BAREKENG: Journal of Mathematics and Its Applications
March 2024 Volume 18 Issue 1 Page 0615–0626
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol18iss1pp0615-0626

MODELING FACTORS AFFECTING EDUCATED UNEMPLOYMENT ON JAVA ISLAND USING GEOGRAPHICALLY WEIGHTED POISSON REGRESSION MODEL

Ditto Satrio Wicaksono[1], Sinta Nuriyah[2], Rahajeng Fajritia[3], Ni Putu Nita Yuniarti[4], Priatmadani[5], Laeli Amelia[6], Sarni Maniar Berliana[7*]

[1,2,3,4,5,6] Diploma 4 Statistics Study Program, Politeknik Statistika STIS
[7]Diploma 3 Statistics Study Program, Politeknik Statistika STIS
[7]Sustainable Development Goals (SDGs) Research Unit, Politeknik Statistika STIS

Otista Street No.64C, Bidara Cina, Jatinegara, Jakarta, 13330, Indonesia

[*]Corresponding author's e-mail: ***sarni@stis.ac.id***

Article History:

Received: 18th September 2023
Revised: 25th January 2024
Accepted: 20th February 2024

Keywords:

Educated Unemployment; Exposure; Geographically Weighted Regression; Poisson Regression; SDG Goal 8; Unemployment.

ABSTRACT

The eighth goal of the SDGs, which aims to encourage sustained, inclusive and sustainable economic growth, full and productive employment, and decent work for all, addresses the problem of unemployment. Indonesia, the fourth-largest country in the world, keeps on dealing with unemployment and its negative consequences. Three provinces on the Java island have higher unemployment rates for educated people than any other provinces. The purpose of this study is to examine the variables affecting educated unemployment in Java. This study uses cross-sectional data published from BPS-Statistics Indonesia website and the Indonesia Investment Coordinating Board (BKPM) for 119 regencies/cities across six provinces on Java Island in 2021. The predictor variables are Gross Regional Domestic Product (GRDP), investment, labor force participation rate, mean years of schooling, regency/city minimum wage, and inflation. The number of working-age population is used as an exposure measure. Four models were used to analyze the factors affecting educated unemployment on Java Island: Poisson regression model and Geographically Weighted Poisson Regression (GWPR) model, both with and without an exposure. Based on the smallest AIC/AICc, the best model is the GWPR model with an exposure. This model creates 11 groups of locations based on the same predictor variables that significantly affect educated unemployment

How to cite this article:

D. S. Wicaksono, S. Nuriyah, R. Fajritia, N. P. N. Yuniarti, Priatmadani, L. Amelia and S. M. Berliana., "MODELING FACTORS AFFECTING EDUCATED UNEMPLOYMENT ON JAVA ISLAND USING GEOGRAPHICALLY WEIGHTED POISSON REGRESSION MODEL," *BAREKENG: J. Math. & App.*, vol. 18, iss. 1, pp. 0615-0626, March, 2024.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Unemployed individuals are those who are willing and capable of working but unable to find employment. The impacts of unemployment on people and the economy have been the subject of several research projects. Unemployment negatively impacts a person's physical and mental health, increasing the risk of chronic illnesses, depression, anxiety, suicide, and substance misuse [1] [2] [3]. Life satisfaction, which is the overall assessment of one's quality of life, can be impacted by unemployment. A longitudinal study conducted in East Germany[3] found that unemployment had a lasting negative impact on life satisfaction, with those who experienced unemployment reporting poorer life satisfaction compared to those who did not for more than 20 years. In addition to lowering tax collections and government spending, unemployment may also raise levels of poverty and inequality in a society [4].

The issue of unemployment and the negative consequences that it can have remain an issue even in Indonesia, the fourth-largest nation in the world. According to the Badan Pusat Statistik (BPS-Statistics), Indonesia's unemployment rate in February 2023 was 5.45%, which is higher than countries with a larger population, such as China (5.2%) and the United States (3.7%). The Sustainable Development Goals (SDGs) related to the unemployment issue is stated in the eight SDGs that aim to promote sustained, inclusive and sustainable economic growth, full and productive employment, and decent work for all. The specific target for unemployment is to substantially reduce the proportion of youth not in employment, education or training. Furthermore, this goal also implicitly relates to the proportion of educated individuals who are unemployed.

Based on data on the number of unemployed for August 2021 published by BPS-Statistics, we calculate the educated unemployment rate in Indonesia at 8.82%. In fact, 62.06% of the total unemployed in that year were educated unemployed, namely those with a high school education or above. DI Yogyakarta Province, as one of the provinces on the island of Java, has the lowest level of educated unemployment (4.22%) followed by two provinces outside Java, namely West Nusa Tenggara Province (4.79%) and Central Sulawesi Province (4.84%). Meanwhile, the highest level of educated unemployment is occupied by two provinces on Java Island, namely West Java Province (12.99%) and Banten Province (11.15%) followed by Riau Islands Province (10.16%) outside Java Island. The island of Java, which is home to more than half of Indonesia's population, makes it attractive as a research area, including research on educated unemployment. The level of educated unemployment in the other three provinces on the island of Java is DKI Jakarta Province (9.6%), East Java Province (8.09%), and Central Java Province (7.96%). This shows that the level of educated unemployment varies between provinces in Java. This variation is interesting to investigate further regarding what variables influence the educated unemployment rate in each location.

Several studies have examined factors affecting unemployment among educated people. Putri [5] analyzed educated unemployment in Central Java using the panel data method by applying linear double log regression with Generalized Least Square (GLS) and found that inflation, economic growth, and wages together have a significant effect on educated unemployment in Central Java Province. Another study by Veronika & Mafruhat [6] showed that there is a significant effect of economic growth, investment, and inflation on educated unemployment in West Java Province. Their study used multiple linear regression with series data from 2006 to 2020. Adriani, Hamzah, & Zakaria [7] using panel data regression with regencies/municipalities as the unit of observations found that Gross Regional Domestic Product (GRDP), minimum wages, and education level have a significant effect on educated unemployment in Central Sulawesi Province. All studies mentioned earlier used location as the unit of analysis in their models.

Spatial effects must be modeled into the relationships between the explanatory variables and the response variables when locations are the units of analysis. When there are spatial components in the sample data, spatial dependency and spatial heterogeneity are two potential problems [8]. The more well-known and widely accepted of these two is spatial dependency or spatial autocorrelation. When an event's intensity at one area affects the intensity in nearby locations, this is known as spatial dependency. Meanwhile, the distribution of an event and the distribution of a relationship across a landscape are related to spatial heterogeneity. Geographically Weighted Regression (GWR) model is a statistical method frequently used to handle spatial effects [9]. The GWR model allows spatial variation in the relationships between the explanatory and response variables, which causes spatial variation in the regression coefficients. This approach estimates distinct regression models for each location unit.

To our best knowledge, there is no study on educated unemployment that took into account spatial effects in their models. The resulting regression model can produce misleading interpretations if the research analysis unit is location, but the modeling has not been adjusted to include potential spatial effects [10]. Simpson's paradox provides a simple illustration of this phenomenon, where the result is reversed when separately analyzed data groups are combined. In this study, we apply a model that takes into account spatial effects in modeling the relationship between the explanatory variables and response variable for count data, which is called Geographically Weighted Poisson Regression (GWPR). To control for different population sizes in each study area, we need to include an exposure measure in the regression model. An exposure is a variable that measures the amount of opportunity for an event to occur [11]. For example, demographers may model death rates in geographic areas as the count of deaths divided by person−years.

The following discussion in this paper is presented on three main topics. We discuss the Poisson regression model and the GWPR model, two regression models that are appropriate for count data, in Section 2. We run two different kinds of models for each one, which are models with and without an exposure. Furthermore, we provide. Additionally, several measurements for model evaluation. Using the Poisson regression model and the GWPR model, respectively, we evaluate factors that affect educated unemployment both globally and locally in Section 3. We provide conclusions in the last section.

2. RESEARCH METHODS

2.1. Poisson Regression Model

A discrete random variable Y is said to have a Poisson distribution with parameter λ if and only if the probability function is [12]

$$P(Y=y|\lambda)=\begin{cases}\dfrac{e^{-\lambda}\lambda^{y}}{y!}, & y=0,1,\ldots;\ \lambda\geq 0\\ 0, & \text{others.}\end{cases} \tag{1}$$

The probability function in Equation (1) can be written as a canonical function of the exponential family with natural parameters $\ln(\lambda)$ as follows

$$f(y)=\exp\left(y\ln(\lambda)-\lambda-\ln(y)\right). \tag{2}$$

Based on Equation (2), the Poisson distribution is an exponential family with the natural parameter being (λ) or with the link function being the natural logarithm then the Poisson regression model can be written as the natural logarithm of the dependent expectation value Y_i that depends on the independent variable x_i as follows

$$\begin{aligned}\ln\left(E\left(Y_i\right)\right)&=\mathbf{x}_i^T\boldsymbol{\beta},\\ E\left(Y_i\right)=\lambda_i&=\exp\left(\mathbf{x}_i^T\boldsymbol{\beta}\right),\end{aligned} \tag{3}$$

where

$$\mathbf{x}_i=\begin{bmatrix}1 & x_{1i} & x_{2i} & \cdots & x_{ki}\end{bmatrix}^T,$$
$$\boldsymbol{\beta}=\begin{bmatrix}\beta_0 & \beta_1 & \beta_2 & \cdots & \beta_k\end{bmatrix}^T.$$

The probability function of Poisson distribution with q as an exposure measurement corresponding to Equation (1) is

$$P(Y=y|\lambda(q))=\begin{cases}\dfrac{e^{-\lambda(q)}\lambda(q)^{y}}{y!}, & y=0,1,\ldots;\ \lambda(q)\geq 0\\ 0, & \text{others}\end{cases} \tag{4}$$

and the Poisson regression model that includes an exposure can be written as follows [13]

$$\lambda(q_i) = q_i \exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right). \tag{5}$$

2.2. Geographically Weighted Poisson Regression (GWPR)

Equation (3) yields a single regression model, which we refer to as the global model, that has the same regression coefficients for all observations (locations). Meanwhile, in the local regression model, each location has its own regression coefficients since the local regression model generates n regression models. The location-specific regression coefficients used in this study are based on vectors of two-dimensional coordinates that represents longitude and latitude for the location i, denoted by $\mathbf{u}_i = \left(u_{1i}, u_{2i}\right)$.

Let n random samples $\left(\mathbf{x}_i, y_i\right), i = 1, 2, \ldots, n,$ from the random variables $y_i \sim P\left(\lambda\left(\mathbf{u}_i\right)\right)$ be given, then the GWPR model based on Poisson regression in **Equation (3)** is written as

$$\lambda_i\left(\mathbf{u}_i\right) = \exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\left(\mathbf{u}_i\right)\right), \tag{6}$$

where

$$\boldsymbol{\beta}\left(\mathbf{u}_i\right) = \left[\beta_0\left(\mathbf{u}_i\right) \quad \beta_1\left(\mathbf{u}_i\right) \quad \beta_2\left(\mathbf{u}_i\right) \quad \cdots \quad \beta_k\left(\mathbf{u}_i\right)\right]^T.$$

When n random samples are taken from the random variables $y_i \sim P\left(\lambda\left(q_i, \mathbf{u}_i\right)\right)$ then the GWPR model based on **Equation (5)** can be written as follows

$$\lambda\left(q_i, \mathbf{u}_i\right) = q_i \exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\left(\mathbf{u}_i\right)\right), \tag{7}$$

where q_i is the exposure for location i.

Each location in the global regression model has equal weights when estimating the parameters. Depending on how close the data points are near to the regression point *i*, different weights are assigned to the data in the GWPR model. The GWPR model assigns greater weights to data points that are closer to the regression point *i*. In Fotheringham, many spatial weighting functions are discussed. Here, we present the adaptive bisquare weighting function that will be used in the application later.

Each location in the global regression model has equal weights when estimating the parameters. Depending on how close the data points are to the regression point *i*, different weights are assigned to the data in the GWPR model. Greater weights are given to data points that are closer to the regression point *i*. In **[9]**, many spatial weighting functions are discussed. Here, we present an adaptive bisquare weighting function in accordance with what will be used in the application later,

$$w_{ij} = \begin{cases} \left(1 - \left(\dfrac{d_{ij}}{b_i}\right)^2\right)^2 & ;\ \mathrm{d}_{ij} \le b \\ 0 & ;\ \mathrm{d}_{ij} > b_i; \quad i, j = 1, \ldots, n \end{cases} \tag{8}$$

where $d_{ij} = \left(\left(u_{1i} - u_{1j}\right)^2 + \left(u_{2i} - u_{2j}\right)^2\right)^{1/2}$, d_{ij} is the Euclidian distance between location i and location j, while b_i is a smoothing parameter or referred to the bandwidth of location i. The selection of optimum bandwidth can be achieved using Generalized Cross Validation (*GCV*) method (see **[9]** for details).

We do the Breusch-Pagan test **[14]** for heteroscdasticity to determine whether or not spatial heterogeneity is present in actual data before we apply the GWPR model. The null hypothesis of this test is the homoscedasticity is present and the alternative is the heteroscedasticity exists. We reject the null hypothesis when the p-value < 0.05, indicating that the data exhibit heteroscedasticity.

2.3. Model Evaluation

To evaluate the goodness of the models, we calculate Akaike's Information Criterion (AIC) and a second-order variant of (AIC_c) for each model as follows [15]:

$$AIC = -2\log\left(L\left(\hat{\boldsymbol{\theta}}\right)\right), \tag{9}$$

$$AIC_c = AIC + \frac{2K(K+1)}{n-K-1}, \tag{10}$$

where $\hat{\boldsymbol{\theta}}$ is a vector of estimated parameters, $L\left(\hat{\boldsymbol{\theta}}\right)$ is the likelihood function of the estimated model parameters, K is number of estimated parameters, and n is sample size. The smaller the AIC or AIC_c the better the model.

2.4. Operational Definition of Variables

The response variable (Y) in this study is the number of educated unemployed, defined as those who have a high school level of education or more and were unemployed at the time of enumeration [16]. The predictor variables consist of six variables, namely Gross Regional Domestic Product (X_i), investment (X_2), regency/city minimum wage (X_3), mean years of schooling (X_4), labor force participation rate (X_5), and inflation (X_6). The measurement of an exposure in this study is the number of working age population. We used the coordinates of district/city government offices as the regression points because the government centers are usually the center of public facilities and services that are relevant as a central point for measuring distances between locations.

Tabel 1. Variable Operational Definition

Variable	Unit	Definition
Number of educated unemployed (Y)	people	Number of people who are unemployed or job seekers with high school education or higher.
Gross Regional Domestic Product (X_1)	trillion rupiah	The total added value generated by all business units in an area at a constant price.
Investment (X_2)	trillion rupiah	Due to the availability data of investment, we only include domestic investment made by investors in Indonesia for business activities.
Regency/city minimum wage (X_3)	million rupiah	Regency/city minimum wage refers to the lowest wage (including regular allowances but excluding overtime pay) paid to employees (per kind of position/job), as determined by a particular regency/city.
Mean years of schooling (X_4)	years	Number of years of studied for people aged ≥ 15 years who have completed formal education (excluding repeat years).
Labor force participation rate (X_5)	percent	The percentage of the labor force in the population aged ten years and over.
Inflation (X_6)	percent	Inflation is obtained through the GRDP deflator or the results of a comparison between GRDP at current prices and GRDP at constant prices.
Number of working age population (q_i)	thousand people	Total population aged 15 years and over.

We used data that was published by BPS-Statistics and the Indonesia Investment Coordinating Board (BKPM) in 2021 for 119 regencies/cities in six provinces on Java Island. The GWR4 and RStudio software are used for data processing and modeling.

3. RESULTS AND DISCUSSION

3.1. Characteristics of Educated Unemployment on Java Island

The characteristics of the educated unemployed in Java in 2021 are presented in **Table 2**, which show the summary statistics for every variable included in this study. The lowest standard deviation is found in the variable of regency/city minimum wage followed by the mean years of schooling. This shows that data on minimum wages and educational attainment are relatively similar between locations with an average wage of 2.65 million rupiah per month and an average length of schooling of 8.44 years. Meanwhile, the variable with the largest standard deviation was the number of educated unemployment. This is probably due to the fact that the population of each location varies greatly. Therefore, including an exposure measure in the model is appropriate to control for variation in population size.

Table 2. Summary Statistics of Variables

Variable	Minimum	Maximum	Mean	Standard Deviation
Number of educated unemployment	8	1,907	295.60	335.03
Gross Regional Domestic Product (trillion rupiah)	3.37	460.02	55.91	85.49
Investment (trillion rupiah)	0.00	43.27	3.78	7.75
Regency/city minimum wage (million rupiah)	1.77	4.80	2.65	0.95
Mean years of schooling (years)	4.86	11.82	8.44	1.59
Labor force participation rate (%)	56.86	80.57	68.40	4.32
Inflation (%)	-4.67	29.09	2.16	3.67
Number of working age population	18	4,454	995.72	692.47

The distribution of educated unemployment across Java's regencies and cities is shown in **Figure 1**. It reveals that the western part of Java Island, particularly the DKI Jakarta Province, West Java Province, and Banten Province, has a relatively high number of educated unemployment, with Bogor Regency, West Java Province, accounting for the highest number of educated unemployed, for 190,715 people. The lowest number of educated unemployment is in Kepulauan Seribu (8) and the highest number is in Bogor Regency (1,907). Meanwhile, Surabaya City (1,103) and Sidoarjo Regency (1,007), both in East Java Province, have substantially higher number of educated unemployment in the eastern part of Java Island. Regencies/cities in Central Java, the southern part of East Java, as well as DI Yogyakarta Province, have a considerably lower number of unemployed educated people.

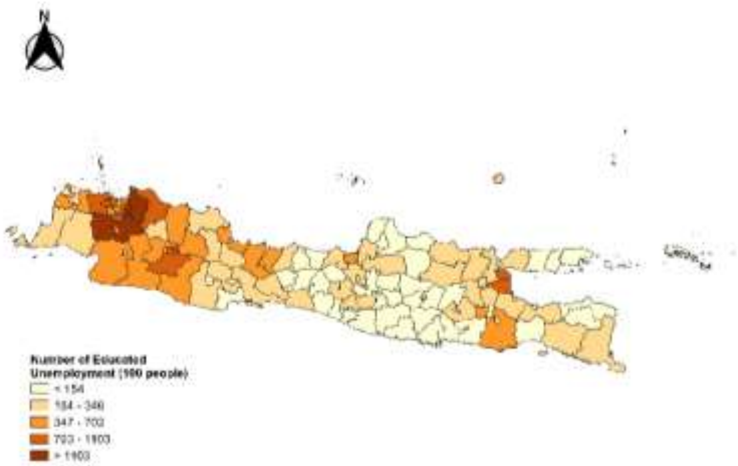

Figure 1. The Distribution of Educated Unemployment in Java, 2021

3.2 Analysis of Educated Unemployment Using Poisson Regression Model

Calculating Variance Inflation Factor (VIF) values enables us to perform multicollinearity identification prior to running the regression models discussed in this paper. The presence of multicollinearity is indicated by a VIF value more than 10, whereas the absence of multicollinearity is shown by a VIF value of 1 [17]. The VIF values for each variable are listed in Table 3. Low VIF values are displayed, demonstrating no multicollinearity in the data.

Tabel 3. Multicollinearity Detection Using VIF Values

Predictor Variable	VIF
Gross Regional Domestic Product (X_1)	3.45
Investment (X_2)	3.05
Regency/city minimum wage (X_3)	2.59
Mean years of schooling (X_4)	1.85
Labor force participation rate (X_5)	1.69
Inflation (X_6)	1.05

The upper panel of Table 4 displays the estimated parameters of the Poisson regression models without an exposure. Every predictor included in the models has a significant effect on the number of educated unemployment in Java at the 5% level of significance. The Poisson regression formula without an exposure in Equation (3) can then be written as follows:

$$\hat{\lambda}_i = \exp\left(7.0978 + 0.0019X_{1i} - 0.0060X_{2i} + 0.5629X_{3i} - 0.0757X_{4i} - 0.0375X_{5i} - 0.0241X_{6i}\right) \quad (11)$$

where i = 1,2,...,119. As an example of how the coefficient shown in Equation (11) or in Table 4 should be interpreted is the average number of educated unemployent decreases by 0.9271 times the initial average number of educated unemployment for every one-year increase in the mean years of shooling (X_4), given that the other predictor variables in the model are held constant. As has been shown, other regression coefficients can be interpreted in the same manner.

Tabel 4. Parameter Estimation of Poisson Regression Models on Educated Unemployment

Variable	Estimation	SE	*z*-statistic	Odds Ratio
Poisson regression without an exposure				
Intercept	7.0978	0.1346	64.3243***	1209.3416
Gross Regional Domestic Product (X_1)	0.0019	0.0001	25.9515***	1.0019
Investment (X_2)	-0.0060	0.0008	-7.6322***	0.9941
Regency/city minimum wage (X_3)	0.5629	0.0080	70.6571**	1.7558
Mean years of schooling (X_4)	-0.0757	0.0042	-17.8900***	0.9271
Labor force participation rate (X_5)	-0.0375	0.0017	-21.8236**	0.9632
Inflation (X_6)	-0.0241	0.0016	-15.3745***	0.9762
Poisson regression with an exposure				
Intercept	-0.3240	0.1415	-2.2894**	0.7233
Gross Regional Domestic Product (X_1)	-0.0007	0.0001	-9.1854***	0.9993
Investment (X_2)	0.0052	0.0008	6.6055***	1.0052
Regency/city minimum wage (X_3)	0.1993	0.0087	22.8758***	1.2205
Mean years of schooling (X_4)	0.0932	0.0043	21.6488***	1.0976
Labor force participation rate (X_5)	-0.0351	0.0019	-18.6107***	0.9656
Inflation (X_6)	0.0162	0.0020	8.0397**	1.0163

* *p*-value<10% ** *p*-value<5% *** *p*-value<1%

Based on the coefficient regressions shown in the lower panel of Table 4, the Poisson regression model when employing an exposure in (5) can be written as follows

$$\hat{\lambda}_i = q_i \exp\left(-0.3240 - 0.0007X_{1i} + 0.0052X_{2i} + 0.1993X_{3i}\ 0 + 0.0932X_{4i} - 0.0351X_{5i} + 0.0162X_{6i}\right) \quad (12)$$

where i = 1, 2, ..., 119. At the 5% level of significance, all predictor variables have a significant effect on educated unemployment. As an illustration of how the Poisson regression model with an exposure should be interpreted, we use the coefficient of mean years of education (X_4) . While holding all other variables in the model constant, the rate of educated unemployment would rise by a factor of 1.0976 for a one year increase in the mean years of schooling.

We used the Breusch Pagan (*BP*) test for heterochedasticity to examine geographical heterogeneity. The result of the *BP* test statistics was 31.87 with *p*-value < 0.000. This implies that the null hypothesis is rejected, indicating the existence of spatial heterogeneity. Therefore, it is recommended to use a spatial regression model that takes spatial variability into account, such as a geographically weighted regression model. The adaptive bisquare weighting function was used because it produced the least *GCV* among a variety of weighting functions, including fixed/adaptive Gaussian weighting functions and fixed/adaptive bisquare weighting functions [9].

3.3 Spatial Analysis of Educated Unemployment Using GWPR without an Exposure

The simultaneous testing on parameters was carried out to examine the significance of parameters simultaneously and gave *p*-value < 0.05, which means that at least one predictor has significant effect on educated unemployment. The calculated parameters of the GWPR models without an exposure are compiled in Table 5. From all 119 regression models that were obtained, it shows variations in estimated parameter values. While Gross Regional Domestic Product (X_1) has a positive effect on educated unemployment in all research locations, investment (X_2), regency/city minimum wage (X_3), mean years of schooling (X_4), labor force participation rate (X_5), and inflation rate (X_6) have varied effects on educated unemployment in each location.

Table 5. Summary of Parameter Estimation of the GWPR Model without an Exposure

Variable	Mean	SD	Min	Median	Max
Intercept	5.9323	3.2898	-0.4857	7.3289	10.4230
Gross Regional Domestic Product (X_1)	0.0068	0.0042	0.0009	0.0076	0.0138
Investment (X_2)	-0.0728	0.0388	-0.1821	-0.0629	0.0159
Regency/city minimum wage (X_3)	0.4481	0.2213	-0.0974	0.6037	0.6341
Mean years of schooling (X_4)	-0.0202	0.0489	-0.0741	-0.0404	0.0743
Labor force participation rate (X_5)	-0.0128	0.0427	-0.0948	-0.0057	0.0523
Inflation (X_6)	0.0057	0.0478	-0.0613	-0.0009	0.1559

Nine location groups were created based on the same predictor variables that significantly influence educated unemployment. The distribution of the nine groups based on the same predictor variables that have significant effects on educated unemployment in each regency/city in Java Island, is shown in Figure 2. In general, at least four predictor variables significantly affect educated unemployment at each location.

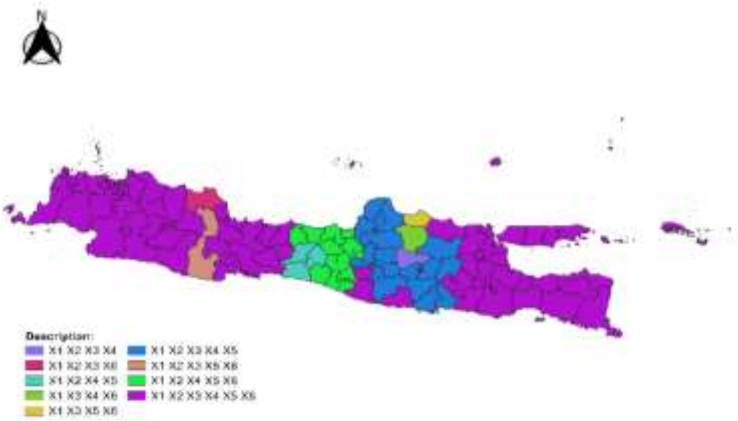

Figure 2. Group of Locations Based on the Same Significant Variables Using the GWPR Model without an Exposure

There appear to be 76 locations where all predictor variables have significant effects on educated unemployment and this group has the most location members. This group is mainly composed of regencies/cities located on the west and east sides of Java Island. Meanwhile, in several locations, there are only 4 or 5 predictor variables that significantly influence educated unemployment, totaling 7 locations and 36 locations, respectively. Almost all of the regencies/cities in this group are found in the central part of Java Island. Table 6 provides a complete list of district/city groups in accordance with Figure 2.

Table 6. Groups of Regencies/Cities on Java Island by Significant Variables Using the GWPR Model Without Exposure

Group	Significant Variables	Number of Regencies/Cities	Regency/City
1	X_1 X_2 X_3 X_4	1	Ngawi
2	X_1 X_2 X_3 X_6	1	Indramayu
3	X_1 X_2 X_4 X_5	3	Banjarnegara, Kebumen, Wonosobo
4	X_1 X_3 X_4 X_6	1	Blora
5	X_1 X_3 X_5 X_6	1	Rembang
6	X_1 X_2 X_3 X_4 X_5	19	Boyolali, Sukoharjo, Wonogiri, Karanganyar, Sragen, Grobogan, Pati, Kudus, Demak, Surakarta City, Ponorogo, Trenggalek, Tulungagung, Nganjuk, Madiun, Magetan, Bojonegoro, Madiun City
7	X_1 X_2 X_3 X_5 X_6	2	Tasikmalaya, Majalengka
8	X_1 X_2 X_4 X_5 X_6	15	Purworejo, Magelang, Semarang, Temanggung, Kendal, Batang, Pekalongan, Magelang City, Salatiga City, Semarang City, Pekalongan City, Kulon Progo, Bantul, Sleman, Yogyakarta City
9	X_1 X_2 X_3 X_4 X_5 X_6	76	Kep. Seribu, Jakarta Barat City, Jakarta Timur City, Jakarta Selatan City, Jakarta Pusat City, Jakarta Utara City, Bogor, Sukabumi, Cianjur, Bandung, Garut, Ciamis, Kuningan, Cirebon, Sumedang, Subang, Purwakarta, Karawang, Bekasi, Bandung Barat, Pangandaran, Bogor City, Sukabumi City, Bandung City, Cirebon City, Bekasi City, Depok City, Cimahi City, Tasikmalaya City, Banjar City, Cilacap, Banyumas, Purbalingga, Klaten, Pemalang, Tegal, Brebes, Tegal City, Gunung Kidul, Pacitan, Blitar, Kediri, Malang, Lumajang, Jember, Banyuwangi, Bondowoso, Situbondo, Probolinggo, Pasuruan, Sidoarjo, Mojokerto, Jombang, Tuban, Lamongan, Gresik, Bangkalan, Sampang, Pamekasan, Sumenep, Kediri City, Blitar City, Malang City, Probolinggo City, Pasuruan City, Mojokerto City, Surabaya City, Batu City, Pandeglang, Lebak, Tangerang, Serang, Tangerang City, Cilegon City, Serang City, Tangerang Selatan City

3.4 Spatial Analysis of Educated Unemployment Using GWPR with an Exposure

Similar to the GWPR regression model without an exposure, the GWPR regression model with an exposure was performed simultaneous testing on parameters to assess the significance of parameters simultaneously. This test provided a p-value<0.05 that indicates all predictor variables are simultaneously affecting the unemployment rate for educated people. In Table 7, the significance of each independent variable from 119 models is summarized. According to this table, regency/city minimum wage (X_3) and the mean years of schooling (X_4) have positive effects on educated unemployment across all research locations. The rest of the predictor variables, however, have different effects on educated unemployment. Figure 3 shows the distribution of location groups based on the same predictor variables that significantly influence educated unemployment.

Table 7. Summary of Parameter Estimation of the GWPR Model with an Exposure

Variable	Mean	SD	Min	Median	Max
Intercept	-1.9811	1.1531	-3.6930	-2.5262	0.8577
Gross Regional Domestic Product (X_1)	-0.0004	0.0012	-0.0017	0.0003	0.0022
Investment (X_2)	-0.0018	0.0173	-0.0296	0.0020	0.0431
Regency/city minimum wage (X_3)	0.2628	0.1312	0.1186	0.2144	0.6646

Variable	Mean	SD	Min	Median	Max
Mean years of schooling (X_4)	0.1314	0.0549	0.0637	0.1184	0.2193
Labor force participation rate (X_5)	-0.0176	0.0167	-0.0521	-0.0165	0.0071
Inflation (X_6)	0.0270	0.0302	-0.0137	0.0121	0.1102

According to **Figure 3**, there are 11 location groups formed based on the same predictor variables that have significant effects on educated unemployment. There is one group in which three predictor variables have a significant effect on educated unemployment, five groups in which there are four significant predictor variables, four groups in which there are five significant predictor variables, and one group in which all predictor variables have a significant effect. All predictor variables significantly affect educated unemployment in 36 regencies/cities. These regencies/cities are clustered in the central and eastern parts of Java Island. Only 3 predictor variables had a statistically significant effect on educated unemployment across 14 locations. These locations include several regencies/cities in Central Java Province as well as all of the regencies/city in DI Yogyakarta Province. There are 4 and 5 predictor variables, respectively, in the remaining 32 districts/cities and 37 districts/cities that significantly affect educated unemployment. The complete list of district/city groups in accordance with Figure 2 is presented in **Table 8**.

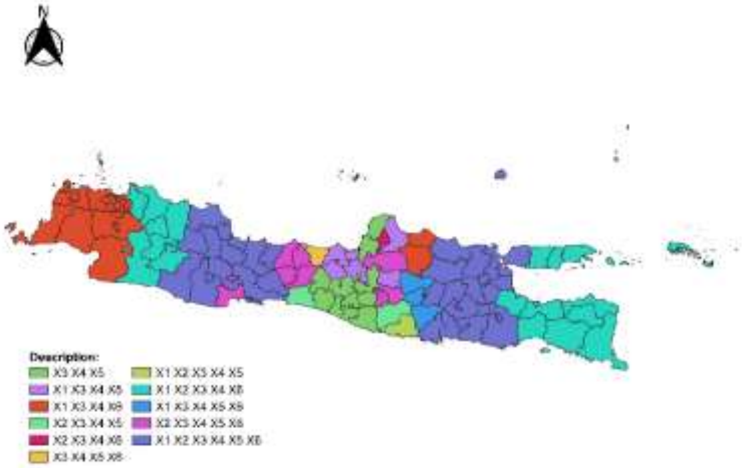

Figure 3. Group of Locations Based on the Same Significant Variables Using the GWPR Model with an Exposure

Table 8. Groups of Regencies/Cities on Java Island by Significant Variables Using the GWPR Model With Exposure

Group	Significant Variables	Number of Regencies/Cities	Regency/City
1	X_3 X_4 X_5	14	Purworejo, Wonosobo, Magelang, Boyolali, Klaten, Jepara, Demak, Magelang City, Salatiga City, Kulon Progo, Bantul, Gunung Kidul, Sleman, Yogyakarta City
2	X_1 X_3 X_4 X_5	6	Sragen, Pati, Semarang, Temanggung, Kendal, Semarang City
3	X_1 X_3 X_4 X_6	21	Kep. Seribu, Jakarta Selatan City, Jakarta Timur City, Jakarta Pusat City, Jakarta Barat City, Jakarta Utara City, Bogor, Sukabumi, Bogor City, Bekasi City, Depok City, Blora, Rembang, Pandeglang, Lebak, Tangerang, Serang, Tangerang City, Cilegon City, Serang City,
4	X_2 X_3 X_4 X_5	2	Kebumen, Wonogiri
5	X_2 X_3 X_4 X_6	1	Kudus
6	X_3 X_4 X_5 X_6	2	Batang, Pekalongan City
7	X_1 X_2 X_3 X_4 X_5	1	Pacitan
8	X_1 X_2 X_3 X_4 X_6	22	Cianjur, Bandung, Subang, Purwakarta, Karawang, Bekasi, Bandung Barat, Sukabumi City, Bandung City, Cimahi City, Lumajang, Jember, Banyuwangi, Bondowoso, Situbondo, Probolinggo, Pasuruan, Sampang, Pamekasan, Sumenep, Probolinggo City,
9	X_1 X_3 X_4 X_5 X_6	3	Ponorogo, Magetan, Ngawi

Group	Significant Variables	Number of Regencies/Cities	Regency/City
10	X_2 X_3 X_4 X_5 X_6	11	Pangandaran, Banjar City, Purbalingga, Banjarnegara, Sukoharjo, Karanganyar, Grobogan, Pekalongan, Pemalang, Surakarta City,
11	X_1 X_2 X_3 X_4 X_5 X_6	36	Garut, Tasikmalaya, Ciamis, Kuningan, Cirebon, Majalengka, Sumedang, Indramayu, Cirebon City, Tasikmalaya City, Cilacap, Banyumas, Tegal, Brebes, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Sidoarjo, Mojokerto, Jombang, Nganjuk, Madiun, Bojonegoro, Tuban, Lamongan, Gresik, Bangkalan, Kediri City, Blitar City, Malang City, Mojokerto City, Madiun City, Surabaya City, Batu City

3.5 Comparison of Models

Four models—two global regression models, or Poisson regression models without and with an exposure, and two local regression models, or GWPR models without and with an exposure—have been created in this study. Using the AIC or AIC_c value as stated in **Equation (10)**, it was determined which of the four created models was the best. The AIC or AIC_c values for all the models are presented in **Table 9**.

Table 9. Comparison of Global and Local Models

Model	*AIC*	*AIC*$_c$
Poisson regression without an exposure	12395.46	12396.47
Poisson regression with an exposure	1677.72	1678.73
GWPR without an exposure	8430.40	8444.49
GWPR with an exposure	835.48	848.98

The model with the lowest AIC or AIC_c is the best one. The local regression model with an exposure yields the minimum AIC or AIC_c, as shown in **Table 9**. It can be concluded that taking into account spatial heterogeneity in modeling the relationship between predictor variables and educated unemployment and including an exposure measure is the best modeling. In this study, it also seems that the global regression model with an exposure measure yields a smaller AIC or AIC_c than the global regression model without an exposure.

4. CONCLUSIONS

The highest distribution of educated unemployment in Java Island is dominated by large and developed regencies/cities such as Jakarta Metropolitan Area (Jabodetabek), the city of Bandung, and the city of Surabaya. This study reveals that there is a significant spatial heterogeneity that a spatial regression model needs to be applied in modeling the relationship between the predictor variables and educated unemployment. An exposure measure is also included in the model in order to control population size differences among locations. Based on the smallest AIC or AIC_c, the best model for analyzing factors affecting educated unemployment are the GWPR models with an exposure. These models create 11 groups of locations based on the same predictor variables that significantly affect educated unemployment.

The GWPR model produced by this study equals the number of research observation units, which in this case are 119 regencies/cities on Java Island. The GWPR models provide a multitude of information, which we are unable to fully explore in this paper. This study's analysis is restricted to groups of locations based on the same predictor variables that have significant effect on educated unemployment. More detailed analysis, such as the sign of the estimated parameters and the magnitude of the accompanying standard errors are left for interested readers.

The study’s limitation is that presumably the data meets equidispersion requirement. Violation of this assumption requires different global and local regression models. Therefore, we recommend more advanced modeling on educated unemployment for further research. These models should use local regression models that can handle cases of overdispersion or underdispersion cases, as well as, an exposure measure.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to the Center for Research and Community Service (Pusat Penelitian dan Pengabdian kepada Masyarakat/PPPM) Politeknik Statistika STIS for providing us with the necessary facilities and resources to conduct this research.

REFERENCES

[1] C. Gedikli, M. Miraglia, S. Connolly, M. Bryan dan W. David, "The relationship between unemployment and wellbeing: an updated meta-analysis of longitudinal evidence," *European Journal of Work and Organizational Psychology,* vol. 32, no. 1, pp. 128-144, 2023.

[2] C. Panari dan M. Tonelli, "Future Directions in the Research on Unemployment: Protean Career Orientation and Perceived Employability Against Social Disadvantage," *Frontiers in Psychology,* vol. 21, no. 701861, 2021.

[3] E. P. Richter, E. Brähler, Y. Stöbel-Richter, M. Zenger dan H. Berth, "The long-lasting impact of unemployment on life satisfaction: results of a longitudinal study over 20 years in East Germany," *Health and Quality of Life Outcomes,* vol. 18, no. 361, pp. 1-7, 2020.

[4] S. D. Simpson, "The Cost of Unemployment to the Economy," 14 April 2022. [Online]. Available: https://www.investopedia.com/financial-edge/0811/the-cost-of-unemployment-to-the-economy.aspx. [Diakses 17 September 2023].

[5] R. F. Putri, "Analisis Pengaruh Inflasi, Pertumbuhan Ekonomi dan Upah Terhadap Pengangguran Terdidik," *Economics Development Analysis Journal,* vol. 4, no. 2, pp. 175-181, 2015.

[6] S. Veronika dan A. Y. Mafruhat, "Pengaruh Pertumbuhan Ekonomi, Investasi dan Inflasi terhadap Pengangguran Terdidik di Provinsi Jawa Barat," *Jurnal Riset Ilmu Ekonomi dan Bisnis,* vol. 2, no. 2, p. 139–146, 2022.

[7] D. Adriani, N. Hamzah dan J. Zakaria, "Pengaruh Produk Domestik Regional Bruto, Tingkat Pendidikan dan Upah Minimum Terhadap Pengangguran Terdidik," *CESJ: Center Of Economic Students Journal,* vol. 2, no. 3, pp. 1-17, 2019.

[8] L. Anselin, Spatial Econometrics: Methods and Models, Dordrecht: Kluwer Academic Publishers, 1988.

[9] A. S. Fotheringham, C. Brundson dan M. Charlton, Geographically Weighted Regression: The analysis of spatially varying relationship, England: John Wiley & Son Ltd, 2002.

[10] S. Mardalena, Purhadi, J. D. T. Purnomo dan D. D. Prastyo, "Bivariate poisson inverse gaussian regression model with exposure variable: infant and maternal death case study," *Journal of Physics: Conference Series,* vol. 1752, no. 012016, pp. 1-10, 2021.

[11] E. L. Frome, "The Analysis of Rates Using Poisson Regression Models," *Biometrics,* vol. 39, no. 3, pp. 665-674, 1983.

[12] A. C. Cameron dan P. K. Trivedi, Regression analysis of count data, 2nd penyunt., New York: Cambridge University Press., 2013.

[13] R. Winkelmann, Econometric analysis of count data, 5 penyunt., Berlin: Springer, 2008.

[14] T. S. Breusch dan A. R. Pagan, "A Simple Test for Heteroscedasticity and Random Coefficient Variation," *Econometrica,* vol. 47, no. 5, pp. 287-1294, 1979.

[15] H. Akaike, "Information theory and an extension of the maximum likelihood principle," dalam *Second International Symposium on Information Theory*, Budapest, 1973.

[16] BPS-Statistics, "Statistik Pengangguran 2001-2006," Jakarta, 2007.

[17] X. Yan dan X. G. Su, Linear regression analysis: Theory and computing, Singapore: World Scientific Publishing Co. Pte. Ltd., 2009.