Gamma Regression for Modeling the Education Index of Cities in Java

Anita Rahayu, M. Aldiki Febriantono
Computer Science Department
School of Computer Science
Bina Nusantara University
Jakarta 11480, Indonesia
anitarahayu@binus.ac.id, m.aldiki@binus.ac.id

Dedy Dwi Prastyo, Purhadi, Sutikno
Statistics Department
Institut Teknologi Sepuluh Nopember
Surabaya 60111, Indonesia
dedy-dp@statistika.its.ac.id, purhadi@statistika.its.ac.id, sutikno@statistika.its.ac.id

Abstract

Many phenomena do not follow a symmetrical distribution. In order to obtain accurate conclusions, the analysis used for this phenomenon can be adjusted to the nature of the data distribution that is not symmetrical, for example the gamma distribution. The gamma distribution can be extended into a regression model. The gamma regression relates a positive response variable following a gamma distribution to one or more predictor variables. In this study, we explore a gamma regression with the gamma distribution has three parameters. This paper also presents the parameter estimation, test statistics, and hypothesis testing for the significance of the parameter. The method used for parameter estimation is Maximum Likelihood Estimation (MLE), which is optimized using numerical iteration, namely Berndt-Hall-Hall-Hausman (BHHH) algorithm. Maximum Likelihood Ratio Test (MLRT) is used for simultaneous test, whereas the Wald test is used for partial test. The developed gamma regression model is applied to the education index with three predictor variables, among others, the percentage of households to private toilet ownership, population density, and the percentage of poor people. The unit of observation is regency/city in Java, Indonesia, of the year 2018. The results show that modeling using multiple predictors is better in terms of accuracy of prediction as well as the interpretation. The future work of this research is how to involve geographical factors as a spatial effect in the model.

Keywords
Education index, Gamma regression, Maximum likelihood estimation, Maximum likelihood ratio test, Wald test

1. Introduction
The gamma distribution is a continuous probability distribution that has a positively skewed (Sallay et al., 2021). Its three parameter consists of shape, scale, and location parameters. Gamma regression is a type of non-linear regression that contains at least one non-linear parameter (Bravo et al., 2021; Pan et al., 2019). Chakraborty and Chakravarty (2012) developed properties and parameter estimations method for discrete gamma distributions. Song et al. (2021), who conducted a study of parameter estimation for four-parameter exponential gamma distribution and asymptotic variance of its quantiles. Nagar et al. (2021) studied some properties of a bimatrix variate Kummer-gamma distribution, like the moment generating function, marginal and conditional distributions, and the moments.

The gamma regression model relates the response variable that follows gamma distribution to the predictor variables. The gamma regression model proposed in this study is the based on the Trivariate Gamma Regression (TGR) presented by Rahayu et al. (2019) and Multivariate Gamma Regression (MGR) presented by Rahayu et al. (2020), which describes the parameter estimation and hypothesis testing. In this study, the parameter estimation method employs the MLE. However, the closed-form solution cannot be obtained. Then requires a numerical method; for example, the BHHH algorithm used in this work.

Gamma regression has been developed by several researchers, including Wang et al. (2021), conducted a study of a dynamic Remaining Useful Life (RUL) prediction and Optimal Maintenance Time (OMT) determination approach using a gamma process model. Tzougas (2020) presents the Poisson-Inverse Gamma regression model with varying dispersion for approximating heavy-tailed and overdispersed claim counts. Marchand et al. (2021) estimating the mean under weighted squared error losses in presence of a lower-bound restriction for one parameter exponential families.

The article is structured as follows. Section 1 presents an introduction to this study. The materials and methods are presented in section 2. We present the results and discussion in Section 3. Conclusions and further research are presented in Section 4.

1.1 Objectives

Based on several theories above, the goals of this study include: (i) construct the gamma regression model, (ii) parameter estimation, (iii) parameter significance test, partially and simultaneously, and (iv) apply the proposed gamma regression model to data of the education index. This gamma regression is implemented to data of the education index as a response variable, along with the percentage of households to private toilet ownership, population density, and percentage of poor people are considered as predictor variables. The observation unit is the regency/city in Java, Indonesia.

2. Literature Review

2.1 Gamma Regression Model

Gamma regression is one of the regressions that can see the relationship between random variable Y as response variable follows a gamma distribution with $X_1, X_2, \ldots, X_p$ as predictor variables (Mohammadpour et al., 2020). According to the requirement of the gamma distribution, the response variable is positive continuous data and has a constant Coefficient of Variation (CoV). According to Chen and Kotz (2013) and Balakrishnan and Wang (2000), a random variable Y with a gamma distribution has three parameters, among others shape (δ), scale (η), and location (φ), denoted by $Y \sim Gamma(\delta,\eta,\varphi)$, with probability density function (pdf) is:

$$f(y) = \begin{cases} \dfrac{(y-\varphi)^{\delta-1} e^{-\frac{y-\varphi}{\eta}}}{\eta^{\delta}\Gamma(\delta)}; & \delta,\eta > 0,\ \varphi < y < \infty \\ & ;\ \text{otherwise.} \end{cases} \tag{1}$$

According to Ewemoje and Ewemooje (2011) and Bono et al. (2020), if $Y \sim Gamma(\delta,\eta,\varphi)$ so:

$$\mu = E(y) = \eta\delta + \varphi, \mathrm{Var}(y) = \delta\eta^2, \text{and } \mathrm{Stdev}(y) = \sqrt{\delta\eta^2} = \sqrt{\delta}\eta.$$

Equation (2) presents gamma regression model with δ,η,φ as paramaters.

$$E(Y) = \eta\delta + \varphi = e^{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}} = e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p}. \tag{2}$$

Based on equation (2), the formula for δ is:

$$\delta = \frac{e^{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}} - \varphi}{\eta}. \tag{3}$$

By substituting the formula δ in equation (3) into equation (1), the pdf obtained for y_l is:

$$f(y_l) = \frac{(y_l-\varphi)^{\frac{e^{\mathbf{x}_l^T\boldsymbol{\beta}}-\varphi}{\eta}-1} e^{-\frac{y_l-\varphi}{\eta}}}{\eta^{\frac{e^{\mathbf{x}_l^T\boldsymbol{\beta}}-\varphi}{\eta}}\Gamma\left(\frac{e^{\mathbf{x}_l^T\boldsymbol{\beta}}-\varphi}{\eta}\right)}; \eta > 0,\ \varphi < y_l < \infty. \tag{4}$$

Parameter estimation can be done using MLE by maximizing the likelihood function to obtain the parameter estimator (Usman et al., 2018). The likelihood function for n random variables $y_1, y_2, \ldots, y_n$ is defined as the joint density function of n random variables. If it is a random sample of the density function $f(y;\theta)$, then the likelihood function is $f(y_1;\theta)f(y_2;\theta)\cdots f(y_n;\theta)$.

Based on the results of the first partial derivative of the log-likelihood function for each parameter, the form is not closed form. To obtain the values $\hat{\eta},\hat{\varphi},\hat{\boldsymbol{\beta}}$ requires numerical optimization. BHHH algorithm is used in this study. The first step to start the BHHH numerical method is to set the initial value for $\hat{\boldsymbol{\theta}}^{(0)} = \left[\hat{\eta}^{(0)} \quad \hat{\varphi}^{(0)} \quad \hat{\boldsymbol{\beta}}^{T(0)}\right]^T$. The initial value $\hat{\boldsymbol{\beta}}^{T(0)}$ obtained from the Generalized Linear Models (GLMs), and the initial value $\hat{\eta}^{(0)}$ is positive according to the pdf statement of the gamma distribution in equation (1). The Hessian $\mathbf{H}(\hat{\boldsymbol{\theta}})$ is approximated as the negative of the sum of the outer products of the gradients of individual observations. The tolerance limit used in this study is $\varepsilon = 10^{-8}$. Next is to do the BHHH iteration process. According to Cameron and Trivedi (2005), the iteration process stops if $\left\|\hat{\boldsymbol{\theta}}^{(p^*+1)} - \hat{\boldsymbol{\theta}}^{(p^*)}\right\| \le \varepsilon$.

Testing the regression parameters of the gamma regression model consists of simultaneous and partial test. Testing the hypothesis of simultaneous regression parameters aims to determine the significance of the regression coefficient in the model simultaneously, and the test statistics used is the statistical likelihood ratio derived based on the MLRT method. Hypothesis testing of partial regression parameters aims to find out which parameters have a significant influence on the model partially, and the test statistics used are the Wald test (Pawitan, 2001). In detail, testing the regression parameters of the gamma regression model has been explained in the previous research, written by Rahayu et al. (2019).

2. Data and Methods

This research uses secondary data obtained from Statistics Indonesia. The education index as the response variable, with three predictor variables: percentage of households to private toilet ownership, population density, and percentage of poor people. The data are calculated from 119 regencies/cities in Java - Indonesia, 2018.

Statistical analysis used is gamma regression, along with the MLE and the BHHH algorithm. MLRT is used for simultaneous test and Wald test is used for partial test. All the hypothesis testing are presented in Rahayu et al. (2019).

3. Results and Discussion

The response variable used for modeling with gamma regression must follow distribution of gamma. With the KS test statistics and the significance level $\alpha = 5\%$, the test statistics value for education index data is $D_n = 0.107$, and the KS table value is $D_{(0.05)} = 0.124$. Due to the value of $D_n < D_{(0.05)}$, so do not to reject the null hypothesis. It means that the education index data follows a gamma distribution.

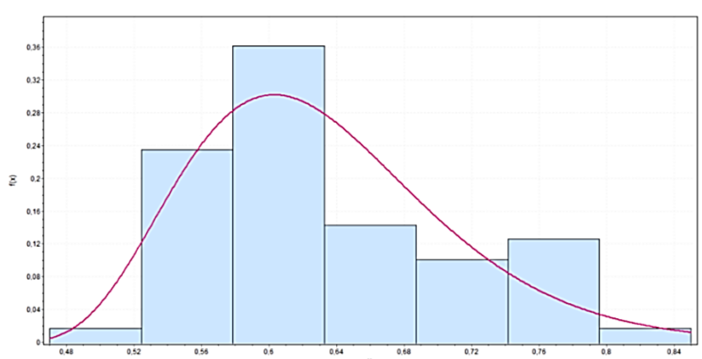

Figure 1. The pdf plot of the education index data

To support this statement, Figure 1 shows a pdf plot for the education index used as a response variable. The education index is the average of the expected years of schooling and mean years of schooling. Expected years of schooling is defined as the length of schooling (in years) that is expected to be felt by children at a certain age in the future. Then, mean years of schooling is the number of years used by the population in undergoing formal education.

Based on Figure 1, the most significant opportunity of last education that can be taken by the population of regency/city in Java is junior high school level. Population opportunities with last education until elementary school level are fewer than last education until junior high school level. Population opportunities with last education until senior high school and higher education level are fewer than last education until junior high school level but still higher than the opportunity for the last education until elementary school level. Therefore the distribution curve for the education index tends to have a longer tail to the right. This evidence is consistent with the characteristics of the gamma distribution, which has positive skewness.

Besides the response variable follows distribution of gamma, the second assumption that must be fulfilled is no multicollinearity. With the VIF formula, the VIF value for each predictor is 1.100 (for X_1), 1.524 (for X_2), and 1.642 (for X_3). Based on these results, it can be seen that each predictor variable has a VIF value of less than ten, so it can be concluded that there is no multicollinearity. Based on these conclusions, the percentage of households to private toilet ownership (X_1), population density (X_2), and percentage of poor people (X_3) can be used as predictor variables in this study.

Table 1. Data Description

Variable Names	**Mean**	**Standard Deviation**	**CV**	**Minimum**	**Maximum**
Y	0.632	0.077	12.210	0.470	0.850
X_1	80.215	9.710	12.100	37.820	98.010
X_2	3298	4493	136.250	278	19757
X_3	9.623	4.211	43.750	1.680	21.210

Table 1 presents data descriptions for response and predictor variables. The average value of the education index in Java, Indonesia, in 2018 is 0.632. This number increased by 0.005 points compared to 2017. The increase in education index was caused by the rise in the expected years of schooling and mean years of schooling. Various efforts have been taken to improve access and quality of education services. Vocational education is also continuous to be strengthened to enhance the productivity and competitiveness of the nation. In addition, strengthening the nation's character is also done with a cultural approach. Improving access to education services is done by improving and providing physical infrastructure for classrooms and school buildings. In 2018, the lowest education index was measured for Sampang Regency at 0.470 and the largest one for Yogyakarta at 0.850.

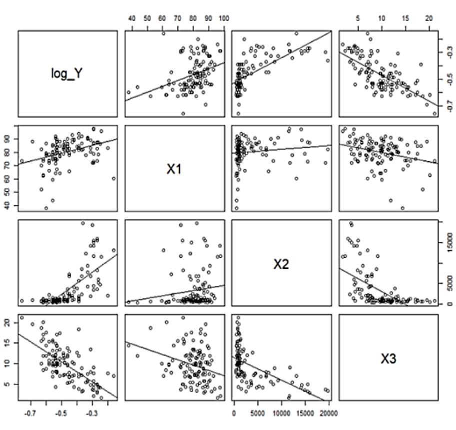

Figure 2. The matrix plot of the response and predictor variables

The matrix plot in Figure 2 can show the relationship between response and predictor variables. Education index (Y) has a positive relationship with the percentage of households to private toilet ownership (X_1) and population density (X_2), but the education index (Y) has a negative relationship with the percentage of poor people (X_3). This finding means that the greater the percentage of households to private toilet ownership (X_1) and population density (X_2), then the education index (Y) is also getting greater. The smaller the percentage of poor people (X_3), then the education index (Y) is getting greater. The correlation between log(Y) with X_1 is 0.362, log(Y) with X_2 is 0.705, and log(Y) with X_3 is -0.680. So the log correlation of response variable with population density (X_2) is greatest when compared to the log correlation of response variable with other predictor variables.

The following is a description that explains the relationship between the education index and the percentage of poor people. For a nation that wants to advance, education is a necessity. Same with housing, clothing, and food needs. The smallest nation is family, and education is a major need. Almost all levels of state schools have become commercialization institutions because they no longer speak to the requirements set by the curricular, but rather the amount of entrance fees for schools. In fact, the implementation of compulsory education is hindered, because to enter elementary school level now must pay dearly so that the poor may not be able to pay it. For the community and wealthy parents, their children will be able to attend public schools, while the poor will fail and not attend school. Entering private schools seems impossible for the poor. As a result, many children of the nation will not get the opportunity to get an education. It is truly a concern. Because, in a country that is more than 60 years old, many of its nation's children will be illiterate and left behind due to poverty, and this country will be left behind because the quality of its human resources cannot compete with other countries. The impact of poverty on education is tremendous. If poverty is not immediately overcome, then to achieve quality education very difficult, because, in modern times like today, competition is very tight, everything needs quality resources and can compete. Therefore, one way to improve the education index is to reduce the percentage of the poor population. Or in other words, the percentage of poor people has a negative relationship with the education index.

There are human physiological needs such as owning a house, which includes toilet facility as part of the needs of every family member. The existence of a toilet facility for families is one indicator of a healthy home in addition to ventilation doors, windows, clean water, landfills, sewerage, bedrooms, living rooms, and kitchens. According to Statistics Indonesia, toilet facility is divided into two groups, among others one's own and others (shared facilities are bathing, washing, and toilet for public and there is no / no use of toilet facility). In this study, the toilet facility used is one's own. The household that does not have a toilet facility usually have lower welfare levels so they cannot afford to pay for their children's education to a higher level. This phenomenon is, of course, affects the expected years of schooling and mean years of schooling, where these two indicators will form the education index. Based on the explanation above, the impact of health problems on education is enormous. The percentage of the household that has a toilet facility is positively related to the education index.

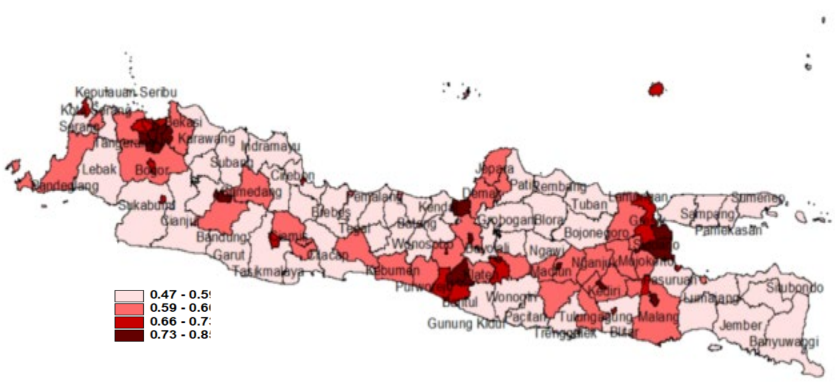

Figure 3. The education index for regency/city in Java

Figure 3 shows the maps of regencies in Java with color express the level value of each response data. Based on Figure 3, it can be seen that there are 21 regencies/cities that have a high education index (0.73 to 0.85). The regencies/cities are Magelang, Bantul, Tangerang, Central Jakarta, Bandung, Bekasi, Depok, Cimahi, Surakarta, Sidoarjo, Kediri, South Jakarta, East Jakarta, Salatiga, Surabaya, Malang, Madiun, Semarang, South Tangerang, Sleman, and Yogyakarta. There are 51 regencies/cities with the lowest level of education index (0.47 to 0.59).

Parameter estimation of the gamma regression model with a single predictor is presented in Table 2. With the significance level $\alpha = 15\%$, population density (X_2) has a significant effect on the education index. The intercepts when the gamma regression model employs X_1 and X_2 as a single predictor are significant.

The following is gamma regression model with a single predictor.

$\hat{\mu}_l = \exp\left(-0.834379 + 0.005310 X_{l1}\right)$ for the percentage of households to private toilet ownership (X_1),

$\hat{\mu}_l = \exp\left(-0.510688 + 0.000023 X_{l2}\right)$ for the population density (X_2), and

$\hat{\mu}_l = \exp\left(-0.196647 - 0.023516 X_{l3}\right)$ for the percentage of poor people (X_3).

Table 2. Parameter Estimation with Single Predictor

Parameter	**Estimated Value**	**Standard Error**	**Z score**	***p*-value**
$Y \sim X_1$				
β_{01}	-0.834379	0.178984	-4.662	0.000*
β_{11}	0.005310	0.004924	1.079	0.281
$Y \sim X_2$				
β_{01}	-0.510688	0.000029	-17397.340	0.000*
β_{21}	0.000023	0.000013	1.830	0.067*
$Y \sim X_3$				
β_{01}	-0.196647	0.646574	-0.304	0.761
β_{31}	-0.023516	0.023998	-0.980	0.327

* Significant at $\alpha = 15\%$

Table 3 presents gamma regression modeling with multiple predictor variables. The estimated value for the scale parameter is 0.017428 and the standard error is 0.000019, while the estimated value for the location parameter is 0.461486 and the standard error is 0.006204. With the significance level $\alpha = 15\%$, the scale parameter γ and the location parameter λ are the significant parameters. The Akaike Information Criterion (AIC) value is -367.837, the Corrected Akaike Information Criterion (AICc) value is -367.087, and the Mean Square Error (MSE) value is 0.003.

Table 3. Parameter Estimation with Multiple Predictors

Parameter	**Estimated Value**	**Standard Error**	**Z score**	***p*-value**
β_{01}	-0.601893	0.000295	-2041.540	0.000**
β_{11}	0.002807	0.003464	0.810	0.418
β_{21}	0.000017	0.000017	0.985	0.325
β_{31}	-0.011417	0.007057	-1.618	0.106**

** Significant at $\alpha = 15\%$

The gamma regression model for the education index is:

$\hat{\mu}_l = \exp\left(-0.601893 + 0.002807 X_{l1} + 0.000017 X_{l2} - 0.011417 X_{l3}\right).$

Hypothesis testing of regression parameters simultaneously can be done using test statistics is formulated in Rahayu et al. (2019), equation (14). The test statistics value is 286.339, and the value of $\chi^2_{0.15;3} = 5.317$. The decision is to reject the null hypothesis, which means that the percentage of households to private toilet ownership, the population density, and the percentage of poor people have a significant effect on the education index simultaneously. Hypothesis testing of regression parameters partially can be done using test statistics is formulated in Rahayu et al. (2019), equation (15).

Based on Table 3, the significant predictor variable that influences the education index is the percentage of poor people (X_3). We present Table 4 to show a comparison of the gamma regression models between single predictor and multiple predictors.

Table 4. The Difference of the Gamma Regression Models Between Single Predictor and Multiple Predictors

Gamma regression	***Y***		
	X_1	X_2	X_3
1. Multiple predictors	+	+	-***
2. Single predictor			
X_1	+		
X_2		+***	
X_3			-

*** Significant at $\alpha = 15\%$

Based on the results of Table 2 and Table 3, there are no sign differences in the regression coefficients for either the predictor variables X_1, X_2, and X_3. In the gamma regression with single predictor and multiple predictors, the regression coefficient on the predictor variables X_1 and X_2 has a positive sign, and the regression coefficient on the predictor variable X_3 has a negative sign. These empirical findings are consistent with the relationship between the predictor and response variables illustrated by the matrix plot in Figure 2. With the significance level $\alpha = 15\%$, population density (X_2) influencing the education index on the gamma regression with single predictor, and on the gamma regression with multiple predictors, the predictor variable is influential on the education index is the percentage of poor people (X_3). The following is the comparison between the actual value and the estimated values for the education index by regency/city in Java, Indonesia, in 2018.

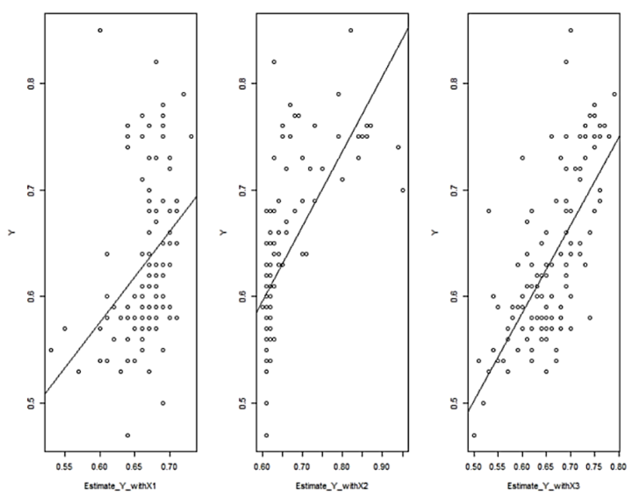

Figure 4. The actual values and the estimated values with a single predictor

Figure 4 shows the spread of education index values with the estimated values for single predictor X_1, X_2, and X_3. It can be seen that there is a difference between Figure 4 and Figure 5. Many estimated education index values for a single predictor are not around the regression line. This empirical result is inversely proportional to the estimated value of the education index for multiple predictors, where the majority of the estimated values are around the regression line. Also, the spread of the education index for single predictor X_1 and X_2 forms a specified pattern. Based on these two reasons, we conclude that gamma regression modeling with multiple predictors is better than gamma regression modeling with a single predictor.

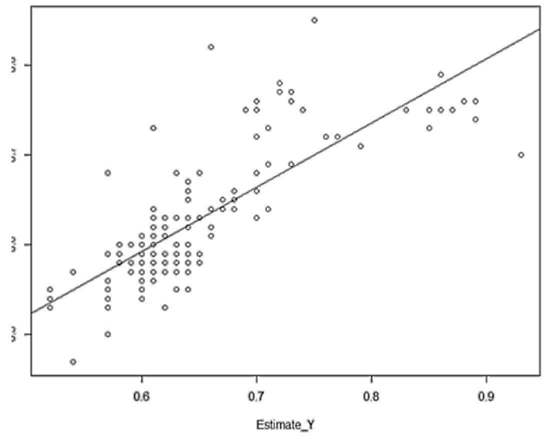

Figure 5. The estimated values and the actual values with multiple predictors

4. Conclusions and Future Research

The proposed gamma regression model in this paper includes hypothesis testing and parameter estimation methods. The method used for parameter estimation is MLE with numerical optimization to get the solution. In this study, numerical optimization employs the BHHH algorithm. The hypothesis testing simultaneously using MLRT and partially using the Wald test. The proposed gamma regression model is applied to model the education index of regency/city in Java with three predictor variables. Besides that, empirical results show that modeling using multiple predictors is better in terms of accuracy of prediction as well as the interpretation, than modeling using single predictor. The future work of this research is how to involve geographical factors as a spatial effect in the model.

Acknowledgment

This work is supported by Bina Nusantara University.

References

Balakrishnan, N. and Wang, J., Simple efficient estimation for the three-parameter gamma distribution, J. *Stat. Plan. Inference*, vol. 85, pp. 115–126, 2000.

Bono, R., Arnau, J., Alarcon, R. and Blanca, M. J., Bias, precision, and accuracy of skewness and kurtosis estimators for frequently used continuous distributions, *Symmetry*, vol. 12, no. 19, 2020.

Bravo, F., Li, D. and TjØstheim, D., Robust nonlinear regression estimation in null recurrent time series, *Journal of Econometrics*, vol. 224, issue 2, pp. 416-438, 2021.

Cameron, A. and Trivedi P., Microeconometrics: methods and applications, 2005.

Chakraborty, S. and Chakravarty, D., Discrete gamma distributions: properties and parameter estimations, *Communications in Statistics-Theory and Methods*, vol. 41, pp. 3301-3324, ISSN: 0361-0926 print/1532-415X online, doi: 10.1080/03610926. 2011.563014, 2012.

Chen, W. W. S. and Kotz, S., The riemannian structure of the three-parameter gamma distribution, *Journal of Applied Mathematics*, vol. 4, no. 3, pp. 514–522, doi: 10.4236/am.2013.43077, 2013.

Ewemoje, T. A. and Ewemooje, O. S., Best distribution and plotting positions of daily maximum flood estimation at Ona river in Ogun-Oshun river Basin, Nigeria *Agricultural Engineering International*, vol. 13, no. 3, pp. 1-13, 2011.

Marchand, E., Rancourt, F. and Strawderman W. E., Minimax estimation of a restricted mean for a one-parameter exponential family, *Journal of Statistical Planning and Inference*, vol. 212, pp. 114-125, 2021.

Mohammadpour, S., Niknam, N., Noughabi, J. J., Yousefi, M., Ebrahimipour, H., Haghighi, H., Kasraei, F., Kargar M. and Sharifi T., The factors associated with direct medical costs in patients with gastric cancer: quantile regression approach compared with gamma regression, *Value in Health Regional Issues*, vol. 21, pp. 127-132, 2020.

Nagar, D. K., Correa, A. R. and Gupta, A. K., Bimatrix variate kummer-gamma distribution, *Results in Applied Mathematics*, vol. 11, 2021.

Pan, J., Mahmoudi, M. R., Baleanu, D. and Maleki, M., On comparing and classifying several independent linear and non-linear regression models with symmetric errors, *Symmetry*, vol. 11, no. 820, 2019.

Pawitan Y., All likelihood: statistical modelling and inference using likelihood, 1st ed., Oxford, UK: Clarendon Press, pp. 41–42, ISBN 9780199671229, 2001.

Rahayu, A., Purhadi, Sutikno and Prastyo, D. D., Trivariate gamma regression, *IOP Conference Series: Materials Science and Engineering*, vol. 546, doi:10.1088/1757899X/546/5/052062, 2019.

Rahayu, A., Purhadi, Sutikno and Prastyo, D. D., Multivariate gamma regression: parameter estimation, hypothesis testing, and its application, *Symmetry*, vol. 12, no. 813, doi: 10.3390/ sym12050813, 2020.

Sallay, H., Bourouis, S. and Bouguila, N., Online learning of finite and infinite gamma mixture models for COVID-19 detection in medical images, *Computers*, vol. 10, no. 6, 2021.

Song, S., Kang, Y., Song, X. and Singh, V. P., MLE-based parameter estimation for four-parameter exponential gamma distribution and asymptotic variance of its quantiles, *Water*, vol 13, no. 15, 2021.

Tzougas, G., EM estimation for the poisson-inverse gamma regression model with varying dispersion: an application to insurance ratemaking, *Risks*, vol. 8, no. 97, 2020.

Usman, M., Zubair, M., Shiblee M., Rodrigues, P. and Jaffar S., Probabilistic modeling of speech in spectral domain using maximum likelihood estimation, *Symmetry*, vol. 10, no. 750, 2018.

Wang, H., Liao, H., Ma, X. and Bao R., Remaining useful life prediction and optimal maintenance time determination for a single unit using isotonic regression and gamma process model, *Reliability Engineering & System Safety*, vol. 210, 2021.

Biographies

Anita Rahayu received bachelor’s, master’s, and doctor’s degrees in Statistics Department from Institut Teknologi Sepuluh Nopember, Surabaya. She is a lecturer at the Department of Computer Science, School of Computer Science, Bina Nusantara University. She is currently Associate Lectuer Specialist S3 in Bina Nusantara University. Her research interests include statistical data analysis, statistical process control, and data science.

M. Aldiki Febriantono is a lecturer at the Department of Computer Science, School of Computer Science, Bina Nusantara University. He obtained bachelor's and master's degrees in electrical engineering at Brawijaya University, Malang. His expertise is in implementing technologies such as automation and IoT using Machine Learning and Artificial Intelligence.

Dedy Dwi Prasetyo graduated from Institut Teknologi Sepuluh Nopember, Surabaya with bachelor’s and master’s degree. He received doctor’s degree from Humboldt University of Berlin. His expertises are statistics, machine learning, business analytics, and financial econometrics.

Purhadi is a lecturer at the Statistics Department, Institut Teknologi Sepuluh Nopember, Surabaya. He obtained bachelor's degree in mathematics at Institut Teknologi Bandung, master’s degree at Waseda University - Japan, and doctor’s degree in mathematics at Gadjah Mada University, Yogyakarta. His expertises are inferential statistics and generalized linear model.

Sutikno received bachelor’s degree in Statistics Department from Institut Teknologi Sepuluh Nopember, Surabaya, master’s degree in Statistics Department from Institut Pertanian Bogor, and doctor’s degree in Climatology from Institut Pertanian Bogor. He is a lecturer at the Statistics Department, Institut Teknologi Sepuluh Nopember, Surabaya. His research interests include spatial statistics, environmental statistics, and extreme value modeling.