Identification of factors affecting the cases of under-age female marriage using geographically weighted panel regression approach in south kalimantan province.

A Rifqi[1,*], **D S Susanti**[1], **Abdurrahman**[2]

[1] Statistics Programe Study, Faculty of Mathematics and Natural Sciences, Lambung Mangkurat University, Banjarbaru, South Kalimantan, Indonesia
[2] Central Statistics Agency, Banjarbaru, South Kalimantan, Indonesia

* Corresponding author's email: abdullsh.rifqi11@gmail.com

Abstract. The topic of this study was chosen because the percentage of underage female marriages in South Kalimantan Province was the highest in Indonesia over the last five years, from 2018 to 2022. This signifies that there are social issues in the local community that the government must address. One possible answer is to identify the factors that contribute to the creation of these conditions in each region. Using the Geographically Weighted Panel Regression (GWPR) method, this study attempts to determine the factors that influence the rise of underage female marriage instances in South Kalimantan Province. The number of poor individuals, population density, average duration of schooling, adjusted per capita expenditure, and total population were chosen as independent variables. Data acquired from South Kalimantan Province's Central Bureau of Statistics' periodic releases. Because there was high spatial heterogeneity between each location in this study, it was quite practical to employ the GWPR approach in developing a conjectural model. The results of evaluating the GWPR model with adaptive Gaussian kernel weights provide significant results and the model can explain the variance of data by 55 percent. Testing the parameters of the GWPR model reveals two (two) regional groupings with distinct influencing variables. The first group consists of ten (ten) regions that are considerably impacted by both the number of impoverished people and the average length of schooling, whereas the second group consists of three (three) regions that are impacted solely by the average length of schooling.

1. Introduction

Marriage is a sacred and very important event in human life, in Indonesia marriage is regulated in Law Number 16 of 2019 where article (7) paragraph 1 explains that marriage is only permitted if the man and woman have reached the age of 19 years so that, marriages that are carried out before the age of the provision are referred to as underage marriages. The law also states that underage marriage may have a negative impact on the growth and development of children and will lead to failure to respect children's fundamental rights such as the right to protection from violence and discrimination, civil rights of childern, health and education rights, and children's social rights.

Marriage must be prepared on the basis of readiness and the importance of the reproductive system, especially for women [4]. This issue is of particular concern because it can have a negative impact on women's health and well-being, including a high risk of maternal and infant mortality and can also have a negative impact on the woman's social life [2].

South Kalimantan is the province with the highest percentage of underage female marriage in Indonesia from 2018 to 2022 [3]. The Central Statistics Agency (BPS) noted that the highest percentage was in 2019 with 45.5% and the lowest in 2022 with 41.5%. This figure is much higher than the national average of 33.5%-36.2%.

Factors that influence underage female marriage can sometimes differ in reference to conditions between one location and another. This is influenced by the geographical or socio-cultural conditions of an area that allows spatial heterogeneity [1]. Therefore, a method is needed that is able to accomodate aspects of spatial heterogeneity, namely by using the Geographically Weighted Regression (GWR) method. However, because the data used is panel data, the GWR method was developed into Geographically Weighted Panel Regression (GWPR). So this research focuses on how much influence the factors that have been determined on underage female marriage in South Kalimantan Province using the Geographically Weighted Panel Regression (GWPR) method.

2. Literature Review

2.1. Panel Data Regression

Panel data regression is a statistical analysis technique or method that aims to produce a model with the effect of independent variabels on the dependent variabel of data set. The general equation of the panel data regression model is as follows [5]:

$$Y_{it} = \beta_{0it} + \sum_{k=1}^{k} \beta_k X_{kit} + \varepsilon_{it} \qquad i = 1, \dots, N; t = 1, \dots, T; k = 1, \dots, K \tag{1}$$

There are three popular approaches to estimate panel data regression models, which are Common Effect Model (CEM), Fixed Effect Model (FEM) and Random Effect Model (REM).

a. Common Effect Model (CEM)

The assumption used in this approach is to ignore all influences from both individual and time units, so that the intercepts and slope are constant across the time and individual unit. The following is the general equation of the CEM approach [6]:

$$Y_{it} = \beta_0 + \sum_{k=1}^{k} \beta_k X_{kit} + \varepsilon_{it} \tag{2}$$

b. Fixed Effect Model (FEM)

The assumption used in this approach is that there are differences between individuals, so the intercept for each individual unit will be different. The following is the general equation of the FEM approach [7]:

$$Y_{it} = \beta_{0i} + \sum_{k=1}^{k} \beta_k X_{kit} + \varepsilon_{it} \tag{3}$$

c. Random Effect Model (REM)

This approach is used to overcome the problems caused in FEM by involving error terms due to changes in individuals and observation times that can cause intercept differences between times. The following is the general equation of the REM approach [8]:

$$Y_{it} = \beta_{0it} + \sum_{k=1}^{k} \beta_k X_{kit} + u_i + \varepsilon_{it} \tag{4}$$

2.2. Selection of Panel Data Regression Model Estimation

a. Chow Test

The Chow test is used to determine which regression model is better to use between CEM and FEM by testing the significance of the β_{0i} intercept using the F test statistic as follows [9]:

$$F = \frac{RSS_n - RSS_{DV}/(N-1)}{RSS_{DV}/(NT-N-K)} \tag{5}$$

b. Hausman Test

The Hausman test is a test used to determine which regression model is better to use between FEM and REM. The test statistic used is chi-squared based the Wald criteria as follows [10]:

$$W = X^2[\boldsymbol{K}] = [\widehat{\boldsymbol{\beta}}_{\boldsymbol{FEM}} - \widehat{\boldsymbol{\beta}}_{\boldsymbol{REM}}]' \left[\boldsymbol{var}[\widehat{\boldsymbol{\beta}}_{\boldsymbol{FEM}} - \widehat{\boldsymbol{\beta}}_{\boldsymbol{REM}}]\right]^{-1} [\boldsymbol{\beta}_{\boldsymbol{FEM}} - \widehat{\boldsymbol{\beta}}_{\boldsymbol{REM}}] \tag{6}$$

2.3. Parameter Significance Test of Panel Data Regression Model

a. Simultaneous Test

The simultaneous test shows whether all the independent variables used in the model have a significant influence together on the dependent variable. This simultaneous test is usually also referred to as the F test statistic as follows:

$$F = \frac{RSS_{OLS} - RSS_{GWR}/v_1}{RSS_{GWR}/\delta_1} = \frac{GWR_{improv}/v_2}{RSS_{GWR}/\delta_2} \tag{7}$$

b. Partial Test

Unlike the simultaneous test, the partial test is used to see the significance of the effect of the independent variable on the dependent variable individually using the t statistical test with the following equation:

$$t = \frac{B_j}{se(B_j)} \tag{8}$$

2.4. Assumption Test

a. Normality

The normality test is used to test the normality of the error values in the regression equation. In this test, the *Jarque-Bera* goodness test is used to measure the skewness and kurtosis of the sample whether it is/was in accordance with the normal distribution. Therefore, the absolute value of these parameters can be a measure of the deviation of the distribution from normal.

Here is the equation for the *Jarque-Bera* test statistic [11]:

$$JB = \frac{n}{6}\left(\frac{S^2(K-3)^2}{4}\right) \tag{9}$$

b. Multicollinearity

To detect whether there is a linear relationship between variabels or multicolinearity, it can be done by calculating the Variance Inflation Factor (VIF) value. The formula is as follows [12]:

$$(VIF)_j = \frac{1}{1-R_j} \tag{10}$$

c. Heteroscedasticity

This test aims to test whether the regression model experinces inequality of variance of error between observations or not. One that can be used to test the homogeneity of residuals is to use the *Breusch-Pagan* test with the following formula [5]:

$$\phi = \frac{1}{2} \times ESS \tag{11}$$

2.5. Weighting

a. Fixed Kernel Function

A fixed kernel function is a kernel function with the same bandwidth value at each observation location. This function is further divided into two, namely the Gaussian and Bisquare Fixed Kernel functions [9].

1) Fixed Gaussian Kernel Function

The equation of this function is as follows:

$$w_{ij} = exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right] \tag{12}$$

2) Fixed Bisquare Kernel Function

The equation of this function is as follows:

$$w_{ij} = \begin{cases} \left(-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right), for\ d_{ij} < b \\ 0 \qquad , for\ d_{ij} \geq b \end{cases} \tag{13}$$

b. Adaptive Kernel Function

The adaptive kernel function is a kernel function with different bandwidth values at each observation location. This function is further divided into two, namely the Gaussian and Bisquare adaptive kernel functions [9].

1) Adaptive Gaussian Kernel Function

The equation of this function is as follows:

$$w_{ij} = exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{b_i}\right)^2\right] \tag{14}$$

2) Adaptive Bisquare Kernel Function

The equation of this function is as follows:

$$w_{ij} = \begin{cases} \left(-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right)^2, for\ d_{ij} < b \\ 0 \qquad , for\ d_{ij} \geq b \end{cases} \tag{15}$$

2.6. Geographically Weighted Panel Regression

Geographically Weighted Panel Regression (GWPR) is a local regression model with repetition of data at location points for each observation at different times. The general model of GWPR is obtained from the combination of Geographically Weighted Regression (GWR) model with FEM panel data regression with *within estimator* as follows [1]:

$$y_{it} = \beta_0(u_{it}, v_{it}) + \sum_{k=1}^{p} \beta_k(u_{it}, v_{it})x_{itk} + \varepsilon_{it} \ \ i = 1, 2, \dots, n\ dan\ t = 1, 2, \dots, T \tag{16}$$

2.7. Model Hypothesis Test

a. Model Fit Testing (goodness of fit)

Model fit testing was performed to determine whether there was a difference between the global regression and the local regression. The test statistic used is the F-test statistic with the following equation [1]:

$$F = \frac{SSE(H_0)/df_1}{SSE(H_1)/df_2} \tag{17}$$

b. Model Parameter Testing

Testing the parameters of this model is carried out to determine which parameters have a significant effect on the dependent variabel. The test statistic used is the T test statistic with the following equation [1]:

$$F = \frac{SSE(H_0)/df_1}{SSE(H_1)/df_2} \tag{18}$$

3. Research Methods

3.1 Data Sources and Variables

The data in this study uses data sourced from routine publications of the Central Statistics Agency (BPS) of South Kalimantan Province (secondary data) consisting of individual units in the form of 13 districts/cities, as well as time units in units of years from 2019-2022 as much as 65 observation data.

Table 1. Researh Variables

No	Variable	Unit
1	Percentage of underage female marriages (Y)	Percent
2	Number of poor people (X_1)	Soul
3	Population density (X_2)	Soul/km^2
4	Average length of schooling (X_3)	Year
5	Adjusted per capita expenditure (X_4)	Thousand Rupiah
6	Total Population (X_5)	Soul

4. Results

4.1 Descriptive Analysis

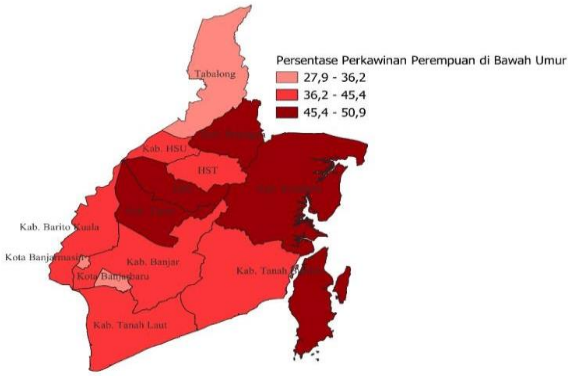

Figure 1. Thematic Map of Percentage of Underage Female Marriage in 2020 of South Kalimantan Province

Figure 1 shows a thematic map of the percentage of underage female marriage in South Kalimantan Province in 2022. In figure 1, cases of underage famele marriage are grouped into three classes. Class 1 shows a low percentage, class 2 shows a medium percentage, and class 3 shows the percentage of underage famele marriage with a high percentage marked in dark red.

Areas that are in class 1 or areas that have a low percentage of underage female marriage consist of Tabalong Regency, Banjarmasin City, and Banjarbaru City. Meanwhile, areas that are dark red or in class 3, namely areas with a high percentage of underage female marriage, are found in Balangan, Kotabaru, Hulu Sungai Selatan, and Tapin districts. And the other 6 regions are in class 2 or areas with a medium percentage of underage female marriage.

4.2 Panel Data Regression Model Identification

Statistically, based on the Chow test and Hausman test performed, the results show that the Fixed Effect Model (FEM) is better than the Common Effect Model (CEM) and Random Effect Model (REM). This means that it is assumed that there are differences between individuals or in this case districts/cities, so that the constant for each district/city will be different.

4.3 Model Parameter Significance Test

a. Simultaneous Test

The F test statistic is used to determine whether the independent variables in this study simultaneously or jointly affect the dependent variable. In the estimation results of the FEM model. The F value is obtained of 10.02 and the value of P-*value* is 0.000 which ≤ α = 0.05 then reject H_0 which means that the independent variable in this study have a simultaneous influence on the dependent variable, which is the percentage of underage female marriage. It is said that the panel data regression model that will be formed is significant and can describe the diversity of sample data as well as for prediction.

b. Partial Test

The t test statistics are used to see whether in this study each independent variabel in the study has an individual or partial influence on the dependen variable used.

Table 2. Partial Test

Variable	Coefficient	*t*	P_{value}
Intercept	137,4637	6,1222	0,0000**
X_1	0,0011	2,7972	0,0074**
X_2	− 0,000188	-0,1664	0,8685
X_3	−11,2	-5,8822	0,0000**
X_4	−0,000712	-0,3473	0,7299
X_5	0,0000139	0,3274	0,7448

Table 2 shows that the P-*value* of the population density variable (X_2), the adjusted per capita expenditure variable (X_4), and the total population variable (X_5) is more than α = 0.05, so H_0 is accepted, which menas that that these variables do not have a significant effect on the percentage of underage female marriage in South Kalimantan Province. While the variable number of poor people (X_1) and the average length of schooling variable (X_3) have a significant effect. With the panel data model as follows:

$$Y^* = 137.46 - 0.001X_1 - 11.2X_2$$

4.4 Model Assumption Test

a. Normality Test

Testing the normality assumption is done using the *jarque-bera* test with the results obtained the value of *P-value* = 0.4127 ≥ α = 0.05 then H_0 is accepted, which means that the residuals are normally distributed. So it can be said that the assumption of normality has been fulfilled.

b. Multicollinearity Test

This test will be used to determine whether there is more than one (multi) linear relationship between the independent variables described in the panel data regression model.

Table 3. Multicollinearity Test

	Variable				
	X_1	X_2	X_3	X_4	X_5
X_1	1	0,80	0,36	0,37	0,83
X_2	0,80	1	0,54	0,58	0,70
X_3	0,36	0,54	1	0,68	0,25
X_4	0,37	0,58	0,68	1	0,49
X_5	0,83	0,70	0,25	0,49	1
VIF	5,629	3,894	2,393	2,849	4,395

From Tabel 3, it can be seen that the correlation value between variables is less than 0.9 and the VIF value between independent variables is < 10 [9]. So it can be said that the panel data regression model obtained is free from multicollinearity symptoms.

c. Heteroscedasticity Test

Heteroscedasticity testing will be carried out using the *Breusch-Pagan Godfrey* (BPG) test to test whether the regression model experiences inequality in the variance of residuals between observations. Based on the results obtained, the BPG value is 15.47 with a value of *P-value* is 0.09 ≤ α = 0.10, so it can be said that the assumption of homogeneity is not fulfilled. Heteroscedasticity usually occurs as a result of different observation areas. There is a heteroscedasticity problem in this study, so the analysis will be carried out to the next stage by paying attention to spatial or location effects using the Geographically Weighted Panel Regression method.

4.5 Geographically Weighted Panel Regression Modeling

Before estimating the parameters of the GWPR model, data transformation (demeaning) is first carried out. The trasnformation of the data will be carried out by applying the within estimator concept. Which transforms the research variables by subtracting them from the average of the corresponding time series.

a. Comparison of Kernel Weights

From the results of research comparing the value of each weight, the optimum weight with the smallest *Cross Validation* (CV) value will be determined [9]. In GWPR modeling, the weight matrix for each year is the same so that the value is repeated for each year period. Furthermore, the weight matrix is used to estimate the parameters in the GWPR model and the model will be different in each location.

b. GWPR Model Significance Test

Simultaneous model testing is to test the *goodness of fit* of the GWPR model. The results of simultaneous model testing in this study show that the F value of 12.59 is greater than $F_{(\alpha;K;n-k)}$ of 2.37 and the P-*value* of 0.000 < 0.05, meaning that H_0 is rejected [9]. Thus, it can be said that GWPR model with *Adaptive Gaussian Kernel* weighting has a better *goodness of fit* than the global regression model. After testing the significance of the model parameters at each observation location and knowing the results show that the independent variables that influence the percentage of underage female marriage in each region in South Kalimantan Province. Two groups of districts/cities are shown in Figure 2 as follows:

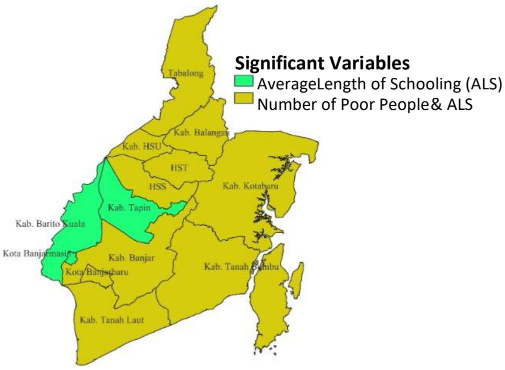

Figure 2. Thematic Map of Significant Independet Variable Groupings

The following is the GWPR model based in the group of significant variables in each observation area:

Table 4. District/City Grouping Based on Significant Variables

Group	District/City	Model
1	Barito Kuala Tapin Banjarmasin	$Y_1^* = 42.08 - 12.473X_3$
2	Tanah Laut Banjar Kotabaru Hulu Sungai Selatan Hulu Sungai Tengah Hulu Sungai Utara Tabalong Tanah Bumbu Balangan Banjarbaru	$Y_2^* = 46.57 + 0.001X_1 - 12.652X_3$

In Figure 2 and Table 4, the results show that efforts can be made to reduce the rate of underage female marriage in South Kalimantan Province by paying attention to influential factors and inplementing policies that are accordance with the characteristics found in each region. So that it is hoped that programs to suppress the percentage of underage female marriage can run effectively.

Then, based on the resulting R^2 value. The GWPR model is a better model than global panel data regression in modeling the percentage of underage female marriage in South Kalimantan Province. GWPR model with *Adaptive Gaussian Kernel* weighting is able to explain the diversity of the percentage of underage female marriage by 55% while the global regression is only 52%. Overall there are two independent variables that affect the percentage of underage female marriage, namely the number of poor people (X_1) and the average years of schooling (X_3).

5. Conclusions

Based on the research results in the discussion, the following conslusions can be drawn:

1. Based on the GWPR model of the five factors studied, two factors were found to influence the percentage of underage female marriage, namely the number of poor people and the average length of schooling where the number of poor people has a positive value, which means that the increasing number of poor people will increase the percentage of underage female marriages.

While the average length of schooling has a negative value, which means that if the average length of schooling increases, percentage of underage female marriage will decrease.

2. Based on the GWPR model, the magnitude of the influence of the number of poor people factor is 0.001 and the average length of schooling factor has the largest influence of -12.652 and the smallest of -12.473.
3. Overall, there are two groups of districts/cities that are grouped based on factors that influence the percentage of underage female marriage. Namely Tapin, Barito Kuala, and Banjarmasin City are only influenced by the average length of schooling factor (X_3). Meanwhile, other district and cities are influenced by the number of poor people (X_1) and the average length of schooling (X_3).

References

[1] N. F. Gamayanti, Junaidi, Fadjryani and Nur'eni., "Analysis of Spatial Effects on Factors Affecting Rice Production in Central Sulawesi Using Geographically Weighted Panel Regression," *Barekeng: J. Math. & App.*, vol. 17, iss. 1, pp. 0361-0370, March 2023.

[2] Badan Pusat Statistik. (2020). Pencegahan Perkawinan Anak Percepatan yang Tidak Bisa Ditunda. Badan Pusat Statistik, 6-10. [3] BPS Provinsi Kalimantan Selatan. (2022). Publikasi BPS Provinsi Kalimantan Selatan.

[3] https://kalsel.bps.go.id/publication.html?Publikasi%5BtahunJudul%5D=&Publikasi%5BkataKunci%5D=Statistik+Kesejahteraan+Rakyat&Publikasi%5BcekJudul%5D=0&yt0=Tampilkan

[4] F. Alfariz, Purhadi., "Pemodelan Jumlah Anak Putus Sekolah Usia Wajib Belajar dan Jumlah Wanita Menikah Dini di Jawa Timur dengan Pendekatan Geographically Weighted Bivariate Negativ Binomial Regression," J. Sains & Seni ITS., vol. 8, 2019.

[5] Baltagi, B. (2005). Econometric analysis of panel data. In Vasa.

[6] Sriyana, J. (2014). Metode Regresi Data Panel: Dilengkapi Analisis Kinerja

[7] Widarjono, A. (2005). Ekonometrika: teori dan aplikasi untuk ekonomi dan bisnis (1st ed., Vol. 2). EKONISIA.

[8] Wooldridge, J. M. (2013). Introductory Econometrics A Modern Approach 5th Edition. In Tolerance Analysis of Electronic Circuits Using MATHCAD.

[9] S. M. Meutuah, H. Yasin, D. Asih et al., "Pemodelan Fixed Effect Geographically Weighted Panel Regression Untuk Indeks Pembangunan Manusia di Jawa Tengah," J. Gaussian., vol. 6, iss. 2, pp. 0241-0250, 2017.

[10] Ghozali, I. (2011). Aplikasi Analisis Multivariat dengan Program IBM SPSS 19 (5th ed.). Badan Penerbit Universitas Diponegoro.

[11] Jarque, C. M., and Bera, A. K. 1987. A Test for Normality of Observation and Regression Residuals. Internasional Statistical Review, vol. 55, pp. 163-172.

[12] D. N. Gujarati, D. Porter,. "Basic Econometrics (4 ed.)," New York: The McGraw-Hill Companies., 2004.