International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 18 (2017) pp. 7392-7396

Explorative Spatial Analysis of Coastal Community Incomes in Setiu Wetlands: Geographically Weighted Regression

Z. Syerrina [1], A.R. Naeim [2], L. Muhamad Safiih [3] and Z. Nuredayu [4]

[1,2,3,4]*School of Informatics & Applied Mathematics, Universiti Malaysia Terengganu, 21300 Kuala Terengganu, Terengganu, Malaysia.*

[1]*ORCID iD: 0000-0002-6456-2219*

Abstract

Regression models can be used to provide a statistically measurable level for each unit change in the independent variables that affect income. In this study, geographically weighted regression (GWR) modellings were extended to evaluate explorative spatial analysis regarding the relationship between coastal community incomes in Setiu Wetlands in Terengganu, Malaysia. We used questionnaire data collected around the coastal line in Setiu Wetlands. The GWR model was compared with a global regression model using several statistical criteria. The GWR model reveals a relationship between coastal community incomes and education, employment status and occupation as fishermen.

Keywords: geographically weighted regression, Setiu Wetlands, coastal community incomes

INTRODUCTION

Regression models can be used to provide a statistically measurable level for each unit change in the independent variables that affect income. However, if the global model applies, it does not take into account spatial effects. When including spatial effects, doing so provides more accurate coefficient estimates and standard errors for variables of interest. By taking into account spatial effects, each study location will have a unique coefficient estimate, which is also known as the local estimate.

Thus, geographically weighted regression (GWR) is introduced as one of the new methods that are able to examine spatial risk factors for different problems. This statistical method adjusts the frame of the global to the local regression model, which enables regression parameter estimation for every single point of the spatial unit (Fotheringham, Brunsdon, & Charlton, 2002). Furthermore, the relationship between the independent variable and dependent variable in a GWR model is different for all regions (Leung, Mei, & Zhang, 2000).

Several previous studies that examined household income factors were conducted in various countries. Chasco et al. (2007) performed studies on household disposable income across Spanish provinces by using GWR. They used global regression analysis to model provincial household income distribution and they extend their studies for spatial effects by re-specifying the basic model using GWR. For their result, the bisquare adaptive GWR estimates found that there is a trade-off between education-qualification and employment activity for explaining per capita household income. As an example, there are provinces with higher education that coexists with lower employment activity.

Meanwhile, Shelnut and Yao (2005) developed an analytical measure of income inequality at the county level for the state of Arkansas in the United States. In Northwest and Central portions of Arkansas, income is unequally distributed, and this pattern is positively correlated with economic growth. In addition, many differences in factors between metropolitan statistical areas (MSAs) and non-MSAs such as educational attainment, sector composition, demographic distribution, and job-market conditions were used to explain this inequality pattern. By using ordinary the least-squares (OLS) model, they found that more jobs are offered at MSAs rather than micropolitan areas due to the quality and quantity of jobs available for commuters.

Kam et al. (2005) investigated the spatial variation of rural poverty in Bangladesh by estimating the household income for 1 million census households. They mapped poverty indices for 415 rural subdistricts. Their results revealed distinct areas with a high poverty incidence that corresponds with ecologically depressed areas. By using geographically weighted regression, the results show that other livelihood-influencing factors such as education, accessibility and services are significantly correlated with poverty.

The study regarding income in Malaysia is still limited because it only focuses on income inequality, income distribution, and poverty. For example, Hasyim (1998) focused on trends existing in income distribution for overall, urban-rural and interethnic inequality for Peninsular Malaysia, Sabah and Sarawak. He also tests for any systematic relationship between inequality and economic development. Nearly 2 decades previously, Kusnic and Davanco (1980) also conducted a study on income distribution in Peninsular Malaysia. They investigated the sensitivity of estimates of income levels, interethnic or urban /rural differences and income inequality.

They used data from 1976 – 1977 from the Malaysian Family Life Survey on a sample of over 1,000 households in Peninsular Malaysia.

From the previous study, Malaysia still lacks research concerned with the factors that affect household income, especially from lower income groups. Under this study, the fishery sector, which is from lower income group, will be examined, and the factors affecting their income will be identified. Fisheries have a long history of being the economic backbone of coastal and island communities. Communities in coastal areas are mostly dependent on marine resources related to the fisheries. Currently, living standards are rising and felt by all segments of society, including fishing sector communities. To help them to improve their income, further study on income from fisheries will be carried out. Results will help the fishermen take necessary action to improve their standard of living.

For that reason, regression analysis is used for this study to test the relationship between weekly income of the coastal community in the coastal line around Setiu Wetlands (Jajaran Merang, Jajaran Setiu and Jajaran Kuala Besut) and factors that are believed to be statistically significant. The objective of this study is to analyse the spatial relationship between income, education, employment status and a job as a fisherman for the study area by using Geographically Weighted Regression (GWR). For comparison purposes, OLS regression was used to measure the relationship between incomes with factors that influence community income. Besides that, this study using regression techniques to identify key characteristics of households affecting both household income and household education and employment background.

Figure 1: Jajaran Kuala Besut

Figure 2: Jajaran Merang and Setiu

MATERIALS AND METHODS

Study Area

The study was conducted in Setiu Wetlands (Jajaran Merang, Jajaran Setiu and Jajaran Kuala Besut) in Terengganu Malaysia. Setiu Wetlands is a part of the Setiu River Basin and the large Setiu-Chalok-Bari-Merang basin wetlands complex. The Setiu Wetland is in the district of Setiu in the state of Terengganu. The size of Setiu Wetlands is 23,000 hectares and is the largest natural wetlands in the East Coast region of Peninsular Malaysia. This study area also continues until Kuala Besut, which is located north of district Setiu. At this area, there are many coastal communities living because of the proximity to the sea.

Data Collection

In order to complete this study, we needed primary data that can be obtained through questionnaires. The data was collected in the 64-village area Jajaran Merang, jajaran Setiu and Jajaran Kuala Besut. The questionnaire collection effort resulted in approximately 1025 responses from 64 villages. The data was collected from house to house in each village. In the questionnaire, the respondents answered questions about their incomes and social, environment and economic status and demographics.

METHOD AND MODEL DESCRIPTION

Objectives

The objectives for this study area analyses the relationship between income and education, employment status and job status by using global regression and Geographically Weighted

Regression (GWR). Then, analysis looks at the factor that affects the incomes reported in the community.

Method

The relationship between coastal community incomes and factors related to incomes can be analysed by comparing the results of global regression and GWR modelling. If the GWR model is more reliable than the global regression model, then regression model between four variables is spatially different for each subdistrict. Subsequently, the global regression model describes the constant relationship for all study regions. The regression modelling was first processed by using the ordinary least squares (OLS) method to examine the global relationship between coastal community incomes (Y) and incomes and social, environment and economic status and demographics. The regression model for this study is:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon \tag{1}$$

Where y is the coastal community income; X_1 is the education; X_2 is employment status, and X_3 is job as a fisherman. β_0 is OLS intercept coefficient; $\beta_1, \beta_2, \beta_3, \beta_4$ is OLS slope coefficient, and ε is an error.

Then, we built a model using Geographically Weighted Regression (GWR) first developed by Brunsdon et al. (1996). One advantage using this technique is that it is based on the traditional regression framework with which most readers will be familiar. Another advantage is that it incorporates the local spatial relationship into the regression framework in an intuitive and explicit manner following the example of Fotheringham et al. (2002). The GWR model we employed was the model that contains temporally correlated error:

$$y = \beta_0(\mu_i, v_i) + \beta_1(\mu_i, v_i) X_1 + \beta_2(\mu_i, v_i) X_2 + \beta_3(\mu_i, v_i) X_3 + \varepsilon \tag{2}$$

Where $\beta_0(\mu_i, v_i)$ and $\beta_1(\mu_i, v_i)$ are GWR coefficients in subdistrict i; location point of subdistrict i is defined by latitude and longitude coordinates (μ_i, v_i) and ε is an error for the GWR equations.

In the GWR modelling, the first step is to define the latitude and longitude coordinates (μ_i, v_i) for each subdistrict. These geographical coordinates were used to specify the Euclidean distance between observed data in subdistrict i in a village and observed data in subdistrict j in the village:

$$d = \sqrt{(\mu_i - \mu_j)^2 + (v_i - v_j)^2} \tag{3}$$

The distance in equation (3) is termed the fundamental background in weighting the data to estimate the GWR model. When the distance is closer between subdistricts, then the larger weight of data will be during the parameter estimation. The weighting was conducted using the Adaptive Gaussian function:

$$\psi = \exp(-d_{ij}^2 / b_{i(k)}^2) \tag{4}$$

Where $b > 0$ is bandwidth constant in which its designation was done by the cross-validation method (Fotheringham et al., 2002). Meanwhile, d_{ij} is the Euclidean distance between i and j and $b_{i(k)}$ is an adaptive bandwidth size defined as the k^{th} nearest neighbourhood distance (Nakaya, T., 2014). The Gaussian function is an exponential function that gives weight 1 for data where the parameter of subdistrict is being estimated and weight with an undeviating decrease value for data in other subdistricts. The weight continues to reduce as the distances among subdistricts increases. The Gaussian function was employed in forming weighted matrix:

$$W(\mu_i, v_i) = diag(\psi_1, \psi_{2,\ldots\ldots\ldots\ldots}\psi_i) \tag{5}$$

where $0 \le \psi_i \le 1$ is weight data for subdistrict to estimate the parameter. Every observed data has one weighted matrix $W(\mu_i, v_i)$ in estimating the parameter. By the algebraic matrix approach, the estimation of parameter $\hat{\beta}(\mu_i, v_i) = (\beta_0(\mu_i, v_i), \beta_1(\mu_i, v_i))^T$ in subdistrict i by means of weighted least squares (WLS) method is express in:

$$\hat{\beta}(\mu_i, v_i) = [X^T W(\mu_i, v_i) X]^{-1} X^T W(\mu_i, v_i) Y \tag{6}$$

(Fotheringham et al., 2002). All observed data in a certain subdistrict has the same estimated parameter. This result is due to the reality that the data weighting only engages the distances among the subdistricts.

The fitness for the global model and gwr model is measured by the Akaike Information Criterion (AIC) and considers the different degrees of freedom. The best model will provide a trade-off between goodness-of-fit and degree of freedom to minimise the AIC. According to Hurvich et al. (1998), the AIC for GWR is defined as:

$$AIC_c = 2n \log_e(\hat{\sigma}) + n \log_e(2\pi) + n\left\{\frac{n + tr(S)}{n - 2 - tr(S)}\right\} \tag{7}$$

where n is the sample size, $\hat{\sigma}$ is the estimated standard deviation error term, and tr(S) denotes the trace of the hat matrix, which is a function of the bandwidth.

RESULT

Result for OLS Regression

The OLS estimation on the transformed data by employing equation (1) presents the global regression model as follows in equation (1):

$$y = 21575.59 + 139.18x_1 + 83.043x_2 + (-196.10)x_3 + \varepsilon$$

where y is income and x_1 is education, x_2 is employment, x_3 is job and ε is an error term for standard deviation.

Table 1: Test of regression coefficient

Variable	Coefficient	Standard Error of Coefficient	T-statistics	P-value	VIF
Intercept	21575.588	2205.679944	9.781831	.000	
Edu	139.177	79.377263	1.753363	.085	1.2
Emp	83.043	36.492317	2.275624	.026	1.1
Job	-196.103	79.058 257	-2.480482	.016	1.1

Table 1 presents the summary of the estimated coefficients for regression specification. The VIF value for these variables is significantly low, indicating that it does not have a multicollinearity problem. All the independent variables had a significant effect on income at the 10% level of significance. The coefficient of education and employment rate had a positive sign, which explains that when education and employment increased, community income also will increase. Meanwhile, the job as fisherman rate indicates the negative sign that describes the fisherman job had an opposite effect on income. Even though the percentage of fishermen is high among the community, it could not be the main contribution to their earnings due to a low standard of living. This required them to find other work in order to increase side-income.

Table 2: The summary of the model estimated

	R^2	Adjusted R^2	F-statistic	P-value	Durbin-Watson
Regression	0.195	0.155	4.841	0.004	0.004

In general, the model describes approximately 15.5% of the variation in the dependent variable, as shown in Table 2. Thus, the results obtained show that the estimated model does not fit the data fairly well with a low adjusted R^2 and low significant F-statistics values. The low value of adjusted R^2 might be caused by other factors not included in this study. Thus, these factors might influence incomes for the fishermen in the study area. The model does not have a serious serial correlation problem because the Durbin-Watson result was 1.742.

GWR MODEL

Local Model (Adaptive Gaussian)

The GWR estimation for the income can be divided into 64 villages. For example,

Village 5

$$y = 21351.35 + 127.53edu + 87.95emp + 213.23job - 10600.8$$

Village 61

$$y = 23519.49 + 97.15edu + 49.88emp + 204.64job - 5647.66$$

In this study, the Gaussian distance function is used to determine the optimum bandwidth. Based on the output obtained by the minimum value of CV given by the optimum bandwidth of 48.89. Bandwidth is used to determine the weight to create the GWR model. Table 2 summarises the standardised local coefficients as it varies depending on the geographical coordinate of the village $u=(x_i,y_i)$.

Table 2: Summary statistics for varying (local) coefficients (n=64)

Coefficient	Minimum	1st Quartile	Median	3rd Quartile	Maximum	Standard Deviation
Intercept	20726.636	21214.385	23449.824	23537.798	23647.845	1195.979
Edu	96.606	97.001	97.601	126.974	157.001	17.004
Empl	47.304	49.442	51.234	90.781	98.632	20.258
Job	-217.916	-209.724	-204.996	-203.17	-196.492	5.512

Table 3: Comparison for global model and local model

Model	Global Model	Local Model (GWR)
AICc	1324.056	1323.735
R square	0.195	0.238
Adjusted R square	0.140	0.157

The summary result of GWR in Table 3 shows that the local model performs better than the global model because the adjusted R-square improved from 14.0% to 15.7% in explaining the variation in the dependent variable and a lower Akaike Information Criterion (AIC) (from 1324.056 to 1323.735). Also, the ability of GWR is to explore the spatial variation of explanatory variables in the model, where the coefficients explanatory variables may vary significantly over geographical space.

Table 4: GWR ANOVA Table

Source	SS	DF	MS	F
Global Residuals	3045589918.382	60.000		
GWR Improvement	163363373.613	2.066	79064923.694	
GWR Residuals	2882226544.768	57.934	49750338.922	1.589234

From Table 4, GWR can provide F-test value, which is used to test whether spatial variation exists in the relationship under study (Brudson et al., 1996). Moreover, it shows the improvement of GWR models and describes the relationship significantly better than the ordinary least square model. This outcome can be demonstrated via an ANOVA test implemented. For the GWR model, the F-value is 1.589. The F-value suggests that the GWR model has a significant improvement over the global model in determining the relationship between incomes and other determinant factors.

CONCLUSION AND DISCUSSION

In this paper, an OLS model and GWR model were studied using the community income in Setiu Weetland. Other variables, such as education, employment and job were included in this study for further investigation of relationships among them. Only employment and job influence fisherman income where employment shows a positive relationship with income and job as fisherman shows a negative relationship with income. The strength of the relationship regarding goodness-of-fit statistics R^2 is higher by using GWR from 0.195 (OLS) to 0.238 (GWR). While the adjusted R^2 value of the global OLS is 0.14, the value for local GWR model is 0.15 across the study area. It can be concluded that GWR can perform better than OLS because it can provide further information for different places.

In terms of future research, we suggest other factors that might influence fisherman's income. This might increase the value of R^2 if the factors selected influence fishermen's incomes. Furthermore, only the global model is able to evaluate the effect of global collinearity between explanatory variables while local collinearity remains are investigated. Other issues on how to reduce the impact of global and local collinearity on tests must be further studied.

ACKNOWLEDGEMENTS

This research was supported by the Niche Research Grant Scheme (NRGS) UMT - P1(R) (Second Phase)-Vote No.: 53131/30, Ministry of Higher Education Malaysia.

REFERENCES

[1] Brunsdon, C., Fotheringham, A. S. & Charlton, M. (1996). Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis, 28*, 281-298.

[2] Brunsdon, C., Fotheringham, A. S. & Charlton, M. (1999). Some Notes on Parametric Significance Tests for Geographically Weighted Regression. *Journal of Regiona lScience, 9*, 497-524.

[3] Chasco, C., García, I., & Vicéns, J. (2007). Modeling spatial variations in household disposable income with Geographically Weighted Regression.

[4] Fotheringham, A. S., Brunsdon, C. & Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships.* Sussex, UK: John Wiley.

[5] Hashim, S. M. (1998). Income inequality and poverty in Malaysia. Rowman & Littlefield.

[6] Kam, Suan-Pheng, Mahabub Hossain, Manik Lal Bose and Lorena S. Villano. 2005. Spatial patterns of rural poverty and their relationship with welfare-influencing factors in Bangladesh. Food Policy. 30(5-6) :551-567.

[7] Kusnic, M. W., & DaVanzo, J. (1980). Income inequality and the definition of income: The case of Malaysia. Santa Monica: Rand.

[8] Leung, Y., Mei, C. L., & Zhang, W. X. (2000). Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environment and Planning A, 32*, 9-32. http://dx.doi.org/10.1068/a3162.

[9] Lin, C. H., & Wen, T. H. (2011). Using geographically weighted regression (GWR) to explore spatial varying relationships of immature mosquitoes and human densities with the incidence of dengue. *International Journal of Environmental Research and Public Health, 8*, 2798-2815. http://dx.doi.org/10.3390/ijerph8072798.

[10] Shelnutt, J. P., & Yao, V. W. (2005). A spatial analysis of income inequality in Arkansas at the county level: evidence from tax and commuting data. Federal Reserve Bank of St. Louis Regional Economic Development, 1(1), 52-65.