Variance Journal of Statistics and Its Applications
P-ISSN: 2685-8738 E-ISSN: 2685-872X
Variance: Journal of Statistics and Its Applications
Volume 7 Issue 1 | April 2025
10.30598/variancevol7iss1page105-112

MODELLING THE NUMBER OF POOR POPULATION IN EAST JAVA USING QUANTILE REGRESSION

Alif Yuanita Kartini[1*], Tisa Dwi Julianti Huda[2], Jauhara Rana Budiani[3]
[1,2,3]Department of Statistics, Faculty of Science and Technology, Universitas Nahdlatul Ulama Sunan Giri
Jalan A Yani No. 10, Bojonegoro, 62115, East Java, Indonesia

***Corresponding Author's Email**: alifyuanita@unugiri.ac.id*

Abstract: *The economic development of East Java continues to increase every year. However, this increase is not directly proportional to a significant decrease in poverty rates. Therefore, research is needed to determine the factors influencing poverty in East Java. This is important because it can be used as a consideration for the East Java Provincial Government in designing strategies to reduce poverty. In the case of the number of poor people in East Java, there are outlier data, so the quantile regression method is used to overcome this. This study uses several quantile values, namely 0.25, 0.50, and 0.75. Based on the quantile regression parameter estimation results, one significant category at all quantile levels is the Average Length of Schooling variable. The quantile regression model obtains four categories of the Poor Population: low, medium, high, and very high. Based on the classification of the Poor Population in East Java in 2023, there are four districts/cities with a low number of poor people, 18 districts/cities with a moderate number of poor people, and 16 districts/cities with a high number of poor people.*

Keywords*: *Economic, Poverty, Quantile Regression.

1. INTRODUCTION

Poverty is a typical problem that occurs in almost every country. Poverty has become a topic of daily debate in the mass media, academia, and the government. Economic, social, and cultural aspects, among others, are part of the poverty problem, so poverty alleviation efforts must be included in various areas of people's lives and implemented comprehensively [1]. Indonesia itself is classified as a developing country with a high poverty rate. Poverty resolution in Indonesia must start from the lowest level, namely the regencies and provincial levels [2]. In the last five years, East Java has had the highest economic growth on the island of Java after DKI Jakarta. However, in 2023, East Java had 4,189 million poor people. This makes East Java the province with the largest number of poor people on the island of Java. According to data from the Central Statistics Agency, East Java's economic development continues to rise yearly. However, this increase is not proportional to a significant reduction in poverty [3]. Therefore, it is essential to research poverty to find out what causes poverty and find solutions to it [4].

Several factors influence the number of poor people, including the Open Unemployment Rate and the Human Development Index, two significant factors in the number of poor people [5]. Meanwhile, other research shows that the education index, health index, and per capita income influence the number of poor people [6]. Furthermore, other research shows that life expectancy and per capita expenditure are significant variables in the number of poor people [7]. Considerable variations in the number of poor people are reflected in the Human Development Index and unemployment [8].

The regression analysis method can be used to find out the factors that affect the poor population in East Java. However, after modeling using regression analysis, it was identified that outliers could affect the analysis results. In addition, outliers will cause errors in concluding if ignored. Therefore, modeling is needed with another robust approach to the existence of outliers, namely, using the quantile regression method. Quantile regression is a method that separates or divides data into specific quantiles, minimizing the weighted absolute remainder that is not symmetrical and assuming a conditional quantile function in a data distribution. In addition, quantile regression can also be an alternative if, with linear regression, there is a violation of assumptions because there is data that is far different from other data points and data diversity [9].

Submitted: 2024-12-17 *Accepted: 2025-06-12*

Several studies on quantile regression have been carried out, including quantile regression used to predict the relationship between internet access and completion of school assignments in the community [10]. Other research uses quantile regression to see how inflation, Gross Regional Domestic Product (GDP), and families receiving social assistance affect household consumption in Indonesia [11]. Based on previous studies, quantile regression can be used in various fields such as the social, health, and economic fields. The main advantage of the quantum regression method is that it is beneficial when used on data with a uniform and asymmetrical distribution. Quantile regression is also not affected by the presence of outliers, so it does not affect data stability [12]. This study offers a quantile regression approach to overcome outliers in poverty data in East Java.

Furthermore, this study uses a quantile regression model to obtain predictive values for the number of poor people in East Java. The results of these predictions are used to group the number of poor people based on the factors that influence them. This study is expected to be a solution and consideration for the East Java provincial government in making policies to handle and reduce cases of poor people in East Java.

2. METHODOLOGY

The data used in this study comes from secondary data on the number of people living in poverty and the factors suspected to affect it, based on regencies /cities in East Java in 2023. The data was obtained from the official website of the Central Statistics Agency of East Java (https://jatim.bps.go.id/). The response variables and predictor variables used in this study are presented in detail in Table 1 as follows:

Table 1. Research Variables

Variable	Operational Definition	Data Scale
Poverty (Y)	The percentage of the population whose income is below the poverty line set by the Central Statistics Agency (BPS).	Ratio
Average Length of Schooling (X_1)	Number of years of study completed in formal education by people aged 15 years and over, excluding repeated or incomplete years	Ratio
Open Unemployment Rate (X_2)	Percentage of the number of unemployed to the number of the workforce. The workforce is the working age population (15 years and over) who are working or have jobs but are temporarily unemployed, and the unemployed.	Ratio
Labor Force Participation Rate (X_3)	The proportion of the working age population (15 years and over) who are economically active, namely those who are working or looking for work, to the total working age population.	
Life Expectancy (X_4)	The average number of years a person is expected to live from birth, based on the mortality conditions prevailing in a given year.	Ratio
Per Capita Expenditure (X_5)	Average consumption costs per person in a region or group during a certain period, calculated by dividing total household consumption expenditure by the population for a year or a certain time period.	Ratio

Furthermore, the analysis steps used in this study are as follows:

1. Make a line diagram to determine the characteristics of the poor people in East Java.
2. Using regression analysis, model the relationship pattern between the number of poor people in East Java and the causal factors. The general form of the regression model equation is as follows [13],[14],[15]:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_K x_{Ki} + \varepsilon_i\ , \quad i = 1,2, \dots, n \tag{1}$$

Where:

y_i : response variable at the i-th observation.

x_{Ki} : the k-th predictor variable at the i-th observation.

β_0 : intercept.

β_K : slope variable predictor i-th.

ε_i : error at the i-th observation.

K : the number of predictor variables.

3. Detect the presence of outliers in regression by creating a boxplot from the residual regression model that has been obtained. Visually, a Boxplot shows data points that deviate significantly from normal values in a data set. A Boxplot provides a clear picture of the distribution of the data, including the median, quartiles, and potential outliers that appear as points outside the whiskers (upper or lower limits) [16],[17],[18],[19].
4. Modeling the number of poor populations in East Java using quantile regression with three quantile levels, namely the 0.25-th, 0.5-th, and 0.75-th quantiles. The general forms of quantile regression are as follows [20],[21],[22],[23]:
$$q_\tau(Y_i|X_i) = X_i^T \beta(\tau_j) \tag{2}$$
for all quantiles in the interval (0,1).
5. Obtain a prediction value for the number of poor populations in East Java based on the quantile regression model that has been obtained.
6. Grouping the number of poor populations based on the predicted value with the following criteria [24],[25],[26]:
   a) Low category if the observation value is below the regression line at the 0.25th quantile or if $y < \hat{q}_{0.25}(y|x)$.
   b) Medium category if the observation value is between the regression line at the 0.25th and 0.50th quantiles or if $\hat{q}_{0.25}(y|x) \leq y < \hat{q}_{0.25}(y|x)$.
   c) High category if the observation value is between the regression line at the 0.50th and 0.75th quantiles or if $\hat{q}_{0.50}(y|x) \leq y < \hat{q}_{0.75}(y|x)$.
   d) Very high category if the observation value is between the regression line at the 0.75th quantile or if $y > \hat{q}_{0.75}(y|x)$.
7. Grouping regency/city areas in East Java Province based on the results of grouping based on prediction values in the form of maps.

3. RESULTS AND DISCUSSION

Before conducting an in-depth analysis using quantile regression, a descriptive statistical analysis was conducted to see the general picture of the number of poor people in East Java Province in 2023. The descriptive statistics regarding the number of poor people in East Java Province are shown in Table 2.

Table 2. Descriptive Statistics of the Number of Poor People in East Java Province

Minimum	Maximum	Average	Deviation Standard
7.10	251.36	110.23	68.31

Table 2 shows that the lowest number of poor people in East Java Province in 2023 is 7.10 thousand. Meanwhile, the highest number of poor people is 251.36 thousand people. Furthermore, the average value of the number of poor people is 110.23 thousand people. And the standard deviation from the number of poor people was found to be 68.31 thousand people. Furthermore, the characteristics of the number of poor people in East Java Province can also be seen in the line diagram shown in Figure 1.

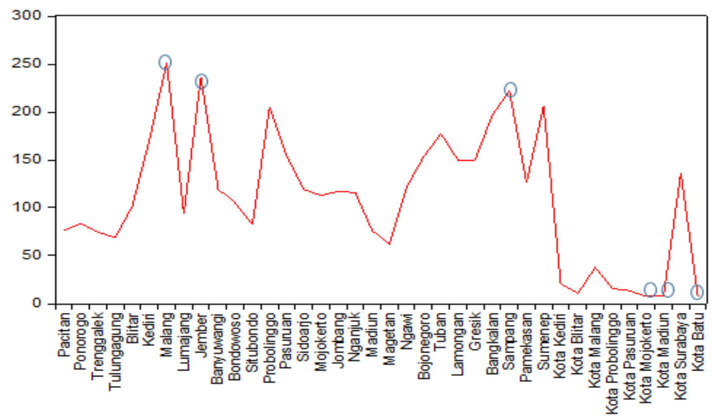

Figure 1. Graph of the Number of Poor Populations

Based on Figure 1, it is seen that the least number of poor populations is found in Mojokerto City with a total of 7.65 thousand people, followed by Madiun City with 8.46 thousand people, and Batu City with 7.1 thousand people in the area. For the most significant number of poor populations, there is the Sampang Regency with a total of 221.71 thousand people living in the area, followed by Jember Regency with 236.46 thousand people living in poverty, and Malang Regency with 251.36 thousand people living in poverty.

The relationship between the predictor variables used in this study and the poor population in East Java Province can be known using multiple linear regression analysis. Based on the results of parameter estimation, a multiple linear regression model is obtained as follows:

$$\hat{y} = 140.562 - 52.902X_1 + 10.008X_2 - 4.641X_3 + 8.385X_4 + 0.008X_5 \quad (3)$$

Based on the regression model obtained, the presence or absence of outliers in the model was identified. The presence or absence of outliers can be identified through the residual boxplots of the model. The following is the residual plot box of the multiple linear regression model as shown in Figure 2.

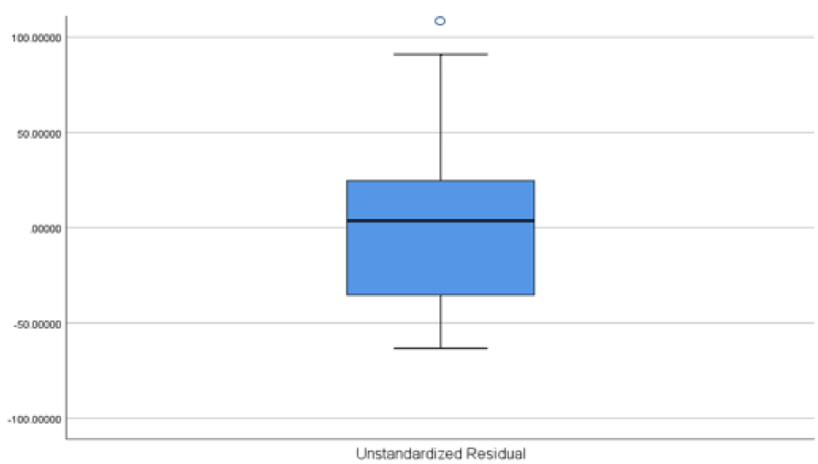

Figure 2. Multiple Linear Regression Residual Boxplot

Based on Figure 2, it can be seen that modeling the number of poor populations using multiple linear regression shows the existence of outliers. The outliers can be seen from the residual boxplot of the model, which displays a rounded point at the top of the vertical line indicating the existence of an outlier. Therefore, quantile regression can be used as an alternative in modeling the number of poor populations in East Java because it can accommodate outliers.

In this study, three levels of quantiles will be used, namely quantile 0.25, quantile 0.50, and quantile 0.75, to create a quantile model for each quantile. The results of the parameter estimation on each quantile are shown in Table 3.

Table 3. Estimation of Quantile Regression Parameters

Parameters	Quantile (θ) 0.25		0.50		0.75	
	Value	**P-Value**	**Value**	**P- Value**	**Value**	**P- Value**
β_0 (θ)	-192.368	0.775	-240.625	0.730	-175.9024	0.802
β_1 (θ)	-499.223	**0.032**	-65.111	**0.003**	-64.87985	**0.000**
β_2 (θ)	-2.055	0.905	4.814	0.680	13.88015	0.244
β_3 (θ)	-3.126	0.252	-5.416	0.169	-3.218724	0.444
β_4 (θ)	10.552	0.244	15.123	0.183	12.16036	0.284
β_5 (θ)	0.012	0.506	0.014	0.154	0.011851	0.220

Table 3 shows that at the 0.25th quantile, one significant variable is the Average School Duration. Meanwhile, in the 0.50th quantile, one considerable variable is the Average School Duration. And at the 0.75th quantile, there is only one significant variable, namely the Average Length of School. Therefore, based on the significance of the parameters' significance at several quantile levels, this study will use one predictor variable, namely the Average School Length, to obtain the predicted value of the Number of Poor Population in East Java. Moreover, the prediction value will group the Number of Poor Populations in East Java. The results of estimating the quantile regression parameters of the number of poor populations based on the average length of schooling at each quantile level are shown in Table 4 below.

Table 4. Estimation of Quantile Regression Parameters of the Poor Population Based on Average Length of School

Variable	Quantile 0.25		0.50		0.75	
	Value	**P-value**	**Value**	**P-value**	**Value**	**P-value**
Intercept	380.814	0.001	402.750	0.000	259.197	0.004
X_1	-23.060	0.018	-35.708	0.000	-29.413	0.001

Based on Table 4, a quantile regression model for the Number of Poor Populations in East Java is formed based on the Average School Length as follows:

$$\hat{Q}_{0.25}\,(y \mid x) = 380.814 - 23.060X_1 \tag{4}$$

$$\hat{Q}_{0.50}\,(y \mid x) = 402.750 - 35.708X_1 \tag{5}$$

$$\hat{Q}_{0.75}\,(y \mid x) = 259.197 - 29.413X_1 \tag{6}$$

From the above quantile regression model, an estimation curve for the Number of Poor Populations in East Java will be formed as shown in Figure 3.

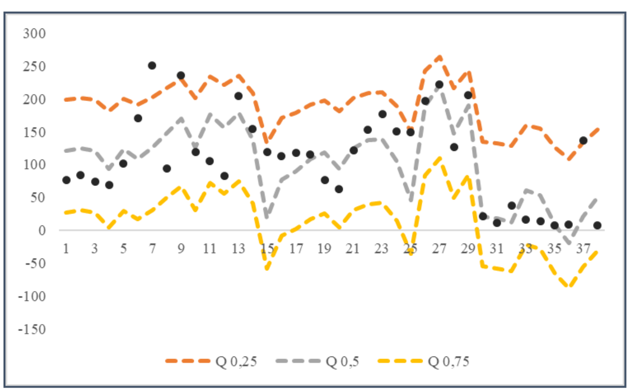

Figure 3. Quantile Regression Estimation Curve for Poor Population Based on the Average Length of School

Based on the above estimates and quantile regression models, it can be used to group the number of poor populations in East Java into several categories, such as the low category if the observation value is below the regression line at the 0.25-th quantile or if $y < \hat{q}_{0.25}$ (y | x), then for the medium category if the observation value is between the regression line at the 0.25-th and 0.50-th quantile, or when $\hat{q}_{0.25}$ $(y \mid x) \leq y < \hat{q}_{0.50}$ $(y \mid x)$, for the high category if the observation value is between the regression line at the 0.50-th and 0.75-th quantile or when $\hat{q}_{0,50}$ $(y \mid x) \leq y < \hat{q}_{0.75}$ $(y \mid x)$, and for the very high category if the observation value is above the regression line at the 0.75-th quantile or if $y > \hat{q}_{0.75}$ $(y \mid x)$ [23]. The results of the calculation based on the estimation of the quantile regression model for the grouping of the Number of Poor Populations are shown in Table 5:

Table 5. Grouping of the Number of Poor Populations in East Java Based on the Quantile Regression Model

Number of Poor People	Number of Regencies/Cities	Regency/City
Low	4	Malang, Malang City, Jember, Probolinggo City.
Medium	18	Kediri, Pasuruan, Pasuruan City, Sidoarjo, Mojokerto, Mojokerto City, Jombang, Nganjuk, Bojonegoro, Tuban, Lamongan, Gresik, Bangkalan, Sumenep, Kediri City, Madiun, Madiun City, Surabaya.
High	16	Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Blitar City, Lumajang, Banyuwangi, Bondowoso, Situbondo, Magetan, Ngawi, Sampang, Pamekasan, Probolinggo, Batu.
Very High	0	Nothing in the very high category

From the results of the grouping in Table 5, it can be seen that there are 4 Regencies/cities classified as low poor population, 18 regencies/cities classified as medium poor population, and 16 regencies/cities classified as high poverty category. Meanwhile, none of them are in the very high category. Furthermore, the grouping of Regencies/Cities in East Java based on Table 4 will be visualized in Figure 4.

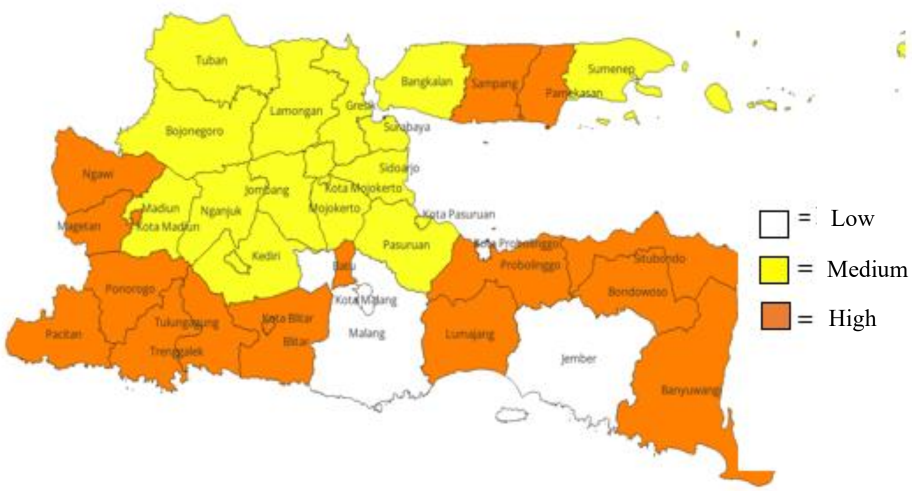

Figure 4. Mapping the Number of Poor Populations in East Java Based on a Quantile Regression Model

In this research, we apply the model in Equation 4, Equation 5, and Equation 6 to classify the poor people in East Java Province based on the average number of schooling in 2023. The number of poor people is grouped into four: the number of low poor people, the number of medium poor people, the number of high poor people, and the number of very high poor people. In Figure 4, it can be seen that districts that have a low number of poor people are marked in white, namely Malang, Malang City, Jember, and Probolinggo City. Districts that have a moderate number of poor people are marked in yellow, namely Kediri, Pasuruan, Pasuruan City, Sidoarjo, Mojokerto, Mojokerto City, Jombang, Nganjuk, Bojonegoro, Tuban, Lamongan, Gresik, Bangkalan, Sumenep, Kediri City, Madiun, Madiun City, and Surabaya. And districts that have a high number of poor people are marked in orange, namely Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Blitar City, Lumajang, Banyuwangi, Bondowoso, Situbondo, Magetan, Ngawi, Sampang, Pamekasan, Probolinggo, Batu. Meanwhile, no district/city has a very high population.

4. CONCLUSION

In this study, there is an outlier in the linear regression residual model, so quantile regression is used to overcome this. The quantile regression analysis in this study used three quantile levels, namely 0.25, 0.50, and 0.75, with the model formed.

$$\hat{Q}_{0,25}\,(y \mid x) = 380{,}814 - 23{,}060X_1$$

$$\hat{Q}_{0,50}\,(y \mid x) = 402{,}750 - 35{,}708X_1$$

$$\hat{Q}_{0,75}\,(y \mid x) = 259{,}197 - 29{,}413X_1$$

The quantile regression model obtained four categories of the Poor Population: low, medium, high, and very high. Based on the classification of the Number of Poor Populations in East Java in 2023, there are four regencies/cities with a low number of poor populations, 18 regencies/cities with a medium number of poor populations, and 16 regencies/cities with a high number of poor populations.

REFERENCES

[1] D. Endrawati, “Determinants of working poverty in Indonesia,” *J. Econ. Dev.*, vol. 24, no. 3, pp. 230–246, 2022.

[2] Z. Hasan, “The effect of economic growth and human development index on poverty in Indonesia,” *J. Econ. Sustain.*, vol. 3, no. 1, p. 12, 2021.

[3] D. Kholifah and H. Sumarsono, “EXAMINING THE DETERMINANTS OF POVERTY IN EAST JAVA.,” *J. Ekon. dan Bisnis Airlangga*, vol. 33, no. 1, 2022.

[4] F. Muslihatinningsih and J. Abidin, “Analisis Kemiskinan Provinsi Jawa Timur,” *J. Ekon. Ekuilibrium*, vol. 4, no. 2, pp. 29–39, 2020.

[5] A. S. Helwani and R. Rusdiansyah, “Analisis Faktor-Faktor yang Mempengaruhi Jumlah Penduduk Miskin di Provinsi Kalimantan Selatan,” *JIEP J. Ilmu Ekon. Dan Pembang.*, vol. 5, no. 1, pp. 217–233, 2022.

[6] E. H. Jacobus, P. Kindangen, and E. N. Walewangko, “Analisis faktor-faktor yang mempengaruhi kemiskinan rumah tangga di Sulawesi Utara,” *J. Pembang. Ekon. dan Keuang. Drh.*, vol. 19, no. 3, pp. 86–103, 2021.

[7] F. P. Utami and I. Lubis, “Analisis Faktor-Faktor yang Mempengaruhi Kemiskinan Di Aceh Bagian Timur,” *J. Samudra Ekon.*, vol. 6, no. 1, pp. 1–9, 2022.

[8] D. Astutik, “Analisis Faktor yang Mempengaruhi Kemiskinan di Jawa Timur (Pendekatan Spasial),” *J. Ilm. Mhs. FEB*, vol. 9, no. 1, 2021.

[9] X. Xu, Q. Wang, and C. Li, “The impact of dependency burden on urban household health expenditure and its regional heterogeneity in China: Based on quantile regression method,” *Front. public Heal.*, vol. 10, p. 876088, 2022.

[10] S. Parwodiwiyono, “Kaitan Akses Internet dan Penyelesaian Tugas Sekolah Berdasarkan Regresi Kuantil Menuju Masyarakat 5.0,” *JARTIKA J. Ris. Teknol. dan Inov. Pendidik.*, vol. 3, no. 2, pp. 342–350, 2020, doi: 10.36765/jartika.v3i2.283.

[11] A. Murwiati and R. Zulkarnain, “Analisis Pengaruh Produk Domestik Regional Bruto , Keluarga Penerima Manfaat , dan Inflasi Terhadap Konsumsi Rumah Tangga Di Indonesia Dengan Regresi Kuantil,” *J. Soc. Sci. Res.*, vol. 3, pp. 8631–8643, 2023.

[12] L. Torossian, V. Picheny, R. Faivre, and A. Garivier, “A review on quantile regression for stochastic computer experiments,” *Reliab. Eng. Syst. Saf.*, vol. 201, p. 106858, 2020.

[13] D. Maulud and A. M. Abdulazeez, "A review on linear regression comprehensive in machine learning," *J. Appl. Sci. Technol. Trends*, vol. 1, no. 2, pp. 140–147, 2020.

[14] S. Rath, A. Tripathy, and A. R. Tripathy, "Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model," *Diabetes Metab. Syndr. Clin. Res. Rev.*, vol. 14, no. 5, pp. 1467–1474, 2020.

[15] G. C. Paul, S. Saha, and T. K. Hembram, "Application of phenology-based algorithm and linear regression model for estimating rice cultivated areas and yield using remote sensing data in Bansloi River Basin, Eastern India," *Remote Sens. Appl. Soc. Environ.*, vol. 19, p. 100367, 2020.

[16] K. Anindya *et al.*, "Effect of multimorbidity on utilisation and out-of-pocket expenditure in Indonesia: quantile regression analysis," *BMC Health Serv. Res.*, vol. 21, no. 1, p. 427, 2021.

[17] A. Boukerche, L. Zheng, and O. Alfandi, "Outlier detection: Methods, models, and classification," *ACM Comput. Surv.*, vol. 53, no. 3, pp. 1–37, 2020.

[18] A. Smiti, "A critical overview of outlier detection methods," *Comput. Sci. Rev.*, vol. 38, p. 100306, 2020.

[19] I. Souiden, M. N. Omri, and Z. Brahmi, "A survey of outlier detection in high dimensional data streams," *Comput. Sci. Rev.*, vol. 44, p. 100463, 2022.

[20] F. Yanuar, T. Abrari, I. R. HG, and A. Zetra, "Spatial autoregressive quantile regression with application on open unemployment data," *Sci. Technol. Indones.*, vol. 8, no. 2, pp. 321–329, 2023.

[21] F. Bilgili, S. Kuşkaya, M. Khan, A. Awan, and O. Türker, "The roles of economic growth and health expenditure on CO2 emissions in selected Asian countries: a quantile regression model approach," *Environ. Sci. Pollut. Res.*, vol. 28, no. 33, pp. 44949–44972, 2021.

[22] F. Bilgili, I. Ozturk, E. Kocak, S. Kuskaya, and A. Cingoz, "The nexus between access to electricity and CO2 damage in Asian Countries: The evidence from quantile regression models," *Energy Build.*, vol. 256, p. 111761, 2022.

[23] F. Liu, M. Umair, and J. Gao, "Assessing oil price volatility co-movement with stock market volatility through quantile regression approach," *Resour. Policy*, vol. 81, p. 103375, 2023.

[24] N. Abid, F. Ahmad, J. Aftab, and A. Razzaq, "A blessing or a burden? Assessing the impact of climate change mitigation efforts in Europe using quantile regression models," *Energy Policy*, vol. 178, p. 113589, 2023.

[25] A. Demir, V. Pesqué-Cela, Y. Altunbas, and V. Murinde, "Fintech, financial inclusion and income inequality: a quantile regression approach," *Eur. J. Financ.*, vol. 28, no. 1, pp. 86–107, 2022.

[26] A. Azimli, "The impact of COVID-19 on the degree of dependence and structure of risk-return relationship: A quantile regression approach," *Financ. Res. Lett.*, vol. 36, p. 101648, 2020.