Available online at http://scik.org
J. Math. Comput. Sci. 2022, 12:118
https://doi.org/10.28919/jmcs/7217
ISSN: 1927-5307

MODELING OF CRIME RATE IN INDONESIA DURING THE COVID-19 PANDEMIC FROM A MACROECONOMIC PERSPECTIVE: USING ROBUST REGRESSION WITH S-ESTIMATOR

TILAS NOTAPIRI[1,2,*], TONI TOHARUDIN[3], YUSEP SUPARMAN[3]

[1]Magister Program in Applied Statistics, Faculty of Mathematics and Natural Sciences, Padjadjaran University, Bandung, Indonesia

[2]Statistics of Kalimantan Tengah Province, Indonesia

[3]Department Statistics, Faculty of Mathematics and Natural Sciences, Padjadjaran University, Bandung, Indonesia

Abstract: Indonesia is the 4th country in Southeast Asia with the highest crime index in 2020. Economic factors are often linked as the main motive for the crime. The purpose of this study is for modeling the crime rate in Indonesia during the Covid-19 pandemic from a macroeconomic perspective. Regression analysis is an analysis that is used to explain and model the relationship between variables. The existence of outliers is often a problem in regression analysis with OLS. To overcome the outlier problem, this research uses robust regression with S-estimation. The results show that was influenced by the unemployment rate, poverty rate, GRDP per capita, population density, and human development index.

Keywords: crime rate; robust regression; s-estimator.

2010 AMS Subject Classification: 62F35.

*Corresponding author
E-mail address: tilas21001@mail.unpad.ac.id
Received January 31, 2022

1. INTRODUCTION

Since the coronavirus officially spread in Indonesia, the government has made various efforts to prevent the spread of Covid-19. One of the efforts is to implement a policy of Large-Scale Social Restrictions (PSBB) in several regions in Indonesia. These restrictions have an impact on the economic stability of the community, especially the lower-class economy. Then it encourages them to commit crimes as a shortcut to meet economic needs [1]. Economic factors are often the main motive for crime during a pandemic [2]. Criminological studies show that more than 50 percent of criminal cases in Indonesia during the Covid-19 pandemic were caused by economic factors [3]. In the end, the Covid-19 pandemic is not only a health problem but also a socio-economic problem.

According to the crime index released by Numbeo, Indonesia ranks 4th in Southeast Asia, or 12th in Asia as the country with the highest crime index in 2020 [4]. Several studies have been conducted to examine the relationship between macroeconomic indicators and the crime rate before the COVID-19 pandemic. The research shows that there was a relationship between the unemployment rate, poverty rate, human development index, GRDP per capita, Gini ratio, population density, and inflation to the crime rate [5][6]. However, studies to see the relationship between crime rates and macroeconomic indicators in Indonesia during the Covid-19 pandemic are still rarely carried out.

Regression analysis is an analysis that is used to explain and model the relationship between variables and can also be used to predict future observations [7]. Regression analysis will be valid for making inferences if it satisfied the assumptions, that is assumptions about the form of the model, errors, predictors, and observations [8]. The violation of assumptions about observations that often occurs is the presence of outliers.

An outlier is an observation that has an extreme value, very different from other observations [9]. Outliers can be evidence that the sample comes from a population that is not normally distributed [10]. Handling outliers can be done by removing observations that are outliers. But in research that uses cross-sectional data of 34 provinces, this is not a good idea, so it is better to use

certain treatments.

Another way is using robust regression, an estimator that is not too strongly influenced by outliers. S-estimation is one of the robust regression estimation methods. This method is more efficient to use in the case of data contaminated with outliers up to 10 percent than other robust estimation methods [11]. The purpose of this study is for modeling the crime rate in Indonesia during the Covid-19 pandemic from a macroeconomic perspective using Robust Regression with the S-estimator.

2. METHODS

2.1. Data Source

The data used in this study is data from 34 provinces sourced from Statistics Indonesia with the period used is 2020. This study uses 8 variables with the following definitions:

- Crime Rate (Y)

  Crime rate is the level of risk of being exposed to a crime per 100,000 population [12]. This figure can show the level of vulnerability of a crime in a certain area at a certain time.

- Unemployment rate (X_1)

  The unemployment rate is the percentage of the number of unemployed to the total labor force (working-age population, 15 years and over) [13].

- Poverty rate (X_2)

  Percentage of poor people who are below the poverty line. The Poverty Line used is the minimum amount of rupiah needed to meet the minimum basic needs of food which is equivalent to 2100 kilocalories per capita per day and non-food basic needs [13].

- Inflation (X_3)

  Inflation is the percentage increase in the price of several goods and services that are generally consumed by households [13].

- GRDP per capita (X_4)

  GRDP per capita is the gross regional domestic product at current market prices divided by

the number of residents in an area in a certain period, in thousand rupiah [13].

- Population density (X_5)

  Population density is the number of people per square kilometer [13].

- Gini Ratio (X_6)

  The Gini Ratio or Gini Index is an indicator that shows the level of inequality in spending as a whole [13].

- Human Development Index (X_7)

  The human development index (HDI) is an indicator that measures basic human abilities, that is longevity, educational attainment, and access to a decent standard of living [14].

2.2. Boxplot

Boxplot is a simple graphic that can identify the shape of data distribution and identify outliers [15]. Boxplot can describe the smallest observation value, largest observation value, first quartile (Q_1), median (Q_2), and third quartile (Q_3) [16]. The steps of the boxplot method are as follows [17]:

1) Calculate the total amount of data.
2) Sort data from the smallest.
3) Calculate Q_1, Q_2, and Q_3.
4) Calculate IQR = $Q_3 - Q_1$.
5) Determining outliers with criteria:

   $Q_3 + (1.5\ x\ IQR) < outlier \leq Q_3 + (3\ x\ IQR)$

   $Q_3 - (1.5\ x\ IQR) > outlier \geq Q_3 + (3\ x\ IQR)$

   $outlier > Q_3 + (3\ x\ IQR)$

   $outlier < Q_3 - (3\ x\ IQR)$

2.3. Difference Fitted Values FITS (*DfFITS*)

The existence of outliers can be calculated with *DfFITS*. *DfFITS* measures the effect of observations on the fitted value [18] by measuring the difference in each observation when certain variables are omitted [19]. *DfFITS* is calculated by the following formula [8]:

(1) $$DfFITS_i = |r_i^*| \sqrt{\frac{h_{ii}}{1- h_{ii}}}$$

where r_i^* is standardized residual and h_{ii} is leverage value. The boundary of data is said to contain outliers if the value of *DfFITS* > $2\sqrt{(\mathrm{m}+1)/\mathrm{n}}$, where m is the number of the independent variable, and n is the number of observations [20].

2.4. S-Estimation

S-estimation is a robust estimation that uses a high breakdown point. The breakdown point is the minimum data size that contains outliers that can be tolerated and handled, which can lead to an inaccurate estimation [21][22]. Breakdown points can identify bad observations and resolve them. S-estimation can reach a breakdown point of up to 50 percent, so it can overcome half of the outlier effects. This method uses the standard error of the residual as a weighted value to overcome the weakness of using the median as a weighted value [23].

The steps of S-estimation are as follows [17]:

1) Estimate $\hat{\beta}^0$ using OLS.
2) Calculate residual $e_i = y_i - \hat{y}_i$.
3) Calculate $\hat{\sigma}_s = \frac{median|e_i - median(e_i)|}{0.6745}$ for the first iteration, $\hat{\sigma}_s = \sqrt{\frac{1}{nK}\sum_{i=1}^{n} w_i e_i^2}$ for the next iteration, with K = 0.1995.
4) Calculate $u_i = \frac{e_i}{\hat{\sigma}_s}$.
5) Calculate weighted value w_i using Tukey's bisquare (tuning constant c = 1.548).
6) Estimate $\hat{\beta}_s$ using WLS with weighted w_i.
7) Repeat steps 2 - 5 to obtain a convergent value of $\hat{\beta}_s$.

3. MAIN RESULTS

3.1. Descriptive Analysis

Crime rate is a number that shows the level of vulnerability of a crime in a certain area at a certain time. The higher the crime rate, the higher the level of vulnerability to crime in an area, and vice versa. This figure can also show the level of risk of being exposed to crime per 100,000

population. However, this figure does not describe the level of risk and seriousness of the crime, only describes the incidence of crime in general. In 2020, the crime rate in Indonesia is 94 crimes per 100,000 population. According to the publication of Criminal Statistics released by Statistics Indonesia, this number has decreased compared to 2019.

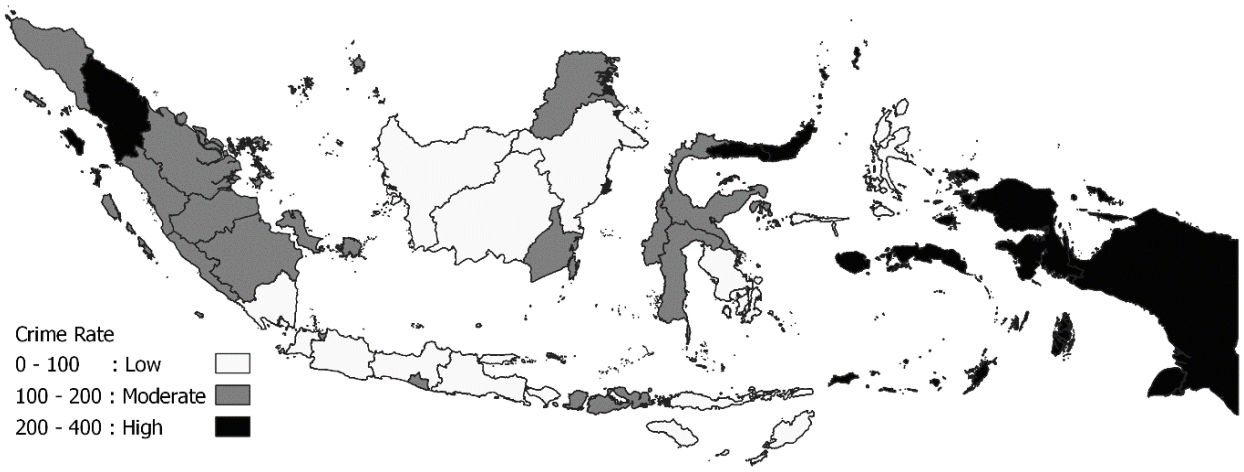

FIGURE 1. Crime Rate in 34 provinces, Indonesia

If analyzed spatially by dividing the crime rate into three categories, is low, medium, and high crime rates, 6 provinces fall into the category of provinces with the most vulnerable crime rates. The provinces are North Sumatra, North Sulawesi, Gorontalo, Maluku, Papua, and West Papua. West Papua is a province that has the highest level of risk of being exposed to crime, 328 crimes per 100,000 population. Meanwhile, 11 provinces fall into the category of low crime rates. The province with the lowest crime rate in West Java, which is 29 crimes per 100,000 population.

3.2. Boxplot Analysis

Before further inferential analysis, early detection of outliers is carried out by looking at the boxplot of each variable. From the results of the boxplot of 8 variables used sequentially, namely the crime rate, unemployment rate, poverty rate, inflation, GRDP per capita, population, Gini ratio, and HDI, it can be seen that each variable has outliers except for the Gini ratio variable (X_6). Boxplots can also be used to understand the distribution characteristics of the data. From the 8 variables, it can be seen that all boxplots have an asymmetrical shape or the median is not in the

middle of the box and one whisker is longer than the other. This indicates that the data may not come from a normal distribution.

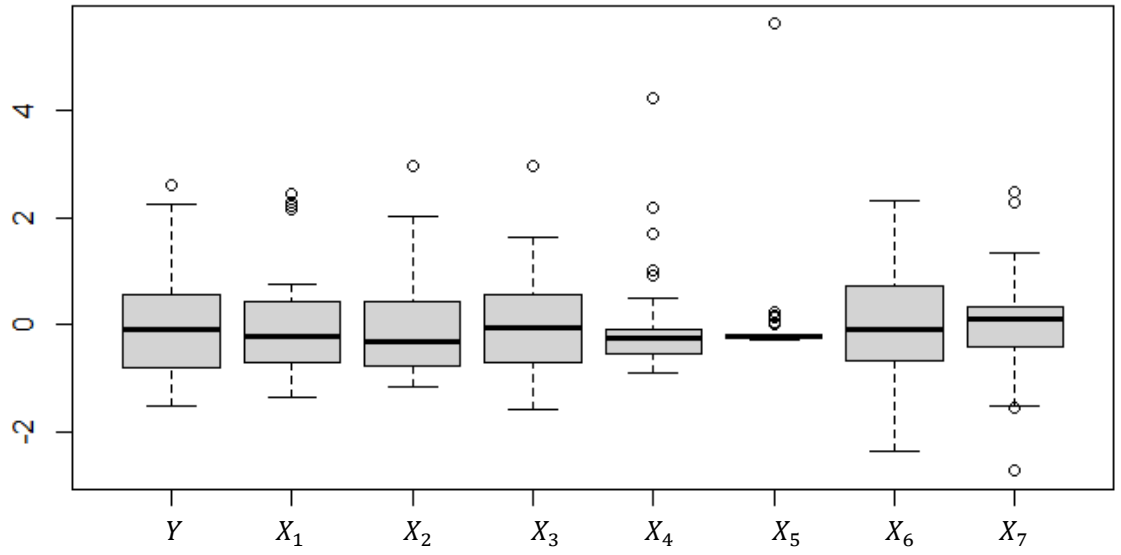

Figure 2. Outlier Identification with Boxplot

If viewed univariately using a box plot, it is indicated that there are outliers in the dependent and independent variables. However, because multiple linear regression will be used in this study, the outlier analysis will also be carried out as a whole through the residual model.

3.3. Analysis of Multiple Linear Regression

The estimation model for the crime rate in Indonesia using the OLS method of least squares is as follows:

$$\hat{Y} = -160.800^{*} + 1.330X_1 + 8.471X_2^{*} - 8.738X_3 + 0.0003X_4 - 0.006X_5 - 218.900X_6 + 3.914X_7$$

*significant (*p-value* < 0.05)

Based on the estimation results, it is found that only one variable has a significant effect, that is the poverty rate (X_2) with adjusted R-square = 0.06789. This figure shows that the ability of the dependent variable explained by the independent variable is not good. From the value of F = 3.304 with a p-value of 0.0119. With a significance level of 0.05, it is found that the p-value < 0.05, so it can be concluded that there is a relationship between the crime rate and the independent variables together.

The estimation model can be used for inference if it meets the standard regression assumptions,

that is linearity, normality, homoscedasticity, non-multicollinearity, and free of outliers. In this study, linearity test using Ramsey reset test [24], normality test using Kolmogorov-Smirnov test [25], homoscedasticity test using Breusch-Pagan test [24], and non-multicollinearity test using variance inflation factor [7]. The results of the assumption test are as follows:

TABLE 1. Regression Assumption Test Results

Assumption	Test Result	Criteria	Conclusion
Linearity	0.3606	satisfied if *p-value* > 0.05	Satisfied
Normality	0.0000	satisfied if *p-value* > 0.05	Not satisfied
Homoscedasticity	0.4304	satisfied if *p-value* > 0.05	Satisfied
Multicollinearity			
X_1	1.515210	satisfied if VIF value < 10	Satisfied
X_2	2.495479	satisfied if VIF value < 10	Satisfied
X_3	1.316429	satisfied if VIF value < 10	Satisfied
X_4	2.631246	satisfied if VIF value < 10	Satisfied
X_5	2.738884	satisfied if VIF value < 10	Satisfied
X_6	1.680057	satisfied if VIF value < 10	Satisfied
X_7	2.603829	satisfied if VIF value < 10	Satisfied

From the results of the assumption test, it is known that the normality assumption is not met. Then outlier detection is carried out using *DfFITS,* it is concluded that there are outliers if $|DfFITS| > 2\sqrt{(8/34)} = 0.9701$. Based on the *DfFITS* test in 34 provinces, 6 provinces had a *DfFITS* value of more than 0.9701. With 6 provinces that have outliers, more than 10 percent, it can be concluded that the data included the outliers so that a Robust regression estimation with S-estimation will be carried out.

3.4. Analysis of Robust Regression with S-Estimator

Robust regression with S-estimator was performed using Tukey's Bisquare (tuning constant = 1.548). The crime rate model in Indonesia using Robust regression with S-estimator is as follows:

$$\hat{Y} = -1321.00^{*} + 15.32X_1^{*} + 11.83X_2^{*} + 9.26X_3 - 0.001X_4^{*} - 0.188X_5^{*} + 621.90X_6 + 15.79X_7^{*}$$

*significant (p-*value* < 0.05)

There are five variables that partially have a significant effect on the crime rate. With a confidence level of 95 percent, the variables that significantly affect the crime rate are unemployment rate (X_1), the poverty rate (X_2), GRDP per capita (X_4), population density (X_5), and HDI (X_7).

The unemployment rate shows a positive relationship to the crime rate with a coefficient of 15.32. This shows that when the unemployment rate increases by 1 percent, the average crime rate in Indonesia will increase by 15.32 points assuming the other independent variables are constant (ceteris paribus).

The poverty rate shows a positive relationship to the crime rate with a coefficient of 11.83. This shows that when the poverty rate increases by 1 percent, the average crime rate in Indonesia will increase by 11.83 points assuming the other independent variables are constant (ceteris paribus).

GRDP per capita shows a negative relationship to the crime rate with a coefficient of -0.001. This shows that when the GRDP per capita increases by one thousand rupiah, the average crime rate in Indonesia will decrease by -0.001 points assuming the other independent variables are constant (ceteris paribus).

Population density shows a negative relationship to the crime rate with a coefficient of -0.188. This shows that when the population density increases by 1 person per kilometer, the average crime rate in Indonesia will decrease by 0.188 points assuming the other independent variables remain (ceteris paribus).

The human development index shows a positive relationship to the crime rate with a coefficient of 15.79. This shows that when the human development index increases by 1 point, the average crime rate in Indonesia will increase by 15.79 points with the assumption that other variables are constant (ceteris paribus).

Based on modeling using Robust Regression with S-Estimator, obtained an adjusted R-squared of 0.8352. This means that the resulting model can explain the crime rate of 83.52 percent and the remaining 16.48 percent is explained by variables not included in the model.

4. CONCLUSION

Crime rate modeling using Robust Regression with S-Estimator, it was found that the crime rate in Indonesia during the Covid-19 pandemic was influenced by the unemployment rate, poverty rate, GRDP per capita, population density, and human development index. This model has a fairly good coefficient of determination, namely 83.52 percent of the crime rate can be explained by the independent variables used.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

[1] S. Maruli and T. U. A. Situmeang, Fenomena kejahatan di masa pandemi Covid-19: Pendekatan kriminologi, Majalah Ilmiah UNIKOM, 19 (2021), 35–43.

[2] N. T. Ummayah and R. Dhiya, Analysis of crime figures in Indonesia due to the Covid-19 pandemic and efforts to overcome it, no. Icsse (2021), 242–248.

[3] S. Somadiyono, Kajian kriminologis perbandingan kejahatan Yang terjadi sebelum pandemi dan saat pandemi Covid-19, J. Belo, 6 (2021), 148–156.

[4] Numbeo, Crime index, https://www.numbeo.com/crime/rankings_by_country.jsp?title=2020 (accessed Dec. 31, 2021).

[5] N. Coomer, America's underclass and crime: The influence of macroeconomic factors, Issues Polit. Econ. 12 (2003), 25.

[6] T. Hidayadi, Sohibunajar, and F. Ulfah, Analisis regresi faktor yang mempengaruhi kejahatan yang dilaporkan di indonesia sebelum pandemi Covid-19, Conference Economic and Business Innovation (2021), 1129-1145.

[7] J. J. Faraway, Practical regression and anova using R, no. July. Bath: University of Bath (2002).

[8] S. Chatterjee and A. S. Hadi, Regression analysis by example 4th edition. John Wiley & Sons, (2015).

[9] D. N. Gujarati and D. C. Porter, Basic Econometrics. Mass: McGraw-Hill (2009).

[10] R. Dawson, How significant is a boxplot outlier?, J. Stat. Educ. 19 (2011), 2.

[11] Ö. G. Alma, Comparison of Robust Regression Methods in Linear Regression, Int. J. Contemp. Math. Sci. 6 (2011), 409-421.

[12] Statistics Indonesia, Statistik Kriminal 2021. Jakarta: Statistics Indonesia (2021).

[13] Statistics Indonesia, Metadata Indikator, 2022. https://sirusa.bps.go.id/sirusa/index.php/indikator/index (accessed Dec. 31, 2021).

[14] Statistics Indonesia, Indeks Pembangunan Manusia. Jakarta: Statistics Indonesia (2021).

[15] D. F. Williamson, R. A. Parker, and J. S. Kendrick, The box plot: A simple visual method to interpret data, Ann. Intern. Med. 110 (1989), 916–921.

[16] M. Y. Darsyah, Penggunaan Stem and Leaf dan Boxplot untuk Analisis Data, J. Karya Pendidik. Mat. 1 (2014), 55–67.

[17] Y. Susanti, H. Pratiwi, S. Sulistijowati, and T. Liana, M-estimation, S-estimation, and MM-estimation in robust Regression, Int. J. Pure Appl. Math. 91 (2014), 349–360.

[18] S. M. A. K. Rahman, M. M. Sathik, and K. S. Kannan, Multiple linear regression models in outlier detection, Int. J. Res. Comput. Sci. 2 (2012), 23–28.

[19] K. Senthamarai Kannan and K. Manoj, Outlier detection in multivariate data, Appl. Math. Sci. 9 (2015), 2317–2324.

[20] M. Mardiana, A. Wibowo, M. Mahmudah, and P. Festi W, A comparison of M-estimation and S-estimation on the factors affecting IR DHF in East Java in 2017, Indones. J. Public Heal. 16 (2021), 349.

[21] S. Sakata and H. White, S-estimation of nonlinear regression models with dependent and heterogeneous observations, J. Econometrics, 103 (2001), 5-72.

[22] F. P. Hidayatulloh, D. Yuniarti, and S. Wahyuningsih, Regresi robust Dengan metode estimasi-S robust regression method to estimate - S, Eksponensial, 6 (2015), 163–170.

[23] E. M. Almetwally and H. M. Almongy, Comparison between M-estimation , S-estimation , and MM estimation, Int. J. Math. Arch. 9 (2018), 55–63.

[24] C. Kleiber and A. Zeileis, Applied econometrics with R. Springer Science & Business Media (2008).

[25] J. S. Jones, Learn to use the kolmogorov-smirnov test in R with data from the opinions and lifestyle survey (well-being module). SAGE Publications (2019).