International Journal of Mathematics and Computer Research
ISSN: 2320-7167
Volume 13 Issue 06 June 2025, Page no. – 5335-5341
Index Copernicus ICV: 57.55, Impact Factor: 8.615
DOI: 10.47191/ijmcr/v13i6.08

Generalized Gamma Regression Model: Simulation Study and Its Application

Hasbi Yasin[1], Sugito[2]
[1,2]Department of Statistics, Universitas Diponegoro, Semarang, Indonesia

ARTICLE INFO
Published Online:
07 June 2025

ABSTRACT
A popular statistical tool that works particularly well in situations when the data distribution is positively skewed and asymmetrical is the generalized gamma (GG) distribution. In many real-world scenarios, multiple factors can influence different outcomes at the same time. The Generalized Gamma Regression (GGR) model, which is intended for answers that have a Generalized Gamma (GG) distribution, is presented in this article. The Berndt-Hall-Hall-Hausman (BHHH) algorithm is used to optimize the Maximum Likelihood Estimation (MLE) technique, which is used for parameter estimation in the GGR model. We employ the Maximum Likelihood Ratio Test (MLRT) and the Wald test for partial testing to evaluate the model's relevance. Thorough simulation validation shows that the GGR model is capable of accurately estimating parameters with little bias. To demonstrate its usefulness, we deployed the GGR model to a real-world case study. We specifically use it to examine the average years of schooling (AYS) in Central Java, Indonesia. In the final analysis, the current research emphasizes the advantages of applying the GGR model for generalized gamma distributed responses and demonstrates the model's resilience in parameter estimation.

Corresponding Author:
Hasbi Yasin

KEYWORDS: Average Years of Schooling (AYS); Generalized Gamma Distribution; Generalized Gamma Regression; Berndt-Hall-Hall-Hausman (BHHH) algorithm; Maximum Likelihood Estimation (MLE).

I. INTRODUCTION

The linear regression framework is a common approach for investigating the relationship between a response value and one or more predictor variables. However, these models rely on a strong normal distribution assumption, which is sometimes difficult to meet. To overcome this, several proposals have been made, especially for asymmetric continuous distributions, including Weibull, log-normal, gamma, exponential, and Rayleigh regression [1]-[9]. These regression models were developed separately.

To model response variables with a positive slope, we are interested in investigating generalized gamma regression. Compared to many other distributions, including gamma, Weibull, exponential, chi-square, Erlang, Rayleigh, and half-normal, generalized gamma has a more general form, which makes it beneficial. [10]. Because of this, our research can be applied to various continuous distribution regression models.

Since the generalized gamma distribution was established in the 1960's [11]-[12], additional research has been suggested by re-parametrizing it to boost flexibility [10], [13]. Yasin et al. [14] proposed univariate generalized gamma regression, which models the mean response determined by numerous predictor variables. In this study, we apply this model to analyse the Average Years of Schooling in Central Java, Indonesia. This indicator is examined with multiple factors that are assumed to affect it. These factors include GDP per capita, Percentage of Poor People, Percentage of Households with Access to Proper Sanitation, Gender Ratio, Labor Force Participation Rate, and Teacher to Student Ratio at the junior high school level.

The remainder of this article is structured as comes next. Section 2 explains the generalized gamma distribution and its accompanying regression model, then provides approaches for parameter estimation and hypothesis testing. In Section 3, we assess our model and procedure using simulated data before applying it to estimate the Average Years of Schooling data in Section 4. Section 5 offers findings.

II. METHOD DETAILS

A. Generalized Gamma Distribution

According to the number of parameters used, the Generalized Gamma distribution can be divided into two types, namely the GG distribution with three parameters or the Generalized Gamma distribution with four parameters (with the addition of the location parameter (threshold)). A continuous random variable Y with GG distribution has the following pdf [10]:

$$f\left(y|\lambda,\tau,\theta,\delta\right)=\frac{\tau}{\theta}\frac{\left(\frac{y-\delta}{\theta}\right)^{\tau\lambda-1}}{\Gamma(\lambda)}\exp\left(-\left(\frac{y-\delta}{\theta}\right)^{\tau}\right) \quad (1)$$

where $y>\delta;\lambda>0;\tau>0;\theta>0;\delta\geq 0$ and $f(y)=0$ for others. λ is the first shape parameter, τ is the second shape parameter, θ is the scale parameter, and δ is the location parameter or written as $Y\sim \text{GG}(\lambda,\tau,\theta,\delta)$. Mean and variance of Generalized Gamma distributed random variables are given in equations (2) and (3).

$$E(Y)=\theta\frac{\Gamma\left(\lambda+\frac{1}{\tau}\right)}{\Gamma(\lambda)}+\delta, \quad (2)$$

$$\text{Var}(Y)=\theta^{2}\left(\frac{\Gamma\left(\lambda+\frac{2}{\tau}\right)}{\Gamma(\lambda)}-\left(\frac{\Gamma\left(\lambda+\frac{1}{\tau}\right)}{\Gamma(\lambda)}\right)^{2}\right), \quad (3)$$

where $\Gamma(\lambda)=\int_0^\infty t^{\lambda-1}e^{-t}dt$, $\Gamma(\lambda+1)=\lambda\Gamma(\lambda)$, and $\Gamma(\lambda)$ is the gamma function.

B. Generalized Gamma Regression (GGR)

In this study, the Generalized Gamma regression model was developed based on the scale parameter as the basic model, which means that the shape and threshold parameters are considered fixed for each observation [14]. The selection of the scale parameter in the Generalized Gamma regression model is done because the scale parameter (θ) has a direct relationship with the mean value of the distribution and can be easily represented using a link function in the Generalized Linear Models (GLMs) framework. Therefore, by substituting the mean of the GG distribution (Eq. (2)) using the "log" link function, the UGGR model is obtained as follows:

$$E(Y)=\exp\left(\mathbf{x}^{T}\boldsymbol{\beta}\right),\text{ or} \quad (4)$$

$$\theta(\mathbf{x})=\frac{\Gamma(\lambda)}{\Gamma\left(\lambda+\frac{1}{\tau}\right)}\left(\exp\left(\mathbf{x}^{T}\boldsymbol{\beta}\right)-\delta\right), \quad (5)$$

where $\theta(\mathbf{x})$ expresses the value of the scale parameter affected by changes in the predictor value.

By substituting Eq. (5) into the Generalized Gamma density equation (Eq. (1)), the pdf of the GGR model for the *i*-th observation is obtained as follows:

$$f(y_i)=\frac{\tau\left(y_i-\delta\right)^{\tau\lambda-1}}{\Gamma(\lambda)}\left(\frac{\left(\exp\left(\mathbf{x}_i^{T}\boldsymbol{\beta}\right)-\delta\right)\Gamma(\lambda)}{\Gamma\left(\lambda+\frac{1}{\tau}\right)}\right)^{-\tau\lambda}\exp\left(-\left(\frac{\Gamma\left(\lambda+\frac{1}{\tau}\right)}{\Gamma(\lambda)}\left(\frac{y_i-\delta}{\exp\left(\mathbf{x}_i^{T}\boldsymbol{\beta}\right)-\delta}\right)\right)^{\tau}\right). \quad (6)$$

1] *Parameter Estimation*: GGR model parameters are estimated using Maximum Likelihood Estimation (MLE). MLE is an instrument to generate parameter estimates by maximising the likelihood function. The flexibility in determining the likelihood function provides ease in estimating parameters using MLE. The likelihood function of *n* random samples $y_1,y_2,\ldots,y_n$ defined as a joined pdf of *n* random variable. If $y_1,y_2,\ldots,y_i,\ldots,y_n$ is a random sample with pdf as in the Equation (1), where $\boldsymbol{\Theta}_{GGR}=\left[\boldsymbol{\beta}^{T}\,\lambda\;\tau\;\delta\right]^{T}$ is the vector parameter of the GGR model then the likelihood function of GGR model is $L(\boldsymbol{\Theta}_{GGR})=\prod_{i=1}^{n}f\left(y_i|\boldsymbol{\Theta}_{GGR}\right)$ and log-likelihood of GGR model is:

$$l(\boldsymbol{\Theta}_{GGR})=\log\left(L(\boldsymbol{\Theta}_{GGR})\right). \quad (7)$$

Maximum value of $l(\boldsymbol{\Theta}_{GGR})$ on the Equation (7) will be obtained when $\frac{\partial l(\boldsymbol{\Theta}_{GGR})}{\partial\boldsymbol{\Theta}_{GGR}^{T}}=\mathbf{0}$ and second derivative of $l(\boldsymbol{\Theta}_{GGR})$ at $\hat{\boldsymbol{\Theta}}_{GGR}$ is a negative definite matrix. However, mathematically no closed form solution is obtained. Therefore, to obtain it, optimization with the BHHH iteration method is used [14]. The use of the BHHH algorithm requires a gradient vector and an approximation of the Hessian matrix obtained from the outer product of the gradient vector of the GGR model parameters [15]. The gradient vector of the GGR model parameters at the i-th observation is compiled based on the first partial derivative equation of each parameter, namely:

$$\mathbf{g}_i(\boldsymbol{\Theta}_{GGR})=\left[\frac{\partial\log\left(f\left(y_i|\boldsymbol{\Theta}_{GGR}\right)\right)}{\partial\boldsymbol{\Theta}_{GGR}^{T}}\right]^{T}. \quad (8)$$

Thus, the gradient vector and Hessian matrix approximation vector for the GGR model parameters are:

$$\mathbf{g}\left(\boldsymbol{\Theta}_{GGR}\right)=\sum_{i=1}^{n}\mathbf{g}_i\left(\boldsymbol{\Theta}_{GGR}\right), \tag{9}$$

$$\mathbf{H}\left(\boldsymbol{\Theta}_{GGR}\right)=-\sum_{i=1}^{n}\mathbf{g}_i\left(\boldsymbol{\Theta}_{GGR}\right)\mathbf{g}_i\left(\boldsymbol{\Theta}_{GGR}\right)^T. \tag{10}$$

where $\mathbf{g}_i\left(\boldsymbol{\Theta}_{GGR}\right)$ is the gradient vector for the i-th observation, and the matrix $\mathbf{H}\left(\boldsymbol{\Theta}_{GGR}\right)$ is a negative definite matrix. Next, the steps to obtain the GGR model parameter estimator using the MLE method and BHHH iteration are presented in Algorithm 1.

Algorithm 1: Parameter Estimation of GGR Model with BHHH Method

1. Input a random sample of n consisting of a response variable $(y_1, y_2, \ldots y_i, \ldots, y_n)$ and a predictor variable $(x_{i1}, x_{i2}, \ldots, x_{ij}, \ldots, x_{ip})$, for $i = 1, 2, \cdots, n$.
2. Determining the initial value of the model parameter vector $\hat{\boldsymbol{\Theta}}^{(0)} = \left[\hat{\boldsymbol{\beta}}^{(0)T} \ \hat{\lambda}^{(0)} \ \hat{\tau}^{(0)} \ \hat{\delta}^{(0)}\right]^T$, in this study it was used $\hat{\lambda}^{(0)}, \hat{\tau}^{(0)}, \hat{\delta}^{(0)}$ to obtain from the estimation of the distribution parameters of the response variable, and $\hat{\boldsymbol{\beta}}^{(0)} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\ln(\mathbf{y})$.
3. Determine tolerance limits for convergence $\varepsilon > 0$ and maximum iteration *of M*.
4. Starting the BHHH iteration starts with $m = 0$ the following steps:
   a. For each $i = 1, 2, \cdots, n$ perform the following process:
      1. Getting the first partial derived value of a function $\log\left(f\left(y_i \middle| \hat{\boldsymbol{\Theta}}_{GGR}^{(m)}\right)\right)$ to each parameter assessed.
      2. Forming individual gradient vectors of i-th observation, $\mathbf{g}_i\left(\hat{\boldsymbol{\Theta}}_{GGR}^{(m)}\right)$ using Equation (8).
   b. Calculate the gradient vector of the GGR model parameter estimator $\mathbf{g}\left(\hat{\boldsymbol{\Theta}}_{GGR}^{(m)}\right)$ using Equation (9).
   c. Getting an approximation of the Hessian matrix estimator parameter of the GGR model $\mathbf{H}\left(\hat{\boldsymbol{\Theta}}_{GGR}^{(m)}\right)$ using Equation (10).
   d. If $\left(\hat{\boldsymbol{\Theta}}_{GGR}^{(m)}\right)^T\mathbf{H}\left(\hat{\boldsymbol{\Theta}}_{GGR}^{(m)}\right)\hat{\boldsymbol{\Theta}}_{GGR}^{(m)} < 0$ then the value of the parameter estimator of the GGR model is updated with the following equation: $\hat{\boldsymbol{\Theta}}_{GGR}^{(m+1)} = \hat{\boldsymbol{\Theta}}_{GGR}^{(m)} - \mathbf{H}^{-1}\left(\hat{\boldsymbol{\Theta}}_{GGR}^{(m)}\right)\mathbf{g}\left(\hat{\boldsymbol{\Theta}}_{GGR}^{(m)}\right)$
   e. If $\left\|\hat{\boldsymbol{\Theta}}_{GGR}^{(m+1)} - \hat{\boldsymbol{\Theta}}_{GGR}^{(m)}\right\| < \varepsilon$ or the iteration has reached the M iteration then the iteration stops, if not then $m = m + 1$, and returns to step 4.a.
   f. If $\left(\hat{\boldsymbol{\Theta}}_{GGR}^{(m)}\right)^T\mathbf{H}\left(\hat{\boldsymbol{\Theta}}_{GGR}^{(m)}\right)\hat{\boldsymbol{\Theta}}_{GGR}^{(m)} \geq 0$ then the value of the parameter estimator of the GGR model is $\hat{\boldsymbol{\Theta}}_{GGR} = \hat{\boldsymbol{\Theta}}_{GGR}^{(m)}$.
5. When the convergent condition is reached and all eigenvalues of the matrix $\mathbf{H}\left(\hat{\boldsymbol{\Theta}}_{GGR}\right)$ are negative, the value of the estimator of the GGR parameter $\hat{\boldsymbol{\Theta}}_{GGR} = \left[\hat{\boldsymbol{\beta}}^T \ \hat{\lambda} \ \hat{\tau} \ \hat{\delta}\right]^T$ and the value of $l\left(\hat{\boldsymbol{\Theta}}_{GGR}\right)$.

2] *Hypothesis Testing*: Testing of regression parameter hypotheses in GGR models simultaneously using Wilk's statistics *likelihood ratio* derived based on Maximum Likelihood Ratio Test (MLRT) and partially tested using the Z test [14].

Parameter testing is carried out with the aim of finding out which model and variables have a significant effect. To determine the importance of model parameters, the Likelihood Ratio Test (LRT) and the Z test for significance are utilized.

Hypothesis:

H_0: $\beta_1 = \beta_2 = \cdots = \beta_k = 0$

H_1: at least one $\beta_j \neq 0$, j = 1, 2, ..., k

Test statistics:

$$G^2 = -2\ln\left(\frac{L(\omega)}{L(\Omega)}\right) = 2\ln L(\Omega) - 2\ln L(\omega), \tag{11}$$

where $\omega = (\beta_0, \lambda, \tau, \delta)$ is the set of model parameters below H_0 (Null Model), and $\Omega = (\boldsymbol{\beta}, \lambda, \tau, \delta)$ is the set of model parameters below H_1 (Full Model). Reject H_0 $G^2 > \chi^2_{(1-\alpha;k)}$ or sig < α.

Then, to partially test the significance of parameters, the Z test is used.

Hypothesis:

H_0: $\beta_j = 0$

H_1: $\beta_j \neq 0$, $j = 1, 2, \cdots, k$

Test statistics:

$$Z = \frac{\hat{\beta}_j}{SE\left(\hat{\beta}_j\right)}, \tag{12}$$

where $SE\left(\hat{\beta}_j\right)$ is the standard error of the parameter $\hat{\beta}_j$, $SE\left(\hat{\beta}_j\right) = \sqrt{Var\left(\hat{\beta}_j\right)}$ and $Var\left(\hat{\beta}_j\right)$ is the main diagonal element of the matrix $(j+2)$ $-\left[\mathbf{H}^{-1}\left(\hat{\Omega}\right)\right]$. Reject H_0 if $|Z| > Z_{\alpha/2}$ or sig < α.

3] *Goodness of Fits*: The coefficient of determination is a statistic developed to score the goodness of fit of a model and identify the degree of its fit to the data. The determination coefficient measures the extent to which a predictor variable impacts a response variable. A greater coefficient of determination suggests a more effective model. Equation (13) describes how to compute the coefficient of determination for a generalized linear model using the likelihood ratio test (LRT) approach [16-17].

$$R^2 = 1 - \exp\left(\frac{2}{n}\left(\log L(\hat{\omega}) - \log L(\hat{\Omega})\right)\right), \qquad (13)$$

where n is the sample size, $\ln L(\hat{\Omega})$ is the log-likelihood of the complete model (*restricted model*), and $\log L(\hat{\omega})$ is the ln-likelihood of the model below H_0 or the incomplete model (*unrestricted model*).

In identifying and evaluating a model, it is necessary to have statistics to obtain the best model. Many methods are used in determining the best model, one of which is *the Akaike Information Criterion* (AIC) developed by Akaike in 1974. One of the main benefits of AIC lies in its convenience which does not require a table for checking. In determining the best model, it is seen from the model with the smallest AIC value. However, AIC may provide less reliable results on small data, so the second order of AIC, called AICc, is the solution. The AICc method is based on the *Maximum Likelihood Estimation* (MLE). The AICc value for the global model can be calculated with the following Equation [18]:

$$\text{AICc} = -2\log L(\hat{\Omega}) + 2q , \qquad (14)$$

where $\log L(\hat{\Omega})$ is the maximum value *of the log-likelihood* model using parameter estimates, and q the number of parameters estimated in the model. The selection of the best model is seen from the smallest AICc value.

III. SIMULATION STUDY

We use a simulation study to check whether our new estimation method works well. In this simulation, we used different sample sizes: 30, 50, 100, 150, 200, and 300 samples, and we repeated each scenario 1000 times. We chose these numbers based on what makes statistical sense, what our computers can handle, and to ensure that our results are reliable and widely applicable. To simulate the GGR model, we need to create data that corresponds to how the predictor variables relate to the response variable, which follows the GG distribution. The steps of the simulation study are outlined in Algorithm 2.

Algorithm 2: Algorithm for GGR Model simulation

1. Specifies the actual parameters for the GGR model to be simulated. The actual parameter consists of the regression coefficient and other parameters of the GG distribution, or $\boldsymbol{\Theta}_{GGR} = \left[\boldsymbol{\beta}^T \lambda\ \tau\ \delta\right]^T$. The true parameters in this simulation study are:
   $\beta_0 = 2.54173; \beta_1 = 0.00475; \beta_2 = -0.00932;$
   $\beta_3 = 0.00586; \lambda = 0.75;\ \tau = 8.5;$ and $\delta = 6.5$.
2. Generate predictor variable from a Uniform distribution:
   $X_1 \sim U(10,90)$, $X_2 \sim U(10,25)$, and $X_3 \sim U(10,50)$.
3. Calculate the scale parameters for each of the k-th response variables and the i-th observation with the following equation:
   $$\theta(\mathbf{x}_i) = \frac{\Gamma(\lambda)}{\Gamma\left(\lambda + \frac{1}{\tau}\right)}\left(\exp\left(\mathbf{x}_i^T\boldsymbol{\beta}\right) - \delta\right),$$
   with $\mathbf{x}_i = \begin{bmatrix}1 & x_{i1} & x_{i2} & x_{i3}\end{bmatrix}^T$.
4. Generate a GG distributed response variable, i.e. $y_i \sim GG\left(\lambda, \tau, \theta(\mathbf{x}_i), \delta\right)$.
5. Estimating GGR model parameters with Algorithm 1.
6. Calculates the bias, variance, and *Root of Mean Squared Error* (RMSE) of each parameter.
7. Interpret the results of the simulation.

We evaluate the accuracy of parameter estimations by computing the mean, standard deviation, and confidence interval for each estimated parameter, as shown in Table I. Moreover, we investigate the bias, variance, and RMSE of these parameter estimates. Figure 1 depicts these details in a visual manner demonstrating the performance of our parameter estimate approach. This complete assessment allows us to assess the GGR model's robustness and dependability in reflecting the linkages driving our simulated data.

Table I. Statistics of the Parameters According to the Sample Size

Size	True Parameter	Mean	Std. Dev	lower (2.5%)	upper (97.5%)
30	2.5417	2.5390	0.0935	2.3609	2.7127
	0.0048	0.0048	0.0008	0.0033	0.0063
	−0.0093	−0.0093	0.0040	−0.0170	−0.0011
	0.0059	0.0059	0.0015	0.0029	0.0087
50	2.5417	2.5415	0.0679	2.3968	2.6720
	0.0048	0.0047	0.0005	0.0037	0.0057
	−0.0093	−0.0093	0.0030	−0.0149	−0.0032
	0.0059	0.0059	0.0011	0.0037	0.0080
100	2.5417	2.5405	0.0460	2.4511	2.6349
	0.0048	0.0047	0.0004	0.0040	0.0055
	−0.0093	–	0.0019	–	–

		0.0093		0.0133	0.0055
	0.0059	0.0059	0.0007	0.0044	0.0073
	2.5417	2.5416	0.0247	2.4955	2.5922
	0.0048	0.0048	0.0002	0.0044	0.0051
150		–		–	–
	− 0.0093	0.0093	0.0011	0.0115	0.0073
	0.0059	0.0059	0.0004	0.0051	0.0067
	2.5417	2.5404	0.0312	2.4758	2.6011
	0.0048	0.0048	0.0003	0.0042	0.0052
200		–		–	–
	− 0.0093	0.0093	0.0014	0.0120	0.0065
	0.0059	0.0059	0.0005	0.0049	0.0068
	2.5417	2.5416	0.0247	2.4955	2.5922
	0.0048	0.0048	0.0002	0.0044	0.0051
300		–		–	–
	− 0.0093	0.0093	0.0011	0.0115	0.0073
	0.0059	0.0059	0.0004	0.0051	0.0067

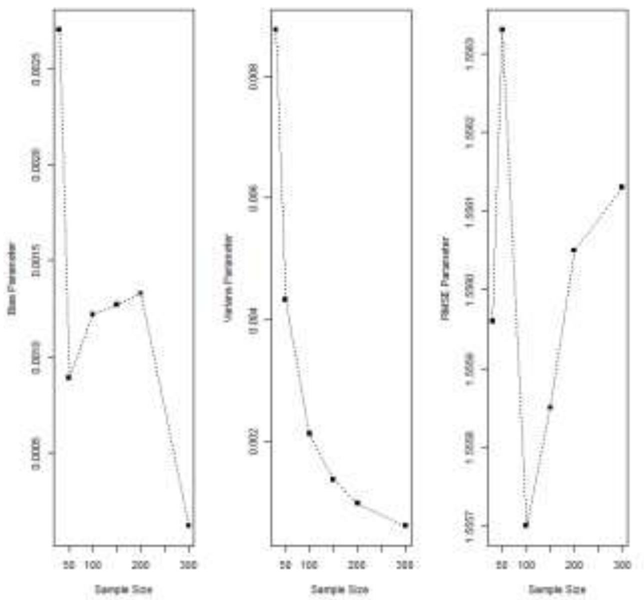

Figure 1. Bias, Variance, and RMSE of GGR Coefficients.

In line with the findings in Table I, our suggested approach is reliable for estimating the parameters of the GGR model correctly regardless of small and high number of samples. The computed parameters' mean is extremely close to the actual parameters. The small standard deviations and tight confidence intervals suggest a high level of accuracy. Furthermore, the pattern identified in Figure 1 shows that the variance associated with the model parameters decreases as the sample size grows. This pattern suggests that as sample size increases, estimated parameter values become more precise and consistent. However, it is important to note that this tendency does not hold true for the estimated parameters' bias and RMSE. Figure 1 does not show a clear relationship between bias and RMSE with the sample size. But it can be concluded that our proposed procedure consistently provides parameter estimates for the GGR model with relatively low bias, offering reliable and accurate results.

IV. APPLICATION OF GGR

In the purposes of this research, we use supplementary information from BPS of Central Java Province covering from 2017 until 2021. The monitoring unit consists of 35 districts and cities of Central Java Province. [19]. The data used relates to education statistics in Central Java, which includes the Average Years of Schooling (AYS) indicator as the response variable. Factors that are thought to influence it include GDP per capita (million Rupiah) (X_1), Percentage of Poor People (X_2), Gender Ratio (X_3), Percentage of Households with Access to Proper Sanitation (X_4), Labor Force Participation Rate (X_5), and Teacher-Student Ratio at junior high school level (X_6). Descriptive statistics for each of the research variables are presented in Table II, which provides an overview of the data characteristics.

Table II. Descriptive Statistics of Research Data

Variable	Mean	SD	Min	Max
Mean Years of Schooling	7.773	1.212	6.180	10.900
GDP per Capita	28.386	17.966	12.370	87.360
Percentage of Poor People	11.245	3.673	3.980	20.320
Gender Ratio	99.373	2.306	93.900	103.940
Percentage of Households Accessing Proper Sanitation	77.954	17.292	9.240	98.070
Labor Force Participation Rate	69.369	3.153	58.730	76.600
Teacher-Student Ratio	18.063	2.934	13.000	39.000

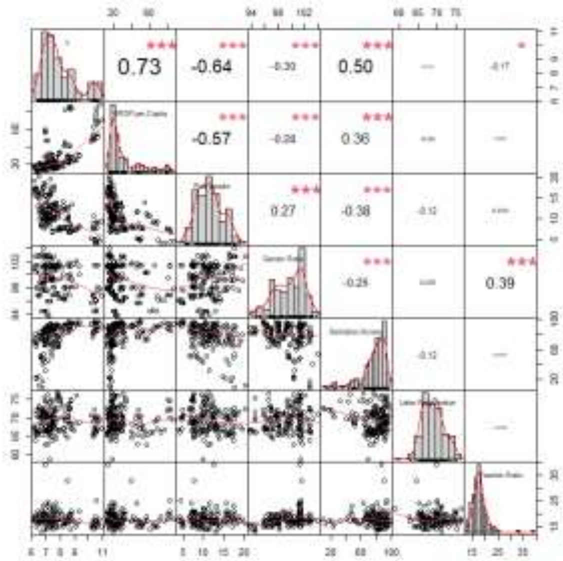

Figure 2. Plot of Correlation Matrix Among Variables

A visual portrait of the correlation between the response variable and the predictor variables is clearly displayed through a matrix plot, as shown in Figure 3. Across the predictor variables, X_1 and X_4 show a positive correlation with the response variable. Inversely, X_2, X_3, X_5, and X_6 indicate a negative correlation with the response variable.

Furthermore, distribution testing was carried out on the AYS variable. This test is through the normality test and the test of the Gamma and Generalized Gamma 3 distributions. This test is carried out with the Kolmogorov-Smirnov (KS) test, which is used to test whether the sample comes from a population with a certain distribution [20].

Table III. Kolmogorov-Smirnov Test

Distribution	Test Statistic	p-value
Normal	0.1552	0.0004
Gamma	0.1393	0.0023
Generalized Gamma	0.0664	0.4241

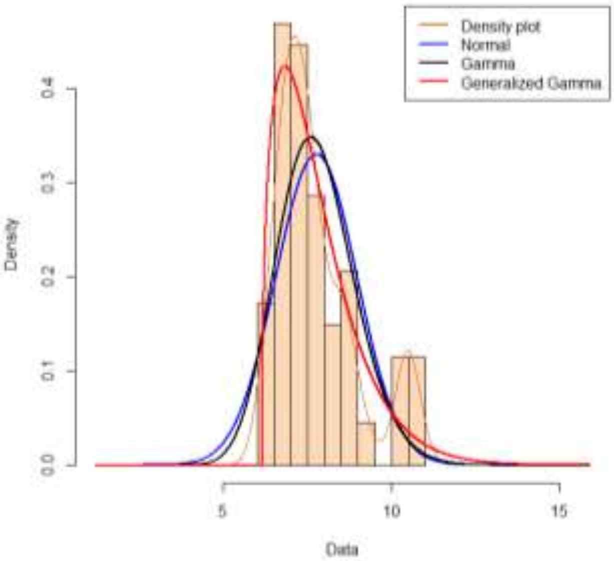

Figure 3. Density Plot of AYS

Based on Table III, it can be concluded that the AYS variable does not fit to a normal distribution or gamma distribution because the p-value are less than 0.05. The Generalized Gamma distribution more accurately describes the distribution of the AYS variable, which is shown by the p-value of 0.4241. This is also supported by visual identification in Figure 3. Therefore, the GGR model is more relevant for AYS modelling compared to linear regression and gamma regression models.

Table IV shows the parameter estimation findings for the GGR model. Based on the model parameters, the test statistic G^2 = 170.428 is calculated with a p-value of 0. Consequently, at a significance level of 0.05, H_0 in the simultaneous testing of the GGR model parameters is rejected, implying that at least one predictor variable influences the response variable in the model. As a result, a partial test is required to establish which factors significantly affect the model using the Z test. According to the Z-test produces in Table IV, the p-value for all parameters is extremely small (sig. at 1%), implying that all predictors have a significant impact on AYS. This regression model has a R^2 value of 62.43%, indicating that the predictor variables analysed can clarify 62.43% of the response variability. The remaining of 37.57% is explained by predictors that not included in the model. Table V gives the overall assessment of model goodness.

Table IV. Coefficient of GGR Model

Parameter	Estimat e	Z Value	Pr. (>|Z|)	Odd
β_0	1.4277	1.65×10^8	0.0000	4.1692
β_1	0.0048	420.3287	0.0000	1.0048
β_2	−0.0060	−204.0319	0.0000	0.9940
β_3	0.0056	2212.7516	0.0000	1.0057
β_4	0.0016	34.3928	0.0000	1.0016
β_5	−0.0010	−413.7003	0.0000	0.9990
β_6	−0.0041	−212.1267	0.0000	0.9960

Table V. Statistics of the GGR Model

Statistic	Output
Null Deviance	494.931 on 171 degrees of freedom
Residual Deviance	349.69 on 165 degrees of freedom
G^2	145.241
p-value	0
AIC	363.69
RMSE	0.7405624
R^2	0.6243359

Based on Table IV, the GGR model for AYS of Central Java Province can be written with the following equation.

$$\hat{y} = \exp(1.4277 + 0.0048X_1 - 0.0060X_2 + 0.0056X_3 + 0.0016X_4 - 0.0010X_5 - 0.0041X_6) \quad (15)$$

Based on Equation (15), it can be interpreted that if GDP per capita (X_1) increases by 1 million rupiah, AYS will increase by 1.0048 times, with other independent variables considered constant. If the percentage of poor people (X_2) increases by 1%, AYS will decrease by 0.9940 times, with other independent variables held constant. If Gender Ratio (X_3) increases by 1%, AYS will increase by 1.0057 times, with other independent variables held constant. If the percentage of households accessing proper sanitation (X_4) increases by 1%, AYS will increase by 1.0016 times, with other independent variables held constant. If the labour force participation rate (X_5) increases by 1%, AYS will decrease by 0.9990 times, with other independent variables held constant. If the Teacher to Student Ratio (X_6) increases by 1 student, AYS will decrease by 0.9960 times, with other independent variables held constant.

V. CONCLUSION

In summary, the Generalized Gamma Regression (GGR) model has proven to be accurate and effective tool for estimating parameters in the context of responses that follow the Generalized Gamma distribution. By using the Maximum Likelihood Estimation (MLE) method and the BHHH algorithm for optimization, this model consistently provides reliable results. The application of the GGR model to analyse the Average Years of Schooling (AYS) in Central Java based on six predictor variables showed the significant results. The inference model results show that the variables of GDP per capita, the gender ratio, and the percentage of households that have access to proper sanitation are factors that can increase the average length of schooling. While the variables of the percentage of poor people, the labour force participation rate and the teacher-to-student ratio are factors that hinder the increase in the average length of schooling in Central Java.

REFERENCES

1. Cepeda-Cuervo, E., Corrales, M., Cifuentes, M.V. and Zarate, H. 2016. On gamma regression residuals, Journal of the Iranian Statistical Society 15(1), 29–44 (2016).
2. Nasution, A.S., Purhadi and Sutikno. 2017. Estimasi Parameter dan Pengujian Hipotesis Pada Model Regresi Gamma (Studi Kasus: Pemodelan Pencemaran Sungai di Surabaya), Paidagogeo: Jurnal Pendidikan 2(2), 17–26.
3. Suyitno, S.: Penaksiran Parameter dan Pengujian Hipotesis Model Regresi Weibull Univariat, Jurnal Eksponensial 8(2), 179–183 (2017).
4. Zhang, Z. 2016. Parametric regression model for survival data: Weibull regression model as an example, Annals of Translational Medicine 4(24), 1–8.
5. Bednarski, T. and Skolimowska-Kulig, M. 2019. On scale Fisher consistency of maximum likelihood estimator for the exponential regression model under arbitrary frailty, Statistics and Probability Letters 150, 9–12.
6. Davidov, O. and Zelen, M. 2000. Exact tests for exponential regression, Journal of Statistical Planning and Inference 88(1), 87–97.
7. Palm, B.G., Bayer, F.M., Cintra, R.J., Pettersson, M.I. and Machado, R. 2019. Rayleigh Regression Model for Ground Type Detection in SAR Imagery, IEEE Geoscience and Remote Sensing Letters 16(10), 1660–1664.
8. Yasin, H., Inayati, S. and Setiawan. 2022. 3-Parameter Gamma Regression Model for Analyzing Human Development Index of Central Java Province, BAREKENG: Jurnal Ilmu Matematika dan Terapan 16(1), 171–180.
9. Diantini, N. L. S., Purhadi, and Choiruddin, A. 2023. Parameter estimation and hypothesis testing on three parameters log normal regression. In AIP Conference Proceedings (Vol. 2554, No. 1, p. 030024). AIP Publishing LLC.
10. Sanchez, R. and Mackenzie, S.A. 2016. Information thermodynamics of cytosine DNA methylation, PLoS ONE 11(3), 1–20.
11. Stacy, E.W. 1962. A Generalization of the Gamma Distribution, The Annals of Mathematical Statistics 33(3), 1187–1192.
12. Stacy, E.W. and Mihram, G. A. 1965. Parameter Estimation for a Generalized Gamma Distribution, Technometrics 7(3), 349–358.
13. Shanker, R. and Shukla, K.K. 2016. On modeling of lifetime data using three-parameter generalized lindley and generalized gamma distributions, Biom Biostat Int J. 4(7), 283–288.
14. Yasin, H., Purhadi and Choiruddin, A. 2022. Estimasi Parameter dan Pengujian Hipotesis Model Geographically Weighted Generalized Gamma Regression, Jurnal Gaussian 11(1), 140–152.
15. Berndt, E.K., Hall, B.H., Hall, R.E. dan Hausman, J.A. 1974. Estimation and Inference in Nonlinear Structural Models, *Annals of Economic and Social Measurement*, Vol. 3, No. 4, 653–665.
16. Magee, L. 1990. R^2 Measures Based on Wald and Likelihood Ratio Joint Significance Tests, *The American Statistician*, Vol. 44, No. 3, 250–253. http://doi.org/10.2307/2685352.
17. Zhang, D. 2017. A Coefficient of Determination for Generalized Linear Models, *The American Statistician*, Vol. 71, No. 4, 310–316. http://doi.org/10.1080/00031305.2016.1256839.
18. Burnham, K.P. dan Anderson, D.R. 2002, *Model Selection and Inference: A Practical Information-Theoretic Approach*, 2nd Ed., Springer-Verlag, New York. http://doi.org/10.2307/3803117.
19. BPS-Statistics of Jawa Tengah Province, LNCS Homepage https://jateng.bps.go.id/, last accessed 2023/02/17.
20. Biu, E. O., Nwakuya, M. T., and Wonu, N. 2019. Detection of Non-Normality in Data Sets and Comparison between Different Normality Test. Asian Journal of Probability and Statistics, Vol. 5, No.4, ISSN: 2582-0230, 1-20.