

Geographical Weighted Regression and Mix Geographical Weighted Regression

Fernando Pongoh^{1,*}, I Made Sumertajaya², Muhammad Nur Aidi²

¹Graduate student of Statistics Department, Bogor Agricultural University, Bogor, Indonesia

²Lecturer of Statistics Department, Bogor Agricultural University, Bogor, Indonesia

Abstract Poverty is a fundamental problem that existed and still occurs in Indonesia. This analysis was conducted to describe the factors that influence the welfare low status in sub-districts of North Sulawesi Province. Global regression has obtained influence variables to poverty status are woman leader in household (X_1), Child not in formal education (X_2), physical defect (X_3), chronic disease victim (X_5), self-owned house (X_6), household using electricity (X_8) and household using its own toilet (X_{10}). Geographical weighted regression analysis (GWR) is a locally model, in order to estimates of parameters at each location. The result of GWR has different parameter estimation at each the location. Mix Geographical weighted regression model (MGWR) has combined local and global parameters. As well as the GWR, MGWR model results has give different parameter estimation at each location. GWR Model is the best model to explain the influence of the low welfare status in sub-districts of North Sulawesi Province which has the smallest value of MSE and AIC and the largest of R^2 between Global Regression and MGWR.

Keywords Global Regression, Geographical Weighted Regression (GWR), Mix Geographical Weigthed Regression (MGWR)

1. Introduction

Poverty is a fundamental problem that existed and still occurs in Indonesian. The Economic crisis in 1998 gave a big problem to the national economy – increasing the number of the poors from 34.01 million (1996) to 49.50 million (1998) [1].

Geographical weighted regression model with the function of normal kernel was the best model in explaining the average expenditure per capita per month in Jember sub-district [2]. Geographical weighted regression (GWR) was something that brought the framework of simple regression model to become weighted regression model [3]. The approach used was point approach in which every parameter value was measured in every geography location, therefore, every geography location spot had different regression coefficient value. In fact, in every situation not all regression coefficients from GWR model varied spaciouly.

The level of spatial diversity in some coefficients could not be significant, or it could be ignored. Consequently, GWR model was developed to the mix geographical weighted regression (MGWR) in which it was the combination of linear regression model and GWR model,

therefore, MGWR model could produce parameter estimation that had global parameter estimation, and other parameter that had local in accordance with its observation location. MGWR was better used to analyze the percentage of poor household in Mojokerto sub-district in 2008 [4].

Tim nasional percepatan penanggulangan kemiskinan (TNP2K) is an institution established as the coordination organization of cross sector and cross functionary in central level to accelerate poverty tackling that has published welfare indicators in Indonesia provinces and integrated basic data in every sub-district, city or province. Based on the data basis, the analysis of the low welfare status in sub-districts of North Sulawesi Province would be performed by using GWR and MGWR analysis.

2. Research Method

The data used in this research were from integrated data basis to Social Protection Program July 2012 in 159 sub-districts of North Sulawesi Province. Moreover, latitude and longitude data were also used in every sub-district.

Response variable (Y) is a low welfare status (%), while the predict variables were:

- X1: Woman leader in household (%)
- X2: Child not in formal education (soul)
- X3: Physical defect (%)
- X4: Patients with chronic diseases (%)
- X5: Chronic disease victim (soul)

* Corresponding author:

fernando.pongoh@gmail.com (Fernando Pongoh)

Published online at <http://journal.sapub.org/statistics>

Copyright © 2015 Scientific & Academic Publishing. All Rights Reserved

X6: self-owned house (%)

X7: Households using protected water as water sources of drinking (%)

X8: Household using electricity (%)

X9: Households using gas cooking fuel/LPG/electricity (%)

X10: Household using its own toilet (%)

X11: Households using the final disposal of feces tank / SPAL (%)

3. Results

3.1. Global Regression

Global regression analysis obtained the influence variable of 5% level significant, i.e., woman leader in household (X_1), Child not in formal education (X_2), physical defect (X_3), chronic disease victim (X_5), self-owned house (X_6), household using electricity (X_8) and household using its own toilet (X_{10}), shown in Table 1. The variable Patients with chronic diseases (X_4), households using protected water as water sources of drinking (X_7), households using gas cooking fuel/LPG/electricity (X_9) and households using the final disposal of feces tank / SPAL (X_{11}) are not significant in that level. The value of R^2 and adjusted- R^2 are 0.969 and 0.967. In anova examination, it was obtained F value of 668.182 and F Table in the real level of 5% - the value of 2.05. This showed that regression model established gave significant influence towards the status data of low welfare in sub-districts of North Sulawesi Province.

Table 1. Parameter Estimation of Global Regression

Variable	Coefficients	Std.Error	p-value
Intercept	0.064598	0.658	0.922
X_1	1.072838	0.235	0.000
X_2	-0.01429	0.003	0.000
X_3	3.921913	1.094	0.000
X_5	0.003374	0.001	0.000
X_6	0.732582	0.030	0.000
X_8	0.244163	0.027	0.000
X_{10}	0.079014	0.035	0.026

The examination of the assumption of the remaining normality that used kolmogorov smirnov test in the variable of low welfare status (Y) produced KS value of 0.49 with the value of p 0.97 at the real level of 5%. It showed that the spread of the remaining data was normal. The examination of multicollinearity was performed by considering the value of Variance Inflation Factor (VIF). The result of VIF showed that the variables had no strong correlation. It indicated that the VIF value in every variable was under 10. The examination of the homogeneity of variation was done exploratively, i.e. by considering the plot between y-assumption and the rest. The plot showed that the data formed the funnel-like pattern that indicated the variation

homogeneity violation.

The test of spatial variation that used Breusch-Pagan (BP) test yielded the BP value of 24.843 with the value-p of 0.000809 at the real level of 5%. This showed that there was spatial variation at the status data of low welfare in sub-district of North Sulawesi Province. The variation of spatial indicated that every sub-district in North Sulawesi Province had its own characteristics, therefore, the local approach was needed to modelize and overcome the variation that happened in low welfare status.

3.2. Geographically Weighted Regression (GWR)

At the analysis of the first step of GWR which was performed determined the value of optimum bandwidth with iteration to get minimum cross-validation. The result of iteration yielded minimum cross-validation of 5.119 with bandwidth value of 42.855. This bandwidth value described that the distance limit in an area where the distance was under 42.855 km gave more influence compared to that of above 42.855 km. The weighted matrix was formed by using the function of adaptive bisquare: $W_{ij} = (1 - d_{ij}^2/42.855^2)^2$.

Table 2. Summary Estimation Parameter of GWR Model

Variable	Coeff			Std
	Min	Max	Mean	
Intercept	29.298	36.893	32.127	1.875
X_1	-0.699	2.104	0.620	0.637
X_2	-3.234	1.824	-0.251	1.125
X_3	-0.760	3.790	1.140	0.983
X_5	-2.304	4.667	0.344	1.482
X_6	-0.947	14.445	7.689	5.307
X_8	-0.507	13.923	5.613	4.757
X_{10}	-0.850	3.316	1.071	0.912

The coefficients parameter of variable X_1 , X_2 , X_3 , X_5 , X_6 , X_8 and X_{10} was shown in Table 2. Model of GWR got the value of R^2 0.994 and adjusted- R^2 0.989. The examination of GWR anova resulted in the value of F-calculation of 6.086, meanwhile the value of table F that was in the real level of 5% was 1.48. This indicated that there was a significant difference between global regression model and GWR model. GWR model give different estimation parameter in every sub-district, Table 3 shown estimation parameter at Posigadan sub-district and Maesa sub-district.

The significant variable ($\alpha = 5\%$) at Posigadan sub-district are physical defect (X_3), self-owned house (X_6), household using electricity (X_8) and household using its own toilet (X_{10}). GWR model $Y = 34.977 + 2.219X_3 + 12.345X_6 + 1.711X_{10}$ and R^2 0.986.

The other, Maesa sub-district significant variable ($\alpha = 5\%$) are woman leader in household (X_1), self-owned house (X_6), household using electricity (X_8) and household using its own toilet (X_{10}). GWR model $Y = 35.208 + 1.678X_1 + 8.468X_6 + 2.597X_8 + 2.988X_{10}$ and R^2 0.983.

Table 3. Estimation Parameter of GWR Model at Posigadan and Maesa

Variable	Posigadan			Maesa		
	Coef	T-test	R ² lokal	Coef	T-test	R ² lokal
Intercept	34.977	58.87	0.986	35.208	60.45	0.983
X ₁	-0.698	-1.26		1.678	1.83*	
X ₂	0.088	0.13		-0.666	-0.78	
X ₃	2.219	2.57*		1.763	1.42	
X ₅	0.823	0.97		0.167	0.18	
X ₆	12.345	13.56*		8.468	8.78*	
X ₈	1.711	2.02*		2.597	2.56*	
X ₁₀	1.873	3.50*		2.988	4.50*	

3.3. Mix Geographically Weighted Regression (MGWR)

The initial step in MGWR that determined the local and global variables yielded the variables of Child not in formal education (X₂), physical defect (X₃), chronic disease victim (X₅), self-owned house (X₆), household using electricity (X₈) and household using its own toilet (X₁₀) as the local variable; and woman leader in household (X₁) variable as the global variable. The value of Bandwith obtained the value of 42.855 with the cross-validation of 4.956. The value of bandwith described that the distance limit in an area in which the distance was under 42.855 km gave more influence compared to that of above 42.855 km. The weighted matrix was formed by using the function of adaptive bisquare: $W_{ij} = (1 - d_{ij}^2/42.855^2)^2$.

The local parameter and global parameter estimation was shown in Table 4. MGWR model obtained the value of R² 0.993 and adjusted-R² 0.989. The examination of anova indicated the value of F-calculation of 6.546, meanwhile the value of table F in the real level of 5% was 1.50. This showed that there was significant difference between the global regression model and MGWR model. MGWR model give different estimation parameter in every sub-district, Table 5 shown estimation parameter at Posigadan sub-district and Maesa sub-district.

Table 4. Summary Estimation Parameter of MGWR Model

Variable	Local				Global	
	Coeff			Std	Coeff	Std
	Min	Max	Mean			
Intercept	29.555	36.898	32.121	1.863	-	-
X ₁	-	-	-	-	0.583	0.275
X ₂	-3.357	1.969	-0.279	1.139	-	-
X ₃	-0.859	3.723	1.123	0.907	-	-
X ₅	-2.279	4.722	0.459	1.456	-	-
X ₆	-0.896	14.552	7.798	5.393	-	-
X ₈	-0.612	13.549	5.556	4.682	-	-
X ₁₀	-1.113	3.440	1.014	0.919	-	-

The significant variable (α = 5%) at Posigadan sub-district are woman leader in household (X₁), physical defect (X₃), self-owned house (X₆), household using electricity (X₈) and

household using its own toilet (X₁₀). MGWR model $Y = 34.456 + 0.583X_1 + 1.486X_3 + 12.643X_6 + 1.230X_{10}$ and R² 0.986.

The other, Maesa sub-district significant variable (α = 5%) are woman leader in household (X₁), self-owned house (X₆), household using electricity (X₈) and household using its own toilet (X₁₀). MGWR model $Y = 35.241 + 0.583X_1 + 8.440X_6 + 3.474 X_8 + 3.141 X_{10}$ and R² 0.982.

Table 5. Summary Estimation Parameter of MGWR Model

Variable	Posigadan			Maesa		
	Coef	T-test	R ² lokal	Coef	T-test	R ² lokal
Intercept	34.456	57.92	0.984	35.241	60.26	0.982
X ₁	0.583	2.12*		0.583	2.12*	
X ₂	0.054	0.08		-1.082	-1.32	
X ₃	1.486	1.74*		1.833	1.46	
X ₅	0.854	0.99		0.679	0.76	
X ₆	12.643	13.69*		8.440	8.65*	
X ₈	1.672	1.96*		3.474	5.48*	
X ₁₀	1.230	2.47*		3.141	4.78*	

3.4. Compare of Regression Global, GWR and MGWR

The global regression model, GWR model and MGWR model were compared with the value of MSE, AIC and R² from every model. The model with the smallest MSE and AIC values and the biggest value of R₂ was the best model. Table 6 showed that the smallest MSE and AIC values were 2.329 and 609.372 at GWR model. The biggest value of R² was also owned by GWR model of 0.984. Therefore, it could be concluded that GWR model was the best model in explaining the influence towards the low welfare status in sub-district of North Sulawesi Province.

Table 6. Compare of Regression, GWR and MGWR

	Model	MSE	R ²	AIC
Regression		4868.253	0.969	776.777
GWR	Adaptive Bisquare	2.329	0.994	609.372
MGWR	Adaptive Bisquare	2.376	0.993	612.824

4. Conclusions

The influencing factors of low welfare status in sub-districts of North Sulawesi Province using global regression analysis are woman leader in household (X₁), the number of Child not in formal education (X₂), physical defect (X₃), chronic disease victim (X₅), self-owned house (X₆), household using electricity (X₈) and household using its own toilet (X₁₀).

In each sub-district could be obtained the different influencing factors of low welfare using GWR analysis. In the case, influence factors at posigadan sub-district is not equal with influence factors at Maesa sub-districts. The influence factors at Posigadan sub-district are physical defect

(X_3), self-owned house (X_6), household using electricity (X_8) and household using its own toilet (X_{10}). At Maesa sub-district, the influence factors are woman leader in household (X_1), self-owned house (X_6), household using electricity (X_8) and household using its own toilet (X_{10}).

REFERENCES

- [1] BPS, Statistik Indonesia: Statistical Year Book of Indonesia, 2014.
- [2] Rahmawati R, "Model Regresi Terboboti Geografis dengan Pembobot Kernel Normal dan Kernel Kuadrat Ganda untuk Data Kemiskinan: kasus 35 desa atau kelurahan di kabupaten jember," thesis. Institut Pertanian Bogor, Bogor, Indonesia. 2010.
- [3] Fotheringham AS, Brunsdon C, Charlton M, Geographically Weighted Regression, the Analysis of Spatially Varying Relationships, John Wiley and Sons, LTD. 2002.
- [4] Purhadi dan Yasin H, 2012, Mixed Geographically Weighted Regression Model, Case Study: the Percentage of Poor Households in Mojokerto 2008, European Journal of Scientific Research, 69(2), 188-196.