Within-Group Estimators for Unbalanced-Panel Data Regression Model of the Open Unemployment Rate Data in East Kalimantan Province

Desi Yuniarti, Dedi Rosadi*, and Abdurakhman

***Abstract*—The COVID-19 pandemic has hit hard the Indonesian economy. Many businesses had to close because they could not cover operational costs, and many workers were laid off creating an unemployment crisis. Unemployment causes people's productivity and income to decrease, leading to poverty and other social problems, making it a crucial problem and great concern for the nation. Economic conditions during this pandemic have also provided an unusual pattern in economic data, in which outliers may occur, leading to biased parameter estimation results. For that reason, it is necessary to deal with outliers in research data appropriately. This study aims to find within-group estimators for unbalanced panel data regression model of the Open Unemployment Rate (OUR) in East Kalimantan Province and the factors that influence it. The method used is the within transformation with mean centering and median centering processing methods. The results of this study may provide advice on factors that can increase and decrease the OUR of East Kalimantan Province. The results show that the best model for estimating OUR data in East Kalimantan Province is the within-transformation estimation method using median centering. According to the best model, the Human Development Index (HDI) and Gross Regional Domestic Product (GRDP) are two factors that influence the OUR of East Kalimantan Province (GRDP).**

***Index Terms*—outliers; panel regression; robust estimators; unbalanced panel data; unemployment.**

I. Introduction

PANEL data combines time series with cross-section data where the same cross-section units are pooled across numerous periods. It has two dimensions: the space dimension, which is represented by cross-section data, and the time dimension, which is represented by time-series data. Thus, it is more informative and allows separate estimation from individual and time characteristics. Panel data with different time-series observations in each cross-section unit is called unbalanced panel data. Unbalanced panel data is more likely to occur in economic data collection. One method of estimating the panel data regression model with a fixed-effects approach is the within-transformation. Hence, the estimator is called within-group estimators [1].

Manuscript received July 09, 2022; revised March 03, 2023. This research is part of research for doctoral studies funded by the Lembaga Pengelola Dana Pendidikan (LPDP) Republik Indonesia.

D.Yuniarti is a doctoral candidate from Department of Mathematics, Universitas Gadjah Mada, Indonesia (e-mail: desiyuniarti@mail.ugm.ac.id). She is also a lecturer from Department of Mathematics, Universitas Mulawarman, Indonesia (e-mail:desi_yuniarti@fmail.unmul.ac.id)

*D.Rosadi (corresponding author) is a professor from Department of Mathematics, Universitas Gadjah Mada, Indonesia (e-mail:dedirosadi@gadjahmada.edu).

Abdurakhman is an associate professor from Department of Mathematics, Universitas Gadjah Mada, Indonesia (e-mail:rachmanstat@ugm.ac.id).

Outliers can cause the within-group estimators to be biased, so it is necessary to find a better parameter estimate. The authors in [2] proposed methods to produce a robust estimator for panel data regression models with a fixed-effects approach, namely within-group generalized M estimator (WGM) and within-group MS (WMS) estimator methods. The initial stage of the WGM and WMS methods by [2] is to center the data on the median of each cross-section unit (median centering) because the median is a more robust estimator than the mean. Within-group estimators using median centering is more preferred over mean centering because it may provide more robust solutions than mean centering. The authors in [3] also explains that the median is more robust to outliers than the mean. The mean is very sensitive to values that deviate between n observations. The breakdown value of the sample mean is $1/n$, so replacing even one observation with a very large value can entirely change the mean. The sample mean breakdown value becomes 0% for large n, which is the worst possible breakdown value. When the median has a breakdown point of around 50%, it can withstand almost 50% of outliers, and this value is the best breakdown value. Research [4] also used the median, which was used in detecting outliers in the classification. Furthermore, research [5] used the median filtering technique to remove noise in the image.

This study aims to determine the estimators for the panel data regression model for the Open Unemployment Rate (OUR) data in East Kalimantan Province using an unbalanced panel data regression analysis with a fixed-effects approach. The COVID-19 pandemic has caused outliers in the data, so the estimation method uses the within-transformation estimation method, with mean centering and median centering. Furthermore, we will compare the Mean Squared Error (MSE) value to determine the best model estimate.

The COVID-19 pandemic has dramatically affected the condition of the Indonesian economy. The government-issued regulations restricting community activities to reduce the number of positive cases of COVID-19 have caused decreased household consumption. Many businesses have to be closed because they cannot cover operational costs, and many workers have been laid off. The number of people working in East Kalimantan in August 2020 decreased by 685 people than in August 2019. The OUR in East Kalimantan was 6.87 percent, which increased by 0.93 percent or 17.92 thousand people compared to August 2019. More than 411 thousand people were affected by COVID-19,

with 30.99 thousand unemployed, 14.34 thousand not in the workforce, 21.20 thousand temporarily unemployed, and 344.85 thousand working reduced hours [6]. Research of [7] also discussed the impact of the COVID-19 pandemic on the economy. A study of [7] reported that an increase in the number of people testing positive for COVID-19 causes Indonesia inflation to decrease. Low inflation indicates a decline in people's purchasing power. For this reason, the government must make various efforts to increase people's purchasing power.

In [6], it is explained that work is economic activities to obtain or help income or profits with a minimum of one hour without interruption in the past week, including activities of unpaid workers who help in a business/economic activity. Unemployment is a term given to a situation where people are not working or looking for work. Many factors cause unemployment, but generally, there is a mismatch between the number of job seekers and the number of available jobs. Unemployment is indeed a severe problem that must be handled with utmost priority. Unemployment is causing the community's productivity and income to decrease, poverty and other social problems.

Overcoming unemployment is therefore very important for the government because not only does unemployment cause poverty and other social problems, buat also affects economic growth. Various studies on unemployment have been carried out, and some of the results of these studies have become the basis for policies for reducing unemployment in Indonesia. We found some studies on unemployment particularly relevant for our study because they analyze unemployment with different methods. These studies analyze unemployment with different methods. Research by [8] used a supply and demand approach in discussing urban unemployment, while the authors in [9] used the spline nonparametric regression. In [10] used descriptive and exploratory methods for the analysis of poverty and unemployment alleviation programs. Research by [11] used the path analysis method.

The government continues controlling the pandemic so that social and economic activities could recover and poverty and unemployment rates could decrease. This study could help suggest factors influencing OUR in East Kalimantan Province.

II. Unbalanced Panel Data Regression Model

Based on [12], observations of one or more variables over several time periods will produce time-series data, while observations of one or more variables from one sample unit or entity at a time will generate cross-section data. Panel data combines time-series and cross-section data, where the same cross-section units are surveyed over several periods. Thus, panel data has the dimensions of space and time. Panel data is balanced or complete if each cross-section unit has the same number of time-series observations. On the contrary, unbalanced or incomplete panel data are panel data with a different number of time-series observations in each cross-section unit.

Panel data has many advantages. According to [1], the advantage of using panel data is that it can control individual heterogeneity, provides more informative, more varied data, greater degrees of freedom, and is be more efficient. It is better able to study the dynamics of adjustments, which allows the estimation of individual's character and characteristics at one point in time to their character and characteristics at another point in time. Panel data can better identify and measure influences that cannot be detected only by cross-section or time-series data. It can minimize bias generated by the aggregation of individuals or companies due to more data units.

A. The One-Way Unbalanced Panel Data Regression Model

The panel data regression model is a regression model that uses panel data. An explanation of the panel data regression model is provided by [1]. In particular, the one-way unbalanced panel data regression model is

$$Y_{it} = \alpha + \mathbf{X}'_{it}\boldsymbol{\beta} + u_{it} \tag{1}$$

where $i = 1, 2, , \cdots, N$, $t = 1, 2, \cdots, T$,

$$\underset{K\times 1}{\mathbf{X}_{it}} = \begin{bmatrix} X_{1_{it}} \\ X_{2_{it}} \\ \vdots \\ X_{K_{it}} \end{bmatrix}, \text{dan} \quad \underset{K\times 1}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}$$

for $n = \sum_{i=1}^{N} T_i$. The model (1) is unbalanced because N individuals are observed over different periods, namely T_i for $i = 1, 2, \cdots, N$. The one-way error component of the model (1) is

$$u_{it} = \mu_i + v_{it} \tag{2}$$

Model (1) with the error component (2) is a one-way unbalanced-panel data regression model with the fixed-effects approach if μ_i is fixed, $\mu_i \sim IIN(0, \sigma_\mu^2)$ and independent of $v_{it} \sim IIN(0, \sigma_v^2)$ and X_{it} independent to v_{it}.

B. Within-Group Estimators of One-Way Unbalanced Panel Data Regression Model with Fixed-Effects Approach

According to [1], one method of estimating a one-way panel data regression model with a fixed-effect approach is to use within-transformation. This method is performed by specifying model (1) in vector form

$$\mathbf{y} = \alpha\boldsymbol{\iota}_n + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{3}$$

where $\mathbf{y}$ is a vector of dimensions $n \times 1$, α is a scalar, $\boldsymbol{\iota}_n$ is a vector of ones of dimension $n \times 1$, $\mathbf{X}$ is a $n \times K$ matriks, $\boldsymbol{\beta}$ is a $K \times 1$ vector and $\mathbf{u}$ is a error vector of dimensions $n \times 1$. The simple form of (3) is

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\delta} + \mathbf{u} \tag{4}$$

where $\mathbf{Z} = (\boldsymbol{\iota}_n, \mathbf{X})$ is a matrix of size $n \times (K + 1)$, and $\boldsymbol{\delta}' = (\boldsymbol{\alpha}', \boldsymbol{\beta}')$ of size $(K + 1) \times 1$. Then the vector form of the one-way error component (2) is

$$\mathbf{u} = Z_\mu\boldsymbol{\mu} + \mathbf{v} \tag{5}$$

where $Z_\mu = \text{diag}(\boldsymbol{\iota}_{T_i})$ is a square diagonal matrix with the vector elements $\boldsymbol{\iota}_{T_i}$ on the main diagonal, ι_{T_i} is a vector of ones of size T_i, $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_N)'$, and $\mathbf{v} = (v_{11}, \ldots, v_{1T_1}, v_{21}, \ldots, v_{2T_2}, \ldots, v_{N1}, \ldots, v_{NT_N})'$.

The ordinary least squares (OLS) solution for the unbalanced panel data regression model with the fixed-effects

approach can be obtained by applying the OLS estimation method to (4), which yields

$$\hat{\boldsymbol{\delta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} \tag{6}$$

provided that the inverse $\mathbf{Z}'\mathbf{Z}$ exists.

The method of estimation with the within transformation called within-group estimation can be done by forming a matrix $\mathbf{Q} = \text{diag}(\mathbf{E}_{T_i})$, that is, the matrix with the main diagonal: $\mathbf{E}_{T_i} = \mathbf{I}_{T_i} - \bar{\mathbf{J}}_{T_i}$ where $\mathbf{I}_{T_i}$ is the identity matrix of dimension $T_i \times T_i$ and $\bar{\mathbf{J}}_{T_i} = \frac{1}{T}\mathbf{J}_{T_i}$, $\mathbf{J}_{T_i}$ denotes a matrix of ones of dimension $T_i \times T_i$. Multiplying (3) by the $\mathbf{Q}$ and $\mathbf{Q}\boldsymbol{\iota}_n = \mathbf{0}$ yields

$$\begin{aligned} \mathbf{Q}\mathbf{y} &= \alpha\mathbf{Q}\boldsymbol{\iota}_n + \mathbf{Q}\mathbf{X}\boldsymbol{\beta} + \mathbf{Q}\mathbf{u} \\ \mathbf{Q}\mathbf{y} &= \mathbf{Q}\mathbf{X}\boldsymbol{\beta} + \mathbf{Q}\mathbf{u} \\ \tilde{\mathbf{y}} &= \tilde{\mathbf{X}} + \tilde{\mathbf{v}} \end{aligned} \tag{7}$$

where $\tilde{\mathbf{y}} = \mathbf{Q}\mathbf{y}, \tilde{\mathbf{X}} = \mathbf{Q}\mathbf{X},$ and $\tilde{\mathbf{v}} = \mathbf{Q}\mathbf{v}$. Applying the OLS method to the (7), then obtained

$$\hat{\boldsymbol{\beta}} = (\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}} \tag{8}$$

for $(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}$ exists.

Furthermore, [1] provided intercept estimates

$$\hat{\alpha} = (\bar{Y}_{..} - \bar{X}_{..}\hat{\boldsymbol{\beta}}) \text{ and } \hat{\mu}_i = \bar{Y}_{i.} - \hat{\alpha} - \bar{X}'_{i.}\hat{\boldsymbol{\beta}} \tag{9}$$

where $\bar{Y}_{..} = \frac{1}{n}\sum_{i=1}^{N}\sum_{t=1}^{T_i} Y_{it}$, $\bar{Y}_{i.} = \sum_{t=1}^{T_i}\frac{Y_{it}}{T_i}$, and $\bar{X}'_{i.} = \left(\bar{X}_{1_{i.}}, \bar{X}_{2_{i.}}, \cdots, \bar{X}_{k_{i.}}\right)$ for $\bar{X}_{k_{i.}} = \sum_{t=1}^{T_i}\frac{X_{k_{it}}}{T_i}$, $k = 1, 2, \cdots, K$. Based on [13] in [1], the within residual $\tilde{\mathbf{u}}$ for an unbalanced-panel data regression model is

$$\tilde{\mathbf{u}} = \mathbf{y} - \tilde{\boldsymbol{\alpha}}\boldsymbol{\iota}_n - \boldsymbol{X}\tilde{\boldsymbol{\beta}} \tag{10}$$

Within transformation by the $\mathbf{Q}$ matrix is a data transformation by determining the deviation of each data from the individual mean or unit cross-section (mean centering).

Outliers lead to efforts to get a robust estimator. In the initial stages of the [2] robust regression method, researchers in [2] proposed data centering on the median (median centering) because the median is a more robust estimator than the mean. The median centering transformation by [2] is

$$\tilde{Y}_{it} = Y_{it} - \underset{1\le t\le T_i}{\text{median}}\, Y_{it} \tag{11}$$

and,

$$\tilde{X}_{k_{it}} = X_{k_{it}} - \underset{1\le t\le T_i}{\text{median}}\, X_{k_{it}} \tag{12}$$

for $i = 1, 2, \cdots, N; \quad t = 1, 2, \cdots, T_i; \quad k = 1, 2, \cdots, K$. After centering, regress $\tilde{Y}_{it}$ against $\tilde{X}_{k_{it}}$ using the OLS method. By using the parameter estimation $\boldsymbol{\beta}$, a fixed-effects estimate can be obtained for each unit cross-section

$$\hat{\alpha}_i = \underset{1\le t\le T_i}{\text{median}}(Y_{it} - \mathbf{X}'_{it}\boldsymbol{\beta}) \quad i = 1, 2, \cdots, N. \tag{13}$$

Research [14] also compared these two data-centering methods. The results of the [14] study concluded that the MSE value of the LS method for the two centers was not much different. The median centering estimators show better results in overcoming vertical outliers for fewer cross-section units in the outliers concentrated within the individual unit scheme.

III. Research Data

A. Open Unemployment

According to [15], unemployment is when someone who belongs to the labor force seeks employment but unable to find it. The main factor that causes unemployment is a decrease in aggregate expenditure. Other factors are (i) being unemployed because they want to find better job, (ii) employers use modern production equipment to reduce labor cost, and (iii) a mismatch between the skills possessed by workers and the skills needed in the industry. Based on its characteristics, unemployment can be divided into open unemployment, hidden unemployment, seasonal unemployment, and underemployement. Open unemployment occurs because the addition of job vacancies is lower than the increase in labor. Open unemployment can also occur due to a decrease in economic activity, the use of technological advances that reduce the use of labor, or as a result of a setback in industry development.

In [15] also explained that unemployment reduces income and the prosperity level of people. It can cause economic and social problems and also interferes with the level of health. Moreover, prolonged unemployment can cause harmful psychological effects for unemployed individuals and their families.

According to [6], the working-age population is aged 15 years and over and considered able to work. Work is an economic activity carried out by someone to obtain or help to get income or profit, at least one hour (uninterrupted) in the past week. These activities include the pattern of unpaid labor activities in business or other economic activities. Furthermore, open unemployment includes people who do not have a job and are looking for work, who do not have a job and are preparing a business, who do not have a job but are not looking for work because they feel it is impossible to get a job, and who already have a job but have not started work. The labor force is an economically active workforce. Labor is the capital of economic development. The number and composition of the workforce will change along with the ongoing democratic process.

B. Data of OUR East Kalimantan Province

The aim of this study is to determine the factors affecting the OUR by regencies/cities in East Kalimantan Province. The data in this study consist of data from ten regencies/cities in East Kalimantan Province as a cross-section unit and 2014-2021 as a time series unit if the time series unit is complete. Table I shows the regencies/cities in East Kalimantan Province. This study uses an unbalanced panel data regression model because some data have different number of time series units in each cross-section unit. Table II shows the number of observations for each cross-section unit. One of the reasons why we use unbalanced panel data regression model is the establishment of a new regency Mahakam Ulu, in 2013. The data analysis in this research is descriptive statistical data and within-group estimators using plm packages of R software [16]-[17]. The research variables in this study consist of dependent and independent variables, as shown in Table III.

TABLE I
REGENCIES/CITIES OF EAST KALIMANTAN PROVINCE

i-index	Regency	i-index	City
1	Paser	8	Balikpapan
2	West Kutai (Kubar)	9	Samarinda
3	Kutai Kartanegara (Kukar)	10	Bontang
4	East Kutai (Kutim)		
5	Berau		
6	North Paser Penajam (PPU)		
7	Mahakam Ulu (Mahulu)		

TABLE II
THE NUMBER OF OBSERVATIONS OF EACH CROSS-SECTION UNIT

i-index	Regency/City	T_i
1	Paser	7
2	West Kutai (Kubar)	7
3	Kutai Kartanegara (Kukar)	7
4	East Kutai (Kutim)	7
5	Berau	7
6	North Paser Penajam (PPU)	7
7	Mahakam Ulu (Mahulu)	3
8	Balikpapan	7
9	Samarinda	7
10	Bontang	6

TABLE III
RESEARCH VARIABLE

Variable	Annoation	Unit
Y	Open Unemployment Rate (OUR)	Percent
X_1	Human Development Index (HDI)	-
X_2	Regional Minimum Wages (RMW)	Rupiahs
X_3	Dependency Ratio (DR)	Percent
X_4	Domestic Investment Project Realization (DIPR)	Million Rupiahs
X_5	Foreign Investment Realization (FIR)	Thousand US Dollars
X_6	Gross Regional Domestic Product (GRDP)	Billion Rupiahs

IV. RESULTS AND DISCUSSION

This section discusses OUR data analysis by regencies/cities in East Kalimantan Province, using descriptive statistics and within-group estimators of an unbalanced-panel data regression model with a one-way fixed-effects model with mean and median centering.

A. Descriptive Statistics of Research Data

Table IV shows descriptive statistics for OUR data and research variables. Based on Table IV, the average of OUR from 10 regencies/cities in East Kalimantan Province in 2014-2021 was 6.63%, meaning that of 100 workforce, seven were unemployed. The average independent variable of those regencies/cities in 2014-2021 is as follows. The average HDI is 73.59 and RMW is 2,637,759 rupiahs. The DR is 46.26, which means that 100 people of working age (considered productive) have a responsibility of 46 people who have not been productive and no longer productive. DIPR is 1.9 trillion rupiahs, FIR is US$ 125,609,200.00, and GRDP is 46,144,984 billion rupiahs.

TABLE IV
DESCRIPTIVE STATISTICS OF RESEARCH DATA

Variable	Amount of Data	Minimum	Maximum	Average
Y	69	2.95	12.44	6.63
X_1	80	64.32	80.76	73.59
X_2	77	1,900,000	3,412,331	2,637,759
X_3	80	37.99	56.23	46.26
X_4	75	1,500	16,773,017	1,937,428
X_5	76	6.4	1,475,211.5	125,609.2
X_6	80	1,426,530	128,610,623	46,144,984

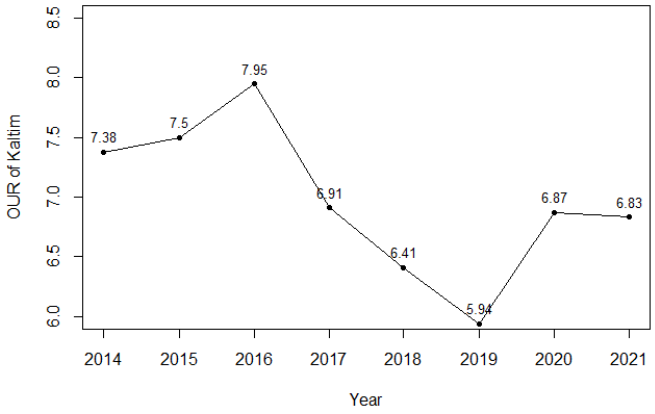

Fig. 1. OUR of East Kalimantan in 2014-2021 (Percent)

Furthermore, Fig. 1 shows the OUR data of East Kalimantan Province from 2014 to 2021. It can be seen that the highest OUR of East Kalimantan Province was in 2016 (7.95%), which means that of 100 workforce, eight were unemployed. OUR of East Kalimantan Province continued to decline until 2019 by 5.94% but increased to 6.87% in 2020. OUR data by regencies/cities in East Kalimantan Province from 2014 to 2021 can be seen in Fig. 2, which shows that the minimum of OUR value was 2.95% in Penajam Paser Utara district in 2021. The highest OUR value was 12.44% in Bontang City in 2017.

The COVID-19 pandemic has hit the economy of East Kalimantan in particular. The increasing pattern of OUR in 2020 could depict this economic condition. In this unusual pattern, outliers occur in our research data, so we examine the possibility of outliers using the Mahalanobis distance. Fig. 3 is the result of the examination using the Mahalanobis distance, which shows that several observations can be predicted as outliers.

B. Unbalanced Panel Data Regression Model of OUR East Kalimantan

Based on (1), the one-way unbalanced panel data regression model for OUR data by regencies/cities in East Kalimantan Province with the fixed-effects approach is

$$\begin{aligned} Y_{it} =& \alpha + \beta_1 X_{1_{it}} + \beta_2 X_{2_{it}} + \beta_3 X_{3_{it}} + \beta_4 X_{4_{it}} + \\ & \beta_5 X_{5_{it}} + \beta_6 X_{6_{it}} + u_{it}; \\ u_{it} =& \mu_i + v_{it}; \\ & i = 1, 2, , \cdots, 10; t = 1, 2, \cdots, T_i \end{aligned} \tag{14}$$

where T_i as in Table II; $X_1, X_2, \cdots, X_6$ are independent variables, μ_i denotes the individual effects, that

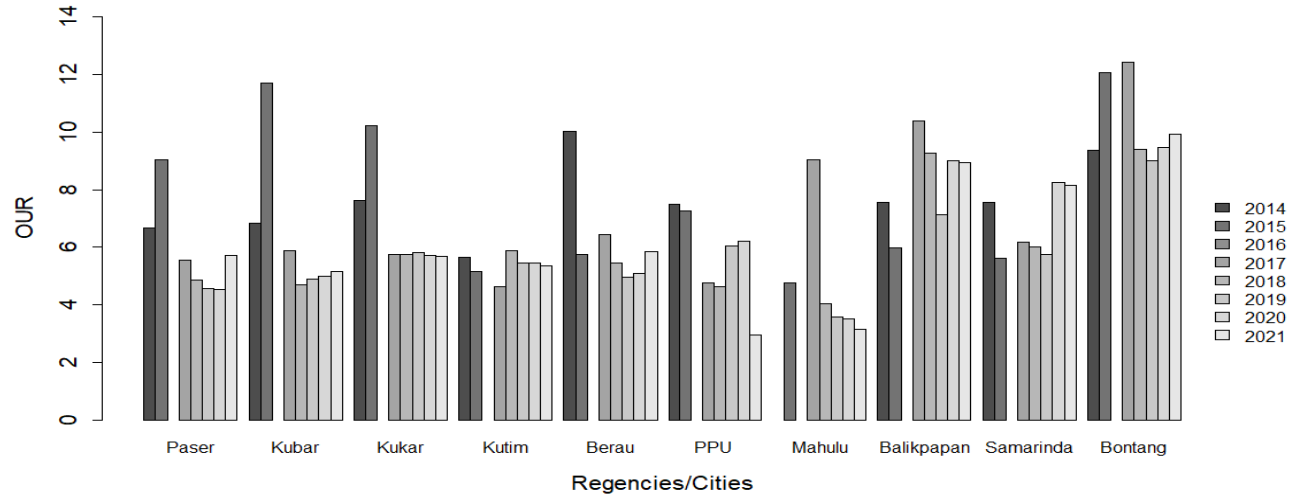

Fig. 2. OUR by Regencies/Cities of East Kalimantan in 2014-2021 (Percent)

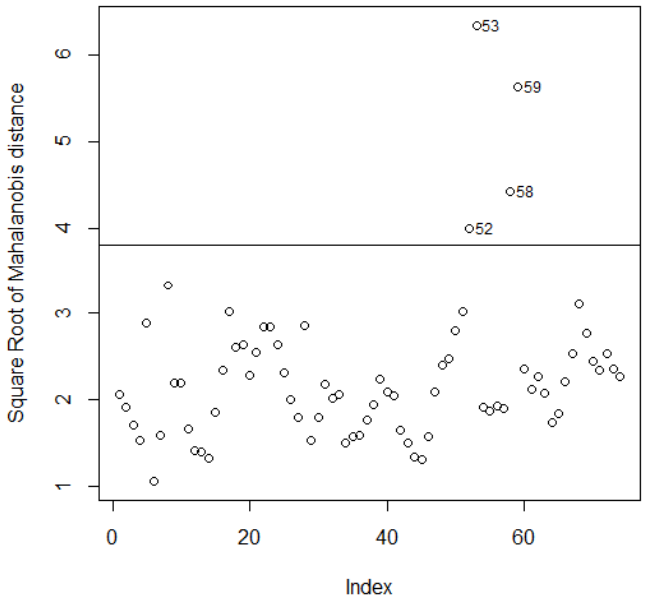

Fig. 3. Plot Square Root of Mahalanobis distance

TABLE V
PARAMETER ESTIMATION OF INITIAL MODEL

Variable	Parameter Estimation	p-value
X_1	-1.1260e+00	0.0442*
X_2	-2.7061e-07	0.8136
X_3	-2.1675e-01	0.3071
X_4	-4.6936e-08	0.5345
X_5	-1.3137e-06	0.2585
X_6	1.6475e-07	0.0560

*significant at $\alpha = 5\%$ level

TABLE VI
PARAMETER ESTIMATION OF MODEL 1

Variable	Parameter Estimation	p-value
X_1	-1.1674e+00	0.0002*
X_3	-1.5369e-01	0.3594
X_4	-4.6979e-08	0.5264
X_5	-1.2854e-06	0.2597
X_6	1.6622e-07	0.0494*

*significant at $\alpha = 5\%$ level

is regencies/cities in East Kalimantan Province; $\mu_i \sim IIN(0, \sigma_\mu^2)$ and v_{it} denotes the remainder disturbances; $v_{it} \sim IIN(0, \sigma_v^2)$. Let $\beta_{0i} = \alpha + \mu_i$, then model (14) can also be written

$$\begin{aligned} Y_{it} =& \beta_{0i} + \beta_1 X_{1_{it}} + \beta_2 X_{2_{it}} + \beta_3 X_{3_{it}} + \beta_4 X_{4_{it}} + \\ & \beta_5 X_{5_{it}} + \beta_6 X_{6_{it}} + v_{it}; \qquad (15) \\ i =& 1, 2, , \cdots, 10; t = 1, 2, \cdots, T_i \end{aligned}$$

Thus, the estimation model of OUR data by regencies/cities in East Kalimantan Province is

$$\begin{aligned} \hat{Y}_{it} =& \hat{\beta}_{0i} + \hat{\beta}_1 X_{1_{it}} + \hat{\beta}_2 X_{2_{it}} + \hat{\beta}_3 X_{3_{it}} + \hat{\beta}_4 X_{4_{it}} + \\ & \hat{\beta}_5 X_{5_{it}} + \hat{\beta}_6 X_{6_{it}}; \qquad (16) \\ \hat{\beta}_{0i} =& \hat{\alpha} + \hat{\mu}_i; \\ i =& 1, 2, , \cdots, 10; \quad t = 1, 2, \cdots, T_i \end{aligned}$$

The following analysis determines parameter estimation using within transformation with mean centering based on (8)-(9) and determines within transformation using median centering.

C. Within-Group Estimators of OUR East Kalimantan Model using Mean Centering

The first within-group estimator we define is within-group estimators of model (16), that is $\hat{\beta}_{0i}, \hat{\beta}_1, \ldots, \hat{\beta}_6$ based on the (8)-(9). Table V reports the parameter estimation of the initial model. Based on Table V, only one HDI variable affects the OUR variable. For this reason, the following analysis is to obtain the best model by eliminating the least influential independent variables (the largest p-value) until we find the best model, the model containing the variables that affect the OUR variable.

The first elimination is the RMW (X_2) variable because it has the largest p-value. Then we determine parameter estimation of the unbalanced panel data model using one-way fixed-effects model for the HDI (X_1), DR (X_3), DIPR (X_4), FIR (X_5), and GRDP (X_6) variables to the OUR variable. The results are in Table VI.

TABLE VII
ELIMINATION STAGE OF VARIABLE INDEPENDENT

Model	Independent Variable	Eliminated Variable
Initial model	HDI (X_1), RMW (X_2), DR (X_3), DIPR (X_4), FIR (X_5), GRDP (X_6)	RMW (X_2)
Model 1	HDI (X_1), DR (X_3), DIPR (X_4), FIR (X_5), GRDP (X_6)	DIPR (X_4)
Model 2	HDI (X_1), DR (X_3), FIR (X_5), GRDP (X_6)	DR (X_3)
Model 3	HDI (X_1), FIR (X_5), GRDP (X_6)	FIR (X_5)
Model 4	HDI (X_1), GRDP (X_6)	-

TABLE VIII
VALUE OF $\hat{\beta}_{0i}$ FOR EACH REGENCY/CITY IN EAST KALIMANTAN PROVINCE USING MEAN CENTERING

i-index	Regency/City	$\hat{\beta}_{0i}$
1	Paser	69.6
2	West Kutai (Kubar)	72.2
3	Kutai Kartanegara (Kukar)	53.8
4	East Kutai (Kutim)	58.8
5	Berau	74.0
6	North Paser Penajam (PPU)	74.4
7	Mahakam Ulu (Mahulu)	70.5
8	Balikpapan	71.0
9	Samarinda	77.0
10	Bontang	80.8

Based on Table VI, it is necessary to eliminate the DIPR (X_4) variable because it has the most significant p-value. The elimination of the least influential variables is carried out one by one until the best model is obtained. The complete elimination stages can be seen in Table VII.

Table VII shows that the best model is Model 4 with the HDI (X_1) and GRDP (X_6) variables affecting the OUR variable. The estimation of Model 4 with the one-way unbalanced panel data regression method using a fixed-effects model approach is

$$\hat{Y}_{it} = \hat{\beta}_{0i} - 0.9918X_{1_{it}} + 2.0483 \times 10^{-7}X_{6_{it}}; \quad (17)$$
$$i = 1, 2, , \cdots, 10; \quad t = 1, 2, \cdots, T_i$$

Table VIII shows $\hat{\beta}_{0i}$ value for each regency/city in East Kalimantan Province. The MSE value of model (17) is 2.0122. We can interpret model (17) as that every increase of one unit of HDI will decrease OUR by 0.9918%, and every increase of one million rupiahs of GRDP will increase OUR by 0.2048%, provided that other variables are constant.

D. Within-Group Estimators of OUR East Kalimantan Model using Median Centering

This section determines within-group estimators of OUR East Kalimantan model using median centering. The first step is to transform data using median centering based on (11)-(12). Then we regress $\tilde{Y}_{it}$ to $\tilde{X}_{k_{it}}$ using the OLS method. Table IX shows the within-group estimators with median centering of the initial model.

Table IX shows that the variable that affects OUR variable is GRDP. In the same way, we eliminate the least influential variables one by one to obtain the best model that contains

TABLE IX
PARAMETER ESTIMATION OF INITIAL MODEL USING MEDIAN CENTERING

Variable	Parameter Estimation	p-value
X_1	-7.8971e-01	0.0718
X_2	-1.1621e-06	0.2465
X_3	-2.5070e-01	0.2042
X_4	-7.2332e-08	0.2759
X_5	-1.2527e-06	0.2295
X_6	1.7983e-07	0.0244*

*significant at $\alpha = 5\%$ level

TABLE X
INDEPENDENT VARIABLE ELIMINATION STAGE FOR PARAMETER ESTIMATION USING MEDIAN CENTERING

Model	Independent Variable	Eliminated Variable
Initial model	HDI (X_1), RMW (X_2), DR (X_3), DIPR (X_4), FIR (X_5), GRDP (X_6)	DIPR (X_4)
Model 1	HDI (X_1), RMW (X_2), DR (X_3), FIR (X_5), GRDP (X_6)	DR (X_3)
Model 2	HDI (X_1), RMW (X_2), FIR (X_5), GRDP (X_6)	RMW (X_2)
Model 3	HDI (X_1), FIR (X_5), GRDP (X_6)	FIR (X_5)
Model 4	HDI (X_1), GRDP (X_6)	-

TABLE XI
VALUE OF $\hat{\beta}_{0i}$ FOR EACH REGENCY/CITY IN EAST KALIMANTAN PROVINCE USING MEDIAN CENTERING

i-index	Regency/City	$\hat{\beta}_{0i}$
1	Paser	64.9
2	West Kutai (Kubar)	67.5
3	Kutai Kartanegara (Kukar)	49.6
4	East Kutai (Kutim)	54.7
5	Berau	69.3
6	North Paser Penajam (PPU)	70.3
7	Mahakam Ulu (Mahulu)	65.9
8	Balikpapan	66.8
9	Samarinda	72.0
10	Bontang	75.9

only the variables that affect OUR variable. Table X shows the complete variable elimination stages for parameter estimation using median centering.

Based on Table X, we get that the best model is Model 4 with HDI (X_1) and GRDP (X_6) variables affecting OUR variable is

$$\hat{Y}_{it} = \hat{\beta}_{0i} - 0.9336X_{1_{it}} + 2.0176 \times 10^{-7}X_{6_{it}}; \quad (18)$$
$$i = 1, 2, , \cdots, 10; \quad t = 1, 2, \cdots, T_i$$

Table XI shows the value of $\hat{\beta}_{0i}$ for each regency/city in East Kalimantan Province based on (13). The model (18) has an MSE of 1.8226, meaning that for every increase in one unit of HDI there is a decrease in OUR by 0.9336%. For every increase in GRDP of one million rupiahs, OUR will increase by 0.2018%, provided that other variables are constant.

V. CONCLUSION

The one-way regression model of unbalanced panel data with a fixed-effects approach to the OUR data of East

Kalimantan Province in 2014-2021 show that the factors affecting the OUR are HDI and GRDP. The best model is parameter estimation using within transformation with median centering with lower MSE results, and not with mean centering.

These within-group estimators with median centering provide a smaller MSE value, with an insignificant difference compared to the MSE value with mean centering. A similar conclusion may be drawn from the results of the parameter estimation.

The result of this study is similar to [14], which concluded that the MSE value of the LS method for the two centers showed small difference. Centering data on the median is the first step in the robust method by [2]. After centering, the authors in [2] use the WGM and WMS methods in determining robust β for panel data regression models with a fixed-effects approach. For this reason, after centering, it is possible to determine the estimation of model parameters using a robust estimation method such as the method used by [2] for better robust parameter estimation. Several researchers have contributed to the development of research [2], including research [14] that proposed robust estimation based on two different data transformations and a study [16] that proposed MM-centering.

REFERENCES

[1] B. H. Baltagi, *Econometrics analysis of panel data*, 3rd ed. New York: John Wiley & Sons Ltd., 2005.

[2] M. C. Bramati and C. Croux, "Robust estimators for the fixed effects panel data model," *Econometrics Journal*, vol. 10, no.3, pp. 521-540, 2007.

[3] P. J. Rousseeuw and M. Hubert, "Anomaly detection by robust statistics," *WIREs Data Mining and Knowledge Discovery*, vol. 8, e1236, 2018.

[4] C. Hsiao and H. Chen, "On Classification from the View of Outliers," *IAENG International Journal of Computer Science*, vol. 37, no.4, pp341-349, 2010.

[5] D. O. Esan, P. A. Owolawi and C. Tu, "Anomalous Detection in Noisy Image Frames using Cooperative Median Filtering and KNN," *IAENG International Journal of Computer Science*, vol. 49, no.1, pp1-10, 2022.

[6] B. P. S. Kaltim, "Employment," 2022. [Online]. Available: https://kaltim.bps.go.id/subject/6/tenaga-kerja.html#subjekViewTab1.

[7] D. Yuniarti, D. Rosadi and Abdurakhman, "Inflation of Indonesia during the COVID-19 pandemic," *Journal of Physics: Conference Series*, 1821(1), 2021.

[8] M. F. R. Ridho, "Pengangguran dan pembangunan perkotaan (studi kasus : Kota Palembang)," *Jurnal Perencanaan Wilayah dan Kota*, vol. 21, no.1, pp. 55-68, 2010.

[9] I. N. Budiantara, R. Diana, Purhadi and S. Darmesto, "Relationship pattern of poverty and unemployement in Indonesia with bayesian spline approach," *International Journal of Basic & Applied Sciences*, vol. 11, no.06, pp. 119-127, 2011.

[10] M. Wardhany, "Unemployement and poverty determinants in Central Java," *Economics Development Analysis Journal*, vol. 6, no.3, pp. 352-365, 2017.

[11] D. P. Puspita, T. Militina and A. S. Effendi, "Employment opportunities and poverty levels in Provinsi Kalimantan Timur," *International Journal of Economics, Business and Accounting Research*, vol. 4, no.1, pp. 141-157, 2020.

[12] D. N. Gujarati and D. C. Porter, *Basic econometrics*, 5th ed., New York: McGraw-Hill Companies, Inc., 2008.

[13] T. Amemiya, "The estimation of the variances in a variance-components Model," *International Economic Review*, vol. 12, no.1, pp. 1-13, 1971.

[14] M. Aquaro and P. Cizek, "One-step robust estimation of fixed-effects panel data models," *Computational Statistics and Data Analysis*, vol. 57, pp. 536-548, 2013.

[15] S. Sukirno, *Makroekonomi teori pengantar*, 3rd ed., Depok: PT. Rajagrafindo Persada, 2011.

[16] Y. Croissant and G. Millo, *Panel data econometrics with R*, New York: John Wiley & Sons Ltd., 2019.

[17] D. Rosadi, *Analisis Ekonometrika & Runtun Waktu Terapan dengan R*, Yogyakarta: ANDI, 2020.

[18] N. M. A. Bakar and H. Midi, "Robust centering in the fixed effects panel data model," *Pakistan Journal of Statistics*, vol. 31, no.1, pp. 33-48, 2015