IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2320-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 9, Issue 9, September 2020

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.512

9940 572 462
6381 907 438
ijirset@gmail.com
www.ijirset.com

Gamma and Lognormal Models in the Generalized Linear Model Perspective

Nusyirwan[1], Netti Herawati[2], Khoirin Nisa[3*]

Assistant Professor, Department of Mathematics, University of Lampung, Bandar Lampung, Indonesia[1]

Associate Professor, Department of Mathematics, University of Lampung, Bandar Lampung, Indonesia[2]

Assistant Professor, Department of Mathematics, University of Lampung, Bandar Lampung, Indonesia[3]

*Corresponding Author

ABSTRACT:The aim of this study is to examine the influence of scale parameters that determine the coefficient of variance of both gamma and log-normal models in perspective of generalizedlinear model (GLM). The gamma and lognormal modelsare compared using the level of accuracy and precision through data simulation with scale parameters c = 0.5, 1.5, 2.5 and two different sample sizes n = 10, 20. The results show that the increasing value of the scale parameter c and sample size n make the estimated coefficient of the gamma regression coefficient more stable and more precise than the log-normal model, as well as the precision.

KEYWORDS:GLM, gamma distribution, log-normal distribution

I. INTRODUCTION

It is often found that the response variable Y is distributed binomial, Poisson, gamma, or exponential. To estimate the parameters in this case, the classical linear model such as linear regression can no longer be used because this model requires the assumption that the response variable Y is normally distributed. Generalized linear model (GLM) is a model developed to solve such problem. Generalized linear model can be used to model the response variable Y that follows an exponential family distribution [1-3].

In GLM, the binomial and Poisson distributions (see e.g. [4-6]) are models with a constant variance if there is no overdispersion. McCullagh and Nelder[7]argued that generally for continuous data, variance is positively correlated with the mean. So a constant variance coefficient is a more realistic assumption than a constant variance.

Gamma distribution as amember of exponential family distributions meets the requirement of GLM. Similar to gamma distribution, lognormal distribution is also suitable to be used in GLM even though it does not belong to the exponential family. Both gamma and lognormal distributions have properties of constant coefficient of variance.It isdifficult to distinguish between these two distributions in case of small values of coefficient variation (CV)< 0.7. For this reason, we study the influence of scale parameters that determine the coefficient of variance in both gamma model and log-normal model in perspective of GLM through data simulation.

II. GENERALIZED LINEAR MODEL (GLM)

GLM is a method for quantifying the relationship between predictor (independent) variables X with response (dependent) variable Y. The dependent variable Y is assumed to be following one of exponential family distributions including normal, Poisson, log-normal, and gamma distribution. The GLM generally can be expressed as follows:

$$E(\boldsymbol{Y}) = \mu = g^{-1}(\boldsymbol{X\beta}),$$

where$\boldsymbol{E(Y)}$ is the expected value of $\boldsymbol{Y}$,$\boldsymbol{X\beta}$ is a linear combination of unknown parameters; g is a link function.[7]

2.1. Gamma Model

Gamma distribution is used to describe a continuous positive random variable. It has two parameters, the shape parameter a and scale parameter c [3]. The probability density function of gamma distribution is as follows,

$$f(y) = \frac{1}{c^a\,\Gamma(a)} y^{a-1} e^{-\frac{y}{c}},$$

where$\Gamma(a)$ is the gamma function (for an integer value of $a$one has $\Gamma(a) = (a-1).(a-2).(a-3) = (a-1)!$). he expectation and variance of a$Y \sim Gamma\ (c, a)$variable are $E(Y) = \mu = a.c$dan$Var(Y) = a.c^2$, so the cofficient of variation is: $CV = \frac{\sqrt{ac^2}}{a.c} = 1/\sqrt{a}$. The parameter $\phi = \frac{1}{a} = CV^2$ is called the dispersion of the gamma distribution.Gamma distributions with the same shape-parameter have the same constant coefficient of variation.

Within the framework of a GLM, the distribution of gamma as an exponential family can be written into the following form,

$$f(y) = \frac{1}{y\Gamma(v)}\left(\frac{vy}{\mu}\right)^v exp\, \exp\left(-\frac{vy}{\mu}\right) \text{for } y\ \geq 0, v > 0, \mu > 0.$$

With variancefunction $V(\mu) = \mu^2$, and dispersion parameter $a(\phi) = \phi = v^{-1}$. While the canonical link function is$\eta = g(\mu) = \frac{1}{\mu}$. The gamma distribution parameter $\phi$is assumed to be, $\hat{\phi} = \frac{\chi^2}{n-p}$ whereχ^2 is a generalized Pearson statistic with $\chi^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}$. Here$V(\hat{\mu}_i) = \mu_i^2dann - p$are the degrees of freedom of residual [8].

2.2. Log-normal Model

Lognormal distribution comes from the normal distribution with the following definition: If X is a random variable with meanand variance σ^2, then the random variable $Y=e^X$ is said to have a lognormal distribution with parameters μ and σ^2.

If Y is lognormal distributed with parameters μ and σ^2, then X = ln(Y) is normal distributed with mean μ and variance of σ^2[4]. The probability density function of the lognormal random variable with parameters μ and σ^2 is as follows,

$$f(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(ln ln(x) - \mu)^2}{2\sigma^2}}, x > 0,$$

whereμ dan σ are the meanand standard deviationfrom the original variable. Then the lognormal random variable has mean $E(X) = e^{\frac{\mu+\sigma^2}{2}}$and variance $Var(X) = \left(e^{\sigma^2} - 1\right) e^{2\mu+\sigma^2}$ [8].

III. METHODOLOGY

The numerical simulation using R programming was conducted to distinguish gamma distribution and lognormal distribution in case of small values of coefficient variation (CV)< 0.7. We determined inverse link function $\mu_i = \frac{1}{\eta_i}$ and set the value of **X** and $\boldsymbol{\beta}$ so that η = $\boldsymbol{X\beta}$(for small and large values of $\frac{1}{\eta}$). We generate $y_i \sim Gamma(\mu_i), i =$ 1, 2, 3, …n, with n = 10, 20 and scale parameters c=0.5, 1.5, 2.5. Then we predict $\boldsymbol{\beta}$ of GLM using gamma model with inverse link function, and using ordinary least square (OLS) regression by log-normal transformation. We repeated the procedure 100 times. Finally, we measure the bias to seek for precision and accuracy of both models. The mean value of the gamma distribution is μ = $a.c$which is directly proportional to the scale parameter c and the shape parameter a. For that, if the value of c is small, then μ will decreases with the condition of constant a, this will cause the coefficient of variation (CV) to be constant.

IV. EXPERIMENTAL RESULTS

The standard error of parameters of the multiple regression model both gamma and log-normal models for n = 10 and c = 0.5, 1.5, 2.5 can be seen in Figures 1-9. Figures 1-3 show the standard error of b_0 for c = 2.5, so the gamma model is more precise than log-normal model. Likewise for the coefficient b_1 and b_2 as shown in figures 4-6 and 7-9, at value of c = 2.5 the precision of gamma model is higher than log-normal.

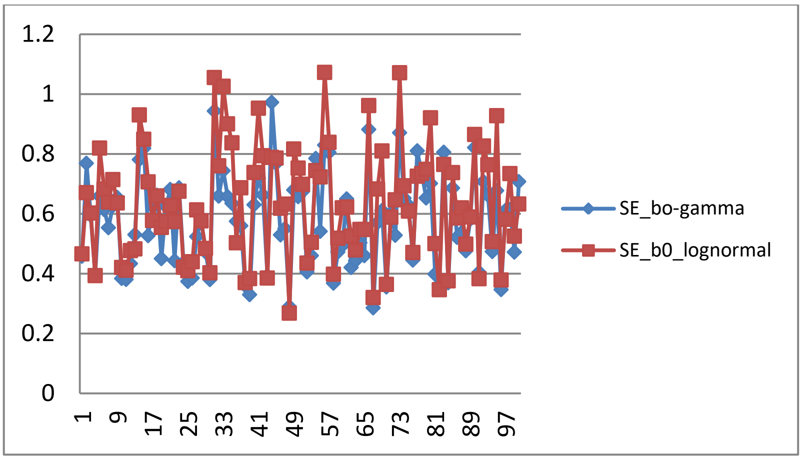


Figure 1. Standard error of b_0 with n=10 and c=0.5

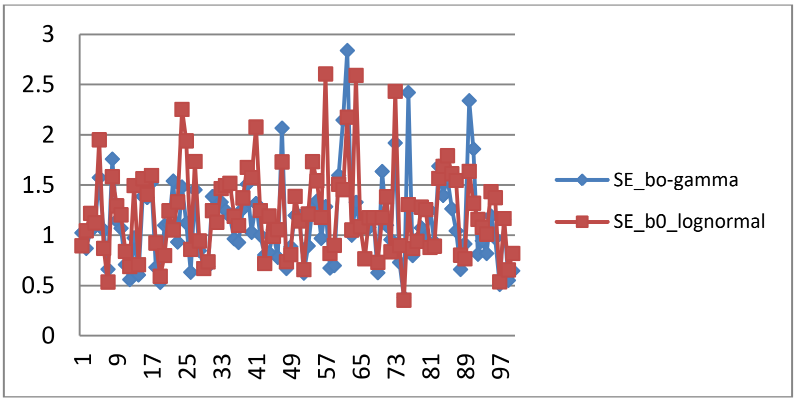


Figure 2. Standard error of b_0 with n=10 and c= 1.5

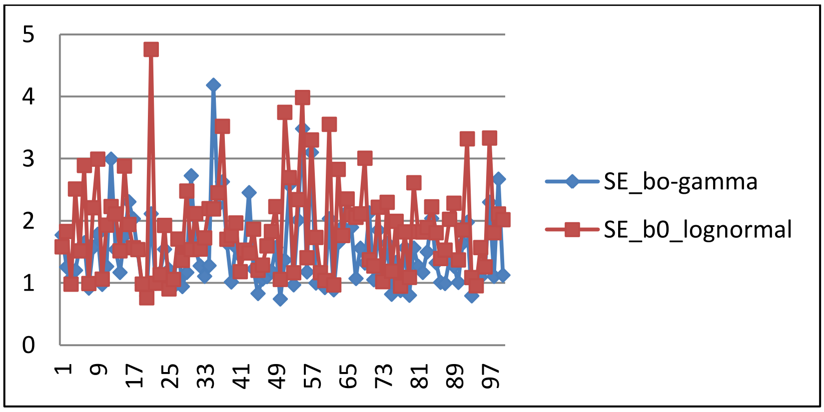


Figure 3. Standard error ofb_0 with n=10 and c=2.5

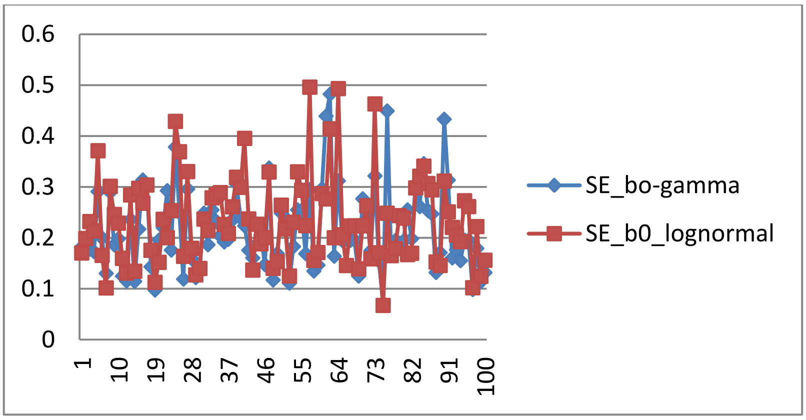


Figure 4. Standard error of b_1 with n=10 and c=0.5

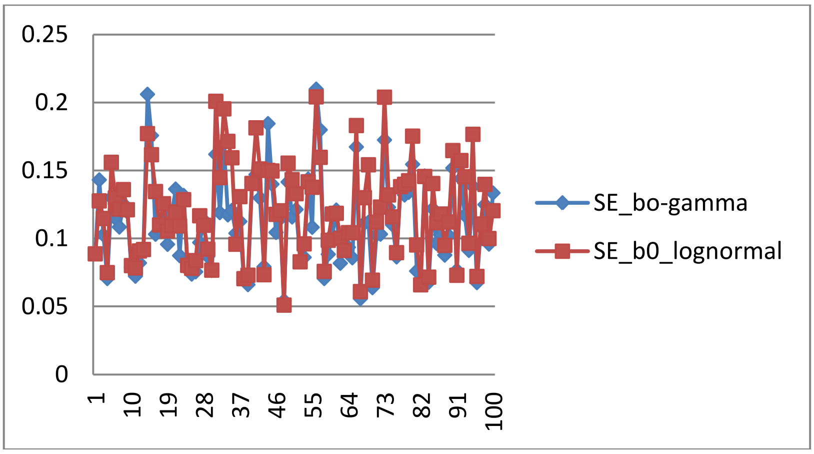


Figure 5. Standard error of b_1 with n=10 and c=1.5

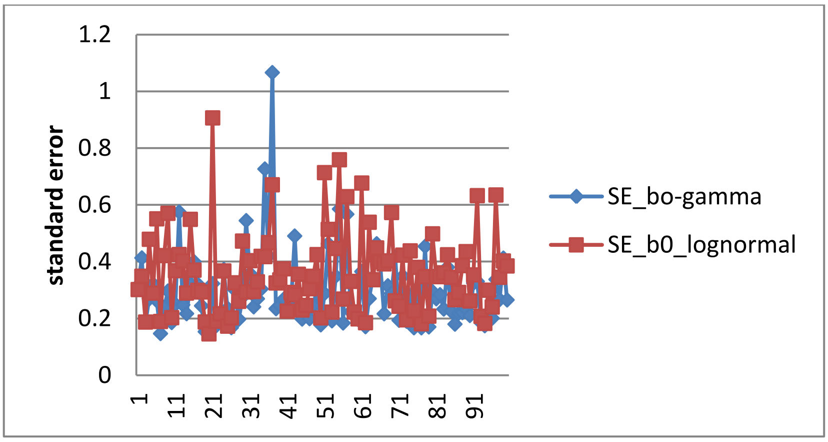


Figure 6. Standard error of b_1 with n=10 and c=2.5

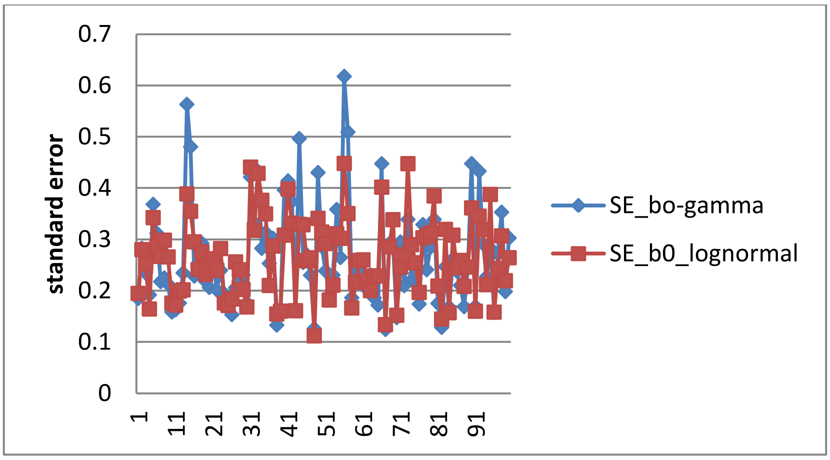


Figure 7. Standard error of b_2 with n=10 and c=0.5

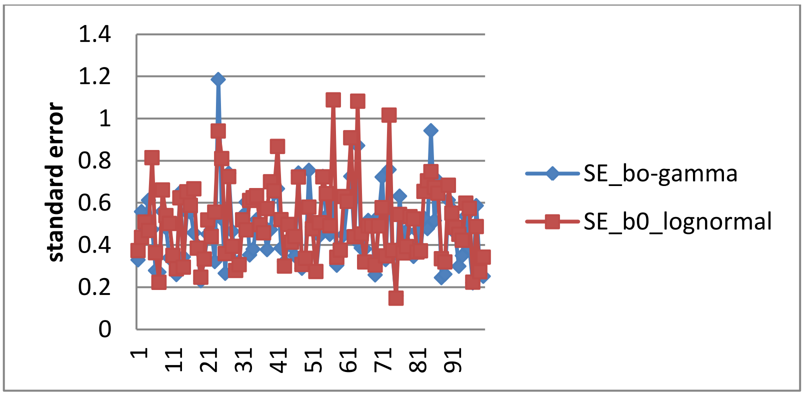


Figure 8. Standard error of b_2 with n=10 and c=1.5

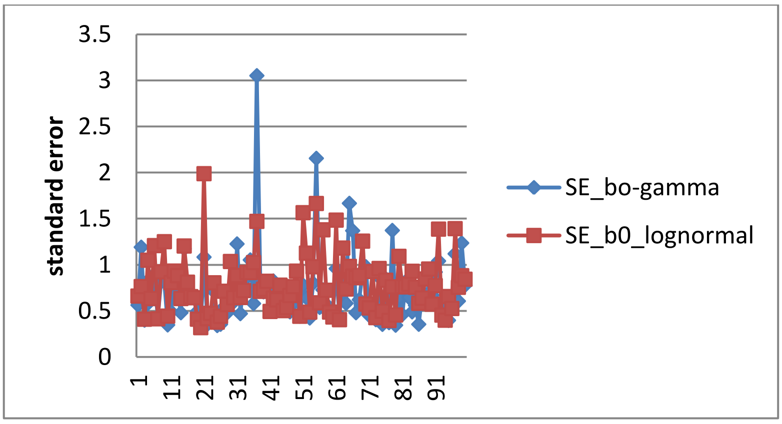


Figure 9. Standard error of b_2 with n=10 and c=2.5

The standard error of parameter estimates of the multiple regression models both gamma and lognormal models for n = 20 and c = 0.5, 1.5, 2.5 are presented in Figure 10-18. In Figure 10-12, the standard error of b_0 for c = 2.5, gamma model is more stable and more precise than the lognormal model. Likewise, for coefficient b_1 and b_2 as shown in Figures 13-18, at c = 2.5, the gamma model has parameter estimates that are considered more stable and more precise than lognormal model.

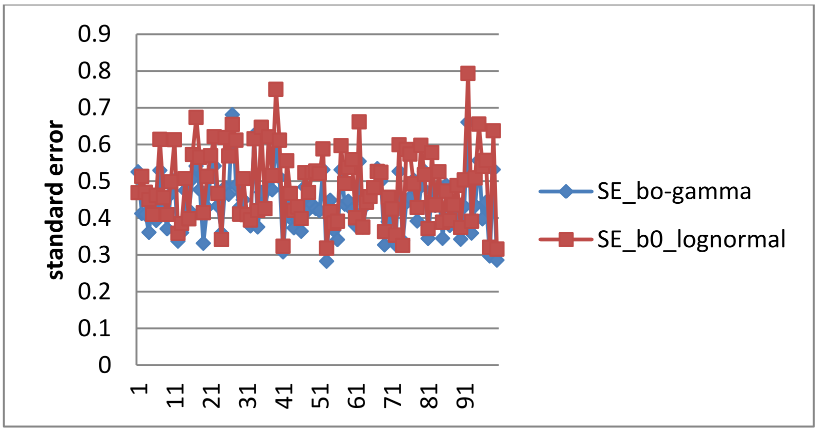


Figure 10. Standard error of b_0 with n=20 and c=0.5

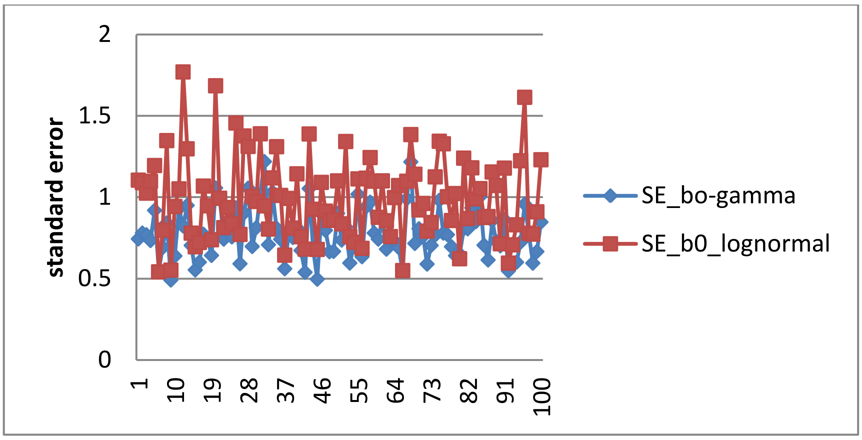


Figure 11. Standard error of b_0 with n=20 and c=1.5

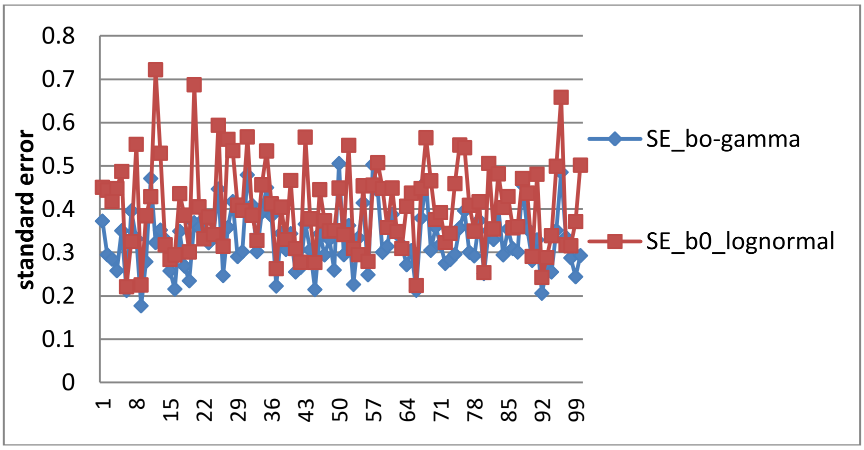


Figure 12. Standard error of b_0 with n=20 and c=2.5

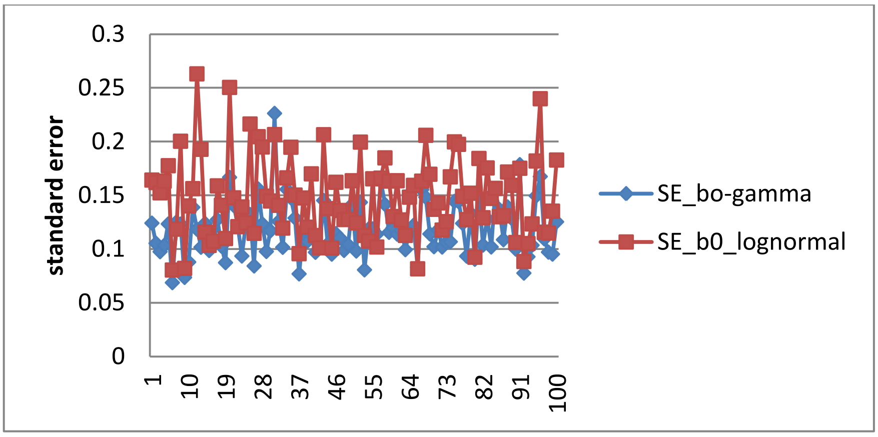


Figure 13. Standard error of b_1 with n=20 and c=0.5

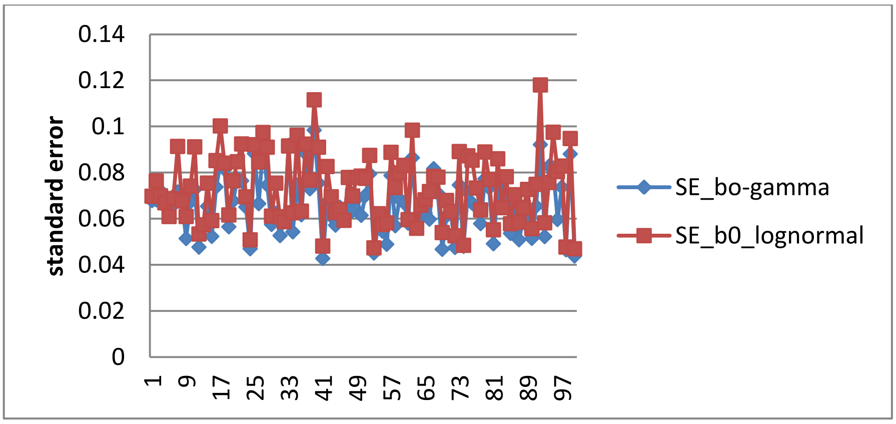


Figure 14. Standard error of b_1 with n=20 and c=1.5.

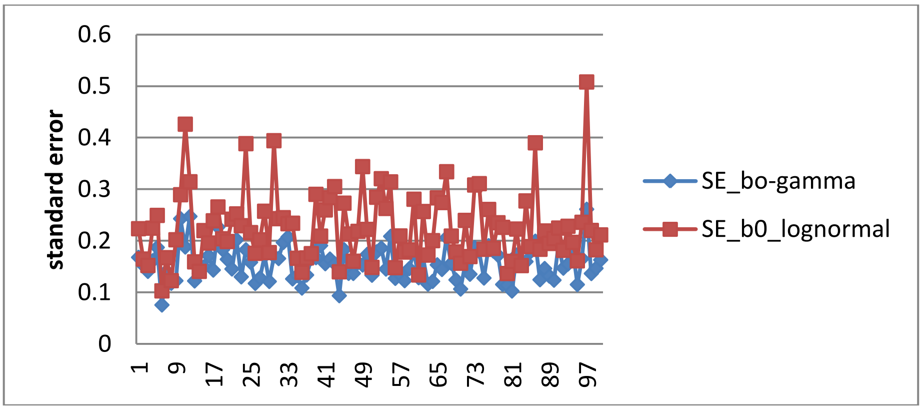


Figure 15. Standard error of b_1 with n=20 and c=2.5

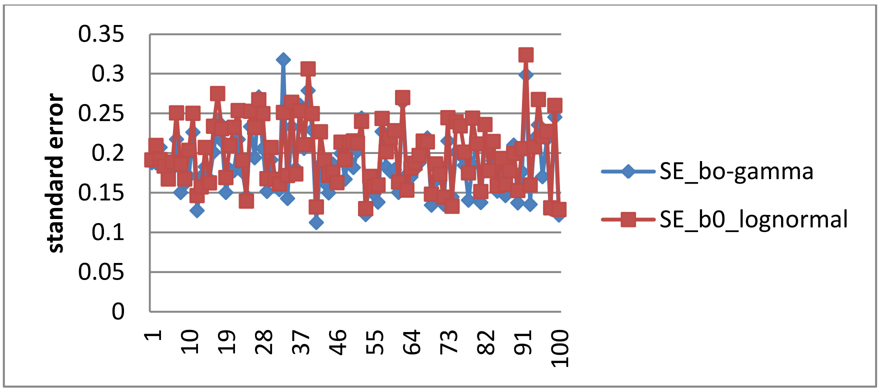


Figure 16. Standard error of b_2 with n=20 and c=0.5

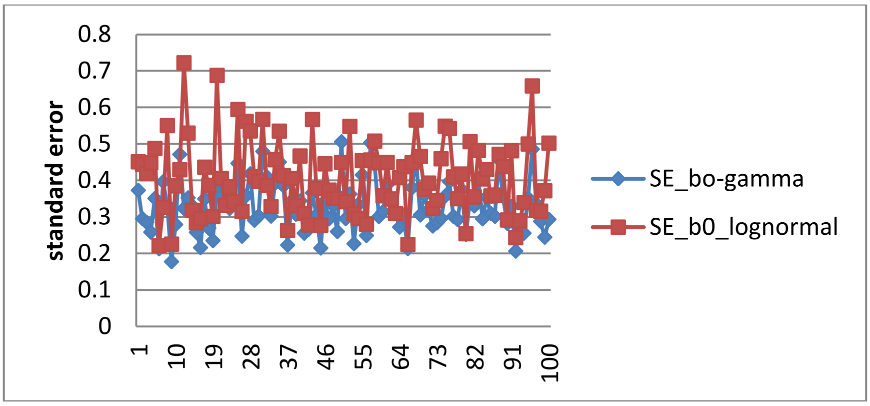


Figure17. Standard error of b_2 with n=20 and c=1.5

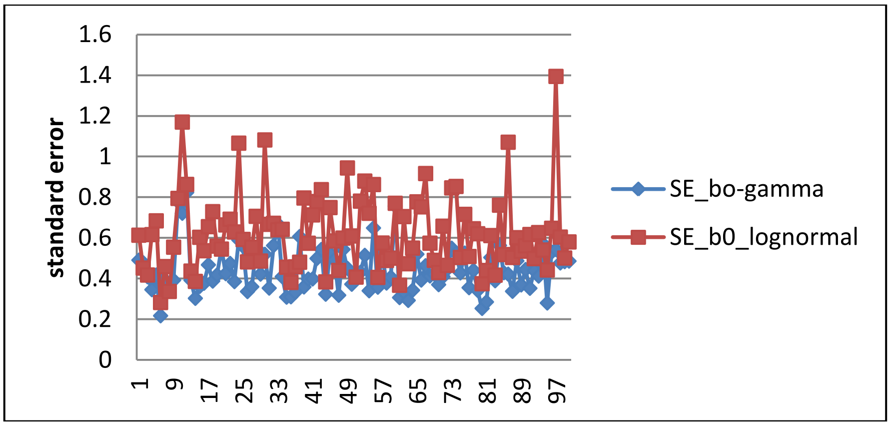


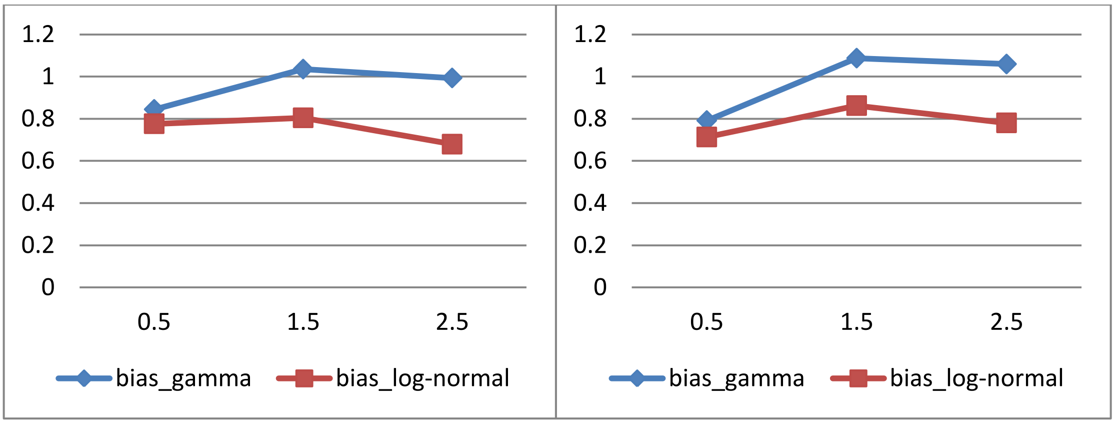
Figure 18. Standard error of b_2 with n=20 and c=2.5

From the simulation results it can be seen that the increasing value of the scale parameter c and sample size n will make the estimated coefficient of the gamma regression coefficient more stable and more precise than the log-normal model.

In addition to precision, the accuracy of the gamma and lognormal estimation parameters is also checked. The bias of the simulation results is presented in Table 1. From Figure 7a-7b, the bias values of parameters b_0 and b_1 for n = 10, the log-normal model has smaller bias than the gamma model, in contrast to figure 7c for n = 10 and the bias value of b_2 of the gamma model is almost the same as lognormal. The pattern ofbias for n = 20 ia almost the same as as$n$ = 10. The effect of the scale parameter c on the estimation accuracy seems clear for n = 10 and n = 20, that the value c = 0.5 has smaller bias value compared to c = 1.5 danc = 2.5 for both the gamma model and the log-normal model.

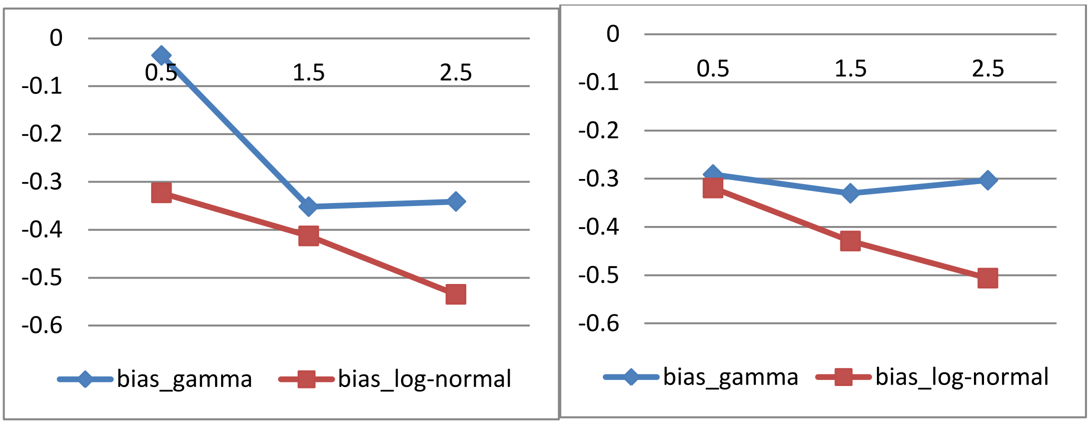
Table 1. Bias of parameter estimates of gamma and lognormal models

n	Parameter estimate	Scale (*c*)	Bias	
			gamma model	log-normal model
10	b_0	0.5	0.844	0.776
		1.5	1.035	0.804
		2.5	0.993	0.679
	b_1	0.5	-0.036	-0.323
		1.5	-0.352	-0.413
		2.5	-0.341	-0.535
	b_2	0.5	-0.094	-0.108
		1.5	-0.171	-0.194
		2.5	-0.251	-0.212
20	b_0	0.5	0.791	0.713
		1.5	1.087	0.862
		2.5	1.059	0.780
	b_1	0.5	-0.291	-0.319
		1.5	-0.330	-0.429
		2.5	-0.303	-0.506
	b_2	0.5	-0.066	-0.068
		1.5	-0.154	-0.155
		2.5	-0.203	-0.218



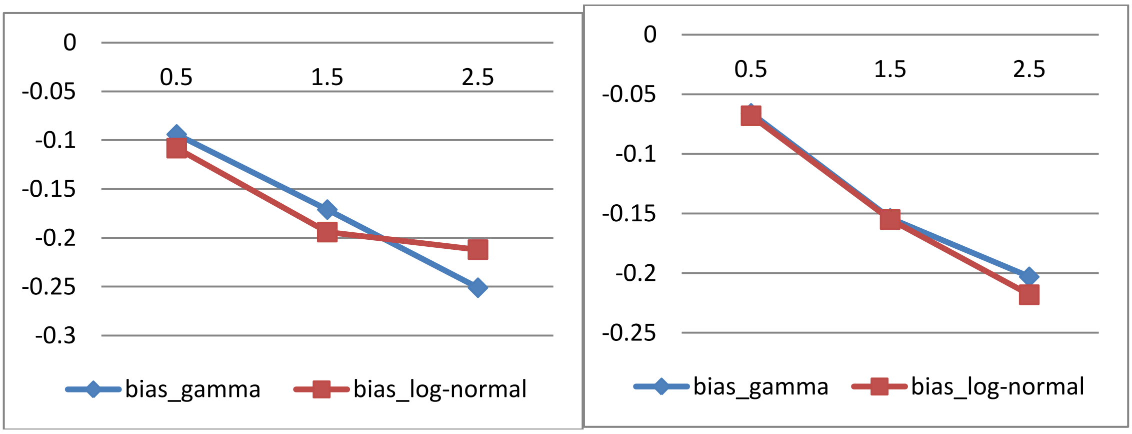
(a). *n=10* (b) *n*=20

Figure 19. Plot of parameter scale *c*vs bias for for parameter estimateb_0



(a). *n=10* (b) n=20

Figure 20. Plot of parameter scale *c*vs bias for for parameter estimateb_1



(a). *n=10* (b) n=20

Figure 21. Plot of parameter scale *c*vs bias for for parameter estimateb_2

V. CONCLUSION

From the results of the study can be concluded both scale parameter(c) and sample size affect the standard error of parameter estimate of the gamma model and lognormal model. Increasing the value of the scale parameter c and sample size will make the estimated parameters of gamma model more stable and more precise than lognormal model. In addition, small scale parameters c will reduce the bias of gamma nad lognormal parameters models so that the model becomes accurate, but does not apply to sample sizes.

REFERENCES

[1] Herawati, N. andNisa, K.“A Robust Procedure for GEE Model”, Far East Journal of Mathematical Sciences: vol102 (03)pp 645-654, 2017

[2] Nisa, K.and Herawati, N. “Robust Generalized Estimating Equation When Data Contain Outliers”, International Series of Interdisciplinary Science and Technology (INSIST),vol02 (01) pp1-5, 2017

[3] Agresti, A. “An Introduction to Categorical Data Analysis Second Edition”, John Wiley & Sons, NewJersey, 2007

[4] Nisa, K.,Kokonendji, C.C., Saefuddin, A., Wigena, A.H., Mangku, I.W. “On Generalized Variance of Normal Poisson Model and Poisson Variance Estimation UnderGaussianity”, ARPN Journal of Engineering and Applied Sciences, vol12 (12)pp 3836-3842, (2017).

[5] Nisa, K.,Kokonendji, C.C., Saefuddin, A., Wigena, A.H., Mangku, I.W. “Empirical Comparison of The Generalized Variance Estimators of Some Normal Stable Tweedie Models: A Simulation Study”, Applied Mathematical Sciences,vol10 (63)pp 3107-3118, 2016

[6] Nisa, K.,Kokonendji, C.C., Saefuddin, A. “Characterizations of Multivariate Normal Poisson Model”, Journal of Iranian Statistical Society, vol14 (2)pp 37-52, 2015
[7] McCullagh, P. and Nelder, J.A. “Generalized Linear Models. 2nd ed”, Chapman &Hall/CRC, Boca Raton, Florida, 1989
[8] Navidi, W. “Statistics for Engineers & Scientist. 4th Edition”, McGraw-Hill, New York, 2015

INNO SPACE

SJIF Scientific Journal Impact Factor

Impact Factor: 7.512

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

Scan to save the contact details