Sigma J Eng Nat Sci, Vol. 43, No. 4, pp. 1495–1506, August, 2025

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2025.00148

Research Article

On the use of different link functions in gamma-pareto regression model: Simulation and application

Nasir SALEEM[1,*], Atif AKBAR[1], Javaria Ahmad KHAN[1]

[1]*Department of Statistics, Bahauddin Zakariya University, Multan, 60800, Pakistan*

ARTICLE INFO

Article history
Received: 21 March 2024
Revised: 13 May 2024
Accepted: 22 July 2024

Keywords:
G-PRM; GLM; Link Functions; SSR; MSE; AMSE; Reaction Rate Data

ABSTRACT

The Gamma-Pareto regression model (G-PRM) is appropriate when the response variable follows Gamma-Pareto distribution (G-PD) is used as the generalized linear model (GLM). For the estimation of the G-PRM the Iterative weighted least squares (IWLS) method is used with a specific link function. In this study, we consider G-PRM under different link functions. However, the researchers do not pay much attention to the selection of suitable link functions. In the context of the G-PRM, three link functions are used, which allow for investigating how inverse, identity, and log link functions perform. A Monte Carlo simulation and a demonstration with real data were used to compare the performance of different link functions in G-PRM. The sum squared residual (SSR), mean squared error (MSE) and average mean squared error (AMSE) are used as the evaluation criterion for suitable link function. Both the simulation and real data findings demonstrate that the G-PRM with the identity link function provides efficient results having minimum SSR, MSE and AMSE. Advantages of the paper, choice of suitable link function always significantly impacts the model`s performance.

Cite this article as: Saleem N, Akbar A, Khan JA. On the use of different link functions in gamma-pareto regression model: Simulation and application. Sigma J Eng Nat Sci 2025;43(4):1495–1506.

INTRODUCTION

In practice, we often meet the case that the response variables are continuous and follow the GP-D. In that case, the G-PRM and GLM are recommended. Application of GP-D in the form of health care economics, medical science, meteorology, occurrence rate and reaction rate etc. For the reliable results of regression analysis link function always play and important role. The regression model explains a phenomenon the (response variable) based on other phenomena (explanatory variables). Classical regression model is developed with the assumption that the response variables are normally distributed. This assumption is used for the validity of the test for both the model and its parameters. Regression modeling is usually based on the probability distribution of the response variable. For exponential family distribution the model is usually in the form of GLM i.e. Gamma, Beta, Pareto, Inverse Gaussian, Weibull etc. Gamma distribution is used in homicide data [1] Beta distribution for modeling reading skills data used by [2] Preto distribution is used in modeling earthquakes, forest fire areas and oil and gas field sizes by [3] presented an application of the Pareto distribution in modeling disk

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

This paper was recommended for publication in revised form by Regional Editor Ahmet Selim Dalkilic

drive sectors errors among others. To add flexibility to the Pareto distribution, various generalizations of the distribution have been derived, including the generalized Pareto distribution [4] the beta-Pareto distribution [5] and the beta generalized Pareto distribution [6].

Motivation of the Study

The purpose of the present study is work free region. Link functions that work with groups are available, as we have seen in the previous section; there must be some proper statistical measure for the comparison. On the basis of having minimum SSR, MSE and AMSE.

Advantages and Dis-Advantages of the Proposed Method

The merits and demerits of the proposed method. MSE is a commonly used metric that measures the average of the squared differences between the predicted and actual values. It gives more weight to large errors and is sensitive to outliers. MSE is useful when the goal is to minimize the overall error in the model. One of the advantages of MSE becomes a disadvantage when there is a bad prediction. The sensitivity to outliers magnifies the high errors by squaring them. MSE will have the same effect for a single large error as too many smaller errors. The residual sum of squares (RSS) measures the difference between observed data and the model's predictions. It is the portion of variability any regression model does not explain, also known as the model's error. Use of RSS to evaluate how well the model fits the data. RSS is a measure of the overall goodness of fit of the regression model, while MSE is a measure of the average distance between the predicted and actual values. RSS is used to compare different regression models, while MSE is used to evaluate the accuracy of a single model. RSS is a sum squared difference, while MSE is an average of squared error. RSS increases as the number of predictors in the model increases, while MSE decreases as the number of predictions increase.

Organized of the Paper

As it is structured, the paper goes in the following order: literature review in sections. In the next section, the gamma-pareto regression model was then used to estimate the parameters of the GLM of the gamma-pareto regression model. In next section discussed the simulation Study and Application real data (Reaction Rate Data). In last section Conclusion.

LITERATURE REVIEW

[7] explained how the link functions were carried out in the beta ridge regression model. In the case of the appropriate connection functions, the criterion to be evaluated is the smallest MSE. [8] discussed a comparison of some used link functions on the school drop-out rates data of East Java that yielded binomial regression models. In the case of the appropriate linking function, the estimates or criteria include Akaike information criterion (AIC), Bayesian information criterion (BIC), log likelihood (LL) and R-squared. [9] used generalized Weibull linear models with various link functions to survival analysis. Under various link functions, fit measures of goodness of fit models include deviance, AIC and BIC. [10] described the significance of Beta regression residuals-based control charts using various link functions. To this end, the data on the thermal power plants were utilized through an application. Moreover, three criteria are employed in checking the performance: the average of the run length, the standard deviation of the run length and the median run length. And also test the performance of the proposed control chart in two ways, namely, by monitoring the intercepts and by monitoring the slope coefficients. [11] examine the influential observation detection of the logistic regression with various link functions and employ pearson residuals. And applied an actual life example to analyse the urine calcium oxalate crystals data. [12] used Deviance and Pearson residuals-based monitoring charts with various link functions in tracking logistic regression profiles, to COVID-19 data. [13] compared the link functions relative to fitting logistic ridge regression as an application to urine data. A Monte Carlo simulation study and a real data set are taken into consideration and with scalar mean squared error as the performance evaluation criterion. Similarly, [14] developed a Gamma-Pareto distribution (G-PD). G-PD is a combination of Gamma and Pareto distribution taking a form of Pareto composed in Gamma, with three parameters (α, β, γ). [15] showed that GP-D is the member of exponential family as a condition to develop GLM based on G-PD. They provided some attributes for developing GLM such as mean, variance, and dispersion parameters of response variable defined the link functions for GLM, G-PD. They also provided deviance and AIC for assessing the goodness and parameter estimation of GLM, G-PD. [16,17] invent a simulation scheme for G-PRM used GLM gamma to analyze the relationship between simulated G-PD response variable with explanatory variable. The result showed that goodness of the model only depends on the goodness of fit the response variable to G-PD and the strength of the relationship of response and explanatory variable. [18] discussed about the modeling aims to analyze whether Tropical Rainfall Measuring Mission (TRMM) satellite data is a good estimator for unobserved station's data. Integrating G-PD data with TRMM data to estimate monthly rainfall using GLM. The Truncated Gamma-Pareto Distribution Used to Study Cosmic Rays, Novel Probability Distributions in Astrophysics. [19] a study Gamma- Pareto distributions (IV) and their importance and application and near to relatives and generalizations provide most flexible families of heavy-tailed distributions. It can be used to model income distributions as well as a wide variety of economic distributions. [14] invent a new distribution, namely as a Gamma-Pareto (IV) distribution, after that it called as G-PD (IV) distribution. Many properties of G-PD (IV) are discussed like a limiting behavior, mean, median and mode also moments, skewness and kurtosis. [20] invent the gamma generalized Pareto distribution, a three-parameter model, is used in the current study.

[21] introduced the exponentiated gamma-Pareto distribution. Here the existing distributions, are gamma-Pareto and exponentiated Pareto. The special cases of the exponentiated

gamma-Pareto distribution and its properties, including distribution shapes, limit behavior, hazard function etc. The creation of traffic conflict techniques heavily relies on the safety continuum. Using GGPD, it is possible to explain two crucial model parameters: the threshold and the shifted value. [22] invented a new probability distribution named as weighted gamma Pareto distribution (WG-PD). By using the T-X family and the concept of weighted probability. [23] a study of the new log-gamma-Pareto distribution (LG-PD) it includes as special cases two models such as gamma-Pareto and Pareto distributions. A real-world use of maximum likelihood estimation to estimate model parameters shows its potential. And the properties of LG-PD and its behavior such as mean, median and mode also skewness and kurtosis. [24] describe new q-rung orthopair fuzzy Aczel–Alsina weighted geometric operators under group-based generalized parameters in multi-criteria decision-making problems. [25] discussed the Q-rung orthopair probabilistic hesitant fuzzy hybrid aggregating operators in multi-criteria decision-making problems. [26] explain the importance Bipolar valued probabilistic hesitant fuzzy sets based on Generalized Hybrid Operators in multi-criteria decision-making problems based on TOPSIS. [27] invent Generalized Dice measures of single valued neutrosophic type-2 hesitant fuzzy sets and their application to multi-criteria decision-making problems. [28] discussed hybrid similarity measures of single-valued neutrosophic type-2 fuzzy sets and their application to MCDM based on TOPSIS. [29] discussed the multi-criteria decision making based on vector similarity measures of picture type-2 hesitant fuzzy sets. The applications of G-PD in above mentioned literature used only one variable. Meanwhile, the variable may be affected by other variable(s). To explain the relationship, we need regression model based on G-PD. For non-normal response variable, the regression model is usually in the form of Generalized Linear Model (GLM). [14] mentioned the mathematical relationship between G-PD and gamma distribution (GD). This is reasonable because the G-PD developed from the GD. The existence of this relationship provides a possibility to analyze the G-PD data through GLM Gamma which expand the scope of data distribution which can be modeled. The scope should be wider than Gamma and Pareto distributions individually. In the aforementioned studies, the logit link function is the tool the researchers use to evaluate G-PRM performance. Different link functions, however, are also essential in determining the superiority. In this study, order to get a clear image of the performance of link functions, we take into consideration three different link functions for G-PRM. We suggest several approaches with various link functions for the parameter estimation for the G-PRM. In addition, a thorough Monte Carlo simulation analysis is carried out under various link functions in order to select a suitable link function that achieves a minimum or average mean squared error (MSE). To prove the effectiveness of the Gamma-Pareto regression model under a specific link function. Data preparation, ensure your data meets the assumptions of the G-PRM, such as positive continuous responses. Model formulation, specify the G-PRM with the chosen link function (e.g., inverse, identity and log). Model estimation, estimate the model parameters using maximum likelihood estimation (MLE). Model evaluation, assess the model's goodness of fit using metrics such as, sum squared error (SSE), Mean squared error (MSE) and Average Mean absolute error (AMAE).

MATERIAL AND METHODS FOR GAMMA-PARETO REGRESSION MODEL AND SIMULATION STUDY

[14] state that the G-PD pdf is provided by,

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

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

The mean and variance of G-P distribution are, $E(a(y))= \alpha\beta$, $V(a(y))= \alpha\beta^2$ respectively. According to [16,17] With parameters, Eq. (1) can be modified $\alpha=\frac{1}{\phi}$ and $\beta=\mu\phi$. The Gamma Pareto density for y under these conditions 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 is continuous and non-negative, $\mu>0$ and $\phi>0$. Since the mean and variance of y are $E(y)=\mu$ 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. Following that, the G-PRM for the mean of response variable y is provided by [16].

According to [16,17], Link function g in GLM is $g(\mu_i)=X_i^T\beta=\eta_i$ where $\mu_i=E(a(y_i))$, which are presented in Table 1.

Table 1. Different link function of Gamma Pareto Distribution

Link function	Form of link function	Reference
Inverse link function	$\mu=\frac{1}{X'\beta}$	Hanum et al. [2016]
Identity link function	$\mu=X'\beta$	
Log link function	$\mu=log(X'\beta)$	
	$\mu=e^{X'\beta}$	

Estimation the GLM Gamma-pareto Regression Model Parameters

Finding the likelihood function's derivative with respect to β_j is the first step in estimating the parameter β_j using maximum likelihood and τ_i is the function of β. By Eq. (2)

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

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's link function. So, the score for β_j in GLM Gamma-Pareto is

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

Let l be the log likelihood of the response variable. Lastly, the 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 [30] as a method for estimating β_j.

The resultant estimate by using IWLS is given by,

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

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

Where z_i is the adjusted response variable, $X_i^T = (1, X_{i1}, X_{i2}, \dots, X_{ip})$ and W = diag $(\hat{\mu}_1{}^2, \dots, \hat{\mu}_n{}^2)$ is the weighted matrix and $\hat{\mu}_1 = \eta_i = \frac{1}{X'\beta}$, $\hat{\mu}_2 = \eta_i = X'\beta$ and $\hat{\mu}_3 = \eta_i = \log(X'\beta)$. And now, Using W and $var(U_j)$ for G-P and obtained the iteration for β_j as, i is a number of observations i=1,2,3, …, n. and j are a number of parameter j=1,2,3, …, p.

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

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

And finally fitted three models using above methodology.

$g(\mu_i) = \eta_i = \frac{1}{X'\beta}$ and fitted model inverse link, $\hat{\mu}_1 = \eta_i = \frac{1}{X'\beta}$ (6)

$g(\mu_i) = \eta_i = X'\beta$ and fitted model identity link, $\hat{\mu}_2 = \eta_i = X'\beta$ (7)

$g(\mu_i) = \eta_i = \log(X'\beta)$ and fitted model log link, $\hat{\mu}_3 = \eta_i = \log(X'\beta)$ (8)

Many types are available of GLM residuals in literature. Here we used a squared residual.

The Residuals (R), Squared Residuals (SR) and Sum Squared Residuals (SSR) for inverse link function in the G-PRM is given by

$$R_{(inverse)} = (y_i - \hat{\mu}_1) \tag{9}$$

$$SR_{(inverse)} = (y_i - \hat{\mu}_1)^2 \tag{9.1}$$

$$SSR_{(inverse)} = \sum(y_i - \hat{\mu}_1)^2 \tag{9.2}$$

$$MSE_{(inverse)} = \frac{\sum(y_i - \hat{\mu}_1)^2}{n-p} \tag{9.3}$$

The Residuals (R), Squared Residuals (SR) and Sum Squared Residuals (SSR) for identity link function in the G-PRM is given by

$$R_{(\text{identity})} = (y_i - \hat{\mu}_2) \tag{10}$$

$$SR_{(\text{identity})} = (y_i - \hat{\mu}_2)^2 \tag{10.1}$$

$$SSR_{(\text{identity})} = \sum(y_i - \hat{\mu}_2)^2 \tag{10.2}$$

$$MSE_{(\text{identity})} = \frac{\sum(y_i - \hat{\mu}_2)^2}{n-p} \tag{10.3}$$

The Residuals (R), Squared Residuals (SR) and Sum Squared Residuals (SSR) for log link function in the G-PRM is given by

$$R_{(\log)} = (y_i - \hat{\mu}_3) \tag{11}$$

$$SR_{(\log)} = (y_i - \hat{\mu}_3)^2 \tag{11.1}$$

$$SSR_{(\log)} = \sum(y_i - \hat{\mu}_3)^2 \quad (11.2)$$

$$MSE_{(\log)} = \frac{\sum(y_i - \hat{\mu}_3)^2}{n-p} \quad (11.3)$$

Simulation Study

This section is aimed at illustrating the work of various link functions by using simulation. Monte Carlo simulation of G-PRM involving the inverse, identity and log link functions has been done in this section. [16,17] invent these link function for G-PRM and define by $\eta_i = \frac{1}{X^T\beta}$ and fitted model inverse link, $\hat{\mu}_1 = \eta_i = \frac{1}{X^T\beta}$, $\eta_i = X'\beta$ and fitted model identity link, $\hat{\mu}_2 = \eta_i = X'\beta$ $\eta_i = \log(X'\beta)$) and fitted model log link, $\hat{\mu}_3 = \eta_i = \log(X'\beta)$. We used algorithm and simulation schemes developed by [16,17], to generate the response variable which follow a G-PRM is defined as $y_i \sim G - P(\alpha, \beta, \gamma)$, where inverse model $\hat{\mu}_1 = E(y_i) = \frac{1}{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3}}$ and identity model $\hat{\mu}_2 = E(y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3}$ and log model $\hat{\mu}_3 = E(y_i) = log(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3})$ with $\alpha = E(y_i) = 10$, and β is the arbitrary values $\beta_0 = 0.05$, $\beta_1 = 0.0025$, $\beta_2 = 0.005$, $\beta_3 = 0.0001$ while ϕ =0.04, 0.11, 0.17, 0.33, 0.67 and 2 is the dispersion parameters [31] and [38-43] and γ is minimum value of response variable. Here, the design matrix X generated from normal distribution as $X_{ij} \sim N(-1,1)$, for i=1, 2, …, n, and j=1,2,3. All the generated x-axis are fixed through the whole simulation study. we generate the data set for samples of size n=25, 50, 100, 200. These are simulation results that were done on the R software. Performance of link functions of G-PRM is run through the simulation 10000 times. We discussed their performance on the basis of average mean squared error (AMSE) and also suggest which link function is suitable for G-PRM.

In Table 2, AMSE are presented by using different link functions with different dispersion parameters and different sample sizes.

- In this section, for small dispersion level ϕ=0.04 and sample sizes n = 25, 50, 100 the identity link function provides a minimum AMSE as compared to the inverse and log link functions. When sample size n = 200 the log link function provides a minimum AMSE as compared to the inverse and identity link functions.

Table 2. Average mean squared error (AMSE) by using different links functions, G-PRM

ϕ	Sample size	Link Functions		
	n.	Inverse link	Identity link	Log link
		AMSE	AMSE	AMSE
0.04	25	0.0137	0.0110	0.0137
	50	0.0399	0.0272	0.0398
	100	0.0577	0.0292	0.0342
	200	0.0566	0.0303	0.0280
0.11	25	0.0014	0.0008	0.0014
	50	0.0316	0.0175	0.0239
	100	0.0279	0.0145	0.0164
	200	0.1451	0.0560	0.0550
0.17	25	0.0227	0.0122	0.0201
	50	0.0037	0.0030	0.0037
	100	0.0067	0.0040	0.0058
	200	0.0367	0.0121	0.0142
0.33	25	0.0582	0.0447	0.0597
	50	0.0991	0.0749	0.0956
	100	0.3108	0.1814	0.0041
	200	0.0110	0.0058	0.0017
0.67	25	0.0434	0.0351	0.0137
	50	0.0159	0.0108	0.0130
	100	0.1431	0.0748	0.0677
	200	0.1359	0.0552	0.0110
2	25	0.0629	0.0443	0.0620
	50	0.0574	0.0316	0.0504
	100	0.0822	0.0461	0.0246
	200	0.1833	0.0777	0.0474

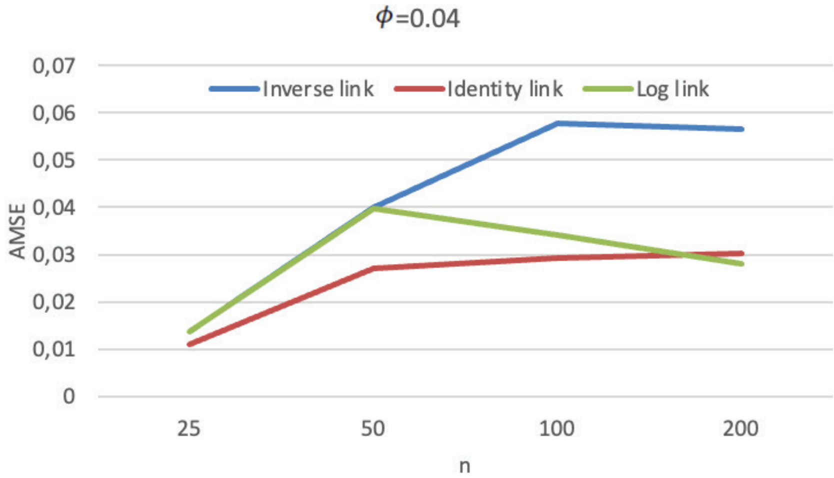

Figure 1. G-PRM under different link functions using AMSE for simulated data, when $\phi = 0.04$.

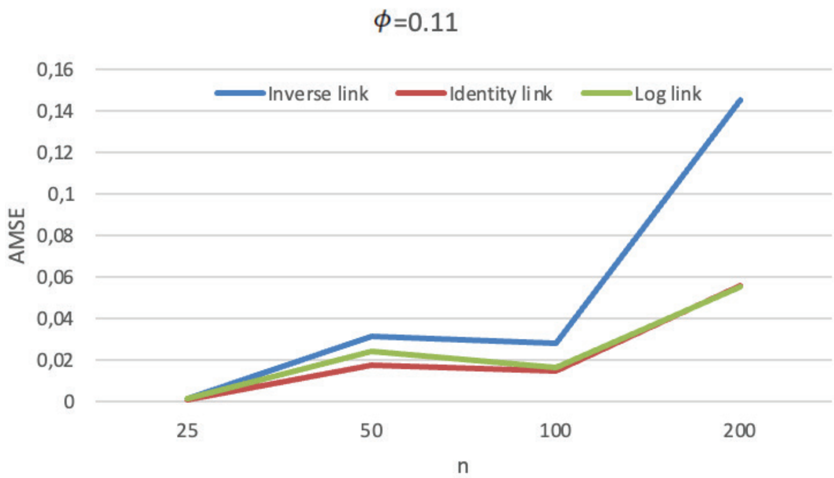

Figure 2. G-PRM under different link functions using AMSE for simulated data, when $\phi = 0.11$.

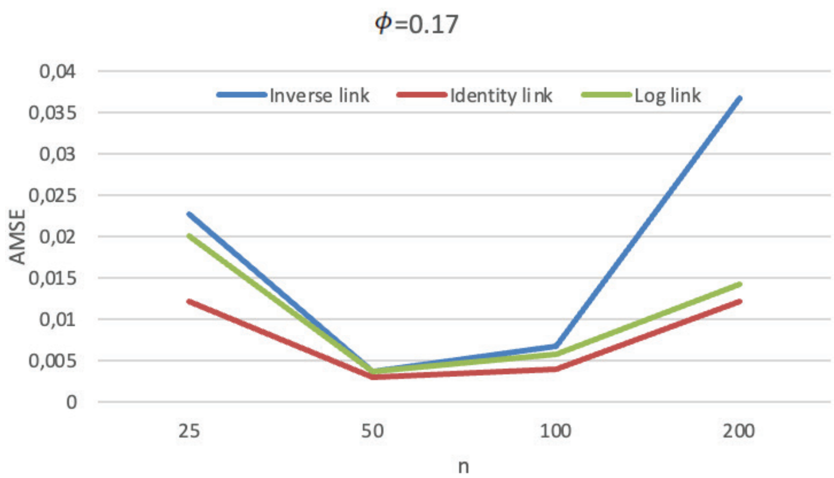

Figure 3. G-PRM under different link functions using AMSE for simulated data, when $\phi = 0.17$.

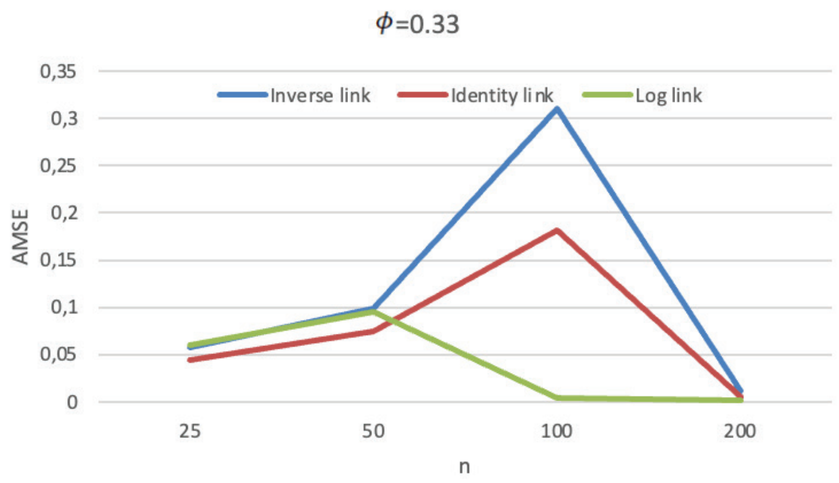

Figure 4. G-PRM under different link functions using AMSE for simulated data, when ϕ = 0.33.

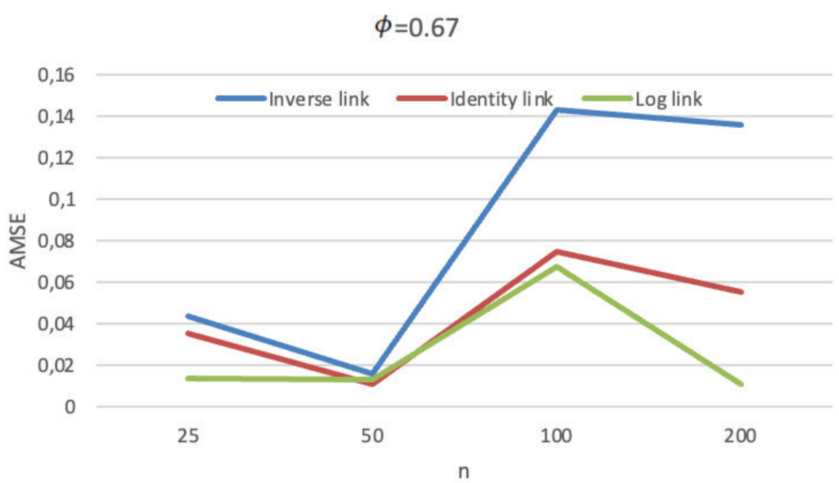

Figure 5. G-PRM under different link functions using AMSE for simulated data, when ϕ = 0.67.

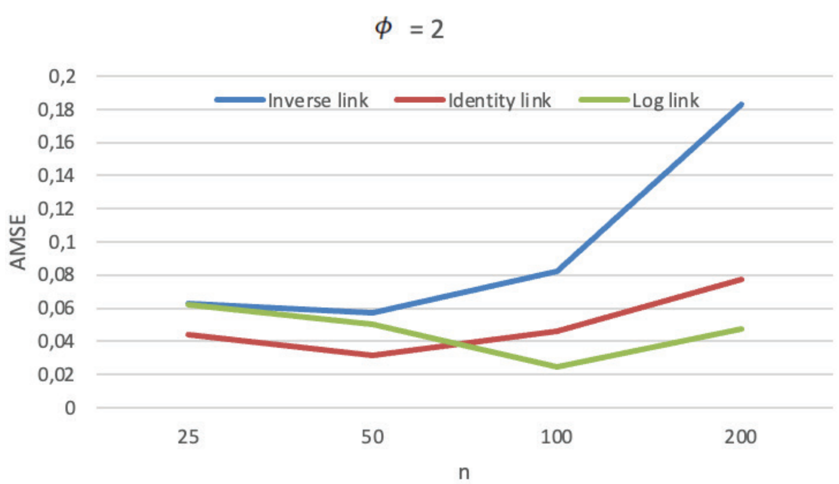

Figure 6. G-PRM under different link functions using AMSE for simulated data, when ϕ = 2.

- When dispersion level ϕ=0.11 and sample sizes n = 25, 50, 100 the identity link function provides a minimum AMSE as compared to the inverse and log link functions. When sample size n = 200 the log link function provides a minimum AMSE as compared to the inverse and identity link functions.
- When dispersion level ϕ=0.17 and for all sample sizes n = 25, 50, 100 and 200 the identity link function provides a minimum AMSE as compared to the inverse and log link functions.
- When dispersion level ϕ=0.33 and sample sizes n = 25, 50 the identity link function provides a minimum AMSE as compared to the inverse and log link functions. When sample sizes n = 100, 200 the log link function provides a minimum AMSE as compared to the inverse and identity link functions
- When dispersion level ϕ=0.67 and sample sizes n = 25 the log link function provides a minimum AMSE as compared to the inverse and identity link functions. When sample sizes n = 50 the identity link function provides a minimum AMSE. For sample size n = 100, 200 the log link function provides a minimum AMSE as compared to the inverse and identity link functions.
- When large dispersion level ϕ=2 and sample sizes n = 25, 50 the identity link function provides a minimum AMSE as compared to the inverse and log link functions. When sample sizes n = 100, 200 the log link function provides a minimum AMSE as compared to the inverse and identity link functions.
- It is evident that results are consistent, as AMSE are decreasing with increase of sample size. On the basis of simulation results the identity link function is best for G-PRM.

The graphical results of G-PRM under different link functions using AMSE for simulated data are reported in figures 1-6.

- The figures 1-6, show the performance of different link functions on the basis of AMSE. It is interesting to note that for small and large dispersion levels and sample size's n = 25, 50 the identity link function provides a minimum AMSE as compared to the inverse and log link functions. We can see in all the graphs when sample size's n = 100, 200 the log link function provides a minimum AMSE as compared to the inverse and identity link functions.

Application: Reaction Rate Data

Next, we shall see how the various link functions will work when applied to G-PRM using a real-life application. We employed the reaction rate data taken from [32,33]. Then [34,35,36,37] 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 response variable follows a gamma distribution (this is required by following [16,17]. However, this data set is not well fitted to the normal distribution since the trend of the dependent variable is positively skewed also from the distribution fitting test, we observed that the GP distribution is well fitted to this data set, the results are reported in Table 3.

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.

$$\hat{y}_i = (0.6700[0.092,\text{S}] - 88.689x_1[0.0007,\text{S}] + 57.982x_2[0.0002,\text{S}] - 138.60x_3[0.0002,\text{N}])^{-1}$$

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

$$\hat{y}_i = (-3.1750[0.3832,\text{N}] + 0.0598x_1[0.0024,\text{S}] - 0.0673x_2[0.0025,\text{S}] + 0.0046x_3[0.0008,\text{S}])$$

Table 3. Distribution goodness of fit tests for Reaction Rate Data

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

Gamma-Pareto Distribution (GPD).

Table 4. Real Data Coefficients by using different link functions

Regression Coefficients and Standard Error	Link Functions		
	Inverse link	Identity link	Log link
β_0	0.6700	-3.1750	1.0028
(S.E)	(0.0932)	(0.3832)	(0.3434)
β_1	-88.689	0.0598	0.9775
(S.E)	(0.0007)	(0.0024)	(0.0016)
β_2	57.982	-0.0673	-0.4075
(S.E)	(0.0002)	(0.0025)	(0.0011)
β_3	-138.60	0.0046	-0.1104
(S.E)	(0.0002)	(0.0008)	(0.0007)

Table 5. Real Data Fitted Models for different link functions

$Y_{i\ response.v}$	$\hat{\mu}_{1inverse}$	$\hat{\mu}_{1identity}$	$\hat{\mu}_{3log}$
3.541	3.6445	3.1277	2.9818
2.397	2.8500	2.7093	2.3065
6.694	4.5243	5.6077	5.7216
4.722	3.6732	5.6695	5.1428
0.593	1.9297	0.7576	1.1971
0.268	1.6179	0.2461	0.8800
2.797	2.1526	3.4517	2.4241
2.451	1.8564	3.1343	1.9287
3.196	2.9239	3.4936	2.9882
2.021	2.1108	2.7643	1.9187
0.896	2.1208	0.7749	1.2649
5.084	2.6867	5.1504	4.0298
5.686	5.6948	5.4398	5.7639
1.193	1.6180	1.1403	1.1053
2.648	2.4622	3.1403	2.4049
3.303	2.4119	3.0862	2.3485
3.054	2.4319	3.0683	2.3477
3.302	2.4421	3.0945	2.3681
1.271	2.4420	1.7354	1.7310
11.648	10.897	7.5470	11.113
2.002	2.8125	2.4733	2.1626
9.604	11.995	9.8081	18.350
7.754	6.1454	5.6315	6.2097
11.59	14.270	9.7023	18.724

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

$$\hat{y}_i = log(1.0028[0.3434, \mathrm{S}] + 0.9775x_1[0.0016, \mathrm{N}] - 0.4075x_2[0.0011, \mathrm{N}] - 0.1104x_3[0.0007, \mathrm{S}])$$

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.

By using above mentioned three link functions (inverse, identity and log) in G-PRM, we have calculated coefficients

Table 6. Real Data Squared Residuals and sum Squared Residuals and Mean Squared Error

Sr.	$SR_{(inverse)}$	$SR_{(identity)}$	$SR_{(log)}$
1	0.0107	0.1708	0.3128
2	0.2052	0.0975	0.0082
3	4.7075	1.1800	0.9455
4	1.0999	0.8978	0.1771
5	1.7868	0.0271	0.3649
6	1.8222	0.0005	0.3745
7	0.4153	0.4287	0.1391
8	0.3535	0.4669	0.2727
9	0.0740	0.0886	0.0432
10	0.0081	0.5524	0.0105
11	1.5000	0.0147	0.1361
12	5.7472	0.0044	1.1113
13	0.0001	0.0606	0.0061
14	0.1806	0.0028	0.0077
15	0.0345	0.2423	0.0591
16	0.7941	0.0470	0.9110
17	0.3870	0.0002	0.4989
18	0.7394	0.0431	0.8723
19	1.3712	0.2157	0.2116
20	0.5633	16.817	0.2861
21	0.6569	0.2222	0.0258
22	5.7209	0.0417	76.493
23	2.5875	4.5051	2.3849
24	7.1826	3.5632	50.899
SSR	97.715	29.691	136.55
MSE	4.8857	1.4845	6.8227

and standard errors of reaction rate data which are presented in Table 4. The impact of link functions can be observed that how link functions change the effect of regressors. In our case, the use of identity link makes the intercept negative but rest of two provide a positive intercept as β_0. Similarly, the use of inverse link for β_1 provide a very large negative co-efficient. For β_2, inverse link provides a large co-efficient but this time with positive effect, other two link functions provide competitively very small negative coefficient. Inverse and log link function generate negative coefficients for β_3 but coefficients generated by inverse link function is very large and other one is very small. Identity link function provide positive small coefficient. Standard errors (S.E) by inverse link functions are smallest then others. In Table 5, we fitted a three G-PRM for inverse, identity and log link functions. And the estimated values of $Y_{response}$, fitted models are $\hat{\mu}_{1inverse}$, $\hat{\mu}_{2identity}$ and $\hat{\mu}_{3log}$. It is evident that G-PRM $\hat{\mu}_{2identity}$ has a minimum SSR and MSE as compare to others $\hat{\mu}_{1inverse}$ and $\hat{\mu}_{3log}$.

CONCLUSION

In this paper, we evaluated the performance of different link functions to estimate G-PRM and identify the best link function. For this purpose, inverse, identity and log link functions are considered. To compare the performance of different link functions, we conducted a Monte Carlo simulation study an empirical application Reaction rate data. It has widespread applications in the fields of industry, medical science Biopsy, rainfall data, chemical science and engineering etc. The Gamma-Pareto regression model is a powerful tool for modeling continuous positive data, and the choice of link function can significantly impact the model's performance. Here are some advantages of our study using different link functions in the Gamma-Pareto regression model. The log link is commonly used in Gamma-Pareto regression as it provides a natural way to model the mean of the response variable. It also ensures that the predicted values are always positive. Advantage easy to interpret, and the estimated coefficients can be exponentiated to obtain relative changes in the mean. The identity link is similar to the log link but does not involve any transformation. It is useful when the response variable has a large range of values. Advantage the model is simpler to implement. The inverse link is useful when the response variable has a large range of values and the relationship between the mean and variance is not strong. Advantage provides more flexible modeling of the mean-variance relationship. The choice of link function in Gamma-Pareto regression depends on the specific characteristics of the data and the research question. Each link function has its advantages, and selecting the appropriate one can lead to more accurate and interpretable results. In the simulation study, we used different dispersion level dispersion levels ϕ=0.04, 0.11, 0.17, 0.33, 0.67 and 2 and different sample sizes n = 25, 50, 100 and 200. we evaluated the performance of the G-PRM under finite sample sizes, dispersion parameters and different link functions. The performance evaluation of these link functions has been done by using SSR, MSE and AMSE. From the simulation results, we observed that the performance of identity link function in G-PRM provides minimum AMSE as compared to inverse and log link functions for all simulation cases. From the reaction rate data identity link function in G-PRM provides a minimum SSR and MSE as compared to inverse and log link function. We also noticed from simulation results that the identity link function gives better performance as compared to the other link functions. Based on real-life and simulated results, we may suggest that identity link function is appropriate whenever practitioners want to apply G-PRM. In general, we recommend the use of identity link function in G-PRM.

Recommendations of research to be conducted in the future, some dimensions still require exploration. In this research, the performance of various link functions is being discussed based on SSR, MSE and AMSE. These can be extrapolated to the influence diagnostics that find the various GLM residuals in various link functions in the G-PRM. Even further generalized to GLM influence diagnostics, where one parameter is estimated in a biased manner (Ridge), modified ridge estimation, Liu estimation, modified Liu estimation and Stein estimation.

FUNDING

No funding resource is available for this study.

DECLARATIONS CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

[1] Altun E, Korkmaz MC, El-Morshedy M, Eliwa MS. The extended gamma distribution with regression model and applications. *AIMS Mathematics*, 2021.

[2] Pammer K, Kevan A. The contribution of visual sensitivity, phonological processing, and nonverbal IQ to children's reading. Scientific Studies of Reading. 2007 ;11(1):33-53.

[3] Burroughs SM, Tebbens SF. Upper-truncated power law distributions. Fractals. 2001; 9(02):209-22.

[4] Pickands III J. Statistical inference using extreme order statistics. the Annals of Statistics. 1975; 119-31.

[5] Akinsete A, Famoye F, Lee C. The beta-Pareto distribution. Statistics. 2008 ;42(6):547-63.

[6] Mahmoudi E. The beta generalized Pareto distribution with application to lifetime data. Mathematics and computers in Simulation. 2011 ;81(11):2414-30.

[7] 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]

[8] 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]

[9] Afsana-Al-Sharmin, Islam MA. Generalized Weibull linear models with different link functions. Adv Appl Stat 2017;50:367–384. [CrossRef]

[10] 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]

[11] 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]

[12] 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]

[13] 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]

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

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

[16] Hanum H, Wigena AH, Djuraidah A, Mangku IW. Developing generalized linear model of Gamma-Pareto Distribution. Far east Mathematical Sciences Journal. 2016.

[17] 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.

[18] 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]
[19] Alzaatreh A, Ghosh I. A study of the Gamma-Pareto (IV) distribution and its applications. Commun Stat Theory Methods 2016;45:636–654. [CrossRef]
[20] 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]
[21] Alzaghal A. The exponentiated gamma-Pareto distribution with application to bladder cancer susceptibility. J Stat Appl Probab Lett 2020;7:39–56. [CrossRef]
[22] Dar AA, Ahmed A, Reshi JA. Weighted Gamma-Pareto distribution and its application. Pak J Stat 2020;36:287–304.
[23] Ashour SK, Said DR, Fahim MA. The log-Gamma-Pareto distribution. Int J Sci Basic Appl Res 2014;16:357–374.
[24] Özlü Ş. New q-rung orthopair fuzzy Aczel–Alsina weighted geometric operators under group-based generalized parameters in multi-criteria decision-making problems. Computational and Applied Mathematics. 2024; 43(3):122.
[25] Özlü Ş. Q-rung orthopair probabilistic hesitant fuzzy hybrid aggregating operators in multi-criteria decision-making problems. Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi. 2023; 27(3):366-73.
[26] Özlü Ş, Al-Quran A, Riaz M. Bipolar valued probabilistic hesitant fuzzy sets based on Generalized Hybrid Operators in multi-criteria decision-making problems based on TOPSIS. Journal of Intelligent & Fuzzy Systems. 2024; 1-20.
[27] Özlü Ş. Generalized Dice measures of single valued neutrosophic type-2 hesitant fuzzy sets and their application to multi-criteria decision-making problems. International Journal of Machine Learning and Cybernetics. 2023;14(1):33-62.
[28] Özlü Ş, Karaaslan F. Hybrid similarity measures of single-valued neutrosophic type-2 fuzzy sets and their application to MCDM based on TOPSIS. Soft Computing. 2022; 26(9):4059-80.
[29] Özlü Ş. multi-criteria-decision making based on vector similarity measures of picture type-2 hesitant fuzzy sets. Granular Computing. 2023; 8(6):1505-31.
[30] Dobson AJ. An introduction to generalized linear models. Boca Raton: Chapman and Hall/CRC; 2002. [CrossRef]
[31] 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]
[32] 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.
[33] 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; 2006. p. 233.
[34] Amin M, Amanullah M, Aslam M, Qasim M. Influence diagnostics in gamma ridge regression model. J Stat Comput Simul 2019;89:536–556. [CrossRef]
[35] 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]
[36] Amin M, Qasim M, Yasin A, Amanullah M. Almost unbiased ridge estimator in the gamma regression model. Communications in Statistics-Simulation and Computation. 2022;51(7):3830-50.
[37] McCullagh P. Generalized linear models. London: Routledge; 1989. [CrossRef]
[38] 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. 2025;43(3): 760-776.
[39] Saleem N, Akbar A, Rehman Irshad AU, Ozdemir, B, Ahmad Khan J. Influence Diagnostics in Gamma-Pareto Regression Model with Cook's Distance for Standardized and Adjusted Pearson, Deviance, and Likelihood Residuals: Simulation and Application. Journal of Mathematics. 2025(1);3440237.
[40] Saleem N, Akbar A. Empirical evaluation of the gamma-Pareto regression residuals for influence diagnostics. Journal of Statistical Computation and Simulation. 2024;94(16):3573-3598.
[41] Saleem N, Akbar A, Imon AR, Khan JA. Multicollinearity in Binomial Regression: A comparison between CERES and PR Plots for detection. Journal of Statistical Modeling & Analytics (JOSMA). 2024;6(1).
[42] Saleem N, Akbar A, Imon AR, Al Mamun, ASM. Detection of Outliers in Binomial Regression Using CERES and Partial Residual Plots. Journal of Statistical Modeling & Analytics (JOSMA). 2022;4(2).
[43] Saleem N, Akbar A, Sharif SB, Al Mamun ASM, Imon AHMR, & Ahmad S. Performance Comparison between CERES and PR Plots in Detecting the Heteroscedasticity Problem using Liver Cancer and Simulated Data. Global Journal of Engineering and Technology, 2022;*7*(3).