Vol. 3 No. (1) (June 2025) 11-18

Journal of Data Insights

http://journalnew.unimus.ac.id/index.php/jodi

Panel Data Regression Approach to Identify Factors Affecting Unemployment in East Java Province

Rizka Amalia Putri [1], **Alwan Fadlurohman**[2*], **Mardiyah Mughni**[3]
[1]Universitas Riau, Indonesia
[2]Universitas Muhammadiyah Semarang, Indonesia
[3]PT Riau Petroleum Mahato

DOI: https://doi.org/10.26714/jodi.v3i1.722

Article Information

Article History:
Submitted: 29th May 2024
Revised: 23th June 2025
Accepted: 27th June 2025

Keywords:
East Jawa; Fixed Effect; Open Unemployment Rate; Panel; Poverty.

Abstract

The Open Unemployment Rate (OOP) in East Java Province is a multidimensional problem influenced by economic and social factors, with significant disparities between districts/cities. This study analyses the effect of Poverty Percentage, Labour Force Participation Rate (TPAK), and Economic Growth on the open unemployment rate using a panel data regression approach to accommodate spatial and temporal heterogeneity. Cross-section (38 districts/cities) and time series (2019-2021) data were analysed through three models: Common Effect Model (CEM), Fixed Effect Model (FEM), and Random Effect Model (REM). The results of statistical tests (Chow, Hausman, and Lagrange Multiplier) show the FEM as the best model with a coefficient of determination of 0.555, explaining 55.5% of the variation in the unemployment rate. The FEM estimation reveals that the Poverty Percentage has a significant positive effect on increasing the unemployment rate, while Economic Growth has a negative impact on reducing the unemployment rate. This finding confirms the need for policies focused on poverty alleviation and increasing economic growth based on regional leading sectors. This study enriches the methodological literature through the application of FEM that controls for region-specific heterogeneity, while providing practical recommendations for policy makers in designing precise unemployment reduction interventions, such as skills training based on industry needs and strengthening labour-intensive programmes.

✉ Corresponding Author:
E-mail: alwan@unimus.ac.id

e-ISSN: 2988 - 2109

INTRODUCTION

Unemployment is one of the main problems faced by many developing countries, including Indonesia. High unemployment rates not only have a negative impact on economic growth, but also trigger various social problems such as poverty, crime, and social inequality that can hinder sustainable development [1]. East Java, as one of the provinces with the largest population in Indonesia, has very complex labour dynamics. Based on data from the East Java Central Bureau of Statistics (BPS), the open unemployment rate (TPT) in this province during the period 2019 to 2021 experienced significant fluctuations. In February 2019, the East Java TPT was recorded at 3.83 per cent, then increased to 5.17 per cent in February 2021.

The difference in unemployment rates between districts/cities was also quite striking during the 2019-2021 period. Some urban areas experienced higher unemployment rates than rural areas, indicating disparities in employment opportunities and labour market access. For example, Malang City and Surabaya tend to have higher unemployment rates than more agrarian districts. The phenomenon of unemployment in East Java cannot be separated from various interrelated factors, both from the macroeconomic side such as economic growth, minimum wage, and investment, as well as social factors such as education and urbanisation.

Previous studies have identified that variables such as Poverty Percentage [2], [3], [4], Labour Force Participation Rate (TPAK) [5], [6] and Economic Growth [2], [7], [8] have a significant influence on the unemployment rate. However, most of these studies still use a cross-section or time series data approach separately, so they have not been able to comprehensively describe spatial and temporal dynamics.

In this context, the panel data regression approach becomes very relevant and widely used in economic analysis, especially to study the factors that influence socio-economic phenomena such as unemployment and poverty. Panel data regression combines cross-section data (e.g. between districts/cities) with time series data (a certain period of time), thus allowing researchers to observe the dynamics of changes in the same units simultaneously over several periods [9], [10].

The panel data regression method provides the advantage of capturing inter-regional variations as well as temporal changes, which cannot be obtained if only using cross-section or time series data separately. In addition, panel data regression is able to control for unobserved heterogeneity across observation units and time, such as district-specific characteristics that are fixed or external factors that change over time. This allows for more accurate and bias-free estimates as unmeasured but influential variables can be controlled through fixed effect or random effect models [11], [12].

Therefore, based on the above framework, the purpose of this research is to model the open unemployment rate in East Java Province using a Panel Data regression approach. This approach is expected to provide a deeper understanding of the pattern of unemployment in East Java so that it can assist the government in formulating more targeted and effective policies. This paper consists of, background explained in section 1. Section 2 describes the dataset and methodology details. We apply panel data regression method to model the Open Unemployment Rate (OOP) in East Java, and discuss the results in Section 3. Section 4 concludes this paper with some final notes.

METHODS

In this section we divide it into three parts, first we explain the scope of the research, second we explain the methods used, and third the research procedure.

Research Scope

In this case we collected data on the Open Unemployment Rate (TPT) from the website https://opendata.jatimprov.go.id/. The purpose of the analysis in this study is to analyse the influence between the TPT variable and several independent variables that are thought to influence it. The

independent variables that will be used include Poverty Percentage (PK), Labour Force Participation Rate (TPAK), and Economic Growth (PE) obtained from the BPS website https://jatim.bps.go.id/.

Panel Data Regression

The panel data regression model is expressed in the form of an equation (1) [5].

$$Y_{it} = \boldsymbol{X}_{it}^{'}\boldsymbol{\beta} + \boldsymbol{Z}_{i}^{'}\boldsymbol{\alpha} + \varepsilon_{it}, i = 1,2,\dots,K; t = 1,2,\dots,T \quad (1)$$

where i denotes K cross-section units, while t denotes T time series. Thera are p independent variables in X_{it}, excluding the constant. Individual specific effects are where $\boldsymbol{Z}_i'\boldsymbol{\alpha}$ consists of constant and individual specific effects, both observable and unobservable. $\boldsymbol{\beta}$ is a slope matrix of size $px1$.

General Structure of the Model

In estimating panel regression models, the method to be used is highly dependent on the assumptions made about intercepts, slope coefficients and errors [13]. In terms of various assumptions and formation factors, the model structure is divided into 3, namely Pooled Regression, Fixed Effect and Random Effect [14].

Pooled Regression

This model assumes that $\boldsymbol{Z}_i$ consists of a constant only or it can be interpreted that there are no specific individual effects. This model structure is often also called the Common Effect Model (CEM) and is expressed in the form of equation (2) [15].

$$Y_{it} = \alpha + \boldsymbol{X}_{it}^{'}\boldsymbol{\beta} + \varepsilon_{it}, i = 1,2,\dots,K; t = 1,2,\dots,T \quad (2)$$

where is the intercept coefficient (constant) which is a scalar number, $\boldsymbol{\beta}$ is a slope matrix of size $px1$ and X_{it} is the $i-th$ observations and $t-th$ time on the p explanatory variables [15].

Fixed Effect

The fixed effect model structure is a model that takes into account the diversity of independent variables by individual. The Fixed Effect model is expressed in the form of equation (3) [15],

$$Y_{it} = \alpha_i + \boldsymbol{X}_{it}^{'}\boldsymbol{\beta} + \varepsilon_{it}, i = 1,2,\dots,K; t = 1,2,\dots,T \quad (3)$$

where $\alpha_i = \boldsymbol{Z}_i^{'}\boldsymbol{\alpha}$, represents all observed effects and specifies the conditional mean that can be estimate. α_i is an unknown fixed parameter that will be estimated. $\boldsymbol{Z}_i$ is assumed to be unobserved and correlated with the independent variables. ε_{it} is a stochastic and independently and identically distributed error with mean 0 and variance σ_ε^2. The independent variables X_{it} is assumed to be independent of the error ε_{it} for all i and t [15].

Random Effect

If the individual effect $\boldsymbol{Z}_i$ has no correlation with the independent variables, then this model structure is known as the Random Effect Model whose model is expressed in the form of equation (4) [15].

$$Y_{it} = \alpha + \boldsymbol{X}_{it}^{'}\boldsymbol{\beta} + \eta_{it}, i = 1,2,\dots,K; t = 1,2,\dots,T \quad (4)$$

where

$$\eta_{it} = \varepsilon_{it} + u_i$$
$$\alpha = E[\boldsymbol{Z}_i'\boldsymbol{\alpha}]$$
$$u_i = \left\{\boldsymbol{Z}_i^{'}\boldsymbol{\alpha} - \left[\boldsymbol{Z}_i^{'}\boldsymbol{\alpha}\right]\right\}$$

There are p independent variables including a constant. α is the average of the unobserved individual effects. u_i is a random effect specific to the i-th observation. In this model u_i is assumed to be independent with ε_{it}, in addition it is also assumed that the independent variable X_{it} with u_i and ε_{it} is independent [15].

Research Steps

The analytical procedures conducted in modelling the Open Unemployment Rate in East Java Province using the panel data regression approach are as followed:

1. Collecting data on factors suspected of influencing TPT in East Java province.

2. Collect data on factors that are thought to affect the open unemployment rate in East Java province;
3. Perform descriptive analysis;
4. Detect the presence or absence of multicollinearity by looking at the Variance Infllation Factor (VIF) value;
5. Modelling TPT in East Java Province with Panel data regression using the Common Effect approach;
6. Modelling TPT in East Java Province with Panel data regression using the Fixed Effect approach;
7. Modelling the TPT in East Java Province with Panel data regression using the Random Effect approach;
8. Selecting panel data regression models with CEM, FEM and REM approaches with chow test, hausman test, and lagrange multiplier test.;
9. Selection of the Best Model with Coefficient of Determination;
10. Testing parameter estimates of the best model using the T test and F test.
11. Conclusions based on the analysis steps.

RESULTS AND DISCUSSION

In this section, we present the results of modelling the TPT in East Java Province using a panel data regression approach.

1. Descriptive Analysis

The percentage of TPT is presented in Figure 1. The district/city with the highest TPT value in 2019 was Malang City at 5,880% while the lowest TPT value in 2019 was Pacitan Regency at 0,910%. In 2020, the district/city with the highest TPT value was Sidoarjo Regency at 10,970% while the lowest TPT was Pacitan Regency at 2,280%. In 2021, the district/city with the highest TPT value was Sidoarjo Regency at 10,870% while the lowest TPT was Pacitan Regency at 2,040%.

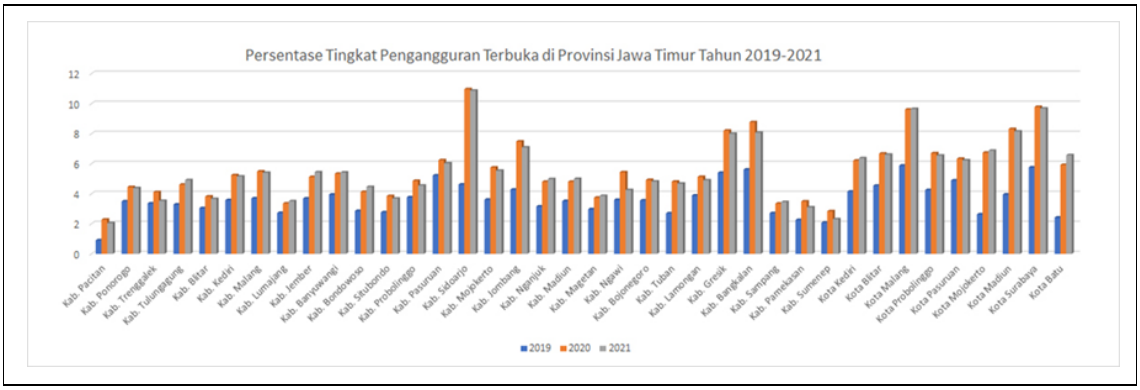

Figure 1. Percentage of Open Unemployment Rate in East Java Province in 2019-2021

Table 1. Descriptive Statistics of Predictor Variables

Variable	Mean	Variance	Max	Min	Range
PK	10.88	20.19	23.76	3.81	19.95
TPAK	70.22	10.58	80.57	63.44	17.13
PE	1.79	13.52	6.51	-6.46	12.97

Table 1 showed that the PK variable has an average value of 10.88, which is smaller than the variance value of 20.19, meaning that there is a large distribution of data between variables. The

TPAK variable has an average value of 70.22 greater than the variance value of 10.58, which means that the variables have a small data distribution. The PE variable has an average value of 1.79 smaller than the variance value of 13.52, which means that the variables have a large data distribution.

2. Multicollinearity Test

Before modelling panel data regression, the assumption that must be seen is multicollinearity. The multicollinearity test can be seen from the VIF (Variance Inflation Factor) value. If the VIF value is more than 10 then there is multicollinearity in the independent variables.

Table 2. Multicollinearity Test

Variable	VIF
PK	1.118532
TPAK	1.095152
PE	1.027810

Table 2 concludes that the VIF value of all independent variables shows a value of less than 10, meaning that there is no multicollinearity in the independent variables.

3. Panel Data Regression

There are several steps to perform modelling, namely determining the three panel data regression model estimates, selecting the best estimated model, testing the significance of the parameters, and testing the residual assumptions of the selected model.

3.1 Panel Data Regression Estimation

There are three approaches to estimating panel data regression parameters, namely the Common Effect Model (CEM), Fixed Effect Model (FEM), and Random Effect Model (REM).

A. CEM

The CEM model is the simplest panel estimation approach, where all data are combined without regard to individuals and time

$$\hat{Y}_{it} = 21{,}493 - 0{,}149X_{1it} - 0{,}207X_{2it} - 0{,}218X_{3it}$$

The CEM model has a coefficient of determination of 0.408, where the model is able to explain 40.8% of the variation in TPT.

B. FEM

The FEM model is a panel data regression estimation method with the assumption that the intercept value of the cross-section or time series unit is different, but the slope coefficient is fixed.

Variation Between Individuals

In the FEM estimation model between individuals, variations are located in individuals, namely 38 districts / cities in East Java Province, where the time factor is ignored and the estimated intercept value for each district / city. The equation of the FEM model with variation between individuals is as follows:

$$\hat{Y}_{it} = \hat{\alpha}_i + 1{,}024X_{1it} + 0{,}048X_{2it} - 0{,}127X_{3it}$$

where $\hat{\alpha}_i$ is a combination of slope and intercept for each district/city in East Java Province. The FEM model of variation between individuals has a coefficient of determination of 0.555, where the model is able to explain the variation in TPT by 55.5%.

Variation Between Times

In the FEM estimation model between time, the variation lies in time, where individual factors are ignored and the intercept value is estimated for each time. The equation of the FEM model with intertemporal variation is as follows:

$$\hat{Y}_{it} = \hat{\alpha}_t - 0{,}155X_{1it} - 0{,}223X_{2it} - 0{,}166X_{3it}$$

where $\hat{\alpha}_t$ is a combination of slope and intercept for each times. The FEM model of in tertemporal variation has a coefficient of determination of 0.423, where the model is able to explain the variation in TPT by 42.3%.

C. REM

The REM model is a model where the intercept on the observation (α_i) is assumed to be a random variable. The REM model is obtained as follows:

$$\hat{Y}_{it} = \hat{\eta}_{it} + 13{,}024 - 0{,}122X_{1it} - 0{,}091X_{2it} - 0{,}199X_{3it}$$

where $\hat{\eta}_{it}$ has different values for each district/city in East Java Province. The REM model has a coefficient of determination of 0.311, where the model is able to explain the variation in TPT by 31.1%.

Best Model Selection

Furthermore, after estimating the CEM, FEM, and REM models, determining the best model of the three models. The first step is to find the best model between CEM and FEM, using the Chow test with the following hypothesis

HHypothesis:

H_0 : The appropriate model is CEM

H_1 : The appropriate model is FEM

The result obtained is the p-value (0.000) < α (0.05) which has a decision to reject H_0, so the FEM estimation method is more suitable than the CEM estimation method. The next step is to conduct the Hausman test, to find out which is better the FEM estimation model or the REM estimation model.

Hypothesis:

H_0 : The appropriate model is REM

H_1 : The appropriate model is FEM

The result obtained is the p-value (0.000) < α (0.05) which has a decision to reject H_0, so the FEM estimation method is more suitable than the REM estimation method. Therefore, it can be concluded that modelling TPT in East Java Province is more appropriate using the Fixed Effect model.

3.2. Significance Test of Best Model Parameters

The panel data regression equation consists of simultaneous test (F test) and partial test (t test) after selecting the FEM estimation with individual effects.

A. F-test

The simultaneous test (F-test) aims to determine the influence of all dependent variables on independent variables. The hypothesis in the F test is as follows:

H_0 : $\beta_1 = \beta_2 = \cdots = \beta_k = 0$

H_1 : at least one $\beta_k \neq 0$, for k = 1,2, and 3

Table 3. Uji F

Model	F-Statistics	p-value
FEM Individu	30,294	0,000
FEM Waktu	26,378	0,000

Table 3 presents the results of the F test of the individual and time FEM panel data regression model showing a p-value of 0.000 < α (0.05), so it can be concluded that rejecting H0 means that the variables of Poverty Percentage, TPAK, and Economic Growth have a significant effect on the Open Unemployment Rate variable in East Java.

B. T-test

The partial test (t-test) aims to determine the significance of individual dependent variables to independent variables. The hypotheses in the t-test are as follows:

H_0 : $\beta_k = 0$

H_1 : $\beta_k \neq 0$, for k = 1,2, and 3

Table 4. Uji T

Model	T-Statistics	p-value
FEM Individu		
$\mathbf{X_1}$	5,228	0,000***
$\mathbf{X_2}$	0,606	0,546
$\mathbf{X_3}$	-4,829	0,000***
FEM Waktu		
$\mathbf{X_1}$	-4,971	0,000***
$\mathbf{X_2}$	-5,384	0,000***
$\mathbf{X_3}$	-1,928	0,056

Table 4 shows that for the individual and time FEM models, there are 2 significant variables, namely the Poverty Percentage and Economic Growth variables that have a significant effect on the Open Unemployment Rate variable in East Java for the Individual FEM model. Meanwhile, for the time FEM, it is the Poverty Percentage and TPAK that have a significant effect on the Open Unemployment Rate variable in East Java.

C. Coefficient of Determination

Based on the results of the determination coefficient, the value as shown in table 5 is obtained:

Table 5. Koefisien Determinasi

Model	Coefficient of Determination
FEM Individual	55,5%
FEM Time	42,3%

The results of the determination coefficient showed that the individual FEM model was better than the time FEM model because it had a higher determination coefficient value, which was 55.5%. Thus, the final model formed is:

$$\hat{Y}_{it} = \hat{\alpha}_i + 1{,}024X_{1it} - 0{,}127X_{3it}$$

Based on the model above, it can be seen that every increase in the percentage of poverty by 1 percent, the TPT increases by 1.024 percent plus $\hat{\alpha}_i$ for each district/city assuming the other variables are constant. For every increase in Economic Growth by one percent, the TPT rate decreases by 0.127 percent plus for $\hat{\alpha}_i$ each district/city assuming other variables are constant.

CONCLUSION

In modeling using panel data regression, there are three model estimation methods that can be used, namely the Common Effect Model, the Fixed Effect Model and the Random Effect Model. These three models have their own assumptions that must be met in order to obtain an accurate model estimate. In determining the most appropriate regression estimation model of the three available models, a series of tests were carried out, namely the Chow Test which was used to choose between the CEM and FEM models; Langrange Multiplier (LM) test used to choose between CEM and REM models; and the Hausman test used to choose between FEM and REM models.

The best panel data regression model for the Open Unemployment Rate in East Java as a dependent variable is the Fixed Effect model with the following model estimation results:

$$\hat{Y}_{it} = \hat{\alpha}_i + 1{,}024X_{1it} - 0{,}127X_{3it}$$

where is the amount of intercept value $\hat{\alpha}_i$ is different for each district/city in East Java Province. This model has a determination coefficient value of 0.555, which means that 55.5% of the variation of the TPT variable can be explained by the Poverty Percentage and Economic Growth variables.

REFERENCES

[1] S. Smith and M. P. Todaro, *Economic Development, 12th Edition (The Pearson Series In Economics)*. 2003.

[2] E. F. Putri, L. O. Arafat, A. Utami, and F. Nisa, "Pengaruh Tingkat Kemiskinan dan Tingkat Pengangguran Terbuka Terhadap Pertumbuhan Ekonomi di Provinsi Jawa Timur," *Jurnal Ilmiah Wahana Pendidikan*, vol. 10, no. 20, pp. 292–304, Oct. 2024, doi: 10.5281/zenodo.14286304.

[3] Amanda Septa, Arista Fairuzani, Camila Yasmin, Giga Hutagaol, and Vanny Hutabarat, "Analisis Variabel Kemiskinan di Indonesia dengan Model Linear Regresi dan Algoritma K-Means pada Tahun 2020-2023," *Profit: Jurnal Manajemen, Bisnis dan Akuntansi*, vol. 4, no. 2, pp. 28–52, Feb. 2025, doi: 10.58192/profit.v4i2.3077.

[4] D. Indriyani, W. Nugroho, and A. E. Sudjianto, "PENGARUH HUMAN DEVELOPMENT INDEX (HDI) DAN TINGKAT PENGANGGURAN TERBUKA (TPT) TERHADAP POVERTY GAP INDEX-P1," *Journal of Management : Small and Medium Enterprises (SMEs)*, vol. 17, no. 2, pp. 311–321, Jul. 2024, doi: 10.35508/jom.v17i2.14425.

[5] S. I. Salsabila, S. Nabila, R. A. Rianti, D. Priyoga, and A. E. Sujianto, "Pengaruh Upah Minimum Dan Tingkat Partisipasi Angkatan Kerja Terhadap Tingkat Pengangguran Terbuka Di Jawa Timur Periode 2018-2022," *ournal of Administrative and Sosial Science (JASS)*, vol. 5, no. 1, pp. 92–107, 2024.

[6] Y. Rahmawati, R. N. Izzati, and F. N. Luthfiyyah, "Pengaruh Tingkat Partisipasi Angkatan Kerja, Indeks Pembangunan Manusia, dan Upah Minimum Terhadap Pengangguran di Banten ," *Independent: Journal of Economics*, vol. 4, no. 3, pp. 49–59, 2024.

[7] I. Hasan, Z. Zunawanis, and R. K. Zahra, "ANALISIS PENGARUH PERTUMBUHAN EKONOMI TERHADAP TINGKAT PENGANGGURAN TERBUKA DI KABUPATEN NAGAN RAYA," *EKOMBIS: JURNAL FAKULTAS EKONOMI*, vol. 4, no. 2, p. 133, Nov. 2019, doi: 10.35308/ekombis.v4i2.5243.

[8] R. F. Yoertiara and N. Feriyanto, "Pengaruh pertumbuhan ekonomi, IPM, dan tingkat pengangguran terbuka terhadap ketimpangan pendapatan provinsi-provinsi di pulau Jawa," *Jurnal Kebijakan Ekonomi dan Keuangan*, pp. 92–100, Aug. 2022, doi: 10.20885/JKEK.vol1.iss1.art9.

[9] M. R. N. Risalah and A. Hapsery, "REGRESI DATA PANEL UNTUK MEMODELKAN FAKTOR-FAKTOR YANG MEMPENGARUHI KEMISKINAN DI JAWA TIMUR TAHUN 2019-2021 ," in *SeminarNasionalHasilRisetdanPengabdian(SNHRP)*, Surabayat: Universitas PGRI Adi Buana Surabaya, 2023, pp. 410–425.

[10] S. Fatonah, F. Fitriani, and A. Indrasetianingsih, "Penerapan Regresi Data Panel pada Permasalahan Tingkat Pengangguran Terbuka Di Jawa Barat," *Unisda Journal of Mathematics and Computer Science (UJMC)*, vol. 9, no. 1, pp. 21–30, Jun. 2023, doi: 10.52166/ujmc.v9i1.4232.

[11] J. Wooldridge, "Introductory Econometrics: A Modern Approach," vol. 20, May 2012.

[12] B. H. Baltagi, *Econometric Analysis of Panel Data*. Cham: Springer International Publishing, 2021. doi: 10.1007/978-3-030-53953-5.

[13] C. Hsiao, *Analysis of Panel Data*. Cambridge University Press, 2014. doi: 10.1017/CBO9781139839327.

[14] W. Greene, "Econometric Analysis Fifth Edition," May 2003.

[15] M. Srihardianti, M. Mustafid, and A. Prahutama, "METODE REGRESI DATA PANEL UNTUK PERAMALAN KONSUMSI ENERGI DI INDONESIA," *Jurnal Gaussian*, vol. 5, no. 3, pp. 475–485, 2016, doi: 10.14710/J.GAUSS.5.3.475-485