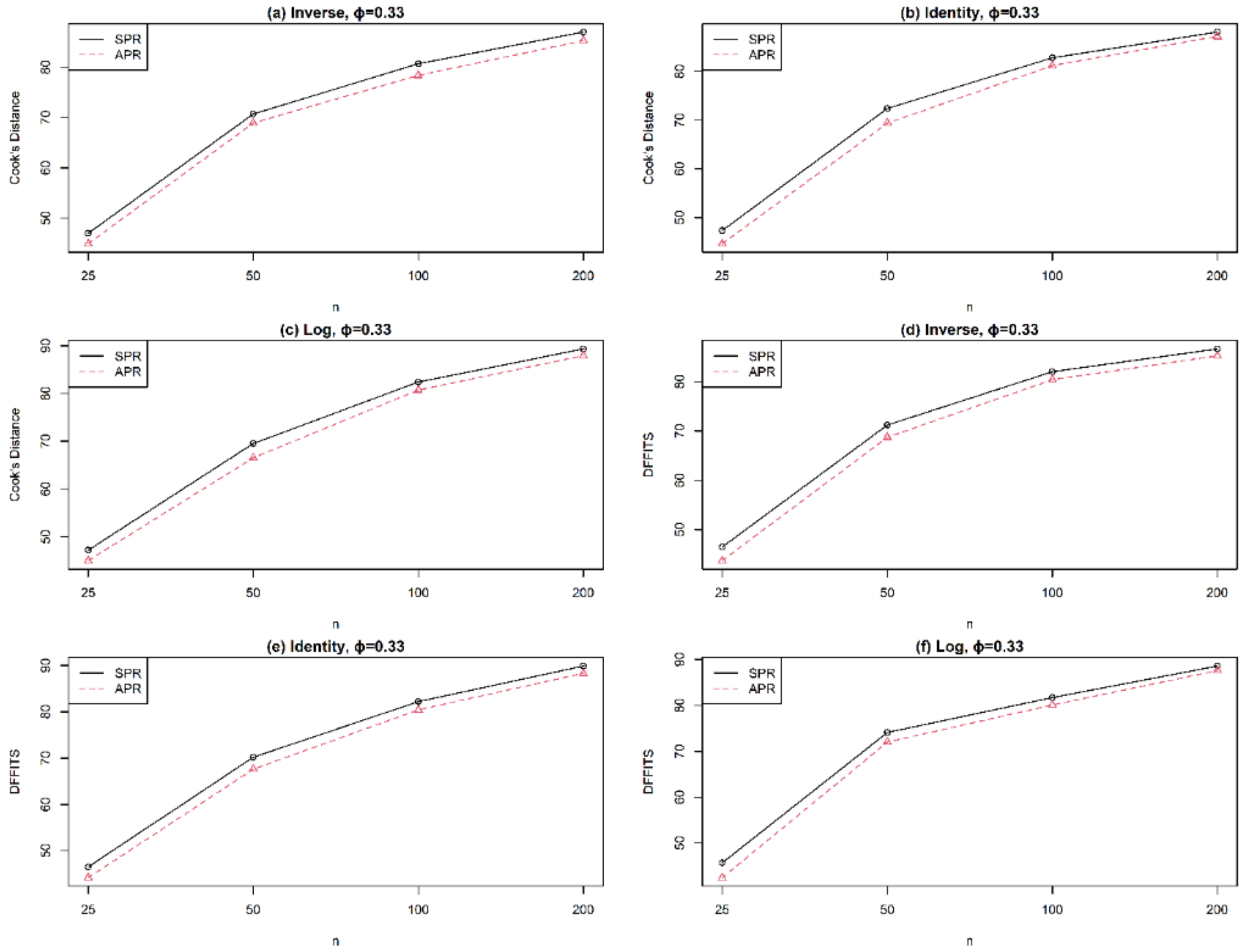

Fig. 4. Index plots of Cook`s distance and DFFITS under different link functions with ϕ =0.33

Table 5. Performance of different link functions with standardized and adjusted pearson residuals for the detection of influential observations when ϕ =0.67

Sample size n	Influential observation	Cook's Distance						DFFITS					
		ϕ =0.67						ϕ =0.67					
		Inverse		Identity		Log		Inverse		Identity		Log	
		SPR	APR	SPR	APR	SPR	APR	SPR	APR	SPR	APR	SPR	APR
25	**10**	83.7	78.3	84.5	79.1	86.3	80.7	85.4	80.2	83.9	79	83.8	79
	15	73.6	69	74.2	69.6	73.2	69.2	73.5	69.8	73.3	69.6	74.3	70.1
	20	59.2	55.9	60.9	56.8	59	55.8	61.3	56.7	62	56.5	60.4	56.6
	25	50.2	47.1	45.9	43	50.8	47.2	45.9	43.4	48.7	46	47	44.2
50	**10**	89.2	84.6	88.8	84.8	86.9	83.2	90	85.5	86.7	82.9	88.2	84.7
	15	81.2	77.1	83.6	81.3	83.4	80.1	80.5	78.1	82.9	79.3	82.6	79.8
	20	77.7	73.9	77.3	74.9	76	73	78.2	75.6	78.9	75.5	79.1	76.1
	25	71.9	70.3	72.3	69.5	71.4	69.2	69.4	66.7	71.2	70	68.6	67
100	**10**	90.5	87	92.6	89.5	90.4	88.3	93.7	90.7	92	89.3	91	88
	15	88.6	86.9	88.2	86.1	89.5	86.5	89	86.7	88.1	85.9	91.1	88.2
	20	84.5	82.8	86.1	83.6	84.9	83	84.1	82.2	85	82.8	87.3	85.7
	25	83.8	82.8	81.7	79.6	83	81.8	81	79.1	81.7	80.3	82.2	79.9
200	**10**	93.3	90.6	93.9	91.5	94.2	91.7	93.1	91.6	93.3	91.8	92.7	90.2
	15	90.9	89.4	90.6	89	91.4	89.3	92.5	91.1	91.2	89.2	89.9	88.3
	20	90.3	89	90.2	88.6	90	89.2	88.8	87.4	90.2	88.9	89.6	88.6
	25	87.5	85.7	88.9	88	89.4	88.2	88	86.7	87.7	86.6	88.2	87.3

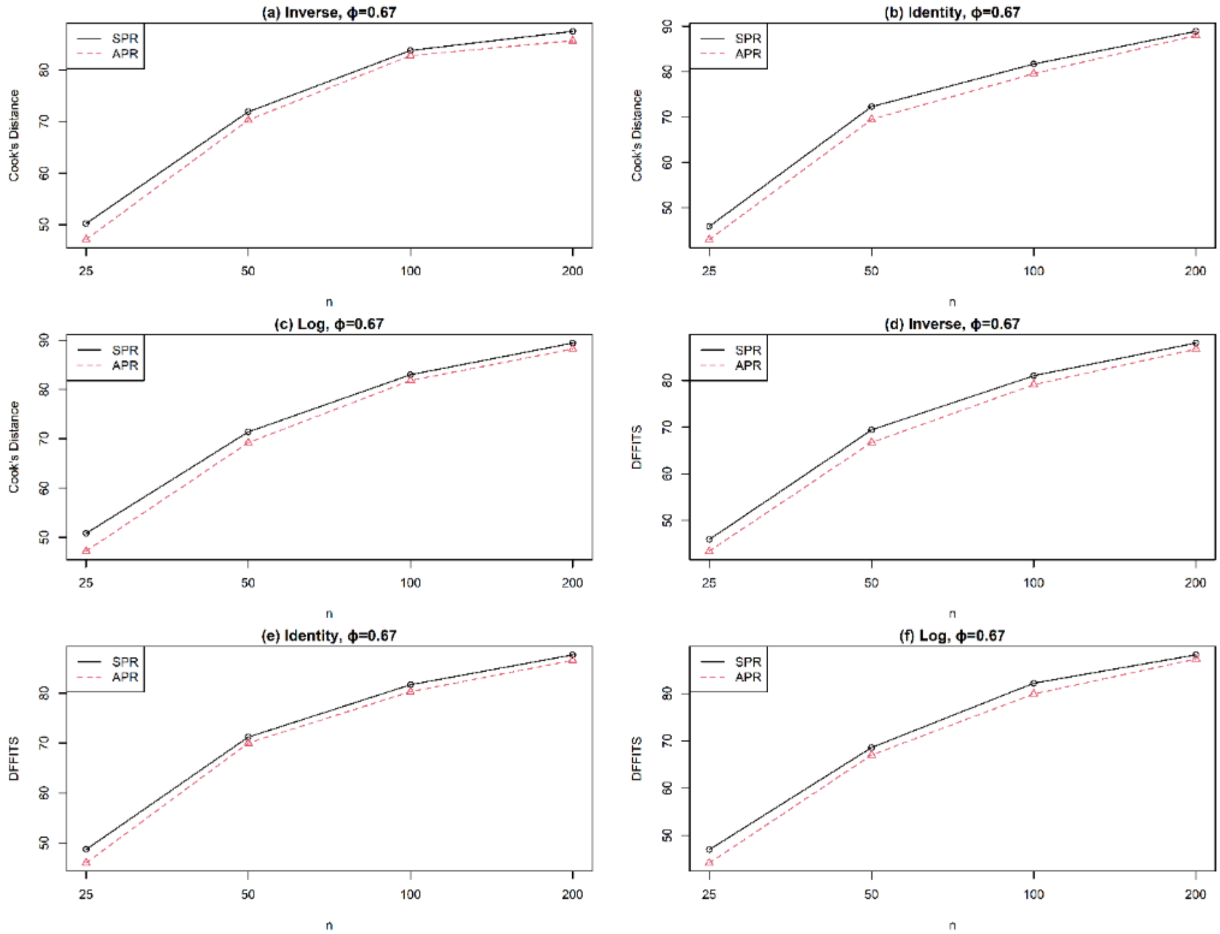

Fig. 5. Index plots of Cook`s distance and DFFITS under different link functions with ϕ =0.67

Table 6. Performance of different link functions with standardized and adjusted pearson residuals for the detection of influential observations when ϕ =2

Sample size n	Influential observation	Cook's Distance						DFFITS					
		ϕ =2						ϕ =2					
		Inverse		Identity		Log		Inverse		Identity		Log	
		SPR	APR	SPR	APR	SPR	APR	SPR	APR	SPR	APR	SPR	APR
25	**10**	85.2	80.7	84.1	80.3	83.6	78.6	84	78.9	82.5	78.9	85.4	80.6
	15	75	70.1	73.3	69.4	76.2	72.4	75.2	72.3	71.9	68.2	75	68.3
	20	60.1	56.6	59.8	56.8	59.3	55.9	59.9	56.3	60.7	57.8	60.9	56.4
	25	46	43.4	47.1	44	46.5	42.5	47.5	45.7	48.3	45.4	48	47.8
50	**10**	89.5	85	89.2	84.8	86.2	82.7	87.6	83.5	88	84.2	89.1	85.5
	15	82.1	78.4	83.5	80.5	85.3	82.1	83.2	80.9	84.6	81.2	83	79
	20	77.6	74.5	76.8	74.8	79.7	76.7	79.1	76.3	77.4	75.1	78.9	75
	25	72.8	70.1	72.9	70.6	72.1	69.9	73	70.5	73.7	70.5	75.1	70.9
100	**10**	92.7	90.5	91.7	89.5	90.4	87.3	90.3	87	89.9	86.7	90.4	87.8
	15	88.4	85.5	90	87.5	87.8	85.6	88.2	86.3	87.4	84.9	89.1	87
	20	85.2	83.2	85.2	83.6	86.8	85	85.2	83.1	84.2	82.4	83.9	82
	25	82.2	80.3	80.7	79	81.2	79	81.3	80	82	80.9	82.1	79.3
200	**10**	93	90.6	94.4	91.8	93.3	90.5	95.3	93.2	93	90.4	93	90.3
	15	91.6	89.8	92.7	91	90.7	89.5	92.5	90.9	90.8	89.3	90.7	89.6
	20	90.4	89.6	89.7	88.2	90.8	89.5	89.4	88.1	89.6	88.4	89.9	88.5
	25	87.1	85.1	88.7	87.5	88	86.2	89.9	88.9	88.3	87.4	86.3	85

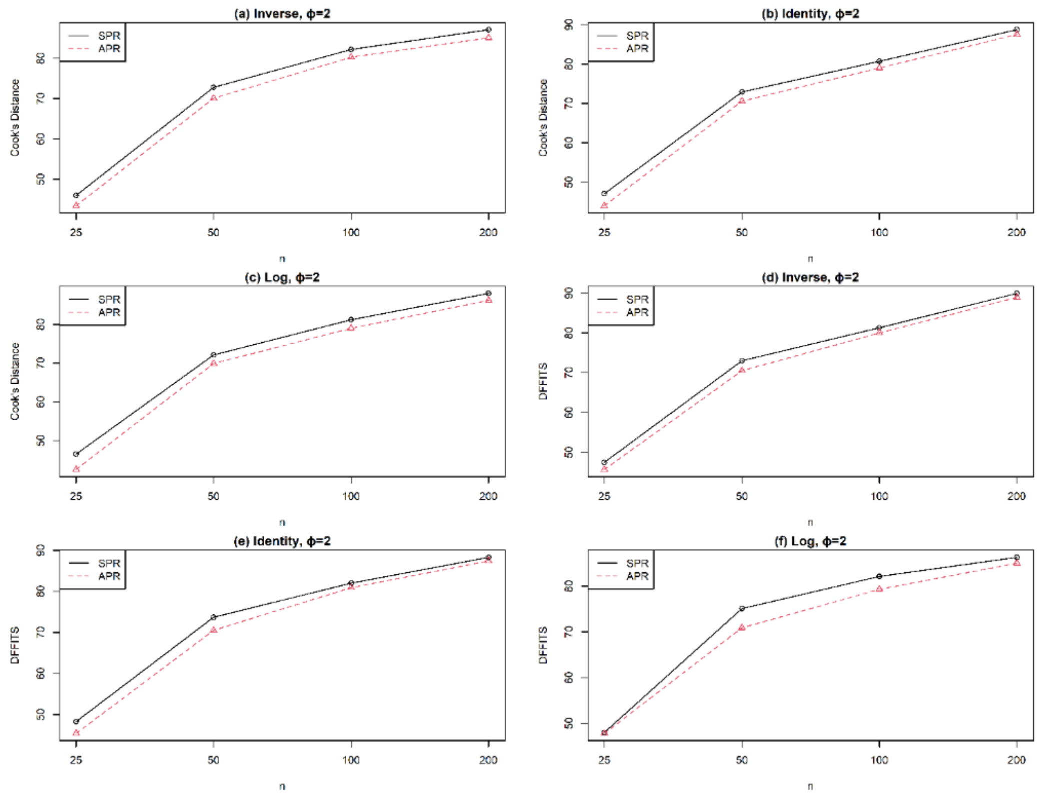

Fig. 6. Index plots of Cook`s distance and DFFITS under different link functions with ϕ =2

Table 7. Performance of different link functions with standardized and adjusted pearson residuals for the detection of influential observations when ϕ =5

Sample size n	Influential observation	Cook's Distance						DFFITS					
		ϕ =5						ϕ =5					
		Inverse		Identity		Log		Inverse		Identity		Log	
		SPR	APR	SPR	APR	SPR	APR	SPR	APR	SPR	APR	SPR	APR
25	**10**	82.1	76.7	84.8	79.6	84	78.5	84.4	79.2	85.2	79.6	82.4	77.1
	15	75.7	70.4	76.1	71.7	73	68.8	74.4	70.7	73.4	69.9	73.5	69.9
	20	61.7	58.1	60	57.1	63.5	58.9	61.8	57.9	61.3	58.1	60.8	57.2
	25	48.6	45.1	50.3	46.7	45.8	42.5	44.8	41.6	47.2	44	46.1	42.5
50	**10**	89.4	85.4	88.3	84.8	88.3	85	90.2	85.7	88.2	83.4	90.9	87.7
	15	83.4	80.8	81.8	78.6	85.1	82.5	84.5	81.4	81.8	78.4	82.9	80
	20	77.9	74.7	75.8	72.9	75	71.6	78	74.4	77.2	74.7	77.1	74.5
	25	72.7	69.9	70.6	68.7	72.8	70.9	74.5	71.1	72.6	70.1	74	72.2
100	**10**	92.7	89.2	89.2	86.7	91.7	88.7	90.2	87	90.5	88	90.1	86.7
	15	89	87	87.2	85	89.2	87.1	87.9	85.5	87.9	86	87.6	85.5
	20	85.3	83.5	83.5	82.1	84.3	82.3	84.8	82.7	85.4	84	86	83.4
	25	82.2	80.7	82.9	80.5	82.7	81.4	84.3	81.7	83.3	81.2	81.9	80.2
200	**10**	94.3	91.6	92.9	90.8	94	92.1	91	90.1	93.7	91.1	92.2	89.6
	15	91.7	89.8	89.8	88.4	93.1	91.1	91	89.5	89.8	87.9	92	90.3
	20	89.2	88.5	88.7	87.4	90.5	89.1	91.6	89.9	89.4	87.6	90.4	89.2
	25	88.4	87.7	88.2	86.6	86.6	85.3	89.8	88.2	89.4	88.4	88.2	86.7

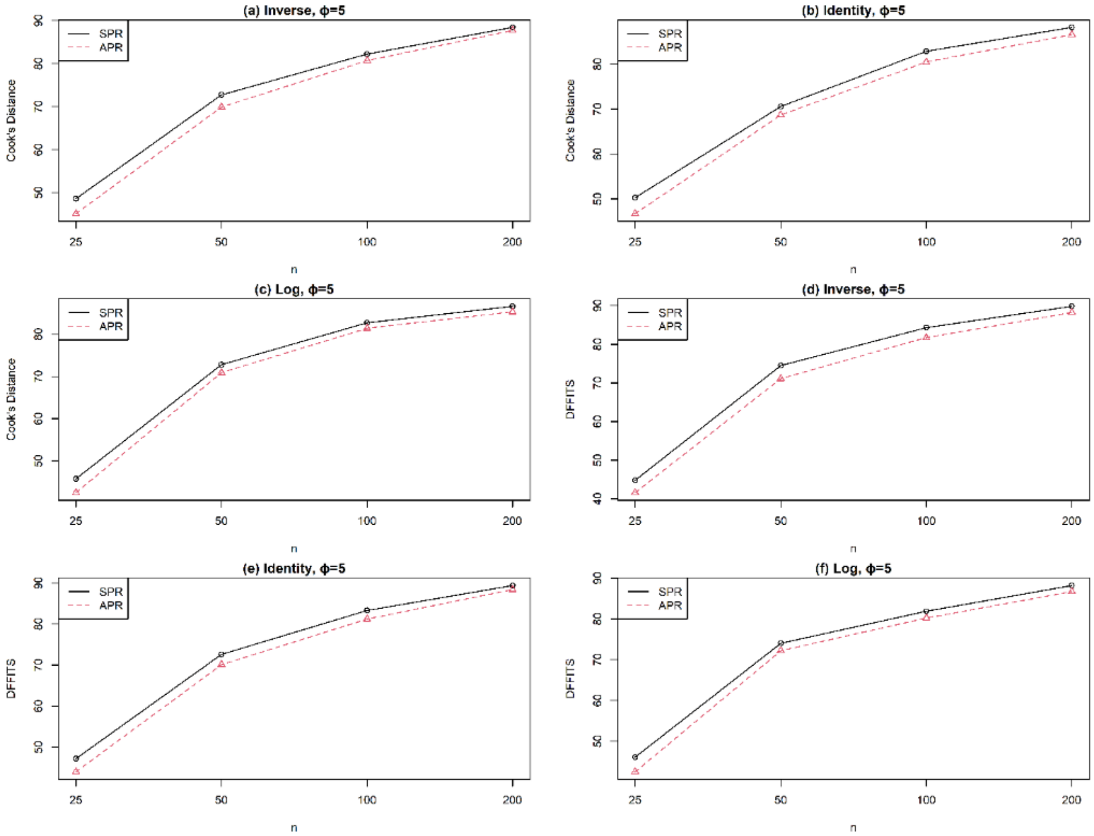

Fig. 7. Index plots of Cook`s distance and DFFITS under different link functions with ϕ =5

Table 8. Performance of different link functions with standardized and adjusted pearson residuals for the detection of influential observations when ϕ =10

Sample size n	Influential observation	Cook's Distance						DFFITS					
		ϕ =10						ϕ =10					
		Inverse		Identity		Log		Inverse		Identity		Log	
		SPR	APR	SPR	APR	SPR	APR	SPR	APR	SPR	APR	SPR	APR
25	**10**	93.8	91.9	82.4	76.2	84.7	79.3	85	81.9	82.9	79.4	82.5	77.4
	15	91.7	89.7	75.7	72.7	75.2	71.3	72.9	68.4	73.2	68.5	75	71.3
	20	87.8	85.4	60.7	57.4	60.4	57.1	61.4	58.1	61.8	58.3	60.1	55.4
	25	89.8	88.7	47.8	45.2	47.7	44.6	46.4	43.7	46.5	43.3	50	47
50	**10**	94.9	93.3	88.1	84.4	88.1	84.9	88.6	84.1	88	84.8	89.2	84.1
	15	91.2	89.8	82.5	79.1	83	79.5	83	79.9	81.7	79.5	81.8	79.1
	20	91	89.3	75.9	73.1	78.8	75.9	77.2	74.1	79.3	76.1	76.5	74.3
	25	89.9	89.4	73	70.8	70.8	68.1	69.3	66.7	72.9	70.3	69.7	66.1
100	**10**	93.6	90.9	90.9	87.7	91.3	88.3	92	88.6	90.2	86.8	91.8	88.4
	15	92.4	91	88.4	86.1	87.6	85.8	89.6	88	87.2	85.2	89.3	87.1
	20	91.7	90.6	84.9	82.6	84.1	82.4	85.4	83.8	88.1	85.9	86.2	84
	25	90.3	89.4	82	80.3	79.4	77.4	80.8	78.5	81.7	80	83.3	81.3
200	**10**	93.3	91.7	93	90.6	92.5	90.1	94.7	91.6	93.2	90.7	94.8	93.2
	15	92.6	91.4	91.9	89.9	91.2	89.4	90.2	88.8	91.7	89.4	91.7	90.4
	20	92.1	91.4	90.3	89.4	88.6	87.4	90.3	89.4	89.9	88.4	89.8	89
	25	88.5	87.9	90.5	89.3	88	87.2	88.2	86.8	88.2	87	87.8	86.6

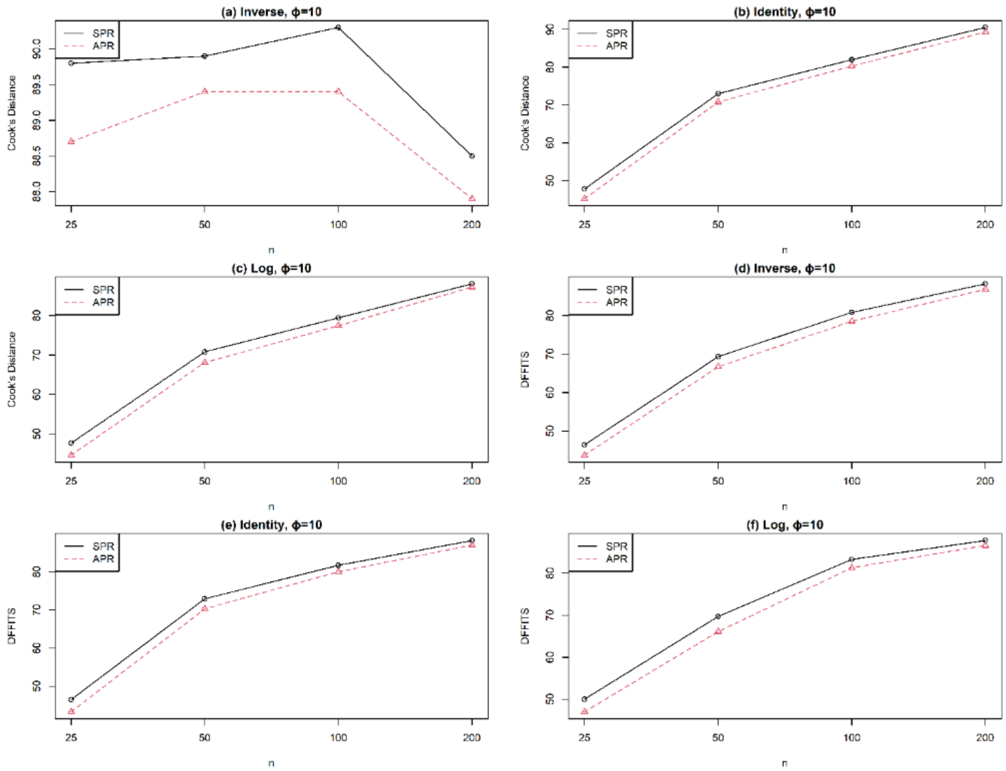

Fig. 8. Index plots of Cook`s distance and DFFITS under different link functions with ϕ =10

IV. 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. We applied reaction rate data set is taken from Huet et al. (2004) and (2006). Then Amin et al. (2019 a, b & 2022) and Yasin et al. (2022) utilized this data set. Data set consist of 24 observations and provide the reaction rate (y) as dependent variable. Three independent variables (p= 3) are used to speed up the reaction rate, which are partial pressure of hydrogen (x_1), partial pressure of n-pentane (x_2) and partial pressure of iso-pentane (x_3). As it is mentioned by Yasin et al. (2022), response variable follows a G-PD is required by following Hanum et al. (2016). However, because of the positively skewed trend of the dependent variable, this data set is not well fitted to the normal distribution. We found that the G-PRM fits this data set well based on the fitting test distribution the findings are shown 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 SPR and APR respectively. we observe that the Cook`s distance with SPR under inverse link function diagnosed 5,6,11 is influential observations. while, on the other hand the Cook`s distance with APR under inverse link function diagnosed 5,6 is influential observations. Similarly, we observe that the Cook`s distance with SPR under identity link function diagnosed only 22 is influential observation while, on the other hand the Cook`s distance with APR under identity link function does not diagnosed any influential observations. It is interesting to note that the Cook`s distance with SPR and APR under log link function diagnosed same observation 5,6,22,24 is influential observation. Now we discussed second diagnostic measure is DFFITS under different link functions and pearson residual form such SPR and APR. For DFFITS with SPR under inverse link function diagnosed only 10 is influential observation. while, on the other hand DFFITS with APR under inverse link function diagnosed 14,19 is influential observations. Similarly, for DFFITS with SPR under identity link function diagnosed only 19 is influential observation while, on the other hand the DFFITS with APR under identity link function does not diagnosed any influential observations. It is interesting to note that the DFFITS with SPR under log link function diagnosed observation 20,21,23 is influential observation. But DFFITS with APR under log link function diagnosed 22, 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 sixth 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 β_o and β_3 are impacted by this finding. The fifth 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 β_o and β_1 are impacted by this finding. Likewise, under various link functions 24, 22, 11 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 tenth observation holds the most influence. The G-PRM estimate of β_1 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.

Table 9. Goodness of fit distribution tests for Reaction Rate Data

Goodness of fit tests (GFT)		Probability Distribution						
		Gamma	Exponential	Gamma-Pareto	Weibull	Uniform	Normal	Log-Normal
Anderson-Darling	Statistic	0.2519	3.1872	0.3299	0.2943	0.9912	1.2462	0.8055
(AD)	P-value	0.7538	0.5881	**0.8103**	0.6288	0.2033	0.0027	0.0390
Cramer-von Mises	Statistic	0.0432	0.4570	0.2797	0.0521	0.2392	0.2127	0.1865
(CVM)	P-value	0.6259	0.0067	**0.6922**	0.4772	0.3143	0.0033	0.0051
Pearson chi-	Statistic	2.0000	14.880	24.774	6.0000	9.6522	10.667	17.995
square (PCS)	P-value	0.8491	0.0033	**0.9605**	0.3062	0.2071	0.0584	0.0949

Gamma-Pareto Distribution (G-PD).

Table 10. Detect influential points (IP) with Gamma-Pareto regression model (G-PRM), Residuals (Under different link functions with diagnostic methods)

IP detection methods	G-PRM		
	Inverse link function	**Identity link function**	**Log link function**
Index plots standardized Pearson residuals **(Cook's Distance)**	5,6,11	22	5,6,22,24
Index plots adjusted Pearson residuals **(Cook's Distance)**	5,6	--	5,6,22,24
Index plots standardized Pearson residuals **(DFFITS)**	10	19	20, 21, 23
Index plots adjusted Pearson residuals **(DFFITS)**	14, 19	--	22, 24

Influential points (IP), Gamma-Pareto regression model (G-PRM)

Table 11. The G-PRM summary with and without influential points. (Inverse link function)

Variables	Full data				After deleting, IP			
	Estimate	SE	Z	P-value	Estimate	SE	Z	P-value
Constant	-3.175	0.382	4.375	0.000	1.695	0.383	4.429	0.000
X_1	0.059	0.011	-3.039	0.006	-0.002	0.001	-2.705	0.014
X_2	-0.067	0.005	13.489	0.000	0.034	0.003	13.630	0.000
X_3	0.004	0.002	-12.333	0.000	-0.031	0.003	-11.723	0.000

Table 12. The G-PRM summary with and without influential points. (Identity link function)

Variables	Full data				After deleting, IP			
	Estimate	SE	Z	P-value	Estimate	SE	Z	P-value
Constant	0.178492	0.093265	1.913822	0.070063	0.1846	0.087571	2.108005	0.048538
X_1	0.000407	0.000219	1.85882	0.077832	0.000462	0.000223	2.074656	0.051847
X_2	-0.00109	0.000262	-4.17186	0.000471	-0.00118	0.000278	-4.24206	0.000441
X_3	0.003005	0.000756	3.97718	0.000742	0.002872	0.000786	3.651584	0.001697

Table 13. The G-PRM summary with and without influential points. (Log link function)

Variables	Full data				After deleting, IP			
	Estimate	SE	Z	P-value	Estimate	SE	Z	P-value
Constant	0.946834	0.343443	2.75689	0.012161	0.87562	0.330174	2.651997	0.015733
X_1	-0.00143	0.00075	-1.90964	0.070629	-0.0016	0.00073	-2.19714	0.040614
X_2	0.009494	0.001181	8.035421	0.710090	0.01065	0.001257	8.474713	0.002501
X_3	-0.01139	0.001656	-6.88204	0.611011	0.01181	0.001594	-7.41159	0.041191

Table 14. Absolute percentage relative change in the G-PRM estimates after deleting IP and R^2_{Efron} (Under different link functions)

Influential points	**Percentage relative change in the G-PRM estimates**				R^2_{Efron}**(0.7856)**
	β_0	β_1	β_2	β_3	
5	**39.257**	**79.556**	4.789	8.483	0.3723
6	**45.567**	6.406	2.906	**66.850**	0.1622
10	3.7850	**45.208**	4.621	9.012	0.8357
11	**24.858**	10.717	5.091	3.210	0.3701
14	0.5774	2.8475	0.385	3.875	0.9845
19	0.3876	0.8677	-0.384	0.783	0.1274
20	0.2158	0.0459	-0.123	0.943	0.2737
21	4.8348	12.478	0.784	3.478	0.3478
22	**55.2314**	0.0002	-0.001	0.003	0.6754
23	0.8934	0.0004	-0.001	0.002	0.7437
24	**62.1846**	0.0007	-0.001	0.002	0.8876

(0. 7856) represent the R^2_{Efron} of the full data
Influential points (IP), Gamma-Pareto regression model (G-PRM)

V. Conclusion

Influential observations influence the G-PRM estimates just like they do for other models. We estimate the G-PRM under different link functions. Therefore, in order to identify an appropriate diagnostic method and a link function in the G-PRM, we compare the performance of two influence diagnostic methods with different link functions in this study. We take into consideration the inverse, identity, and log link functions for these reasons. We diagnose the influential observations with taken into account link functions using the two influence diagnostic methods, namely Cook's distance and DIFFITS. The G-PRM diagnostic techniques' effectiveness with various link functions should be assessed. We employ a real application along with the Monte Carlo simulation. The results of the simulation indicate that the cook's distance and DFFITS with log link functions perform better than the inverse and identity link functions for small dispersion and all sample sizes. The cook's distance and DIFFITS with inverse link function perform better than the identity and log link function for large dispersion and small sample sizes. Similarly, the cook's distance and DFFITS with all link functions yield nearly identical results of influence diagnostics performance for large dispersion and all sample sizes.

Funding. No funding resource is available for this study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

1. Alzaatreh, A., & Ghosh, I. (2016). A study of the Gamma-Pareto (IV) distribution and its applications. Communications in Statistics-Theory and Methods, 45(3), 636-654.
2. Alzaatreh, A., F., Famoye, & C. Lee, (2012). Gamma-Pareto distribution and its applications. Journal of Modern Applied Statistical Methods, 11(1), 7.
3. Alzaghal, A. (2020). The exponentiated gamma-Pareto distribution with application to bladder cancer susceptibility. J. Stat. Appl. Probab. Lett, 7, 39-56.
4. Amin, M., S., Afzal, M. N., Akram, A. H., Muse, , A. H., Tolba & T. A. Abushal, (2022). Outlier detection in gamma regression using Pearson residuals: Simulation and an application. AIMS MATHEMATICS, 7(8), 15331-15347.
5. Amin, M., M., Amanullah, & M. Aslam, (2016). Empirical evaluation of the inverse Gaussian regression residuals for the assessment of influential points. Journal of Chemometrics, 30(7),394-404.
6. Amin, M., M., Amanullah, & G. M. Cordeiro, (2017). Influence diagnostics in the gamma regression model with adjusted deviance residuals. Communications in Statistics-Simulation and Computation, 46(9), 6959-6973.
7. Amin, M., M., Amanullah, M., Aslam, & M. Qasim, (2019a). Influence diagnostics in gamma ridge regression model. *Journal of Statistical Computation and Simulation*, *89*(3), 536-556.
8. Amin, M., M., Qasim, & M. Amanullah, (2019b). Performance of Asar and Genç and Huang and Yang's two-parameter estimation methods for the gamma regression

model. *Iranian Journal of Science and Technology, Transactions A: Science*, *43*, 2951-2963.

9. Ashour, S. K., D. R., Said, & M. A. A Fahim, (2014). The Log-Gamma-Pareto Distribution.

10. Atkinson, A. C. (1981). Two graphical displays for outlying and influential observations in regression. Biometrika, 68(1), 13-20.

11. Atkinson, A. C. (1985). Plots, transformations, and regression: an introduction to graphical methods of diagnostic regression analysis.

12. Belsley,D. A., E., Kuh, R. Welsch, (1980). RegressionDiagnostics: Identifying InfluentialData and Sources of Collinearity. New York:Wiley.

13. Chatterjee S, Hadi AS (1988). Sensitivity Analysis in Linear Regression. Wiley: New York,.

14. Cook, R. D. (1977). Detection of influential observation in linear regression. Technometrics, 19(1), 15-18.

15. Cook, R. D. (1986). Assessment of local influence. Journal of the Royal Statistical Society Series B: Statistical Methodology, 48(2), 133-155.

16. Cook, R. D., & S. Weisberg, (1982). Residuals and influence in regression.

17. Cordeiro, G. M. (2004). On Pearson's residuals in generalized linear models. Statistics & probability letters, 66(3), 213-219.

18. Cox, D. R., & E. J. Snell, (1968). A general definition of residuals. Journal of the Royal Statistical Society: Series B (Methodological), 30(2), 248-265.

19. Dar, A. A., A., Ahmed, & J. A. Reshi, (2020). weighted gamma-pareto distribution and its application. Pakistan Journal of Statistics, 36(4).

20. De Andrade, T. A., L. M. Z., Fernandez, F. G., Silva, & G. M. Cordeiro, (2017). The gamma generalized pareto distribution with applications in survival analysis.

21. Dobson, A. J., & A. G. Barnett, (2002). An introduction to generalized linear models. CRC press.

22. Hanum, H. (2015). Modeling Extreme Rainfall with Gamma. Applied Mathematical Sciences, 9(121), 6029-6039.

23. Hanum, H., A. H., Wigena, A., Djuraidah, & I. W. Mangku, (2016). Developing generalized linear model of Gamma-Pareto Distribution. Far east Mathematical Sciences Journal. Link finction

24. Hanum, H., A. H., Wigena, A., Djuraidah, & I. W. Mangku, (2016). Modeling Gamma-Pareto distributed data using GLM Gamma. Glob. J. Pure Appl. Math, 12(4), 3569-3575.

25. Hanum, H., A. H., Wigena, A., Djuraidah, & I. W. Mangku, (2017). The application of modeling gamma-Pareto distributed data using GLM gamma in estimation of monthly rainfall with TRMM data. Sriwijaya Journal of Environment, 2(2), 40-45.

26. Hardin, J. W., J. M. Hilbe, (2012). Generalized Linear Models and Extensions. 3rd ed. USA: Stata PressPublication.

27. Huet S, A, MA, Bouvier Poursat E, Jolivet Bouvier AM. Statistical tools for nonlinear regression: a practical guide with S-PLUS and R examples. New York: Springer; 2004 and 2006.

28. Lee AH (1986). Assessing partial influence in generalized linear models. Biometrics; 44: 71–77.

29. McCullagh, P., J.A. Nelder, (1989). Generalized Linear Models. London: Chapman & Hall.

30. Pierce, D. A., & D. W. Schafer, (1986). Residuals in generalized linear models. Journal of the American Statistical Association, 81(396), 977-986.

31. Pregibon, D. (1981). Logistic regression diagnostics. The annals of statistics, 9(4), 705-724.

32. Preisser, J. S., & B. F. Qaqish, (1996). Deletion diagnostics for generalised estimating equations. Biometrika, 83(3), 551-562.

33. Simas, A. B., & G. M. Cordeiro, (2009). Adjusted Pearson residuals in exponential family nonlinear models. Journal of Statistical Computation and Simulation, 79(4), 411-425.

34. Thomas, W., & R. D. Cook, (1989). Assessing influence on regression coefficients in generalized linear models. Biometrika, 76(4), 741-749.

35. Ullah, M. A., G. R. Pasha, (2009). The origin and developments of influence measures in regression. *Pakistan Journal of Statistics* 25:295–307.

36. Williams, D. (1987). Generalized linear model diagnostics using the deviance and single case deletions. Journal of the Royal Statistical Society: Series C (Applied Statistics), 36(2), 181-191.

37. Yasin, A., M., Amin, M., Qasim, A. H., Muse, & A. B. Soliman, (2022). More on the ridge parameter estimators for the gamma ridge regression model: simulation and applications. *Mathematical Problems in Engineering*, *2022*.\

38. Zheng, L., K., Ismail, & X. Meng, (2014). Shifted Gamma-Generalized Pareto Distribution model to map the safety continuum and estimate crashes. Safety Science, 64, 155-162.