Modeling Indonesia Inflation Data: Gamma-Normal or Normal

By Herlina Hanum Herlina Hanum

Modeling Indonesia Inflation Data:

Gamma-Normal or Normal

Herlina Hanum

Department of Mathematics, Sriwijaya University
Kampus Inderalaya Km 32, Ogan Ilir
South Sumatra, Indonesia

Novi Rustiana Dewi

Department of Mathematics, Sriwijaya University
Kampus Inderalaya Km 32, Ogan Ilir
South Sumatra, Indonesia

Dwi Setyawan

Department of Soil Science, Sriwijaya University
Kampus Inderalaya Km 32, Ogan Ilir
South Sumatra, Indonesia

Abstract

Indonesia inflation consists of positive and negative values with almost symmetrical distributions. In general, the data better fit a gamma-normal distribution. In some cases, it also fits a normal distribution. Regression analysis for data which follows gamma-normal might be done, through certain transformation and link function, using generalized linear model (GLM) gamma. On the other hand, there is an easy and famous ordinary linear regression for normally distributed data. For this set of data, both analyses give a very good estimate with MAPE less than 5 %.

1 Introduction

Gamma-normal distribution is developed by Alzaatreh et al. [1]. It is one of gamma based distribution. This distribution combined the goodness of gamma and normal distribution. Since the two parameters α and β provide a great deal of flexibility, gamma distribution is used to model many nonnegative random variables of the continuous type [6]. On the other hand, a normal distribution is usually used for modeling symmetry distributed data. Since it is defined in (-∞, ∞), it can be used for data with negative values. As the consequent, the gamma-normal distribution might be used for modeling asymmetry data which contains some negative values.

Hanum et al. [4] note that regression model for Gamma-Pareto distributed is GLM. The analysis of GLM gamma-Pareto, with certain transformation, may use GLM gamma. Alzaatreh et al. [1] note that $Z = -\log(1 - \Phi(Y))$ transform a gamma-normal random variable X into a gamma random variable Y, where $\Phi(.)$ cumulative distribution function of normal distribution. Accordingly, regression analysis for gamma-normal distributed data may also use GLM gamma. Meanwhile, it has been known that the regression model for normally distributed data is ordinary linear regression model. The model assumes that the response variable is normally distributed [3]. So does the error term. Both GLM gamma and ordinary regression model have been available in some statistical software i.e. SAS, Minitab, and R.

Indonesia inflation data represent monthly inflation in Indonesia. The data are provided by Indonesian Central Bureau of Statistics (Indonesia Badan Pusat Statistik-BPS). The inflation will be positive if money value decreases. In contrary, the inflation is negative. The data consist of general inflation, foodstuff, transportation, etc. In this research, Indonesia inflation data are analyzed using models for gamma normal dan normal distributions. Firstly, data are tested for goodness of fit for two distributions. Subsequently, regression analysis is conducted using best-fitted regression models for individual distributions. Independent variable is foodstuff inflation.

2 Gamma-Normal distribution and parameter estimation

Gamma-normal distribution has pdf

$$g(y) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \phi(y) \left(-\log(1 - \Phi(y))\right)^{\alpha-1} (1 - \Phi(y))^{\frac{1}{\beta}-1}, -\infty < y < \infty.$$

Parameters estimation of gamma-normal use the log likelihood function

$$\log L(\alpha, \beta, \mu, \sigma)$$

$$= -n\log\Gamma(\alpha) - n\alpha\log\beta - \frac{n}{2}\log(2\pi) - n\log(\sigma)$$

$$- \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2$$

$$+(\alpha - 1)\sum_{i=1}^{n} \log(-\log(1 - \emptyset(y_i))) + (\beta^{-1} - 1)\sum_{i=1}^{n} \log\log(1 - \emptyset(y_i)).$$

Initial values of μ and σ are the mean and variance of Y. Meanwhile, initial values of α and β are the parameters of $Z = -log(1 - \Phi(Y))$ which follows a gamma distribution. The initial value of α and β respectively are $\alpha_0 = \bar{z}^2/\text{var(z)}$ and $\beta_0 = \frac{\text{var(z)}}{\bar{z}}$ as defined by Alzaatreh et al. [1].

3 Research Methods

The data are general inflation and raw food inflation. The data collected by BPS from January 2006 until August 2017.

All computation in this research is using R software.

1. Provide initial value of $\alpha, \beta, \mu,$ and σ.
2. Fitting data Indonesian inflation (Y) with gamma-normal distribution using mle2 function in bbmle package. The function will provide log likelihood and
   $AIC = -2log(\hat{\theta}|x) + 2k$ with $log(\hat{\theta}|x)$ is the log-likelihood value and k independent variables as defined by Burnham and Anderson [2].
3. Test the goodness of fit using Kolmogorov-Smirnov (KS)
   $$D_{n1,n2} = \frac{Max}{x}|F_{X,n1}(x) - F_{Y,n2}(x)|$$ as shown by Hassani & Silva [5].
4. If Y follows gamma-normal distribution then regress Y with *bm*, the raw food inflation.
5. Determine the goodness of estimation $MAPE = \frac{1}{n}\left(\sum_{1}^{n}\left|\frac{actual-estimate}{actual}\right|\right)100\%$ as defined by Moreno et al. [7], between y and ŷ for the gamma-normal model.
6. Fitting data Y with normal distribution using fitdistr function in MASS package, test the goodness of fit with KS
7. If Y fits normal distribution then Analyze Y dan *bm* with linear regression.
8. Get the fitting values (ŷ)
9. Determine MAPE between y and ŷ for linear regression

4 Results and discussion

Fitting Indonesia Inflation to Gamma-Normal

The general inflation of Indonesia has minimum -0.45, maximum 3.29, median 0.385, mean 0.4747, and standard deviation 0.5509. The data are right skew with skewness 0.54 (Figure 1). Since the data likely did not have symmetry

distribution, the data may not be fit to a normal distribution. But it also could not fit gamma distribution since it has some negative values. These are the reason we fit it with gamma-normal distribution.
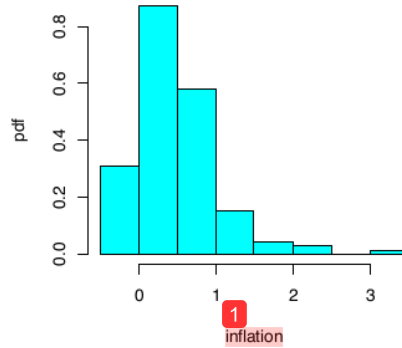

Figure 1. Distribution of general inflation in Indonesia.

Here we use two types of Gamma-Normal following Alzaatreh et al [1]. They are gamma-normal with 2 parameter (denoted as GN-2) which is gamma-standard normal with parameters $\mu = 0$ and $\sigma = 1$, and 4 parameters gamma-normal (denoted as GN-4). Parameter estimation for GN-2 and GN-4 starts with determining initial values of α, β, μ, and σ. In GN-2 μ and σ are fix with values 0 and 1 respectively. Meanwhile, the initial value for parameter μ and σ in GN-4, which are the mean and variance of Y, are 0.4747 and 0.5508 respectively. In order to provide an initial value for α and β, transform Y with $Z = -log(1 - \Phi(Y))$. The initial value of α and β respectively are $\alpha_0 = \bar{z}^2/var(z) = 0.3563$ and $\beta_0 = \frac{var(z)}{\bar{z}} = 2.9233$.

Table 1. Characteristics of distributions using 2 and 4 parameters

Distributions	Gamma-Normal 2 Parameter	Gamma-Normal 4 Parameter
Parameter estimate	$\alpha = 3.7384564$ $\beta = 0.3378258$ $\mu = 0$ $\sigma = 1$	$\alpha = 0.85173$ $\beta = 2.62665$ $\mu = 0.18216$ $\sigma = 0.38083$
Log-likelihood	-114.57	-110.03
AIC	233.1323	228.0501
K-S p-value	0.5821	0.7805

The result of parameter estimation is presented in Table 1. Both types of gamma-normal distribution have the p-value of K-S test larger than the level of significant 0.05. This result leads to the conclusion that the data fit both types of gamma-normal distribution. The table also shows that AIC of GN-4 smaller than AIC of GN-2. On the other hand, the p-value of K-S test of GN-2 larger than the p-value of GN-2. It means that the data better fit GN-4.

Regression model of Gamma-normal data

Following Hanum et al. [4] and Alzaatreh et al. [1] the gamma-normal distributed data may be analyzed using glm gamma. The transformation is $Z = -log(1 - \Phi(Y))$. Theoretically, Z will follow the gamma distribution. The food commodities inflation (bm) is used as an independent variable. Variable Z with *bm*, are analyzed using glm gamma. Here we used logarithmic link function. The analysis provides the estimate of Z that is $\hat{z}$. The estimate of Y is yielded by transformation $\hat{y} = \Phi^{-1}\left(1 - e^{-\hat{z}}\right)$.

Log link function is the best log link for the data. Analysis the data using glm gamma gives a highly significant estimation of parameters in the model of glm gamma with p-values 2e-16. The regression model of gamma-normal is

$$\hat{y} = \Phi^{-1}\left(1 - e^{-(\exp(-0.87+0.6902\,bm))}\right).$$

MAPE 4.94 of this estimation indicates that the gamma-normal regression model is very good in estimating general inflation of Indonesia.

Modeling inflation with Normal distribution

Despite the histogram of the data is seemingly not symmetry, we try to fit the data with normal distribution. The estimated parameter are the mean and variance of the data which are $\mu = 0.4747$ and $\sigma = 0.5509$. This estimation has log likelihood, AIC, and the p-value of KS test respectively are -115.163, 234.326, and 0.6826. The p-value indicates that the data also fit the normal distribution.

The value of log likelihood and AIC of the normal distribution are less than those of GN4. On the other hand, the p-value of KS test of GN4, that is 0.7805, is larger than the p-value of the normal distribution. These facts note that the data less fit normal than GN4.

Since the data still fit to normal distribution, we analyzed Y with *bm* using the famous ordinary linear regression. The result is copied below

```
Coefficients:
        Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.23754   0.02767  8.584 1.58e-14 ***
bm          0.34175   0.01823 18.751  < 2e-16 ***

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2951 on 140 degrees of freedom
Multiple R-squared: 0.7152,   Adjusted R-squared: 0.7132
F-statistic: 351.6 on 1 and 140 DF,  p-value: < 2.2e-16
```

The p-value of F and t statistics say that the test is significant at a level less than 0.001. This result is supported by the MAPE 0.8163. The tests and MAPE conclude that the ordinary linear regression is very good in estimating general inflation.

The comparison

As one point of view, here we represent the plot of estimated quantile of normal and four parameters gamma-normal distribution for general inflation (Figure 2). The two quantile give an almost overlap line. Furthermore both normal and gamma-normal could not estimate 6 out of 140 values of the inflation those upper 1.5. These mean there is no significant difference between normal and gamma-normal in modeling this data, even though the data better fit gamma-normal.
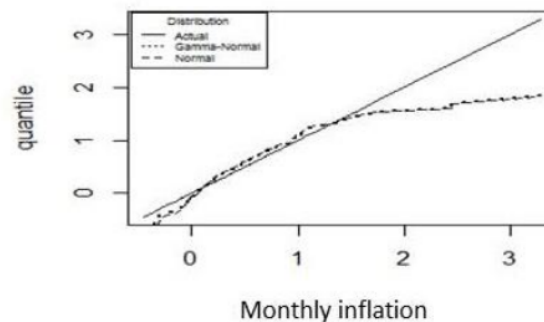

Figure 2. Quantile plot of gamma-normal and normal for general inflation.

Secondly, the regression model using normal distribution has a better estimate than regression of gamma-normal. It can be seen from MAPE value 0.8163 of normal compared to 4.9464 of gamma-normal. Although both models are very good to estimate general inflation, analyzing data using ordinary linear regression is much easier than gamma-normal regression. It is not necessary to transform to gamma then back transform to get the estimated value of Y. We also

do not need to choose the link function that suitable for the relationship between response and independent variable. Most of all, the regression model of gamma-normal is not easy to interpret.

The question is when we should use gamma-normal model. When the distribution of the data almost symmetry (the skewness close to zero) the data will be normally distributed. The data of breaking stress of carbon fiber and Strength of 1.5 cm glass fiber data which is used by Alzaatreh et al. [1] is very good fit to gamma-normal distribution. Unfortunately, similar with general inflation data, the data also very good fit to a normal distribution with the p-value of KS.test 0.9708. p-value = 0.1108 On the other hand, when the data have a quite long right tail the data did not fit gamma-normal. Examples of this situation are data of Inflation of transportation with skewness 3.4261 and Inflation of processed food 2.2538. Since Inflation of processed food data only contains positive values, it fits gamma distribution. It has been known that gamma distributed data is usually analyzed using glm gamma.

Conclusion

Inflation data which is gamma-normal distributed also follow a normal distribution. For this type of data, ordinary linear regression is much easier than gamma-normal regression. It may be better to fit data to normal distribution first, before testing the fitness to gamma-normal distribution.

References

[1] A. Alzaatreh, F. Famoye and C. Lee, The gamma-normal distribution: Properties and applications, *Computational Statistics and Data Analysis*, **69** (2014), 67–80. https://doi.org/10.1016/j.csda.2013.07.035

[2] K.P. Burnham and D.R. Anderson, *Model Selection and Multimodel Inference -A Practical Information-Theoretic Approach*, Springer-Verlag, New York, 2002. https://doi.org/10.1007/b97636

[3] J.J. Faraway, Practical Regression and Anova using R. Retrieved from https://cran.r-project.org/doc/contrib/Faraway-PRA.pdf, 2002.

[4] H. Hanum, A.H. Wigena, A. Djuraidah and I.W Mangku, Modeling Gamma-Pareto distributed data using GLM Gamma, *Global Journal of Pure and Applied Mathematics*, **12** (2016), 3569–3575.

[5] H. Hassani and E.S. Silva, A Kolmogorov-Smirnov Based Test for Comparing the Predictive Accuracy of Two Sets of Forecasts, *Econometrics*, **3** (2015), 590–609. https://doi.org/10.3390/econometrics3030590

[6] R.V. Hogg, A.T. Craig and J.W. McKean, *Introduction to Mathematical Statistics*, (7th ed.), Pearson Education, Inc, Boston, 2013.

[7] J.J.M. Moreno, A.P. Pol, A. S. Abad and B.C. Blasco, Using the R-MAPE index as a resistant measure of forecast accuracy, *Psicothema*, **25** (2013), 500–506.

Modeling Indonesia Inflation Data: Gamma-Normal or Normal