Integra: Journal of Integrated Mathematics and Computer Science
Journal Homepage: https://integrajimcs.com/index.php/integra/index

integra
Journal of Integrated Mathematics and Computer Science

Research Paper

Robust Panel Data Regression Analysis using the Least Trimmed Squares (LTS) Estimator on Poverty Line Data in Lampung Province

Windi Lestari[1*], Widiarti[1], Bernadhita Herindri Samodera Utami[1], Mustofa Usman[1], Vitri Aprilla Handayani[2]

[1] *Department of Mathematics, Faculty of Mathematics and Natural Science, Universitas Lampung, Lampung, 35145, Indonesia*
[2] *Faculty of Information Technology, Institut Teknologi Batam, Batam, 29425, Indonesia*
[*] **Corresponding author**: windilestari095@gmail.com

Keywords
Poverty Line, Robust Regression, Least Trimmed Squares (LTS), Panel Data

Abstract
Robust regression is an alternative method in regression analysis designed to produce stable parameter estimates, even when the data contain outliers or deviate from classical assumptions. One of its estimation techniques, the Least Trimmed Square (LTS), works by minimizing the smallest squared residuals, thereby assigning smaller weights to extreme data points. This method serves as a solution when classical approaches, such as Ordinary Least Squares (OLS), fail to meet the assumptions, especially in socio-economic data that are often complex and prone to outliers. This study employs robust regression with the LTS estimator on panel data to examine the impact of population size , population density , and registered job vacancies on poverty lines in Lampung Province. The data cover 15 districts and cities from 2019 to 2023. The analysis results show that the model obtained has a coefficient of determination of R^2=0.8909. This means that the three predictor variables can explain 89.09% of the variation in the poverty line.

Received: 3 February 2024, Accepted: 12 April 2024
https://doi.org/10.26554/integra.20241210

1. INTRODUCTION

Regression analysis is a statistical method used to explore patterns and measure the relationship between two or more variables that are causally related [1]. One of the main goals of regression analysis is to estimate the coefficients in a regression model to understand the influence of independent variables on the response variable. The most commonly used method is the Ordinary Least Squares (OLS), which requires the fulfillment of several classical assumptions, including the normality of residuals. However, in practice, these assumptions are often violated due to the presence of outliers [2]. Outliers are observations that deviate significantly from the rest of the data and may result from input errors, unrepresentative sampling, or actual extreme events.

Panel data, which combines cross-sectional and time-series elements, allows for a more in-depth analysis of variable dynamics over time [3]. Panel data enhances estimation efficiency, reduces multicollinearity, and enriches the information available. However, like other types of data, panel data is also vulnerable to outliers, which can distort the accuracy of model estimations. Research using panel data regression [4, 5, 6, 7] demonstrates the advantages of panel data in capturing both temporal and spatial variations. These studies highlight that panel data models are effective for analyzing complex phenomena, offering richer insight compared to purely cross-sectional or time-series data alone.

Removing outliers is not always appropriate, as they may contain valuable information. Therefore, a more robust estimation approach is needed, such as robust regression. Robust regression is designed to produce stable and reliable parameter estimates, even when the data deviate from classical assumptions or contain extreme values. This method assigns smaller weights to outliers, making the model more resistant to their influence. Several studies that apply robust regression methods

[8, 9, 10]demonstrate the effectiveness of robust regression in producing accurate models under data irregularities. Further applications by Husain [11], Aristiarto [12], and Azzahro [13], reinforce the advantages of robust regression. These studies indicate that robust regression offers greater resilience against data anomalies, making it a valuable method in socio-economic and environmental analyses where extreme or influential observations are likely to occur.

One of the techniques in robust regression is the Least Trimmed Squares (LTS) estimator, which minimizes the sum of the smallest squared residuals. Several previous studies have demonstrated the effectiveness of the LTS method in various contexts, such as agricultural production data [14], blood pressure modelling [15], and poverty analysis [16]. These studies reinforce the reliability of LTS in producing robust results under challenging data conditions.

In addition, the study by Sari [17] showing the robustness of LTS in epidemiological modeling. Similarly, Muamalah [18] revealing that LTS was effective in handling outliers. Another study by Amni [19] confirming the stability of LTS in multivariate economic contexts. Nurmulyati [20], and Lestari [21]. These various studies consistently show that LTS is a reliable and versatile method for obtaining robust parameter estimates across a wide range of data and application fields.

Based on this, the present study applies robust regression using the LTS estimator to analyze the factors influencing poverty lines in Lampung Province. The independent variables include population size, population density , and registered job vacancies, while the dependent variable is the poverty line. The p overty line represents the minimum income required to meet basic needs. By applying robust regression, this study aims to develop a more accurate model in identifying key determinants of poverty and to offer a more resilient methodological alternative for socio-economic data analysis.

2. METHODS

The data used in this study includes poverty line figures at the regency/city level in Lampung Province. The data were obtained from the official website of BPS Lampung. The data spans the years 2019 to 2023 and covers 15 regencies or cities: Lampung Barat, Tanggamus, Lampung Selatan, Lampung Timur, Lampung Tengah, Lampung Utara, Way Kanan, Tulang Bawang, Pesawaran, Pringsewu, Mesuji, Tulang Bawang Barat, Pesisir Barat, Bandar Lampung, and Metro. The poverty line (Y) serves as the dependent variable, while the three independent variables include total population (X_1), population density (X_2), and registered job vacancies (X_3).

This study aims to estimate parameters and apply a robust regression model using the Least Trimmed Square (LTS) estimator to poverty line data in Lampung Province for the 2019-2023 period. Data analysis was conducted using R Studio software. The steps undertaken in this research are as follows:

a. Conduct a multicollinearity test to examine the Variance Inflation Factor (VIF) values and determine whether any predictor variables are highly correlated, which can affect the stability of parameter estimates.
b. Perform an initial estimation of panel data regression parameters using the Ordinary Least Squares (OLS) method to gain a preliminary understanding of the relationship between the response and predictor variables.
c. Identify the most appropriate panel data model by conducting the Chow Test to compare the common effect and fixed effect models, followed by the Hausman Test to determine whether the fixed effect or random effect model is more suitable based on the consistency of estimator.
d. Test the classical assumptions on the selected panel data model, including:
    - Checking residual normality using the Jarque-Bera Test.
    - Testing residual variance homogeneity using the Breusch-Pagan Test.
    - Checking for autocorrelation in the residuals using the Breusch-Godfrey Test.
e. Detect outliers using Leverage Value and TRES (Studentized Deleted Residual) on both predictor and response variables to identify influential observations.
f. Estimate the robust regression parameters using the LTS method through the following steps:
    - Determine the initial estimator β_0 Using OLS.
    - Calculate residuals and then compute the squared residuals e_{it}^2.
    - Compute the sum of squared residuals for the first hhh observations using the formula $\frac{NT+p+1}{2}$.
    - Calculate $\Sigma_{i=1}^{h}\ e_i^2$
g. Perform parameter testing to assess the effect of predictor variables on the response variable, both partially and simultaneously.
h. Validate the LTS estimator model using the coefficient of determination R^2 And assess the relative efficiency of the model.

Table 1. Descriptive Analysis of the Research Tables

	GK	JP	KP	LK
Min	341012	154895	53,28	16
Q1	408406	290529	140,31	306
Median	443313	453921	230,58	602
Mean	457914	593272	790,12	1092
Q3	483785	1015538	481,20	1188
Max	761790	1522426	6585,77	8196
StDev	74519.40	4.395135	196914	1456.32

3. RESULTS AND DISCUSSION

In this study, a robust regression approach using the Least Trimmed Squares (LTS) estimator was employed to analyze the effect of total population, population density, and registered job vacancies on the poverty line in Lampung Province. This model

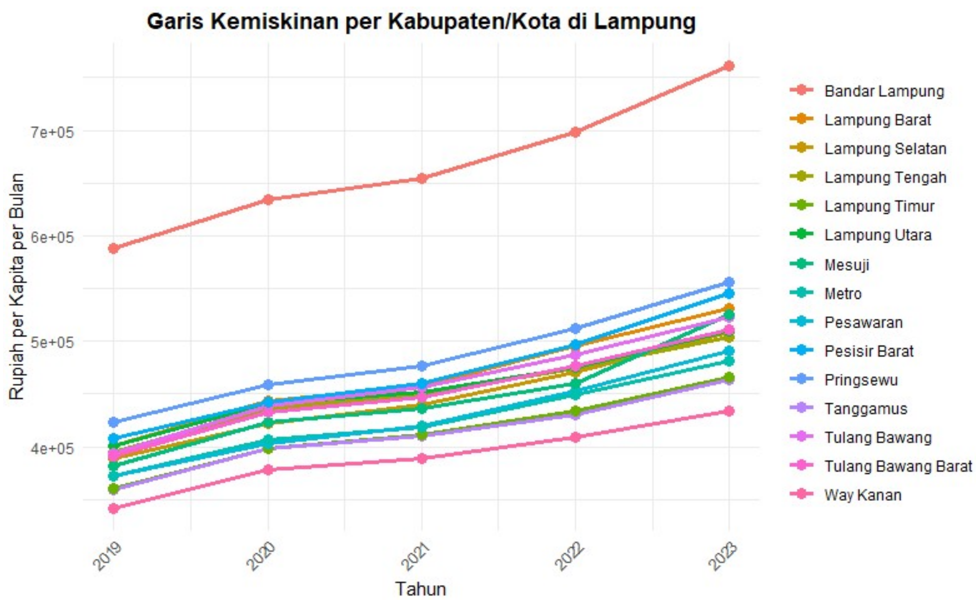

Figure 1. Poverty Line Graph in Lampung Province 2019-2023

was selected due to its capability to handle outliers and provide more reliable estimations.

3.1 Descriptive Analysis

Descriptive analysis is a statistical technique used to provide a general overview or summary of data. Its main purpose is to understand the fundamental characteristics of the observed data and to present them in a simpler and more comprehensible form.

Figure 1 displays the graph of poverty line in Lampung Province. Bandar Lampung had the highest poverty line. Larger and more populated areas like Bandar Lampung tend to have a higher poverty line, while smaller or remote areas such as Way Kanan and West Coast have lower values. This may be due to differences in living costs, minimum wages, and access to services. Table 1 below show the descriptive analysis of the research table.

a. Poverty Line ranges from a minimum of 341,012 (Way Kanan, 2019) to a maximum of 761,790 (Bandar Lampung, 2023). The mean is 457,914, the median is 443,313, the lower quartile is 408,406, the upper quartile is 483,785, and the standard deviation is 74,519.40.
b. Total Population ranges from 154,895 (Pesisir Barat, 2019) to 1,522,426 (Lampung Tengah, 2023). The mean is 593,272, the median is 453,921, the lower quartile is 290,529, the upper quartile is 483,785, and the standard deviation is 439,513.5.
c. Population Density ranges from 53.28 (Pesisir Barat, 2019) to 6,585.77 (Bandar Lampung, 2022). The mean is 790.12, the median is 230.58, the lower quartile is 140.31, the upper quartile is 481.20, and the standard deviation is 1,969.14.
d. Job Vacancies range from 16 (Pesisir Barat, 2022) to 8,196 (Lampung Tengah, 2019). The mean is 1,092, the median is 602, the lower quartile is 306, the upper quartile is 1,188, and the standard deviation is 1,456.32.

3.2 Multicollinearity Test

The multicollinearity test aims to assess whether there is a significant correlation among the independent variables in a regression model. An ideal regression model should be free from multicollinearity. One common method to detect multicollinearity is by calculating the Variance Inflation Factor (VIF). The hypotheses for this test are:

$H_0 : r_{x_i,x_j} = 0$ (No multicollinearity exists) $H_1 : r_{x_i,x_j} \neq 0$ (Multicollinearity exists)

The VIF statistic :

$$VIF_J = \frac{1}{1 - R_j^2}$$

Table 2. Result of Multicollinearity Test

Variable	X_1	X_2	X_3
VIF	1.819351	1.172254	1.649538

Table 2 displays the results of multicollinearity test. Based on Table 2, the VIF values for the independent variables are 1.819, 1.172, and 1.650, respectively. Since all VIF values are below the threshold of 10, there is no indication of multicollinearity in the regression model. Therefore, the multicollinearity assumption is satisfied, and the regression model can proceed to the next stage of analysis.

3.3 Panel Data Regression Modeling

Panel data regression model selection involves a series of tests to determine the most appropriate model among the Common

Effect Model (CEM), Fixed Effect Model (FEM), and Random Effect Model (REM).

3.3.1 Common Effect Model (CEM)

The CEM assumes that both intercepts and slopes remain constant over time and across observational units. Table 3 show the estimation results for the CEM.

Table 3. Output of CEM

Variabel	Estimate	$p - value$
Intercept	42315	2.2e-16
X_1	0.31314	0.1525
X_2	31.511	8.448e-10
X_3	-7.9804	0.1524

Based on Table 3, it is evident that X_2 Is statistically significant, as its p-value is below the significance level of $\alpha = 0.05$. In contrast, X_1 and X_3 are not significant since their p-values exceed 0.05. The resulting CEM equation is:

$Y_{it} = 4231 + 0.031X_{1it} + 31.51X_{2it} - 7.980X_{3it}$

From this equation, it can be concluded that population density. (X_2) has a significant positive influence on the poverty line. This implies that for every one-unit increase in population density, the poverty line is expected to rise, assuming other variables remain constant.

3.3.2 Fixed Effect Model

The Fixed Effect Model (FEM) assumes that the intercept varies across observational units (regencies/cities), while the slope remains constant over time and across units [3]. Table 4 displays the FEM estimation results.

Based on Table 3, it is evident that X_2 Is statistically significant, as its p-value is below the significance level of $\alpha = 0.05$. In contrast, X_1 and X_3 are not significant since their p-values exceed 0.05. The resulting CEM equation is:

$Y_{it} = 4231 + 0.031X_{1it} + 31.51X_{2it} - 7.980X_{3it}$

From this equation, it can be concluded that population density. (X_2) has a significant positive influence on the poverty line. This implies that for every one-unit increase in population density, the poverty line is expected to rise, assuming other variables remain constant.

Table 4. Output of FEM

Variable	Estimate	$P - value$
Intercept	41638	2.2e-16
X_1	0.036519	0.2253
X_2	30.247	1.032e-05
X_3	-3.6938e	0.4729

According to Table 4, the variable X_1 is statistically significant as it has a p-value below the significance level of $\alpha = 0.05$. Meanwhile, X_2 and X_3 are not significant, indicated by their p-values exceeding 0.05. The resulting FEM equation is as follows:

$Y_{it} = \beta_{0it} + 0.633X_{1it} + 3.824X_{2it} + 5.806X_{3it}$

Based on the Equation, it can be concluded that population size has a significant effect on the poverty line. In FEM, the common intercept is replaced with individual effects for each regency/city. The individual effects are calculated using the within transformation approach, as shown in the following equation:

$\hat{\alpha} = \bar{Y}_i - \beta\bar{X}_i$

Table 5. Individual Effect of FEM

District/Cities	Individual Effect
Bandar Lampung	-90715
Lampung Barat	271719
Lampung Selatan	-233876
Lampung Tengah	-463461
Lampung Timur	-305587
Lampung Utara	52587
Mesuji	303387
Metro	306036
Pesawaran	123181
Pesisir Barat	367544
Pringsewu	221226
Tanggamus	5777
Tulang Bawang	177404
Tulang Bawang Barat	268184
Way Kanan	87535

Table 5 shows that Pesisir Barat District has the highest individual effect (367,544), indicating a strong fixed impact on the poverty line compared to other regions. These differences highlight the presence of significant time-invariant regional characteristics affecting the poverty line that are not captured by the model's explanatory variables.

Table 6. Output of REM

Variable	Estimate	$P - value$
Intercept	41638	2.2e-16
X_1	0.036519	0.2253
X_2	30.247	1.032e-05
X_3	-3.6938e	0.4729

3.3.3 Random Effect Model

The Random Effect Model (REM) assumes that differences in characteristics across time periods and units are accounted for through the model's error term. In this approach, unobserved effects are assumed to be random and uncorrelated with the explanatory variables.

Table 6 shows that X_2 is statistically significant, with a p-value less than the 0.05 significance level. On the other hand, X_1

and X_3 are not statistically significant, as their p-values exceed 0.05. The REM estimation equation is as follows:

$Y_{it} = 4163 + 0.036X_{1it} + 30.24X_{2it} - 3.69X_{3it}$

Based on the Equation, it can be concluded that population density has a significant positive effect on the poverty line. In other words, as population density increases, the poverty line also tends to rise, holding other factors constant.

3.4 Panel Regression Model Selection

Panel regression model selection involves several standard tests to determine the most appropriate model among the Common Effect Model (CEM), Fixed Effect Model (FEM), and Random Effect Model (REM).

3.4.1 Chow Test

The Chow test is used to determine whether the Fixed Effect Model (FEM) provides a significantly better fit than the Common Effect Model (CEM). It evaluates whether individual-specific effects significantly explain the variation in the panel data. Hypotheses:

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_n = 0$ (CEM is appropriate)

H_1 at least one $\beta_i \neq 0$, (FEM is appropriate)

Test statistic:

$$F = \frac{\frac{RSS_1 - RSS_2}{(n-1)}}{\frac{RSS_2}{nT-n-K}}$$

Table 7. Output of the Chow Test

Test	$F - Statistic$	$p - value$
Cross-Section	5.202	3.527e-06

Table 7 shows that the $p - value$ is 3.527e-06, which is significantly lower than the significance level of 0.05. Therefore, H0 is rejected, and it is concluded that the FEM is more appropriate than the CEM for the panel data analysis.

3.4.2 Hausman Test

The Hausman Test is used to determine the most appropriate model between the Fixed Effect Model (FEM) and the Random Effect Model (REM) in panel data analysis [3]. The test assesses whether the individual effects are correlated with the regressors. Hypotheses:

$H_O : E(\mu_i, e_{it}) = 0$ (REM is appropriate)

$H_1 : E(\mu_i, e_{it}) \neq 0$ (FEM is appropriate)

Test statistic: $W = \hat{q} Var(\hat{q})^{-1} \hat{q}$

Table 8. Output of The Hausman Test

Test	$F - Statistic$	$p - value$
Chi-Square	21.393	8.723e-05

Table 8 shows that the $p - value$ is 8.723e-05, which is lower than the 0.05 significance level. Therefore, H_0 is rejected, and it can be concluded that the FEM is more appropriate than the REM for this analysis.

Model Assumption Testing After selecting the best panel regression model, which in this case is the Fixed Effect Model (FEM), it is necessary to test the classical assumptions of regression, including the normality of residuals.

3.4.3 Normality Test

The normality test aims to evaluate whether the residuals of the regression model follow a normal distribution. A well-specified regression model generally assumes normally distributed errors. In this study, the Jarque-Bera (JB) test is used to assess the distribution of residuals.

According to Mardiatmoko [22], the hypotheses are as follow

$H_0 : \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ (residuals are normally distributed)

$H_1 : \varepsilon_i \nsim \mathcal{N}(0, \sigma^2)$ (residuals are not normally distributed)

The Jarque–Bera statistic is calculated as:

$$JB = \frac{n}{2}\left(S^2 + \frac{(K-3)^2}{4}\right)$$

Table 9. Output of Normality Test

Test	Statistic Test	$p - value$
Jarque-Berra	17.616	< 2.2e-16

Table 9 shows that the $p - value$ is less than 2.2e-16, which is lower than the 0.05 significance level. Thus, H0 is rejected, indicating that the residuals of the FEM model are not normally distributed, and the normality assumption is not fulfilled.

3.4.4 Heteroskedasticity Test

The heteroskedasticity test aims to evaluate whether the residuals of the regression model have constant variance. A good regression model is expected to exhibit homoscedastic residuals, meaning the variance of the errors remains consistent across observations . In this study, the Breusch-Pagan (BP) test is used to detect heteroscedasticity.

Hypotheses =

$H_0 = Var(\varepsilon_i) = \sigma^2$ (no Heteroskedasticity)

$H_1 = Var(\varepsilon_i) \neq \sigma^2$ (presence of heteroskedasticity)

Test statistic:

$BP = \frac{1}{2} f^T Z(Z^T - Z)^(- 1) Z^T f$

Table 10. Output of Heterskedasticity Test

Test	Statistic Test	$p - value$
Breusch-Pagan	7.8159	0.04997

Table 10 shows that the $p - value$ is 0.04997, which is slightly less than the 0.05 significance level. Therefore, H0 is rejected, indicating that heteroskedasticity is present in the model. Thus, the assumption of homoscedastic residuals is not fulfilled.

3.4.5 Autocorrelation Test

The autocorrelation test is conducted to detect any correlation between residuals in a regression model. Autocorrelation often occurs in time series or panel data, where the residual in the current period may be influenced by the residual in the previous period. In this study, the Breusch-Godfrey (BG) test is applied to detect autocorrelation. Hypotheses:

$H_0 = Cov(\varepsilon_i, \varepsilon_{i-1}) = 0$ (no autocorrelation)

$H_1 = Cov(\varepsilon_i, \varepsilon_i - 1) \neq 0$ (autocorrelation exists)

Test statistic:

$BG = N \times R^2$

Table 11. Output of Autocorrelation Test

Test	Statistic Test	$p - value$
Breusch-Godfrey	0.82533	0.3636

Table 11 shows that the $p - value$ is 0.3636, which is greater than the 0.05 significance level. Therefore, H0 fails to be rejected, indicating that no autocorrelation is present in the model. Hence, the autocorrelation assumption is fulfilled.

3.5 Outlier Detection

Outliers are extreme observations that deviate significantly from the overall pattern of the data. They can potentially influence the regression estimates and reduce model reliability.This study applies two methods for detecting outliers: Leverage Value and Studentized Deleted Residual (TRES).

3.5.1 Leverage Value

According to Kutner et al. [23], the leverage value measures how far an observation is from the mean of the independent variables. Table 12 displays the outlier detection using Leverage value. It is derived from the diagonal of the hat matrix (H), calculated as follows.

$H = X(X'X)^{-1}X'$

Where $X = [1, x_{i1}, x_{i2}, \cdots, x_{np}]$ is the vector of variables for observation i , and the diagonal h_{ii} represents the influence of observation iii on the regression model.

3.5.2 TRES

TRES is used to detect outliers in the response variable by computing the residual after excluding the observation and then standardizing it. Table 13 displays the outlier detection using TRES. Kutner et al. [23] define TRES as:

$$TRES_i = \frac{d_i}{s_{d_i}} = e_i \left[\frac{n - p - 2}{JKS(1 - h_{ii}) - e_i^2} \right]^{\frac{1}{2}}$$

Where:

$$e_i = y_i - \hat{y}_i$$
$$d_i = y_i - \hat{y}_{i(i)}$$
$$s_{d_i} = \text{Standard deviation}(d_i)$$
$$h_{ii} = X_i(X'X)^{-1}X_i'$$

Table 12. Outlier Detection Using Leverage Value

Observation	Leverage Value h_i	Outlier
2	0.19750948	Yes
3	0.21118636	Yes
4	0.22171883	Yes
5	0.21724340	Yes
12	0.41824930	Yes
16	0.39367241	Yes
18	0.14044956	Yes
19	0.16317914	Yes
20	0.11082075	Yes
25	0.37539370	Yes

Table 13. Outlier Detection TRES

Observation	TRES |$TRES_i $|	Outlier
1	2.085601651	Yes
18	3.632668574	Yes
30	2.746896295	Yes
34	4.894368133	Yes
37	2.037292463	Yes
59	2.150973564	Yes
70	2.324113125	Yes
73	2.787568948	Yes

3.6 Robust Regression Least Trimmed Squares (LTS)

Robust regression is an alternative approach used to obtain more accurate regression models when the dataset contains outliers. This method is designed to produce reliable estimates despite the presence of extreme data points that may significantly affect the analysis results.

There are several robust regression approaches developed to handle outliers that may affect the accuracy of parameter estimates. These robust estimation methods include: Maximum Likelihood Type Estimator (M-estimator), Least Median of Squares (LMS) Estimator[, Method of Moment (MM) Estimator, Least Trimmed Squares (LTS) Estimator , and Scale (S) Estimator.

The Least Trimmed Squares (LTS) method minimizes the sum of the smallest squared residuals from a selected subset of observations, making it highly resistant to outliers.

The estimator of the LTS regression is formulated as follows:

$$\hat{\beta}_{LTS} = \arg\min_{\beta} Q_{LTS}(\beta)$$

With

$$Q_{LTS}(\beta) = \sum_{i=1}^{h} e_{(i)}^2, \quad \text{where } e_{(1)}^2 \leq e_{(2)}^2 \leq \cdots \leq e_{(n)}^2$$

Here, h represents the number of observations with the smallest residuals included in the estimation, while the remainin n-h observations with the largest residuals are excluded (assigned zero weight). When h=n, the LTS method reduces to the traditional Ordinary Least Squares (OLS) method. The optimal value of h can be calculated using the following formula:

$$h = \left\lfloor \frac{n}{2} \right\rfloor + \left\lfloor \frac{p+1}{2} \right\rfloor \quad \text{atau} \quad \left\lfloor \frac{n}{2} \right\rfloor + \left\lfloor \frac{p+2}{2} \right\rfloor$$

Table 14. Outlier Detection Using Leverage Value

Variable	Coefficient	R^2
Intercept	44650	0.8909
X_1	0.01931	
X_2	1.853	
X_3	-1.5921	

Table 14 displays, the robust regression model estimated using the Least Trimmed Squares (LTS) method is as follows:

$Y_{LTS} = 4465 + 0.019X_{1it} + 1.853X_{2it} - 1.592X_{3it}$

The coefficient determination $R^2 = 0,8909$ indicates that approximately 89.09% of the variation in poverty line can be explained by the independent variables. Therefore, under data conditions involving outliers, the LTS estimator produces a reliable regression model for poverty line estimation in Lampung Province during the 2019-2023 period.

Parameter significance testing was conducted through both simultaneous testing using the F-test and partial testing using the t-test, to assess the influence of each independent variable individually.

Table 15. Simultaneous Significance Test Results (F-test)

Method	F Statistic	$p-value$
LTS	25.63759	2.398696e-11

Table 16. Partial Significance Test Results (t-test)

Variable	T Test	$P-value$
Intercept	42373.12	3.951793e-45
X_1	0.02725534	1.700492e-01
X_2	32.44655	9.957101e-09
X_3	-7.267381	1.261435e-01

Table 15 shows that the $p-value$ of 2.398696e-11 is far below the significance level. Therefore, it can be concluded that the independent variables jointly have a significant effect on the poverty line in Lampung Province. Table 16 below displays the partial significance test results.

According to Table 16, only the variable population density (X_2) has a p-value smaller than α = 5% = 0,05, indicating that population density has a significant partial effect on the poverty line. Meanwhile, the other variables are not statistically significant on their own. Nevertheless, although the two variables are not individually significant, the simultaneous test shows that the overall model remains statistically significant.

4. CONCLUSIONS

Based on the analysis results, it can be observed that the poverty line across regencies/cities in Lampung Province may be influenced by differences in living costs, minimum wages, as well as the level of infrastructure and social services available in each area. From this analysis, the robust regression model estimated using the Least Trimmed Squares (LTS) method is obtained as follows:

$Y_{LTS} = 4465 + 0.019X_{1it} + 1.85X_{2it} - 1.592X_{3it}$

The coefficient of determination R^2 = 0.8909 indicates that the three predictor variables explain 89.09% of the variation in the poverty line. Therefore, under conditions where the data contain outliers, it can be concluded that the LTS estimator yields a reliable regression model for estimating the poverty line in Lampung Province during the 2019–2023 period. Based on the results of simultaneous and partial tests, it can be concluded that at a 5% significance level, total population (TP), population density (PD), and job vacancies (JV) jointly have a significant effect on the poverty line in Lampung Province.

5. ACKNOWLEDGEMENT

The author would like to thank Mathematics and Statistics Research laboratory, Universitas Lampung for the support given.

REFERENCES

[1] D. C. Montgomery, E. A. Peck, and G. G. Vining. *Introduction to Linear Regression Analysis*. John Wiley & Sons Inc., 2012.

[2] S. Shoukat, S. Nawaz, M. M. Rasheed, A. Javaid, and H. A. Sami. Efficiency of ols and huber m estimator in case of outliers. *Journal of Excellence in Social Sciences*, 3(3):55–60, 2024.

[3] D. N. Gujarati and D. C. Porter. *Basic Econometrics*. McGraw-Hill Education, 2009.

[4] B. D. Baltagi. *Econometric Analysis of Panel Data*. John Wiley & Sons Ltd, 2005.

[5] C. Hsiao. *Analysis of Panel Data*. Cambridge University Press, 4 edition, 2024.

[6] I. P. Hutagalung and O. Darnius. Analisis regresi data panel dengan pendekatan common effect model (cem), fixed effect model (fem) dan random effect model (rem) (studi kasus: Ipm sumatera utara periode 2014–2020). *FARABI: Jurnal Matematika Dan Pendidikan Matematika*, 5(2):217–226 (in Indonesia), 2022.

[7] N. A. Salsabila, H. K. Juliarto, A. F. Syawal, and D. A. Nohe. Analisis regresi data panel pada ketimpangan pendapatan

daerah di provinsi kalimantan timur. In *Prosiding Seminar Nasional Matematika, Statistika, dan Aplikasinya*, volume 2, pages 241–253 (in Indonesia), 2022.

[8] T. S. Edriani, A. Rahmadani, and D. M. M. Noor. Analisis hubungan kepadatan penduduk dengan pola penyebaran covid-19 provinsi dki jakarta menggunakan regresi robust. *Indonesian Journal of Applied Mathematics*, 1:51–60 (in Indonesia), 2021.

[9] R. Yousuf and M. Sharma. Applications of mitscherlich baule function: A robust regression approach. *Research in Statistics*, 2(1):1–5, 2024.

[10] D. M. Khan, A. Yaqoob, S. Zubair, M. A. Khan, Z. Ahmad, and O. A. Alamri. Applications of robust regression techniques: An econometric approach. *Mathematical Problems in Engineering*, 2021:1–9, 2021.

[11] A. Husain and S. R. W. Jamaluddin. Pemodelan data angka kematian bayi menggunakan regresi robust. *SAINTEK: Jurnal Sains, Teknologi & Komputer*, 1:1–7 (in Indonesia), 2023.

[12] R. Aristiarto, Y. Susanti, and I. Susanto. Analisis regresi robust estimasi gm pada indeks keparahan kemiskinan provinsi-provinsi di indonesia. In *Seminar Nasional Riset dan Inovasi Teknologi (SEMNAS RISTEK)*, volume 7, pages 205–209 (in Indonesia), 2023.

[13] I. A. Azzahro and A. Sofro. Regresi robust untuk pemodelan deforestasi di indonesia. *MATHunesa: Jurnal Ilmiah Matematika*, 2(3):496–507 (in Indonesia), 2023.

[14] E. Setyowati, R. Akbarita, and R. R. Robby. Perbandingan regresi robust metode least trimmed square (lts) dan metode estimasi-s pada produksi padi di kabupaten blitar. *Jurnal Matematika UNAND*, 1(3):329–341 (in Indonesia), 2021.

[15] I. R. Akolo and A. Nadjamuddin. Analisis regresi robust estimasi least trimmed square dan estimasi maximum likelihood pada pemodelan ipm di pulau sulawesi. *Euler: Jurnal Ilmiah Matematika, Sains, dan Teknologi*, 10(2):211–221 (in Indonesia), 2022.

[16] P. Anggun, Sugito, and H. Yasin. Pemodelan regresi ridge robust s, m, mm-estimator dalam penanganan multikolinearitas dan pencilan. *Jurnal Gaussian*, 10(3):402–412 (in Indonesia), 2021.

[17] Y. N. Sari, R. R. Robby, R. Akbarita, and R. Narendra. Perbandingan analisis regresi quantil dan robust least trimmed square (lts) untuk mengidentifikasi faktor-faktor yang mempengaruhi penyebaran penyakit malaria di jawa timur. In *Prosiding Seminar Nasional Universitas Insan Budi Utomo*, volume 5, pages 182–193 (in Indonesia), 2024.

[18] A. F. Muamalah, P. T. B. Ngastiti, and A. Isro'il. Perbandingan hasil model regresi robust estimasi m (method of moment), estimasi m (maximum likelihood type), dan estimasi lts (least trimmed square) pada produksi padi di kecamatan sekaran. *MATHunesa: Jurnal Ilmiah Matematika*, 12(3):540–548 (in Indonesia), 2024.

[19] W. O. Amni, A. K. Jaya, and Nirwan. Perbandingan analisis komponen utama robust minimum covarian determinant dengan least trimmed square pada data produk domestik regional bruto. *ESTIMASI: Journal of Statistics and Its Application*, 5(2):266–281 (in Indonesia), 2024.

[20] T. W. Nurmulyati, D. Permana, N. Amalita, and Z. Martha. Comparison of estimate method of moment and least trimmed squares in models robust regression. *UNP Journal of Statistics and Data Science*, 2(2):204–248, 2024.

[21] T. E. Lestari and R. D. T. Yuansa. Response surface regression with lts and mm-estimator to overcome outliers on red roselle flowers. *Jurnal VARIAN*, 4(2):91–98, 2021.

[22] G. Mardiatmoko. Pentingnya uji asumsi klasik pada analisis regresi linier berganda (studi kasus penyusunan persamaan allometrik kenari muda *[Canarium indicum L.]*). *Barekeng: Jurnal Ilmu Matematika & Terapan*, 14(3):333–342 (in Indonesia), 2020.

[23] Michael H. Kutner, Christoper J. Nactsheim, J. Neter, and W. Li. *Applied Linear Statistical Models.* McGraw-Hill Education, 2004.

