PENERBIT UMT
Menjana Khazanah Ilmuwan

Journal of Mathematical Sciences and Informatics
Journal Homepage: https://journal.umt.edu.my/index.php/jmsi
eISSN: 2948-3697
DOI: http://doi.org/10.46754/jmsi.2021.12.005

JMSI
Journal of Mathematical Sciences and Informatics

A STATISTICAL INFERENCE ANALYSIS ON CRIME RATES IN PENINSULAR MALAYSIA USING GEOGRAPHICAL WEIGHTED REGRESSION

NUR EDAYU ZAINI[1*], SYERRINA ZAKARIA[1], NUZLINDA ABDUL RAHMAN[2] AND WAN SALIHA WAN ALWI[1]

[1]Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu, 21030 Kuala Terengganu, Terengganu, Malaysia; nuredayu23@gmail.com, syerrina@umt.edu.my, salihaalwi@gmail.com. [2]School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM Penang, Malaysia; nudzlinda@usm.my

**Corresponding author: nuredayu23@gmail.com*

ARTICLE INFO

Article History:
Received June 2021
Accepted July 2021
Available online
December 2021

Keywords:
Geographical Weighted Regression (GWR);
Multiple Linear Regression (MLR);
Violence crime rate;
Statistical inference;

ABSTRACT

Geographical Weighted Regression (GWR) is used to improve decision-making in spatial analysis. Instead of the Ordinary Least Square (OLS) regression method that gives a single estimated parameter, the GWR method can provide unique estimated parameters in each location. This study aims to conduct a formal statistical inferential framework on the violent crime rate using the GWR. This analysis discovers the geographical distribution and pattern of criminal cases in Peninsular Malaysia using the average crime rates from 2000-2009, with focus on on violent crime. The comparison of OLS regression, known as Multiple Linear Regression (MLR) with the GWR method, was done to show that GWR was the best model. The GWR output suggests that about 30% of districts showed a significant correlation between violent crime and non-citizen rates. These findings contradict the result from the MLR model, also known global model. The global model could not create any other connection to explain the lack of parameter-location correspondence. Finally, the importance of local relationships in crime studies is necessary to understand the actual crime rate.

2020 Mathematics Subject Classification:
2020 Computing Classification Codes:
©Penerbit UMT

INTRODUCTION

Crime analysis is the statistical investigation of criminal events, including burglaries, rapes, robberies, assaults, and murders [1]. rom 2000 to 2009 reported crimes increased by about 25.3% [2]. Crimes will have socioeconomic impacts, including loss of property and life. Consequently, the authorities have to step up measures to bring down crime.

Many studies in various countries have done to investigate the causes of crime [3,4,5]. study by [3] found links between fines and violent crimes in the United States at city and county level. They found that fines showed a positive linear relationship with crime rates at the state level. Due to complexity, violence and penalties have a robust significant association at the state level but not at the local level. In another study, routine activity theory was used to investigate the increased population in Toronto associated with the number of criminal activities [4]. It shows that there is a positive connection between crime and population, especially in public areas. Another study use the ordinary least square regression (OLS) and GWR to explore disorder effect and collective effect on crime in the United States [5]. This study shows that the impact of disorder and efficacy varies across an entire neighbourhood. They conclude that crime can be understood from opportunity, highlighting the role of informal social control in communities.

Crime studies are limited in Malaysia, especially in spatial analysis. According to [6], a few reported analyses of Malaysia's crime rate had been carried out. These studies show that the three main groups of deterrence and punishment, criminality, or economic or incentives for crime and social demographics – all of which affect crime in Malaysia. Over a decade, Malaysia's crime rate has risen by 165% (on average, 16.5% a year). On average, there were approximately 685 cases daily, or about 30 cases per hour [6].

Meanwhile, [7] uses multiple regression analysis to investigate the connection between unemployment rate, income per capita, and population density towards violent crime. The data was collected between 1985 and 2014 in Malaysia. This study focused on finding the factors that affect crime by discounting demographics. The result showed that violent crimes are expected to increase for every unit increase in unemployment and population density.

The current research on crime using a statistical model gives the whole picture of the factors that affect crime. A detailed study on the factor affecting crime based on geography should be performed to map the pattern of crime. Meanwhile, they successfully perform specific structural measures to associate crime with spatially varying variable [8]. Besides, these factors should be tested by performing statistical inferences to ensure that the element genuinely affects crime locally. These significant factors might vary according to location.

Previous studies commonly present local parameter estimates and proceed with statistical diagnostics in terms of mapping. A study creates a map by overlapping the local parameter estimates and the local t-values [9]. The GWR software and GIS software were used to improve the traditional mapping strategies. The technique was using masking and transparency techniques on the local t-value layer. This allows only the significant parameter estimate values to be visualised. This shows that the local t-value map does meaningful in-data interpretation, and it encourages the reader to pay attention to the main areas. However, this study just covers mapping , and it does not continue with hypothesis testing.

Recent work provides an overview of the crime trends and their patterns in Malaysia, using the space-time Normal Mixture Models [2]. Most of the crimes arose in urban areas, such as Selangor, Kuala Lumpur, and Johor. The latest study analyses the spatial relationships in Peninsular Malaysia between criminal cases and social, environmental, and economic status through the Geographically Weighted Regression (GWR) [10]. The GWR was shown to be better suited than the OLS model. For this study, the extension will be conducted on statistical inferences to test the significance of affecting variables. The violent crime rates and several independent variables data set will be used, and the analyses are extended until the hypothesis testing process by using the t-test value. Thus, the main objective of this study is to perform statistical inferences on significant variables. The data used in this study is explained in section 1.1.

Data

The violent crime data was obtained from The Royal Malaysian Police (RMP), and independent data was obtained from the Department of Statistics Malaysia (DOSM). The crime rate root is divided into violent crime and property crime. The index of violence included murder, armed group robbery, weapon-free robbery, robbery with firearms, rape, and accidentally causing injuries. The data for the 82 districts in Peninsular Malaysia is the average data from 2000 to 2009 [11]. Violent crime rate (VR) is defined as a standardised value by the total number of crimes. The violent rate formula is:

$$VR = \frac{V}{T} \tag{1}$$

where V is the number of violent crime cases, and T is the total crime. The dependent variable is the violent crime rate. Meanwhile, the independent variables are:

x_1: the unemployment rate;

x_2: the population density;

x_3: the non-citizen rate;

x_4: the basic household index;

x_5: the middle household index;

x_6: the single mother rate.

The unemployment rate refers to the percentage of unemployed people divided by the number of active labors of the same category. The population density is the average number of people living per square mile/km. The non-citizen's rate is measured by the percentage of non-citizens divided by the total population. BASIC and MIDDLE indexes were household-based socioeconomic indexes built using the factor analysis method [12]. The number of single mothers refers to women aged 15 and above who are widowed, divorced or separated permanently.

METHODOLOGY

In this study, two regression models were employed: Multiple Linear Regression (MLR) and GWR model. Regression analysis is a direct approximation of the relationship between independent and dependent variables. The simplest method of regression analysis, called simple linear regression, contains a unique number of independent variables. Multiple linear regression is an application of regression relating to several independent variables used as a global model in the study.

$$y_i = \beta_0 + \sum_{k=1}^{M} \beta_k x_{ki} + u_i \tag{2}$$

where y_i is the observation of the dependent variable y and observation indexed by i=1,..., N, β_0, represent intercept of the regression model, $\beta_\kappa(k=1,2,...,M)$ represents the regression coefficients, $x_{\kappa i}$ is the independent variable of i^{th} value of x_κ and u_i are the independent normally distributed error terms with zero mean and constant variance [13].

The second method is the Geographical Weighted Regression introduced by [14]. It is an extension method from the multiple linear regression model of equation (1), which considers the local variations. The coefficients in the model rather than global estimates are specific to the location i. This model includes local spatial relationships between dependent/response variables and independent/explanatory variables into the regression framework.

$$y_i = \beta_0(\mu_i, v_i) + \sum_{j=1}^{m} \beta_j(\mu_i, v_i)x_{ij} + \varepsilon_i \tag{3}$$

Equation (3) indicates $\beta_0(\mu_i, v_i)$ is a constant function and $\beta_j(\mu_i, v_i)$ that represents as a GWR coefficient in subdistrict i, location point of subdistrict i is defined by latitude and longitude coordinates (μ_i, v_i) and ε is a random error term at the location. In the GWR modeling, the weighting system was conducted through four types of functions: Fixed gaussian, fixed bi-square, adaptive bi-square, and adaptive gaussian. The first step is to define the latitude and longitude coordinates (μ_i, v_i), or each subdistrict. These geographical coordinates were used to specify the Euclidean distance between the observed data in subdistrict and subdistrict in the village:

$$d = \sqrt{(\mu_i - \mu_j)^2 + (v_i - v_j)^2} \tag{4}$$

The distance in equation (4) is a fundamental background in weighting the data to estimate the GWR model. When the distance is closer between subdistricts, the larger weight of data will be during the parameter estimation. The weighting was conducted by employing four types of function, as shown in Table 1.

Table 1: The different types of weighted functions

Type of Weighted	Functions	
Fixed Gaussian	$\psi_{ij} = \exp(-d_{ij}^2 / b^2)$	(5)
Fixed bi-square	$\psi_{ij} = \begin{cases} \left(1 - d_{ij}^2 / b^2\right)^2 & d_{ij} < b \\ 0 & d_{ij} > b \end{cases}$	(6)
Adaptive bi-square	$\psi_{ij} = \begin{cases} \left(1 - d_{ij}^2 / b_{i(k)}\right)^2 & d_{ij} < b_{i(k)} \\ 0 & d_{ij} > b_{i(k)} \end{cases}$	(7)
Adaptive Gaussian	$\psi_{ij} = \exp(-d_{ij}^2 / b_{i(k)}^2)$	(8)

where $b > 0$ is bandwidth constant in which its designation was done by cross-validation method [14], and Ψ_{ij} is the weight value of observation at location j for estimating the coefficient at a location i. Meanwhile, d_{ij} is the Euclidean distance between i and j, and $b_{i(k)}$ is an adaptive bandwidth size defined as the kth nearest neighbourhood distance [15]. A Gaussian function is an exponential function that gives weight 1 for the data, where the subdistrict parameter is estimated and weighted with an undeviating decrease value for the data in other subdistricts. It must be noted that the weight continues to reduce as the distances among subdistricts increase.

The Gaussian function was employed in forming a weighted matrix:

$$W(\mu_i, v_i) = diag\ (\psi_1, \psi_2, \ldots\ldots, \psi_i) \tag{9}$$

where $0 \leq \psi_i \leq 1$ is weight data for the subdistrict to estimate the parameter. For every observed data, it has one weighted matrix $W_{ij} = (u_i, v_i)$ in estimating the parameter. By the algebraic matrix approach, the estimation of a parameter $\beta(u_i, v_i) = (\hat{\beta}_0(u_i, v_i),\ \beta_1(u_i, v_i))^T$ in subdistrict i by means of weighted least squares (WLS) method is expressed in [16]:

$$\beta(\mu_i, v_i) = \left[X^T W(\mu_i, v_i) X\right]^{-1} X^T W(\mu_i, v_i) Y \tag{10}$$

All observed data in a subdistrict has the same estimated parameter. It is due to the reality that the data weighting only engages the distances among the subdistricts. Further, the analysis on hypothesis testing will be investigated in section 2.1.

Hypothesis Testing Procedure on Individual GWR Coefficient

According to [16], finding the significant parameters affecting the response variable is practically the same as multiple regression analysis but slightly different in terms of output. A technique proposed to deal with various numbers of variables ‘p’ and district ‘n’.

The hypothesis testing framework can be retrieved as follows:

$$H_0: \hat{\beta}_{ik}(u_i, v_i) = 0 \tag{11}$$

$$H_1: \hat{\beta}_{ik}(u_i, v_i) \neq 0 \tag{12}$$

for $k = 1,2,\ldots,p$ and $i = 1,2,\ldots,n$

The partial testing of GWR is solved by obtaining the estimated parameter of $\hat{\beta}_{ik}(u_i, v_i)$, and standard error is $SE(\hat{\beta}_{ik}(u_i, v_i))$. The $SE(\hat{\beta}_{ik}(u_i, v_i))$ was used to test the level of significance of each location by using a t-test.

The t-test statistics can be measured by:

$$t_{calculate} = \frac{\hat{\beta}_{ik}(u_i, v_i)}{SE(\hat{\beta}_{ik}(u_i, v_i))} \quad (13)$$

where $t_{calculate}$ will follow t distribution with a significance level α, then reject H_0 if the value of $t_{calculate} > t_{tabulated}$. Meanwhile, the t distribution with n-p degree of freedom (df) for $\alpha/2$ level of significance. This study assumes that α = 5% thus shows that the $t_{tabulated}$ for 76 df, the critical $t_{\alpha/2} = t_{0.025} = \pm 1.992$. Finally, this t-value will be compared and analysed in the next part. In terms of t-value, it can be both positive and negative numbers.

RESULTS AND ANALYSIS

This section explains the output for both models; section 3.1 presents ML regression, section 3.2 presents GW regression, and section 3.3 compares both models.

Results for Multiple Linear Regression for Mean Violent Crime Rate

Table 2 shows the output of ML model. Violent crime rate parameters were analysed using ordinary least square estimation. The variable number is six, and the number of observations is 82, using multiple linear regression analysis.

Table 2: Summary of MLR coefficient value

	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
Intercept	-2.298000	0.345600	-6.651000	0.000000***
X_1	0.002568	0.067680	0.038000	0.969840
X_2	-0.000047	0.000064	-0.731000	0.467130
X_3	**7.384000**	**2.477000**	**2.982000**	**0.00387 *****
X_4	0.005396	0.220500	0.024000	0.980540
X_5	0.188600	0.317100	0.595000	0.553780
X_6	0.173600	0.319200	0.544000	0.588180

*** significant at 0.05 significance level

After retaining the equation (2), the OLS estimation on the transformed data was present the global regression model as follows:

$$y = -2.298 + 0.0026X_1 - 0.00005X_2 + 7.384X_3 + 0.0054X_4 + 0.189X_5 + 0.174X_6 \quad (7)$$

The X_3 (non-citizen) was registered as the significant variable with a positive sign from the multiple linear regression output, as shown in definition 3.1. Thus, this indicates that the number of violent crimes increase as the number of non-citizen increase. Meanwhile, the population density registers a negative sign, which shows that the number of violent crimes increase in a low-density population. Other variables that record a positive sign are unemployment rate, basic household index, middle household index, and single mother rate.

Results for GW Regression

The results of the GW regression will be divided into sections. Section 3.2.1 explains the GW kernel result, section 3.2.2 explains the coefficient value, section 3.2.3 presents the t-value of the GW model, and section 3.2.3 describes the hypothesis testing process.

The GW Kernel Results

GW regression provides a type of kernel available for different kinds of data. According to Table 3, the adaptive bi-square is the best kernel function with a small number AICc of 11.56 and a large number adjusted R square of 52% for violent crime data. Next, the best kernel function for violent crime is adaptive bi-square.

Table 3: Summary of GWR output for different kernel means the violent

	Fixed Gaussian	**Fixed Bi-square**	**Adaptive Gaussian**	**Adaptive Bi-square**
AICc	16.38655	14.67835	37.07832	11.56079
Adjusted R square	0.5443192	0.5754308	0.3126717	0.5215349

The Coefficient Values of the GWR Model

The GWR coefficient value for each variable is formed, as shown in Table 4. Each variable has its range for the GWR coefficient from minimum to maximum.

Table 4: Summary of GWR coefficient value at data points (n=82) means the violent

	Min.	**1st Qu.**	**Median**	**3rd Qu.**	**Max.**
Intercept	-3.030600	-2.496400	-2.240500	-1.847900	-1.071200
X_1	-0.238540	0.000728	0.044901	0.123420	0.177700
X_2	-0.000546	-0.000104	0.000010	0.000045	0.000100
$\mathbf{X_3}$	**-0.771210**	**1.471300**	**4.326900**	**6.720500**	**9.288000**
X_4	-1.138400	-0.357940	-0.214920	0.380920	1.034500
X_5	-0.666520	-0.122010	0.119290	0.374560	0.700200
X_6	-0.734140	-0.189930	-0.114810	0.130520	0.574600

The non-citizen rate was selected for further analysis based on the MLR model since it is the only significant variable out of six. In this paper, this variable will be further discussed in terms of its relationship to violent crime. Figure 1 represents the coefficient estimated value or also known as . The coefficient estimated non-citizen and violent crime rate between -0.771206 and 9.2880, and there is a positive relationship in the study area, as shown in Figure 1. Moreover, a few negative parameters estimated in the local model contradict the MLR model for the non-citizen. A negative parameter indicates that a high non-citizen rate in some areas influences the violent crime rate. Afterwards, the statistical test for spatial variation is required to provide evidence for the importance of exploring spatiality in the statistical model. It is usually necessary to map the parameter estimated for visual inspection, and perhaps the critical findings are to obtain a t-values pattern for parameter estimates. Figure 1 shows the distribution of estimated parameters in the entire study area.

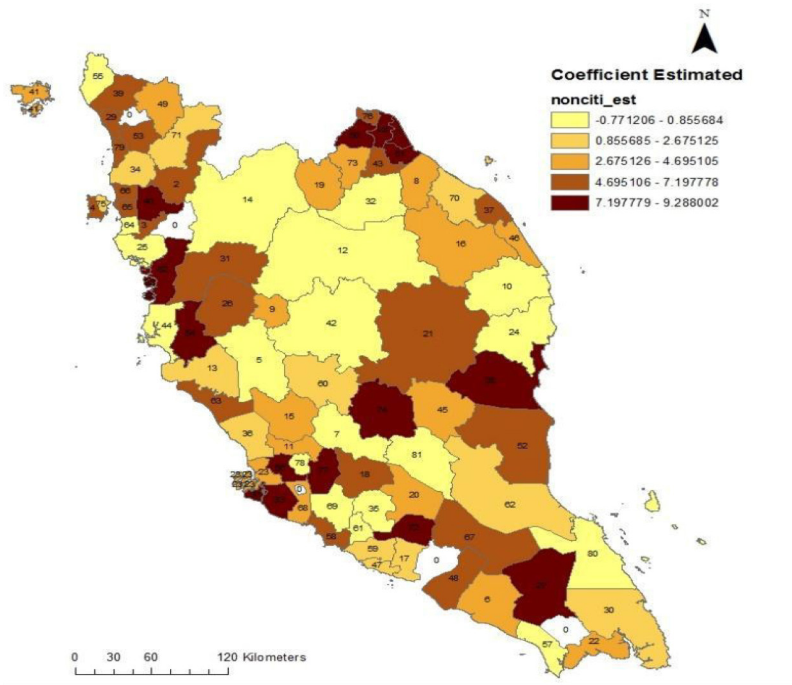

Figure 1: The coefficient estimated values on violent crime rate and non-citizen rate

Local T-Value for GWR Model

The local t-value of the non-citizen rate is generated for each district, as shown in Figure 2. Based on Figure 2, the darkest colours indicate a region with statistically significant positive values while the t-test value is greater than +1.992. The light colour indicates a region with statistically significant negative values while the t-test value is smaller than -1.992 [9]. Thus, it reveals that a highly significant positive is mainly clustered in urban areas like Kota Bharu and Kuantan. Then, industrial areas, such as Kluang, Muar, Ulu Langat, Temerlong, Kulim, Sabak Bernam, and Perak Tengah also indicate a highly significant positive due to their concentration of non-citizen worker. The cluster of negative values is located in rural areas, such as Gua Musang, Lipis, Mersing, and Ulu Perak. About 5% of districts registered a negative parameter, which was far smaller than a positive parameter. This is an exciting finding. It could be hypothesised that there are significant positive relationships concerning violent actions, and the non-citizen rate displays significant variation of 5% across Malaysia's cities. These results suggest that the global model may not encapsulate some other relationship that may explain the lack of correspondence between the parameters in rural areas. Other variables might have an influence on violent crime in rural areas. In conclusion, it is essential to emphasise the significance of local relationships in crime studies.

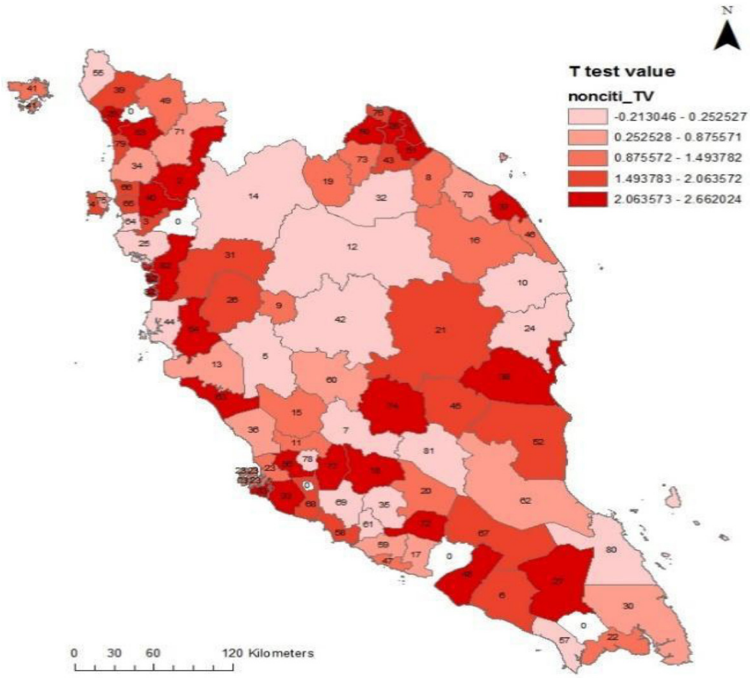

Figure 2: The t-test values on violent crime rate and non-citizen rate

Hypothesis Testing for the GWR Model

Hypothesis testing was conducted on the non-citizen rate to get a better picture of significant accuracy. The results in Figure 3 is obtained from equation (3). Based on Figure 3, decision "0" reject H_0 and decision "1" means not to reject H_0. About 30% of districts have rejected the null hypothesis, and 70% have not rejected the null hypothesis. This explains why only 30% of districts showed that the non-citizen rate influenced the violent crime rate while the other 70% did not show this relationship. The districts that reject the null hypothesis are Kota Bharu, Kuala Terengganu, Pasir Mas, Kuala Selangor, Temerloh, Kluang, Tampin, Petaling, Sabak Bernam, and Perak Tengah. Even though a smaller percentage of districts rejected the null hypothesis, this positively indicates that the non-citizen rate significantly affects the violent crime rate in urban and industrial areas. This result is coherent with the finding by [17]. There is a large amount of news concerning foreign nationals in Malaysia. This section concludes by examining how the number of violent crimes could be diminished if the number of non-citizens decreased.

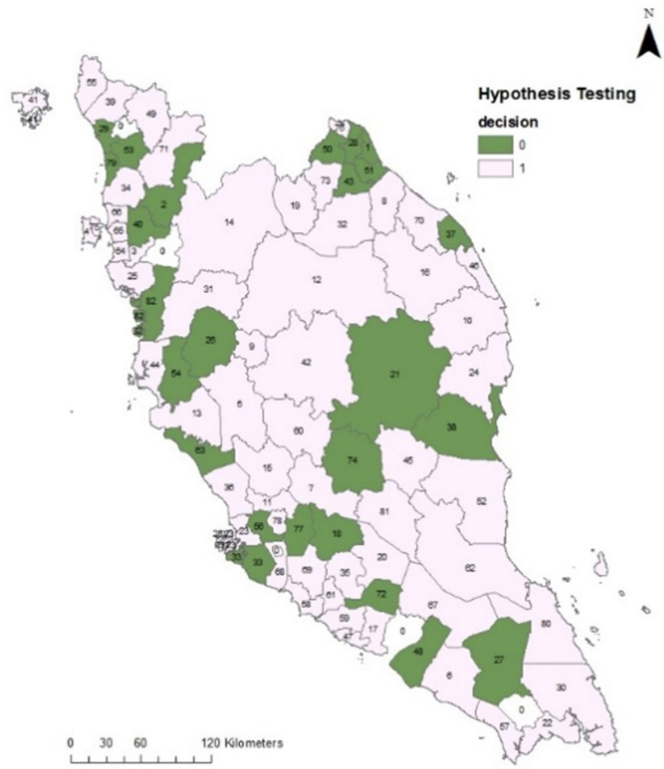

Figure 3: The hypothesis testing on violent crime rate and non-citizen rate

Comparison ***between ML Regression and GW Regression***

Table 5 shows the AIC value and Adjusted R-square for both MLR and GWR models.

Table 5: Comparison of the global model and the local model

	Global Model (Ml Regression)	**Local Model (Gw Regression)**
AICc	49.5694	11.56079
Adjusted R square	0.1167	0.5215349

The result from Table 5 shows that the Akaike information criterion value from the local model is 11.56079. This value is much smaller than the global regression model. The GWR model is more appropriate in describing the relationship of violent crime rate with independent variables compared to MLR model. The adjusted R square results from the GWR model ($\bar{R}^2 > 52.15\%$) obtained are slightly higher than the MLR model. The result is parallel with the result obtained by [11]. These findings support the conclusion that the local model performs better than the global model.

DISCUSSION AND CONCLUSION

This paper is an extension of [11]. It comprises work concerned with applying an MLR model and GWR model using the violent crime rate from 2000 until 2009. Under this study, several variables were included: the unemployment rate, population density, non-citizen rate, primary household index, middle household index, and single mom rate. It was found that the GWR model outperforms the MLR model with a small number of the Akaike information criterion (11.5607) and a higher adjusted R square (52.15%). In this study, further analysis involving the GWR model includes mapping the coefficient determination, a t-test value, and a decision for hypothesis testing. Moreover, the non-citizen variable was significant in the violent crime rate. It can be concluded that the GWR model provides a significant positive relationship between violence measures and the non-citizen rate at a 5% level. It shows that some districts showed that the non-citizens rate significantly affect the violent crime rate.

Several studies also found that non-citizens affect the violent crime rate. he United States also has a problem with rising immigration that led to an increased number of non-citizens under the supervision of the criminal justice system [18]. In Malaysia, the arrival of foreign nationals coincides with more crime being featured on the news [17]. Even though the crime rate is easily affected by the number of non-citizens, the cause and the reason behind the problem could be further investigated. Research by [19] in the Netherlands shared that a lack of conventional life opportunities is the leading cause of many crimes among migrants. Meanwhile, immigrants in Southern California were examine based on new approaches, i.e., racial/ethnic category, region of origin, and where they are located. This approach gives a better explanation of the neighbourhood-crime relationship rather than the traditional approach [20].

In terms of statistical inferences analysis, GWR encourages the analyst to verify their model using t-test value in the R programme. It is possible to conduct hypothesis testing using the GWR model. However, when it comes to hypothesis testing, it mainly involves simulation analysis, and the application analysis is still limited. In the future, more studies could be done on spatial statistical inferences procedure, and more techniques are introduced to overcome the existence of type one error rate. The type one error should be avoided, and it can be done by adjusting the t-test value when testing the significance of local parameter estimates in GWR. This action could avoid excessive false discoveries in the modeling [21].

This article establishes the extension of GWR and demonstrates how it can identify and measure the different kernel bandwidths. Spatial studies give essential guidelines to policymakers, authorities, and researchers on the significant factor affecting crime.

CONFLICT OF INTERESTS

There is no conflict and interest.

ACKNOWLEDGEMENTS

The crime data were obtained from the Royal Malaysia Police (RMP) and the other data obtained from DOSM.

ID	DISTRICT
0	ALOR GAJAH
1	BACHOK
2	BALING
3	BANDAR BAHARU
4	BARAT DAYA
5	BATANG PADANG
6	BATU PAHAT
7	BENTONG
81	BERA
8	BESUT
9	CAMERON HIGHLANDS
10	DUNGUN
11	GOMBAK
12	GUA MUSANG
13	HILIR PERAK
16	HULU TERENGGANU
17	JASIN
18	JELEBU
19	JELI
20	JEMPOL
21	JERANTUT
22	JOHOR BAHRU
24	KEMAMAN
25	KERIAN
26	KINTA
23	KLANG
27	KLUANG
28	KOTA BHARU
29	KOTA SETAR
30	KOTA TINGGI
31	KUALA KANGSAR

ID	DISTRICT
82	LARUT & MATANG
42	LIPIS
43	MACHANG
44	MANJUNG
45	MARAN
46	MARANG
47	MELAKA TENGAH
80	MERSING
48	MUAR
49	PADANG TERAP
50	PASIR MAS
51	PASIR PUTEH
52	PEKAN
53	PENDANG
54	PERAK TENGAH
55	PERLIS
56	PETALING
57	PONTIAN
58	PORT DICKSON
60	RAUB
61	REMBAU
62	ROMPIN
63	SABAK BERNAM
64	SEBERANG PERAI SELATAN
65	SEBERANG PERAI TENGAH
66	SEBERANG PERAI UTARA
67	SEGAMAT
68	SEPANG
69	SEREMBAN
70	SETIU
71	SIK

32	KUALA KRAI	72	TAMPIN
33	KUALA LANGAT	73	TANAH MERAH
34	KUALA MUDA	74	TEMERLOH
35	KUALA PILAH	75	TIMUR LAUT
36	KUALA SELANGOR	76	TUMPAT
37	KUALA TERENGGANU	77	ULU LANGAT
38	KUANTAN	14	ULU PERAK
39	KUBANG PASU	15	ULU SELANGOR
40	KULIM	78	W.P. KUALA LUMPUR
41	LANGKAWI	79	YAN

REFERENCES

[1] A. S. Sidhu. (2005). An academic and statistical analysis. *Journal of The Kuala Lumpur Royal Malaysia Police College*, *1*(4), 1-28.

[2] S. Zakaria, & N. A. Rahman. (2016). The mapping of spatial patterns of property crime in Malaysia: Normal mixture model approach. *Journal of Business and Social Development*, *4*(1), 1- 11.

[3] S. Smith, K. Gangopadhyay, S. Singh Gill, & S. McIntosh. (2018). A Relationship Between Fines and Violent Crimes. *2018 IEEE Fourth International Conference on Big Data Computing Service and Applications (BigDataService)*, 115-120. doi: 10.1109/BigDataService.2018.00025.

[4] B. Remi. (2018). Routine activity, population(s) and crime: spatial heterogeneity and conflicting propositions about the neighborhood crime-population link. *Applied Geography*, *95*(2017), 79-87. doi: 10.1016/j.apgeog.2018.04.016.

[5] R. E. Stein, J. F. Conley & C. Davis. (2016). The differential impact of physical disorder and collective efficacy: A geographically weighted regression on violent crime. *GeoJournal*, *81*(3), 351–365. doi: 10.1007/s10708-015-9626-6.

[6] C. F. Tang. (2011). An exploration of the dynamic relationship between tourist arrivals, inflation, unemployment, and crime rates in Malaysia. *International Journal of Social Economics*, *38*(1), 50-69. doi: 10.1108/03068291111091963.

[7] M. A. R. A. Rahim. (2016). The dependencies of economic indicators towards violent crime: A case study in Malaysia (Bachelor Dissertation).

[8] M. Cahill & G. Mulligan. (2007). Using geographically weighted regression to explore local crime patterns. *Social Science Computer Review*, *25*(2), 174-193. doi: 10.1177/0894439307298925.

[9] S. A. Matthews & T. C. Yang. (2012). Mapping the results of local statistics: Using geographically weighted regression. *Demographic Research*, *26*, 151.

[10] S. Zakaria, & N. A. Rahman. (2017). Explorative spatial analysis of crime rates among the district of Peninsular Malaysia: Geographically weighted regression. In *Proceedings of the International Conference on Computing, Mathematics and Statistics (iCMS 2015)* (pp.145-156). Singapore: Springer. doi: 10.1007/978-981-10-2772-7_15.

[11] S. Zakaria, & N. A. Rahman. (2015). Analysing the violent crime patterns in Peninsular Malaysia: Exploratory spatial data analysis (ESDA) approach. *Jurnal Teknologi*, *72*(1).

[12] N. A. Rahman., S. Zakaria. (2012). The household-based socioeconomic index for every districts Peninsular Malaysia. In *Proceeding of World Academy of Science, Engineering and Technology*.

[13] A. J. Dobson. (1990). *An introduction to Generalised Linear Models*. London: Chapman and Hall

[14] A. S. Fotheringham, C. Brunsdon & M. Charlton. (2002). Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Wiley. https://www.wiley.com/enmy/Geographically+Weighted+Regression:+The+Analysis+of+Spatially+Varying+Relationships+-p-9780471496168.

[15] T. Nakaya. (2014). Geographically Weighted Regression (GWR) Software, GWR 4.0. ASU GeoDa Center website. https.geodacenter.asu.edu/gwrsoftware.

[16] N. E. Zaini & S. Zakaria. (2018). A statistical analysis for geographical weighted regression, *IOP Conference Series: Earth and Environmental Science, 169*(1). doi: 10.1088/1755-1315/ 169/1/012105.

[17] N. Misman, H. Mohd Adnan, A. S. Firdaus & C. M. Ahmad. (2017). Foreign nationals as offenders and victims in Malaysian crime news. *SHS Web of Conferences*, *33*(28). doi: 10.1051/shsconf/20173300028.

[18] N. Demleitner & J. Sands. (2002). Non-citizen offenders and immigration crimes: New challenges in the federal system. *Federal Sentencing Reporter*, *14*(5), 247-254. doi:10.1525/fsr.2002.14.5.247.

[19] A. Leerkes, G. Engbersen & J. van der Leun. (2012). Crime among irregular immigrants and the influence of internal border control. *Crime Law Soc Change, 58,* 15-38. https://doi.org/10.1007/s10611-012-9367-0.

[20] C. E. Kubrin, J. R. Hipp and Y. A. Kim. (2018). Different than the sum of its parts: Examining the unique impacts of immigrant groups on neighbourhood crime rates, J Quant Criminal, *34*, 1-36. https://doi.org/10.1007/s10940-016-9320-y.

[21] A. R. da Silva & A. S. Fotheringham. (2016). The multiple testing issue in geographically weighted regression. *Geographical Analysis*, *48*(3), 233-247.