ICORHESTECH 2024
Vol. 1, No. 1, 2024, pp. 369~374
Journal Homepage: https://journal.ibrahimy.ac.id/index.php/icorhestech

Robust Regression Analysis Of Gm Estimation On The Poverty Gap Index Of Indonesian Provinces

Fauzaan Nabil[1], Yuliana Susanti[2], Etik Zukhronah[3]
[1,2,3]Statistics Departement, Sebelas Maret University, Central Java, Indonesia
*Corresponding author: *fauzaanardi@student.uns.ac.id*

ABSTRACT

Poverty has been a severe problem in Indonesia since the post-independence era until today. One indicator that can be used to measure the poverty level in a region is the poverty gap index, which describes the average size of the gap between each population and the poverty line. The data on the poverty gap index in Indonesia in 2022 contains outliers and misleading data that are not normally distributed, so the least squares method is inappropriate. One method that can be used to overcome the outlier problem is robust regression analysis. This study aims to determine the Generalized M (GM) estimation robust regression model and the factors that affect the poverty gap index in provinces in Indonesia. The estimation used is GM estimation, the development of M estimation when M estimation is less sensitive to outliers. The results showed that the GM estimation robust regression model has a value of 100% R^2_{adj} It was also found that the factors that significantly affect the poverty gap index in Indonesia in 2022 are the percentage of poor people, the Gini ratio, the poverty line, and the percentage of households with a PLN electricity source.

Keywords:

Poverty Gap Index; Robust Regression; GM Estimation.

INTRODUCTION

Various aspects of life can be caused by one crucial factor in life, namely the economy. A poor economy or poverty can impact multiple elements of multidimensional aspects of life, such as health, education, psychology, crime, and various other components. The best program to overcome poverty is based on the concept of poverty experienced and felt by the community (Maipita, 2014).

Poverty has been a severe problem in Indonesia since the post-independence era until now. According to the Badan Pusat Statistic (BPS), poverty can be defined as the inability of people to meet their needs, namely primary and secondary needs, as one of the minimum standards for fulfilling basic needs (Cahyat, 2004). The Poverty Gap Index, which describes the average size of each population's gap against the poverty line, is one indicator that can be used to measure poverty in a region or country.

Regression analysis is a statistical analysis used to investigate and model the relationship between two or more variables. The Ordinary Least Square (OLS) is often used in regression analysis. In applying the OLS method, there is a possibility of cases where there are outliers in the data. One solution that can be used to analyze data containing outliers so that the model can be resistant and robust to these outliers is to use robust regression (Freund et al., 2006).

Several outliers violate the normality assumption in Indonesia's poverty gap index data, so the results will not be optimal if analyzed using regression analysis. In this case, robust regression analysis is the best method (Susanti et al., 2014). In this research, the estimation that will be used is the Generalized M (GM) estimation. GM estimation is a development of M estimation when M estimation is less resistant to outliers in the variables x_i. The basic idea behind GM estimation is to limit the influence of outliers on the variable by using a weighting function that is only resistant to outliers in the variable x_i by using a weighting function that only depends on the outliers x_i (Junus, 2021).

According to Kusuma et al. (2019), in research on modeling the poverty gap index in Indonesia using robust regression analysis of M estimation with Huber weights, the results show that the factors that significantly affect the poverty gap index are the percentage of poor people, the

gini ratio, and the percentage of households using lighting with electricity sources (40% and below).

This study aims to obtain a GM estimation robust regression parameter estimation model on poverty gap index data in Indonesia in 2022. The independent variables used are the percentage of poor people, the Gini ratio, the poverty line, and households with PLN electricity lighting sources. The benefit of this research is that it can obtain a robust regression model of the poverty gap index for each province in Indonesia and help the government and the community implement appropriate policies to reduce the depth of poverty, especially for provinces with a high poverty gap index value.

METHOD

This research uses quantitative methods because hypothesis testing, statistical analysis, and interpretation of the analysis results are used. The object of this research is the poverty gap index data for each province in Indonesia in 2022. The data used is secondary data obtained from the Badan Pusat Statistic (BPS) publication. The data consists of 34 provinces in Indonesia. The variable used is the poverty gap index as the dependent variable. In contrast, the independent variables are the percentage of poor people ($X1$), the Gini ratio ($X2$), the poverty line ($X3$), and the rate of households with a PLN electric lighting source ($X4$).

The analysis method used in this research is robust regression analysis using GM estimation. Wilcox (2005) said, in general, GM estimation can be defined by

$$\hat{\beta}_{GM} = min \sum_{i=1}^{n} w_i \rho \left(\frac{e_i}{v(x_i)} \right)$$

$$= \sum_{i=1}^{n} w_i \rho \left(\frac{y_i - \sum_{j=0}^{k} x_{ij} \beta_j}{v(x_i)\hat{\sigma}} \right)$$

with $v(x_i)$ is the weight function for the variable x_i. Then, the above equation can be solved by using a decrease in β_j, and equated to zero, then obtained

$$n \sum_{i=1}^{n} w_i \rho' \left(\frac{y_i - \sum_{j=0}^{k} x_{ij} \beta_j}{v(x_i)\hat{\sigma}} \right) = 0$$

$$\sum_{i=1}^{n} x_{ij} w_i \psi \left(\frac{y_i - \sum_{j=0}^{k} x_{ij} \beta_j}{v(x_i)\hat{\sigma}} \right) = 0$$

According to Anderson (2008), the weights used are Schweppe weights with $w_i = \sqrt{1 - h_{ii}}$ and $v(x_i) = w_i$, then obtained

$$w_i = w_i \frac{\psi \left(\frac{u_i}{v(x_i)} \right)}{(u_i)}$$

$$= \frac{w_i}{u_i} \psi \left(\frac{u_i}{v(x_i)} \right)$$

$$w_i = \frac{\sqrt{1 - h_{ii}}}{u_i} \psi \left(\frac{u_i}{\sqrt{1 - h_{ii}}} \right)$$

where $u_i = \frac{e_i}{\hat{\sigma}}$, with $\hat{\sigma} = 1{,}48M_i$, and M_i is the median of $(n-k)$ largest of e_i and $\psi(x) = \max\{-K,\ \min(K,x)\}$ is the Huber influence function with $K = 2\sqrt{\frac{k+1}{n}}$, and k is the number of independent variables.

The following is the algorithm used in this research.

1) Estimating regression coefficients on the model using OLS.
2) Testing classical assumptions on regression models.
3) Detect the presence of outliers in the data.
4) Performing robust regression coefficient estimation of GM estimation on the model.
   a) Determining the error $e_i = y_i - \sum_{j=0}^{k} x_{ij}\beta_j$.
   b) Determine the value M_i which is the median of $(n-k)$.
   c) Determining the estimated value of $\hat{\sigma} = 1{,}48M_i$.
   d) Determining the value of $u_i = \frac{e_i}{\hat{\sigma}}$.
   e) Determining the value of the weight $w_i = \frac{\sqrt{1-h_{ii}}}{u_i}\psi\left(\frac{u_i}{\sqrt{1-h_{ii}}}\right)$.
   f) Estimating parameters $\hat{\beta}_{GM}$ using the weight value of w_i.
   g) Repeating steps a - f until the value of $\hat{\beta}_{GM}$ that converges.
5) Testing the significance of the parameters of the GM estimation robust regression model.
6) Conclude from the analysis results and interpret the regression model obtained.

RESULTS AND DISCUSSION

Ordinary Least Square

The regression model obtained from the poverty gap index of provinces in Indonesia in 2022 using OLS is

$$\hat{Y} = 0.08384 + 0.2030X_1 + 2.395X_2 + 0.0000006933X_3 - 0.01688X_4$$

with a value R^2_{adj} of 98.69%, meaning that 98.69% of the poverty gap index variable (Y) can be explained by the variables of the percentage of poor people ($X1$), the Gini ratio ($X2$), the poverty line ($X3$), and the percentage of households with PLN electricity lighting sources ($X4$). In contrast, other variables outside the model explain the other 1.31%.

Classical Assumption Test

The analysis step after obtaining the regression model with OLS is to test the classical assumptions to determine whether the misleading assumptions of the model are met before entering into robust regression analysis. Based on the results of the classical assumption test (normality test, homoscedasticity test, autocorrelation test, and multi-correlation test) on the poverty gap index data of provinces in Indonesia, it can be seen that all classical assumptions are met except the normality assumption, which is caused by outliers in the data.

Outlier Identification

Outlier detection is done using the difference in fitted value (DFFITS). If the comparison value of $|DFFITS|$ is greater than $2\sqrt{\frac{k+1}{n}} = 2\sqrt{\frac{4+1}{34}} = 0.766965$, it can be determined that the data is an outlier.

Based on the identification results, it can be seen that observations 14, 19, 33, and 34 are outliers in the data.

Robust Regression Model GM Estimation

The calculation process begins by determining the initial estimate of the regression coefficient obtained from OLS, namely $\hat{\beta}_0 = (8384 \times 10^{-5}; 2030 \times 10^{-4}; 2.395; 6933 \times 10^{-10}; -1688 \times 10^{-5})$, then after that, based on the GM estimation algorithm, calculates the value of e_i, M_i, $\hat{\sigma}$, u_i, and w_i. After that, the iteration process is carried out using the iteratively

reweighted least squares method with weighting w_i until the value of $\hat{\beta}$ that converges. The iteration results of the GM estimation are presented in Table 1.

Table 1. Iteration Results of GM Estimation

Iterations	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
1	0.1640004	0.1990269	2.406339	6.789845×10^{-7}	-0.01735031
2	0.2468687	0.1978946	2.349776	6.763241×10^{-7}	-0.01790485
3	0.3172657	0.1970392	2.298213	6.660552×10^{-7}	-0.01832474
4	0.3790466	0.1961236	2.273462	6.526383×10^{-7}	-0.01873016
5	0.4360168	0.1951607	2.260157	6.393471×10^{-7}	-0.01912246
6	0.4840314	0.1942884	2.253967	6.283578×10^{-7}	-0.01946747
7	0.519566	0.1936064	2.252478	6.204905×10^{-7}	-0.01973313
8	0.5393484	0.1931508	2.257527	6.168746×10^{-7}	-0.01990001
9	0.551659	0.1929543	2.254904	6.137763×10^{-7}	-0.01998662
10	0.5582214	0.1929173	2.24877	6.110736×10^{-7}	-0.02001457
⋮				⋮	
32	0.570877	0.192934	2.228962	6.051249×10^{-7}	-0.02004104
33	0.5708775	0.192934	2.228962	6.051247×10^{-7}	-0.02004104
34	0.5708778	0.192934	2.228961	6.051246×10^{-7}	-0.02004104
35	0.570878	0.192934	2.228961	6.051245×10^{-7}	-0.02004104
36	0.5708781	0.192934	2.22896	6.051244×10^{-7}	-0.02004104
37	0.5708782	0.192934	2.22896	6.051244×10^{-7}	-0.02004104
38	0.5708783	0.192934	2.22896	6.051244×10^{-7}	-0.02004104
39	0.5708783	0.192934	2.22896	6.051243×10^{-7}	-0.02004104
40	0.5708784	0.192934	2.22896	6.051243×10^{-7}	-0.02004104
41	0.5708784	0.192934	2.22896	6.051243×10^{-7}	-0.02004104

The results of the GM estimation robust regression model that converged at the 41st iteration are as follows:

$$\hat{Y} = 0.5708784 + 0.192934X_1 + 2.22896X_2 + 6.051243 \times 10^{-7}X_3 - 0.02004104X_4$$

with a value of R^2_{adj} close to 100%, meaning that close to 100% of the variable variance of the poverty gap index variable (Y) can be explained by the variables of the percentage of poor people ($X1$), the Gini ratio ($X2$), the poverty line ($X3$), and the percentage of households with a PLN electric lighting source ($X4$).

Parameter Significance Test

a. Simultaneous Test

A simultaneous significance test is conducted using the F to determine whether or not there is an overall influence of the independent variables on the dependent variable in the GM estimation model.

The obtained value of $F_{count} = 1.281 \times 10^{10} > F_{(0.05,4,29)} = 2.701$ well as the value of $pvalue = < 2.2 \times 10^{-16} < \alpha = 0.05$, it can be concluded that at least one independent variable significantly affects the dependent variable.

b. Partial Test

A partial significance test is conducted using the t to determine whether or not there is an influence of independent variables individually on the dependent variable in the GM estimation model. The output obtained from the simultaneous test analysis results is shown in Table 2.

Table 2. Test Result t GM Estimation

Independent Variable	$p - value$	$\lvert t_{count} \rvert$
X_1	$< 2.2 \times 10^{-16}$	74061
X_2	$< 2.2 \times 10^{-16}$	6787
X_3	$< 2.2 \times 10^{-16}$	3092
X_4	$< 2.2 \times 10^{-16}$	25640

The value of $\lvert t_{count} \rvert$ of the four independent variables $> t_{(0.025,29)} = 2.045, as$ well as the value of $p - value < \alpha = 0.05$, thus rejecting H_0. It can be concluded that all independent variables have a significant effect on the dependent variable.

CONCLUSION

Based on the results of the analysis, the robust regression equation model with GM estimation on the poverty gap index data of provinces in Indonesia in 2022 is obtained as follows

$$\hat{Y} = 0.5708784 + 0.192934X_1 + 2.22896X_2 + 6.051243 \times 10^{-7}X_3 - 0.02004104X_4$$

The interpretation of the regression model is that every 1 unit increase in the percentage of poor people ($X1$) will increase 0.192934 poverty gap index (Y), every 1 unit increase in the Gini ratio ($X2$) will increase 2.22896 poverty gap index (Y), every 1 unit increase in the total poverty line ($X3$) will increase 6.051243×10^{-7} poverty gap index (Y), and every 1 unit increase in the percentage of households with a PLN lighting source ($X4$) will decrease 0.02004104 poverty gap index (Y).

The variables that significantly affect the poverty gap index for each province in Indonesia in 2022 are the percentage of poor people, the Gini ratio, the poverty line, and the percentage of households with a PLN electricity source. The value of R^2_{adj} is close to 100%, meaning that close to 100% of the variance of the variable poverty gap index (Y) can be explained by the variables of the percentage of poor people ($X1$), Gini ratio ($X2$), poverty line ($X3$), and the percentage of households with a PLN electric lighting source ($X4$).

Suggestions for those interested in continuing research on the same topic include using or adding other independent variables that can affect the poverty depth index in Indonesia. In addition, further research can also use other estimation methods to overcome the problems that occur.

ACKNOWLEDGEMENTS

Fauzaan Nabil of the Statistics Department, Faculty of Mathematics and Science, Sebelas Maret University (FMIPA UNS), prepared this journal article based on the report Robust Regression Analysis of GM Estimation on The Poverty Gap Index of Indonesian Provinces. The opinions expressed here are those of the authors and do not necessarily reflect the views of the funding agency.

REFERENCES

Ayu, G., Suciptawati, P., dan Ayu, Ida. (2022). Identifikasi Faktor yang Memengaruhi Gini Ratio di Indonesia. *E-Jurnal Matematika*, Vol 11.

Chen, C. (2002). Robust Regression and Outlier Detection with the Robust Procedure. North Carolina: SAS Institute.

Cheng, C. L., Shalabh & Garg, G. (2014). Coefficient of Determination for Multiple Measurement Error Models. Journal of Multivariate Analysis. Vol 126.

Draper, N., and Smith, H. (1992). Applied Regression Analysis 2nd Edition. New York: John Wiley & Sons, Inc.

Freund, R. J., Wilson, W. J. & Sa, P., (2006). Regression Analysis Statistical Modeling of a Response Variable Second Edition. Boston: Academic Press.

Junus, F.D. (2021). Estimasi Parameter Model Mixed Geographically Weighted Regression Menggunakan Metode Generalized M-Estimator [Skripsi]. Universitas Hassanudin.

Kusuma, I., Susanti, Y., dan Subanti, S. (2021). Pemodelan Kemiskinan di Indonesia Menggunakan Analisis Regresi Robust. Seminar Nasional Aplikasi Sains & Teknologi.Jurusan Statistika FMIPA UNS.

Montgomery, D. C., & Peck, E. A. (2012). Introduction to Linear Regression Analysis. New York: John Wiley & Sons, Inc.
Nurdin, N., Raupong, Islamiyati. A., (2014). Penggunaan Regresi Robust Pada Data Yang Mengandung Pencilan Dengan Metode Momen. Jurnal Matematika, Statistika & Komputasi. Vol 10, No. 2.
Rousseeuw, P. J. & Hubert, M., (1997). Recent Developments in Progress. L-1 Statistical Procedures and Related Topics. Vol 31.
Susanti, Y., Pratiwi, H., dan Qona'ah, N. (2021). Regresi Robust (Teori dan Penerapannya). Surakarta : UNS Press.
Wilcox, R., (2005). Introduction to Robust Estimation and Hypothesis Testing Second Edition San Diego: Academic Press.

BIOGRAPHIES OF AUTHORS

	**Fauzaan Nabil Ardi Wijaya**, born in Sragen on September 21st, 2002, is pursuing a bachelor's degree in statistics at the University of Eleven March Surakarta. In addition, the author has also been active in organizational activities and community activities. The author can be contacted via email: fauzaanardi@student.uns.ac.id
	**Yuliana Susanti**, born in Bantul, December 19th 1961. Obtained a bachelor of science degree (S.Si) from Gadjah Mada University in 1986, then a master's degree (M.Si) from Gadjah Mada University in 2001. She serves as Head Lector at Sebelas Maret University and lecturer for the Statistics Department at Sebelas Maret University. In addition, she is also active as a writer with the latest work, "Statistical Methods with R," published by UNS PRESS, Surakarta 2021. The author can be contacted via email: yulianasusanti@staff.uns.ac.id
	**Etik Zukhronah** is a statistics study lecturer and is a Head Lector at Sebelas Maret University. She earned her bachelor of science degree (S.Si) from Gadjah Mada University in 1990, then her master of science degree (M.Si) from Bogor Agricultural University in 1998. In addition, she is also active in making scientific journals, with the last journal entitled "Detection of the Indonesian financial crisis using Markov switching dynamic conditional correlation GARCH" in 2024. The author can be contacted via email: etikzukhronah@staff.uns.ac.id