Gender Inequality Index Modeling in Indonesia Using Geographically Weighted Panel Regression Method

Khaira Umi(✉), Junaidi, and Fadjryani

Statistics Department, Faculty of Mathematical and Natural Sciences, University of Tadulako, Palu 94118, Indonesia
khairakhairum@gmail.com

Abstract. The Gender Inequality Index (GII) is an index that describes the failure of human development achievement due to the inequality of gender that is measured by several aspects such as health, empowerment, and labor market access. The Geographically Weighted Panel Regression (GWPR) is a developed model which amalgamates the GWR model with panel data regression. The GWPR is used to address spatial heterogeneity problems in panel data. This study aims to determine the results of the GII modeling in Indonesia using the GWPR and their several influenced factors. The results of the GII modeling using the GWPR with adaptive kernel bisquare weighted functions produce different model equations at each location. The GWPR model produced an R^2 of 73.97% with factors that significantly affect the GII including the percentage of people aged over 10 years who have no school (X_2), the percentage of non-health facilities of childbirth (X_3), the percentage of women sitting in parliament (X_4), DI (X_5), GDI (X_6), OUR (X_7) and LFPR (X_8). Moreover, given an example, by the assumption of other variables are ignored the GII of Central Sulawesi Province in the t-th year will decrease by 0.0296. Moreover, the increase of 1% of women sitting in parliament (X_4) in the t-th year by the assumption of ignoring other variables, will then decrease the GII by 0.0524. In addition, the increase of 1 unit of HDI (X_5) in the t-year by assuming other variables are ignored, will decrease the GII by 1.7063.

Keywords: Panel Data Regression · GWPR · Gender Inequality Index · Adaptive Kernel Bisquare

1 Introduction

Gender is an idea that mentions a role system and interrelation between men and women which are not shown by the differences in their biological. The difference indicates by the socio-cultural, political, and economic environment. The equality of gender determines the same status among women and men, the same potential in realizing their full rights to come up in developing economic, social political, and national. Gender equality and justice can be realized among women and men if there is no discrimination to participate and access over development. Moreover, they have equality to take benefit from development [4].

N. I. Inda et al. (Eds.): ISST 2022, APR 7, pp. 303–311, 2023.
https://doi.org/10.2991/978-94-6463-228-6_34

The value of Indonesia's Gender Inequality Index (GII) from 2018 to 2020 tends to continue to decline. Indonesia's GII in 2018 of 0.436 or 43.6% continued to decrease until 2020 by 0.400 or 40%. Although in the last three years, the GII in Indonesia has continued to decline, the distribution of the GII numbers in Indonesia is still quite diverse. In 2020 the low GII achievement was occupied by DI Yogyakarta Province while the GII with a high category was occupied by the West Nusa Tenggara Province [1]. The difference value for the GII shows that the achievement of gender development in each region is still not optimal as indicated by the inequality of development in the fields of health, education, politics, and employment.

The differences in the GII category for each region in Indonesia show that the achievements of gender development in these regions are still uneven, consequently, this allows for spatial heterogeneity. The method that can be used for the spatial homogeneity case is the Geographically Weighted Regression (GWR). However, for advanced research, this is not sufficient to conduct observation only at one particular time. We need to make surely observations at various time periods. Based on this issue, Rahayu stated that the panel data regression analysis has to be carried out where this involves cross-section and time series units. In her study, she also considered spatial aspects, as a result, the GWPR method was used [7].

Furthermore, the Fixed Effect Model (FEM) regards that the interception differs for each location while the slope between locations is the same. The within estimator is one of the alternatives used in estimating parameters in FEM models. The purpose of using the within estimator method is to eliminate the unobserved effect (αi). The concept of within estimators is to transform research variables by subtracting them from the corresponding average time series [3]. The general model of FEM panel data regression is [2]:

$$y_{it} = \beta_{0it} + \beta_1 x_{1it} + \ldots + \beta_k x_{kit} + \varepsilon_{it}; \text{ for } i = 1, 2, \ldots 34 \text{ and } t = 1, 2, 3$$

where k is the k-th predictor variable, t is time, i show a location, y_{it} determines the response variables on the i-th observations and t-th time, β_{0it} indicates interception of the i-th and t-th time observations, β_k is the k-th predictor variable parameter, x_{itk} shows the k-th predictor variables on the i-th and t-th time observations and ε_{it} is the errors with the assumption of independent, identical and distributed normally. Moreover, the value of mean and variance are zero and constant, respectively.

The GWPR is a model that was developed by a combination of the GWR model with panel data regression. The GWPR is used to address effect and heterogeneity spatial problems for panel data. The GWPR analysis aims to combine overall location (cross-sectional) and observation [10]. In this research, it is assumed that the conditions of each observation unit were different from each other. As a result, the panel data regression with the FEM is used. The GWPR model is given in the following equation [8]:

$$y_{it} = \beta_0(u_{it}, v_{it}) + \sum_{i=1}^{k} \beta_k(u_{it}, v_{it}) x_{kit} + \varepsilon_{it}$$

where y_{it} determines the response variables on the i-th observation and t-th time, $\beta_0(u_{it}, v_{it})$ shows interception of the the i-th observation and t-th time, $\beta_k(u_{it}, v_{it})$ is parameters of the k-th predictor variable on the i-th observation and t-th time, x_{itk} is the k-th predictor variables on the i-th observation and t-th time, (u_{it}, v_{it}) shows the coordinate point of the location on the i-th observation and t-th time and ε_{it} is the errors. The errors are assumed independent, identical, and distributed normally with a mean (zero) and variance (constant).

Research that used the GWPR method in a case study of poverty in Indonesia was conducted by [6]. They obtained that the FEM with the function of the weighted exponential adaptive kernel is the GWPR model much better to be used. Another research is carried out by Marsono who applied spatial detection in Indonesia's gender inequality index model. In this research, they said that the factors which significantly affect the GII are the percentage of the female population 10 years and over who have not/ have never been to school, women aged 20–24 years who have the status of mating/ living together 18 years, non-health facilities of childbirth, and HDI in women and IDG [5]. Based on several previous researches that had been conducted, in this study the gender inequality index (GII) model in Indonesia will be carried out by using the Geographically Weighted Panel Regression (GWPR).

2 Material and Methods

2.1 Material

The data obtained from the publication website of the Central Statistics Agency, namely the book Calculating the Gender Inequality Index in 2018, 2019, and 2020 are applied in this research. The variables are the Gender Inequality Index (Y), the Proportion of Women aged 20–24 years who were of first marriage age or first cohabitation before the age of 18 years (X_1), the percentage of people aged over 10 years who have no school (X_2), the percentage of non-health facilities of childbirth (X_3), the percentage of women sitting in parliament (X_4), Human Development Index (HDI) (X_5), Gender Development Index (GDI) (X_6), Open Unemployment Rate (OUR) (X_7), and Labor Force Participation Rate (LFPR) (X_8).

2.2 Methods

The GWPR method is used to analyze the data by applying the R 4.1.3 software. The steps of the analysis are conducted as follows: the data collection of gender inequality index along with 8 free variables, to conduct a descriptive statistical analysis, before the panel data regression model is selected, firstly we look for each panel data regression model. These are the Common Effect Model (CEM), Fixed Effect Model (FEM), and Random Effect Model (REM). The Chow test, Hausman test, and Lagrange Multiplier test are then used to select the panel data regression model. Furthermore, the classical assumptions of regression of panel data tests include tests of normality, multicollinearity, and autocorrelation. If the assumptions are not satisfied then data transformation is carried out. This can be done by conducting a heteroskedasticity test, transforming the data within the estimator, and conducting a spatial heteroskedasticity test using the Breuch Pagan test to determine the optimum bandwidth based on the minimum CV.

This is then followed by the determination of the weighting matrix using the adaptive weighting function of the bisquare kernel, Estimation of the GWPR model parameters. Furthermore, model testing for its suitability, testing the significance of the GWPR model parameters finally the Interpretation and conclusions are carried out [9].

3 Result and Discussion

The following is an overview of the characteristics of the GII data from 2018 to 2020.

From Fig. 1, it can be seen during the period 2018–2020 the GII of each province in Indonesia tends to decrease, although it does not show a significant decrease every year. In 2018 the area that has the lowest GII is Bali Province (0.118). Moreover, the lowest GII from 2019 to 2020 is D.I Yogyakarta Province with values of 0.081 and 0.069 respectively. Meanwhile, the region that has the highest GII from 2018 to 2019 is Southeast Sulawesi province with values of 0.531 and 0.527, respectively. Then in 2020 West Nusa Tenggara Province became the province that had the highest GII of 0.531.

The FEM within estimators is used in this study. The FEM model selection is used by assuming that each observation unit has different characteristics. In order to convince the selection of FEM statistically, the Chow Test and Hausman Test were performed. The obtained results by using the Chow test and the Hausman test are then used to conclude that the FEM model is the right model to be used in research. The results of the Chow test and the Hausman test are given in Table 1.

The classical test is conducted to maintain the accuracy of the formed models. The classical test that was carried out includes normality, multicollinearity, autocorrelation, and heteroskedasticity tests. The normality test using the Kolmogorov-Smirnov test yielded a value of p-value = 0.4045. This is greater than 0.04 thus failing to reject. Moreover, it can be concluded that with a significance of 5% residual data, the FEM panel follows a normal distribution. For the multicollinearity test, a VIF value between

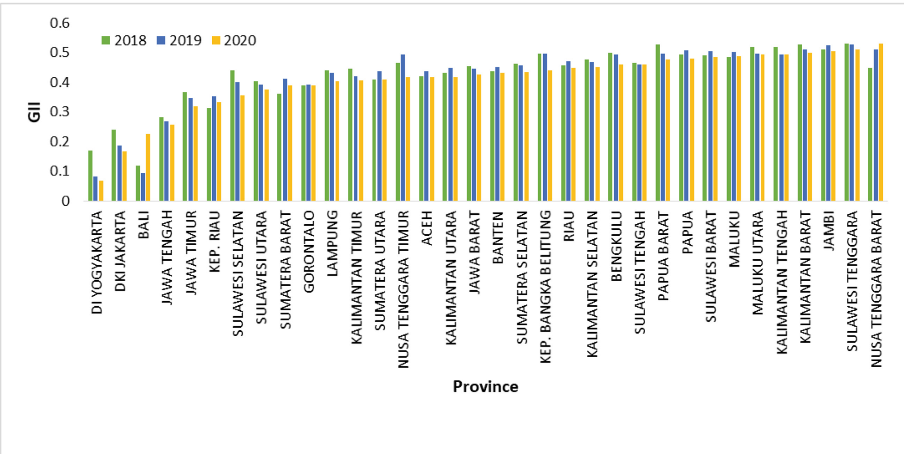

Fig. 1. Gender Inequality Index

Table 1. Results of the Chow Test and Hausman Test

Test	Value	*p-value*	Conclusion
Chow	$F_{count} = 0{,}587$	$2{,}2 \times 10^{-16}$	Reject H_0, using the FEM model
Hausman	$\chi^2{}_{count} = 413{,}42$	$2{,}2 \times 10^{-16}$	Reject H_0, using the FEM model

predictor variables is less than 10 in consequence there is no indication of multicollinearity. It can be said that there is any relationship between predictor variables. One of the tests used to determine whether there has or is no autocorrelation is the Durbin-Watson (DW) test. The DW test value is in the area indicating the decision fails to refuse. This can be said that with a significance level of 5%, there is no autocorrelation occurred. The heteroskedasticity test on the FEM model using the Breusch-Pagan test yielded a BP value of 17.673. This value is greater than 15.507. On the other hand, the p-value is less than 0.05 which indicates the rejection. As a result, it can be concluded that the variance error is not homogeneous or that heteroskedasticity occurred. In order to strengthen the research, the Breusch-Pagan test is also used to determine whether there are any specific influences in the panel data. This is in accordance with research conducted by [9]. The following are the Breusch-Pagan test results for site-specific influences, time-specific influences, or both on FEM model panel data regression (Table 2).

Before estimating the GWPR model, data transformation is firstly carried out in accordance with the concept of within estimators, namely transforming research variables by subtracting them from the corresponding average time series. The spatial heteroskedasticity test using the Breusch-Pagan test is given in Table 3.

Table 3 shows the Breusch-Pagan test result. The BP value is 15.938. This value is greater than 15.507 with the p-value is 0.0432 less than 0.05 which indicates the rejection of the decision. This determines the diversity between observation sites or the spatial

Table 2. Breusch-Pagan Test Results Specific Influence on FEM Model

Breusch-Pagan Test	χ^2 count	*p-value*	Conclusion
Location and time-specific influences	60.825	6.195×10^{-14}	There are location and time-specific influences
Site-specific influence	57.956	2.68×10^{-14}	There are location-specific influences
Time-specific influence	57.956	2.68×10^{-14}	There is a time-specific influence

Table 3. The results of the Breusch-Pagan (BP) test

BP	χ^2 table	*p-value*	Decision
15.938	15.507	0.0432	Reject H_0

heteroskedasticity that occurred. Furthermore, the GWPR analysis is carried out which pays attention to the spatial aspect showing the diversity between observation locations.

The first step in estimating the GWPR model is to account for the Euclidean distance between observation sites based on the coordinate of latitude and longitude. It further determines the optimum bandwidth obtained based on the minimum Cross Validation (CV). The kernel weighting function used is adaptive bisquare with a minimum CV of 0.0274. The optimum bandwidth value varies from province to province. After getting the optimum bandwidth, the next step is to determine the weighting matrix of each i-th location. The spatial weighting matrices can be determined using the adaptive bisquare kernel weighting function. Before determining the weighting matrix, firstly we calculate the Euclidean distance of each location. In the GWPR modeling, the weighting matrix values are the same for each period of the year consequently this produces a weighting matrix of size.

The obtained weighting value will then be used to estimate the parameters of the GWPR model in each province in Indonesia. The difference in parameter values occurs because of the different weighting values between locations of each observation location. This depends on the distance and the generated bandwidth value. The difference in models in the GWPR is due to the presence of spatial elements that describe the homogeneity of the region or differences in geographical location.

Furthermore, the model conformity tests are performed to determine the difference between the panel data regression model and the GWPR. The suitability testing of the GWPR model by using parameters is conducted simultaneously. The results of the model suitability testing is given in Table 4.

From Table 4, we obtained the p-value (0.0003) is lower than 0.05 indicating the rejection of the decision. This can be said that with a significance level of 5% the FEM and GWPR panel data regression models are significantly different. In order to find out the best model between the FEM and GWPR panel data regression models, we can determine from the AIC value and the resulting coefficient of determination (R^2). The AIC and R^2 values from the FEM and GWPR are presented in Table 5.

The GWPR model parameters significance test is carried out to determine the parameters that affect the response variables significantly. The results of the GWPR model

Table 4. Model Conformity Test Results

F count	F table	*p-value*
2.33	0.6816	0.0003

Table 5. Model Conformity Test Results of the FEM and GWPR

Models	AIC	R^2
FEM	−5,0785	0,3838
GWPR	−600,4796	0,7397

parameters significance test for each province in Indonesia at a real level of 5% from a provincial grouping based on predictor variables that have a significant effect. Table 6 is the provinces group that is based on the significant predictor variables.

Furthermore, the spread of variables of significant effect can be presented in the form of spatial maps. Figure 2 shows a map of the provinces group in Indonesia by variables that have a significant effect on the GII.

Table 6. Group of the province by the significance of predictor variables

Group	Significant Variables	Provinces
1	X_4, X_5	Aceh; North Sumatera; West Sumatera; Riau; Jambi; South Sumatera; Bengkulu; Bangka Belitung Islands; Riau Islands; DKI Jakarta; Banten; West Kalimantan; North Sulawesi; Central Sulawesi; Southeast Sulawesi; Gorontalo; Maluku dan North Maluku.
2	X_3, X_4, X_7	Central Java; D.I Yogyakarta; East Java; Bali; Central Kalimantan; South Kalimantan; South Sulawesi dan West Sulawesi.
3	X_3, X_4, X_7, X_8	WNT dan ENT
4	X_5	North Kalimantan; West Papua dan Papua.
5	X_2	West Java
6	X_4	East Kalimantan
7	X_4, X_5, X_6	Lampung

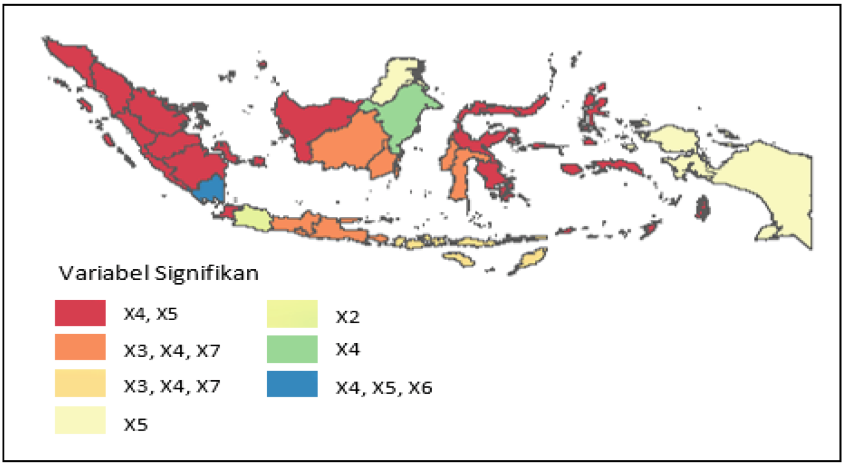

Fig. 2. Significant Variable Grouping Map

The variables that significantly affect the percentage of the GII in Indonesia include the percentage of people aged over 10 years who have no school, the percentage of non-health facilities for childbirth, and the percentage of women sitting in the parliament, HDI, GDI, OUR, and LFPR. This is in accordance with previous research by [5] where the factors that significantly influence the GII include the percentage of people aged over 10 years who have no school, the percentage of non-health facilities for childbirth, and HDI. The following is an example of the GWPR model in Central Sulawesi Province.

$$\hat{y}_{26t} = -0.0296 - 0.0524X_{426t} - 1.7063X_{526t} + \varepsilon_{26t}$$

By the assumption of other variables are ignored, the GII of Central Sulawesi Province in the t-th year will decrease by 0.0296. Any increase of 1% percentage of women sitting in parliament in Central Sulawesi Province in the t-th year by assuming other variables are ignored, will reduce the GII by 0.0524. Moreover, every increase of 1 HDI unit in Central Sulawesi Province in the t-year and other variables are ignored, it will reduce the GII by 1.7063. The GWPR model of Central Sulawesi Province has a determination (R^2) value of 0.7348 or 73.48%. This means that the percentage of women sitting in parliament (X_4) and the Human Development Index (HDI) (X_5) can explain the diversity of the GII in Central Sulawesi Province of 73.48% while 26.52% is explained by factors that exclude in the model.

4 Conclusion

Based on the analysis which has been carried out, we can conclude that the Gender Inequality Index (GII) model in Indonesia by applying the GWPR with adaptive kernel weighting function is better than the FEM of panel data regression, with an R^2 of 73.97%. The GWPR produces different model equations where there are 34 models representing each location. Here is an example of the GWPR model in Central Sulawesi Province $\hat{y}_{26t} = -0,0296 - 0,0524X_{426t} - 1,7063X_{526t} + \varepsilon_{26t}$. Furthermore, several factors which affect the Gender Inequality Index (GII) in Indonesia in the 2018–2020 period includes the percentage of people aged over 10 years who have no school, the percentage of non-health facilities of childbirth, the percentage of women sitting in the parliament, HDI, GDI, OUR and LFPR.

Acknowledgements. I would like to thank my supervisors who have given advice and inclusion in this research.

References

1. Badan Pusat Statistik.: Kajian Penghitungan Indeks Ketimpangan Gender 2021. Badan Pusat Statistik, Jakarta (2020).
2. Gamayanti, N. F., Junaidi.: Pemodelan Hasil Produksi Padi Di Provinsi Sulawesi Tengah Menggunakan Fixed Effect Model (FEM). Barekeng : Jurnal Ilmu Matematika dan Terapan 15 (2), 347–354 (2021).

3. Halim, I. T., Ramadhanty, A. P., Oscarini, D. R., Putra, G. M., Tobing, H. F. B., Nooraeni, R.: Faktor-Faktor yang Mempengaruhi Persentase Penduduk Miskin di Indonesia Tahun 2015-2018 Menggunakan Regresi Data Panel. JURNAL EMACS (Engineering, Mathematics, and Computer Science), 2(2), 55-63 (2020).
4. Hubies, A. V.: Pemberdayaan Perempuan dari Masa ke Masa. IPB Press, Bogor (2010).
5. Marsono : Deteksi Spasial Pada Model Indeks Ketimpangan Gender Indonesia. Buana Gender, 6(1), 49–66 (2021).
6. Martha, S., Yundari., Rizki, S. W., Tamtama, R.: Penerapan Metode Geographically Weighted Panel Regression (GWPR) Pada Kasus Kemiskinan Di Indonesia. Barekeng : Jurnal Ilmu Matematika dan Terapan, 15(2), 241–248 (2021).
7. Rahayu, N. S.: Geographically Weighted Panel Regression Untuk Pemodelan Persentase Penduduk Miskin Di Provinsi Jawa Tengah. Tesis (2017).
8. Sutro, Yundari., Martha, S.: Pemodelan Fixed Effect Geographically Weighted Panel Regression Untuk Indeks Pembangunan Manusia Di Kalimantan Barat. Buletin llmiah Math. Stat. dan Terapannya (Bimaster), 9(3), 413–422 (2020).
9. Wati, DC., Azka, DA., Utami, H.: The Model of Per-Capita Expenditure Figures in Sumatera Selatan uses a Geographically Weighted Panel Regression. Indonesian Journal of Statistics and Its Applications, 5(1), 61-74 (2021).
10. Yu, D.: Exploring Spatiotemporally Varying Regressed Relationships: The Geographically Weighted Panel Regression Analysis. The International Archives of the Photogrammety, Remote Sensing and Spatial Information Sciences, 38(2), 134-139 (2010).