VOL. 13, NO. 1, JANUARY 2018 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences
©2006-2018 Asian Research Publishing Network (ARPN). All rights reserved.

www.arpnjournals.com

SIMULATION AND APPLICATION OF THE SPATIAL AUTOREGRESSIVE GEOGRAPHICALLY WEIGHTED REGRESSION MODEL (SAR-GWR)

I. Gede Nyoman Mindra Jaya[1], Budi Nurani Ruchjana[2], Bertho Tantular[1], Zulhanif[1] and Yudhie Andriyana[1]
[1]Statistics Study Program, FMIPA Universitas Padjadjaran, West Java, Indonesia
[2]Mathematics Study Program, FMIPA Universitas Padjadjaran, West Java, Indonesia
E-Mail: mindra@unpad.ac.id

ABSTRACT

The applications of standard regression analysis on spatial data are not appropriate because of the characteristics of the spatial data. Spatial data has two characteristics are spatial dependence and spatial heterogeneity. Modeling spatial data using standard linear regression model leads to bias, inconsistency and inefficient results. Several models have been developed to accommodate the characteristics of the spatial data. However, the models generally developed to solve only one problem of the spatial data (e.g., spatial dependence or spatial heterogeneity). Four kinds of spatial econometrics models usually used to accommodate spatial dependence are spatial autoregressive (SAR), spatial lagged exogenous variables (SLX), spatial error model (SEM), and spatial Durbin model (SDM). To accommodate the spatial heterogeneity, geographically weighted regression (GWR) or varying coefficient model (VCM) is usually used. Our research proposed to develop a new model to accommodate two characteristics of the spatial data. The model is developed based on the combination SAR and GWR model. We call the model as Spatial Autoregressive Geographically Weighted Regression (SAR-GWR). We used Instrumental Variables (IV) approach and Two Stage Least Square (TSLS) to estimate the parameters of the model. We have done the simulation study by mean Monte Carlo simulation to check the bias and efficiency of the parameter estimates. SAR-GWR model provides better results with small bias and Root Mean Square Error (RMSE) rather than standard GWR. We also found that our method relative robust to the multicollinearity problem. We also applied SAR-GWR model in modeling prevalence rate of the Tuberculosis (TB^+) disease in Bandung and we found the healthy house index gives serious effect in increasing theprevalence rate of TB^+ in Bandung City.

Keywords: GWR, SAR, SAR-GWR, spatial dependence, spatial heterogeneity.

1. INTRODUCTION

The regression analysis is usually used to model the relationship between one dependent variable and several of the independent variables [1]. However, for the spatial data, the regression analysis is not appropriate to use because of the characteristics of the spatial data: spatial dependence and spatial heterogeneity ([2], [4]). The application of the standard regression for spatial data lead to bias, inconsistency, and inefficiency of the parameters estimate ([3], [5], [6]).

Several spatial models have been developed to accommodate the characteristics of the spatial data. However, the model generally developed to solve only one characteristic of the spatial data. The problem of spatial dependence is accommodated using spatial econometrics approach. Four kinds of spatial econometrics models usually used to accommodate spatial dependence are spatial autoregressive (SAR), spatial lagged exogenous variables (SLX), spatial error model (SEM), and spatial Durbin model (SDM) or apply those combinations [7]. However, SAR model is more often used because of the ease of computing process and provide the information of the spill over effects [8].

Spatial heterogeneity can be seen in two points view are spatial heteroscedasticity (non-constant of the error variance) and spatial instability structure relationship (varying coefficients) [3]. For the former condition, the standard regression model with robust standard error can be used to be an alternative solution [9]. The last condition needs a different solution. To accommodate spatial instability structure relationship, the Geographical Weighted Regression (GWR) is commonly used [3].

The literature suggests that both the problems could be considered in the spatial model simultaneously [2]. Despite the possibility of addressing spatial heterogeneity and spatial dependence at the same time, only a few studies have been focused on developing modeling of spatial dependence and heterogeneity simultaneously and develop the parameter estimation for this problem.

In this research, we develop a new model to accommodate the spatial dependence and spatial heterogeneity simultaneously. We call the model as Spatial Autoregressive-Geographically Weighted Regression (SAR-GWR).The model is developed based on the combination SAR and GWR models. We also introduce Instrumental Variable (IV) approach with Two-Stage Weighted Least Square (TSWLS) estimation to estimate the parameters of the SAR-GWR model. Instead the analytical approach, we use Monte Carlo simulation study to evaluate the bias and efficiency the parameters estimate of our model. We also check the robustness of our method from the multicollinearity problem. The

©2006-2018 Asian Research Publishing Network (ARPN). All rights reserved.

multicollinearity problem becomes a serious problem for GWR estimation.

We apply SAR-GWR model to build a model of the prevalence rate of Tuberculosis (TB^+) in Bandung city. Bandung city is a capital city of West Java-Indonesia. The $TB^+$is still a major health problem in Bandung city. To control or manage the increasing and spreading of TB^+, we have to identify the potential risk factors of TB^+. Here m SAR-GWR becomes a promising model to present an accurate estimate of the prevalence rate of TB^+ in Bandung City.

We drive SAR-GWR model in section 2. In section 3, we present the simulation study, and section 4discuss an application and the last section 5 we present the conclusions and further research.

2. SAR-GWR MODEL

SAR Model

SAR model can be written as

$$y_i = \rho \sum_{j=1}^{n} w_{ij} y_j + \beta_0 + \sum_{k=1}^{K} \beta_k x_{ki} + \varepsilon_i \quad (1)$$

Where ρ denotes spatial autoregressive coefficient and w_{ij} is an element of spatial weight matrix **W**. The parameters model can be estimated by mean Instrumental Variable approach or Maximum Likelihood (see [5]).

GWR Model

Now, we are going to discuss the basic concept of the GWR. Let we define the GWR model with K exogenous variables X:

$$y_i = \beta_{0i} + \sum_{k=1}^{K} \beta_{ki} x_{ki} + \varepsilon_i \; ; i = 1,2,\ldots,N \quad (2)$$

where y_i is the dependent variable at location i, x_{ki} is the value of k th independent variable at location i, $\beta_{ki}$is the local regression coefficient of k th independent variable at location i, β_{0i} is the intercept parameter at location i, and ε_i is the random error at location i, which may follow an independent normal distribution with zero mean and homogenous varianceσ^2 [10]. The contrast with standard regression model, the GWR allows regression coefficients are varying from location to location. Note that the model (2) has $n(K+1)$ parameters that must be estimated. The numbers of the parameters are larger than the available degrees of freedom based on the given number of observations. The GWR solve the lack of degrees of freedom by develop parameters estimation that represent variation of the parameters over space [11]. The parameters in equation (2) are estimated by a weighted least squares (WLS) procedure, making the weighting system dependent on the location in geographical space and, therefore, allowing local rather than global parameters to be estimated.

$$\hat{\boldsymbol{\beta}}(\boldsymbol{i}) = [\mathbf{X^T \Psi}(\boldsymbol{i})\mathbf{X}]^{-1}\mathbf{X^T \Psi}(\boldsymbol{i})\mathbf{y} \quad (3)$$

where $\boldsymbol{\Psi}(i) = \text{diag}(\boldsymbol{\omega}_1(i), \boldsymbol{\omega}_2(i), \ldots, \boldsymbol{\omega}_n(i))$ is the diagonal weights matrix that varies for any prediction location i, **X** is the matrix exogenous variable with a first column of 1's for intercept, **y** is the vector dependent variables, and $\hat{\boldsymbol{\beta}}(i) = (\hat{\beta}_0(i), \hat{\beta}_1(i), \ldots, \hat{\beta}_K(i))$ is the vector of $K+1$ local regression coefficients at location i, where $\boldsymbol{\Psi}(i)$ is a matrix of weights specific to location i such that observations nearer to i are given greater weight nearer to i are given greater weight than observations further away.

To estimate the parameters model GWR, first, we have to define weights matrix $\boldsymbol{\Psi}(i)$. The weights matrix is specified as a local kernel function that models a distance decay effect from the N calibration locations to the prediction location i. One of the most commonly used kernel functions, and the one used in this analysis, is the bi-square nearest neighbour function:

$$\boldsymbol{\Psi}(i) = \begin{cases} \left[1 - \left(\frac{d_{ij}}{b}\right)^2\right]^2 & if j \in \{N_i\} \\ 0 & if j \notin \{N_i\} \end{cases} \quad (4)$$

where d_{ij} is the distance between the calibration location j and the prediction location i, b is the threshold distance to the Nth nearest neighbor, and the set $\{N_i\}$ contains the observations that are within the distance range of the threshold Nth nearest neighbour (see [3]). The next Kernel function that usually use too is Gaussian function:

$$\Psi(i) = \begin{cases} \exp\left(-0.5\left(\frac{d_{ij}}{b}\right)^2\right) & if j \in \{N_i\} \\ 0 & if j \notin \{N_i\} \end{cases} \quad (5)$$

SAR-GWR Model

The SAR-GWR model can be written as:

$$y_i = \rho_i \sum_{j=1}^{n} w_{ij} y_j + \mathrm{X}_i \beta_i + \varepsilon_i \quad (6)$$

where y_i is a dependent variable at i th location, ρ_i is a $(N \times 1)$ vector autoregressive parameter, w_{ij} is an element of spatial weight matrix **W**, $\mathbf{X}_i = (1, \mathrm{X}_{1i}, \mathrm{X}_{2i}, \ldots, \mathrm{X}_{Ki})$ is a $(1 \times (K+1))$ vector of independent variable, $\boldsymbol{\beta}_i = (\beta_{0i}, \beta_{1i}, \ldots, \beta_{Ki})'$is a $((K+1) \times 1)$ parameters regression coefficients at i th location, and $\varepsilon_i$is a disturbance at i th location which assume independently and identically distribution. Model (6) can be written as a matrix notation using Hadamard Product.

©2006-2018 Asian Research Publishing Network (ARPN). All rights reserved.

$$\boldsymbol{y} = (\mathbf{X} \circ \boldsymbol{\beta})\boldsymbol{\iota}_p + \boldsymbol{\rho} \circ \mathbf{Wy} + \boldsymbol{\varepsilon} \tag{7}$$

where "∘" is the Hadamard Product.

Parameters estimation

The parameters estimation for SAR-GWR is motivated by combining the parameters estimation for SAR and GWR. The efficient estimator for SAR model is an Instrumental Variable approach which can calculate using two-stage least square. While GWR is estimated using weighted least square. WLS estimation for SAR-GWR leads to the bias and inconsistent estimate in case spatial autoregressive structure. In contrast, IV method allows an unbias and consistent estimation of the parameter spatial autoregressive model [6]. To facilitate estimation, we arrange the **X**-values and the values of the endogenous lagged variable $\mathbf{W}_i\mathbf{y}$ in an $(1 \times (p+1))$ vector $\mathbf{Z}$[12]:

$$\mathbf{Z} = [\mathbf{X} \quad \mathbf{Wy}] \tag{8}$$

and the respective regression coefficients in a $(p+1) \times 1$ parameter vector $\boldsymbol{\theta}$:

$$\boldsymbol{\theta} = [\boldsymbol{\beta} \quad \boldsymbol{\rho}] \tag{9}$$

with the definitions (7) and (8) the extended spatial lag model (7) reads

$$\boldsymbol{y} = ([\mathbf{X} \quad \mathbf{Wy}] \circ [\boldsymbol{\beta} \quad \boldsymbol{\rho}])\boldsymbol{\iota}_{p+1} + \boldsymbol{\varepsilon}$$
$$\boldsymbol{y} = (\mathbf{Z} \circ \boldsymbol{\theta})\boldsymbol{\iota}_{p+1} + \boldsymbol{\varepsilon} \tag{10}$$

An IV estimator is based on the assumption that a set of r instruments, $r = p + 1$, where $p = K + 1$ arranged in an instrument matrix **Q**of size $n \times r$, is asymptotically uncorrelated with the disturbances $\boldsymbol{\varepsilon}$[12]

$$plim \frac{1}{n}\mathbf{Q}'\boldsymbol{\varepsilon} = \mathbf{0} \tag{11}$$

but has to correlate with the original variables stored in matrix Z,

$$plim \frac{\mathbf{1}}{\boldsymbol{n}}\mathbf{Q}'\mathbf{Z} = \mathbf{M}_{\mathbf{QZ}} \tag{12}$$

where $\mathrm{M_{QZ}}$ is a finite nonsingular moment matrix. A number of instruments $r = p + 1$ is equal to the number of exploratory variables including vector one and Wy**.** Let's consider at first the numbers of instruments are equal to the number of original exploratory variables including autoregressive component $(\mathbf{X_0}, \mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_K}, \mathbf{Wy})$.Then both matrices **Z**and **Q** have size $(n \times (p+1))$. We may use the X-variables as their own instruments because they are fixed variables. So that we only need one additional instrument variable which does not has correlation in large sample the the a set with the error term and in the same time has correlation with Wy. However, we have to note that the Wydoes not only have one instrumental variable but all variable in **Q**are the instrument for **Wy**because of the information in X-variable will also be used to approximate Wy[12].

In order to derive an IV estimator for $\boldsymbol{\theta}$ for the case that the number of instruments is equal to the number of exploratory variables, we premultiply (10) by (1/ N)**Q**[12]:

$$\frac{\mathbf{1}}{\boldsymbol{N}}\mathbf{Q}'\mathbf{y} = \frac{\mathbf{1}}{\boldsymbol{N}}\mathbf{Q}'(\mathbf{Z} \circ \boldsymbol{\theta})\boldsymbol{\iota}_{p+1} + \frac{\mathbf{1}}{\boldsymbol{N}}\mathbf{Q}'\boldsymbol{\varepsilon} \tag{13}$$

For large N $(N \to \infty)$ the last term goes in probability to zero. In this case equation (13) reduces to

$$\frac{\mathbf{1}}{\boldsymbol{N}}\mathbf{Q}'\mathbf{y} = \frac{\mathbf{1}}{\boldsymbol{N}}\mathbf{Q}'(\mathbf{Z} \circ \boldsymbol{\theta})\boldsymbol{\iota}_{p+1} \tag{14}$$

Since **Q** and **Z** are both matrices of size $(n \times (p+1))$, the matrix product $\mathbf{Q}'\mathbf{Z}$gives a matrix of size $((p+1) \times (p+1))$.

Here we use instrument variable approach with WLS estimation to estimate SAR-GWR parameters for each location as in (2):

$$\widehat{\boldsymbol{\theta}}_i^{\mathbf{IV}} = (\mathbf{Q}'\boldsymbol{\Psi}_i\mathbf{Z})^{-\mathbf{1}}\mathbf{Q}'\boldsymbol{\Psi}_i\mathbf{y} \tag{15}$$

As (14) converges in probability to $\boldsymbol{\theta}_i$
$plim\widehat{\boldsymbol{\theta}}_i^{\mathrm{IV}} = \boldsymbol{\theta}_i$
$\widehat{\boldsymbol{\theta}}_i^{\mathrm{IV}}$ is a consistent estimator for $\boldsymbol{\theta}_i$
Proof:
$\widehat{\boldsymbol{\theta}}_i^{\mathbf{IV}} = (\mathbf{Q}'\boldsymbol{\Psi}_i\mathbf{Z})^{-\mathbf{1}}\mathbf{Q}'\boldsymbol{\Psi}_i\big((\mathbf{Z} \circ \boldsymbol{\theta}) + \boldsymbol{\varepsilon}\big)$
$= \boldsymbol{\theta}_i + (\mathbf{Q}'\boldsymbol{\Psi}_i\mathbf{Z})^{-\mathbf{1}}\mathbf{Q}'\boldsymbol{\Psi}_i\boldsymbol{\varepsilon}$
$plim\widehat{\boldsymbol{\theta}}_i^{\mathbf{IV}} = \boldsymbol{\theta}_i$

It has been proved that IV estimator is consistent. However, the (14) requires that the number of instruments equals the number of regressors. In a practicalaspect, we need more than one instrument variables for increasing the efficiency of the parameters estimated. For this condition, we can use two-stage least squares (TSLS), the estimator.

The idea IV approach for estimating parameters SAR-GWR is creating an instrument variable $\widehat{\mathbf{Z}}$with dimension $(n \times (p+1))$. The first $p$of$\widehat{\mathbf{Z}}$areidentical with the first p columns of **Q**, since the exogenous variables $\boldsymbol{X_0}, \boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_K}$are instruments for themselves. The $(p+1)$th column of $\widehat{\mathbf{Z}}$contains the ultimate instruments $\widehat{\mathbf{Wy}}$for**Wy**which may be constructed by a linear combination of all instruments $\boldsymbol{X_0}, \boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_K}, \mathbf{W}\boldsymbol{X_1}, \ldots, \mathbf{W}\boldsymbol{X_K}$ in **Q**.

So that, the first stage is obtaining $\widehat{\mathbf{Z}} = [\mathbf{X} \quad \widehat{\mathbf{Wy}}]$by regress the $\boldsymbol{Z}$ on **Q** where:

$$\widehat{\mathbf{Z}} = (\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{Z} \tag{16}$$

©2006-2018 Asian Research Publishing Network (ARPN). All rights reserved.

and the second stage is obtaining the parameter estimate SAR-GWR

$$\widehat{\boldsymbol{\theta}}_i^{\mathbf{IV}} = (\widehat{\mathbf{Z}}'\boldsymbol{\Psi}_i\widehat{\mathbf{Z}})^{-1}\widehat{\mathbf{Z}}'\boldsymbol{\Psi}_i\mathbf{y} \tag{17}$$

Estimation steps

a. Define matrix $\mathbf{Z} = [\mathbf{X} \quad \mathbf{Wy}]$
b. Define matrix $\mathbf{Q} = [\mathbf{X} \quad \mathbf{WX}]$
c. Obtain instrumental variable $\widehat{Z}$by regress the$\mathbf{Z}$ on $\mathbf{Q}$ where $\widehat{\mathbf{Z}} = (\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{Z}$
d. Estimate the parameter SAR-GWR using $\widehat{\boldsymbol{\theta}}_i^{\mathrm{IV}} = (\widehat{\mathbf{Z}}'\boldsymbol{\Psi}_i\widehat{\mathbf{Z}})^{-1}\widehat{\mathbf{Z}}'\boldsymbol{\Psi}_i\mathbf{y}$
e. Estimate the $\mathrm{Cov}(\widehat{\boldsymbol{\theta}}_i^{\mathrm{IV}}) = \hat{\sigma}^2(\widehat{\mathbf{Z}}'\widehat{\mathbf{Z}})'$ where $\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} e_i^2}{n-p}$

3. SIMULATION STUDY

We use Monte Carlo simulation to show that the new method we proposed gives the good results. We used a map of Bandung as a spatial reference in this simulation study. Bandung is composed of 30 districts so that we have 30 spatial units. The spatial weight matrix for the SAR model based on the Queen contiguity structure. While the weight matrix of the GWR based on the Gaussian Kernel.

The scheme of the simulation

a) Generate variable bivariate normal X_1 and X_2 with correlation $(X_1, X_2) = R$

$X_1 = \mu_1 + \sigma_1 Z_1$

$X_2 = \mu_2 + \sigma_2\left[RZ_1 + Z_2\sqrt{1-R^2}\right]$

with

$$\mathrm{Z_1 \text{ and } Z_2 \sim N(0,1), R = \{0, 0.3, 0.5, 0.7\}, \mu_1 = \mu_2 = 1, \sigma_1 = \sigma_2 = 2}$$

b) Generate variable dependent y

$$\mathbf{y} = (\mathbf{I} - \boldsymbol{\rho}\mathbf{W})^{-1}(\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \circ \mathbf{X}_1 + \boldsymbol{\beta}_1 \circ \mathbf{X}_1 + \boldsymbol{\beta}_2 \circ \mathbf{X}_2) + (\mathbf{I} - \boldsymbol{\rho}\mathbf{W})^{-1}\boldsymbol{\varepsilon}$$

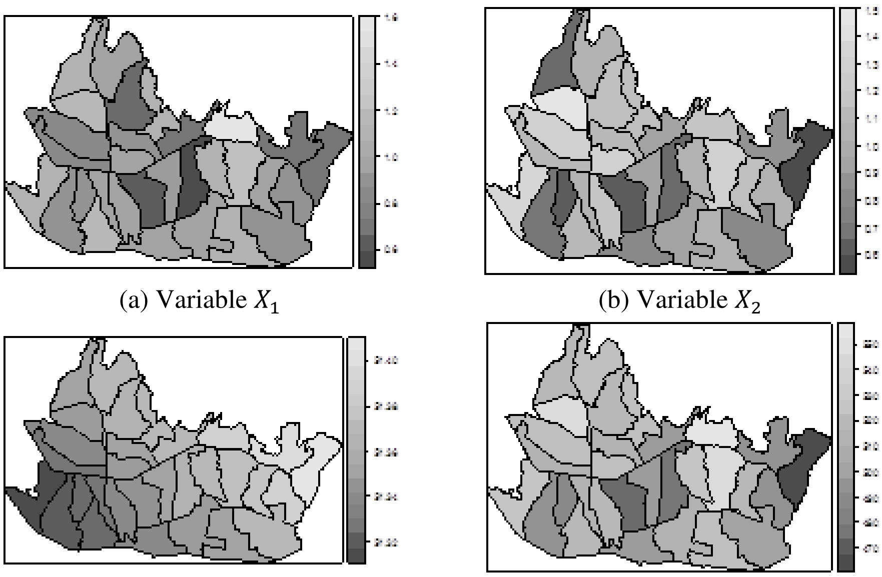
with $\boldsymbol{\beta}_0 = \mathbf{1} + 0.5 \times Lat + 0.5 \times Long$

$\boldsymbol{\beta}_1 = \mathbf{1} + 0.5 \times Lat + 0.5 \times Long$

$\boldsymbol{\beta}_2 = \mathbf{1} + 0.5 \times Lat + 0.5 \times Long$

$\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$

Here we assume that the spatial autoregressive constant over the regions to simplify the simulation process. We set several values of $\rho \in \{0.1, 0.3, 0.5, 0.7\}$ for every simulation. The visualisation values of the variables and the parameters in this simulation can be seen in the map below:



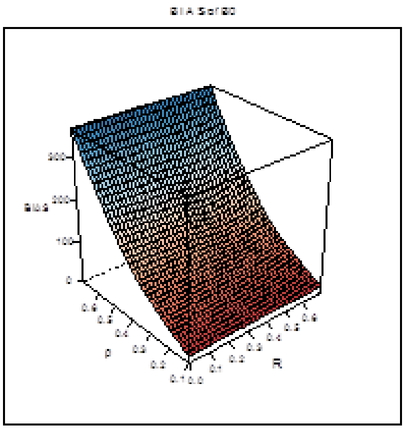
(a) Variable X_1 (b) Variable X_2

(c) Regression Parameter β_0=β_1=β_0 (d) Variable $\boldsymbol{y}$

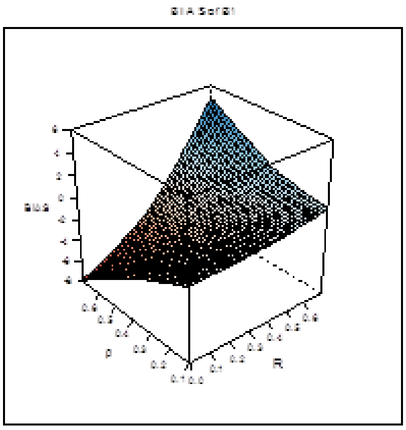
Figure-1. The maps of the simulation value of the variables and parameters.

The Monte Carlo simulation has been done with 1000 time iteration processes. We use surface plot to visualize the Bias and Root Mean Square Error (RMSE) of the parameters estimate

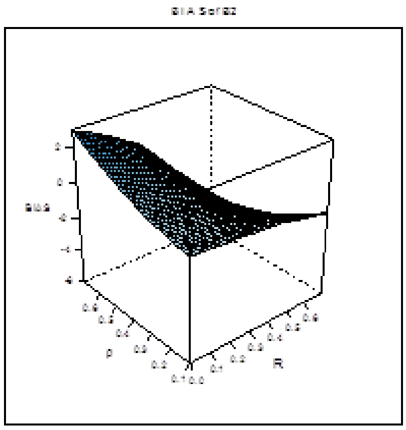
VOL. 13, NO. 1, JANUARY 2018 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences
©2006-2018 Asian Research Publishing Network (ARPN). All rights reserved.
www.arpnjournals.com



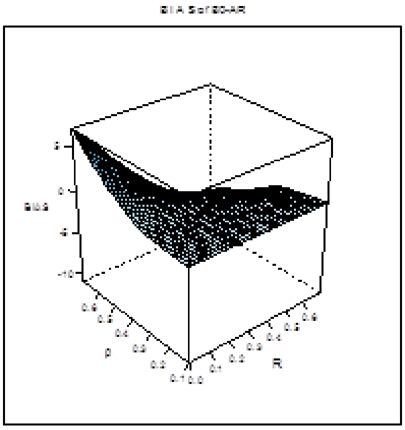
(a) Bias β_0 from GWR



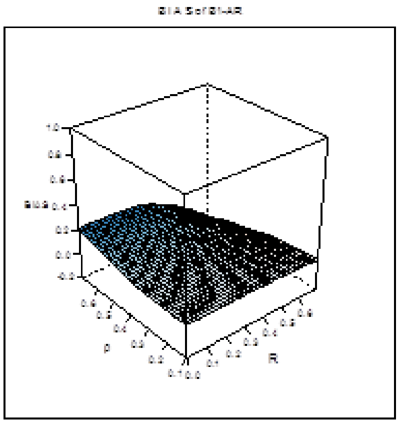
(b) Bias β_1 from GWR



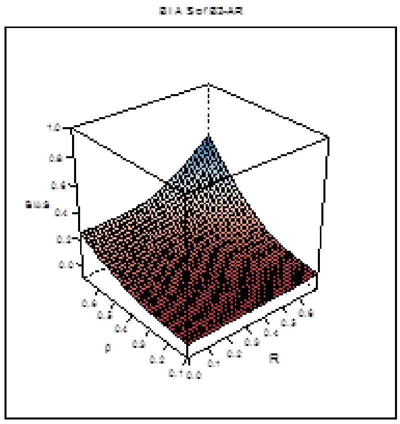
(c) Bias β_2 from GWR



(d) Bias β_0 from SAR-GWR



(e) Bias β_1 from SAR-GWR



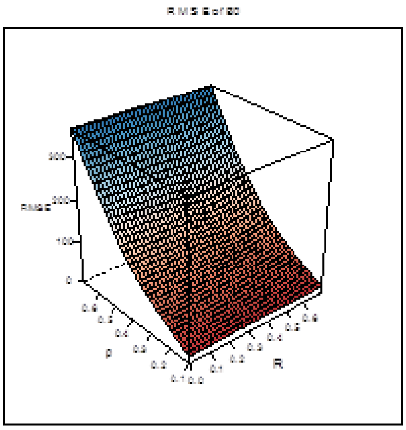
(f) Bias β_0 from SAR-GWR

Figure-2. The Comparison of the Bias of GWR and SAR GWR.

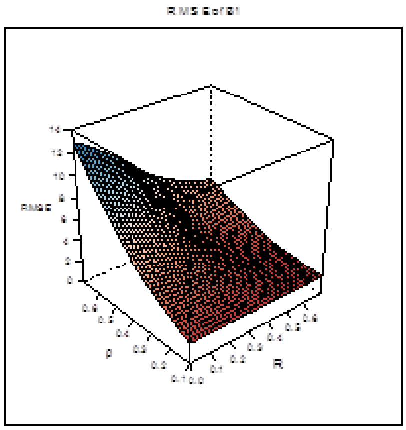
Figure-2 shows that the comparison of the Bias parameters GWR versus SAR-GWR. The GWR model has larger Bias than SAR-GWR. The parameter intercept β_0 based on GWR model has positive Bias. Figure-2(a) shows that the increasing Bias of the intercept β_0 is influenced by the spatial autoregressiveρ. The collinearity does not have the significant effect of the Bias intercept. This is because of the intercept does not directly relate to the covariate. The interesting result comes from Figure-2 (b-c). We can see that the slope β_1 and β_2 have a contrary surface plot. The increasing spatial autoregressive lead to increase negative Bias for β_1 and positive Bias for β_2. The increasing of the multicollinearity does not have a significant effect on the Bias slopeβ_1 and β_2. The minimum Bias of β_1 and β_2 are reached at the spatial autoregressive lower than 0.50. It means that the spatial autoregressive gives a serious problem in context Bias for GWR if the ρ larger than 0.50. The different results come toform SAR-GWR. The parameters model SAR-GWR has small Bias compare than parameters GWR. Figure 2(d) shows only intercept that have high Bias.

ARPN Journal of Engineering and Applied Sciences
©2006-2018 Asian Research Publishing Network (ARPN). All rights reserved.

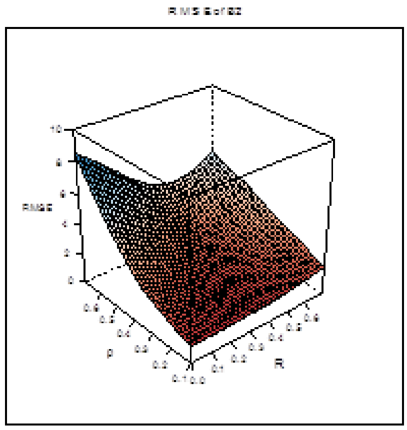
www.arpnjournals.com



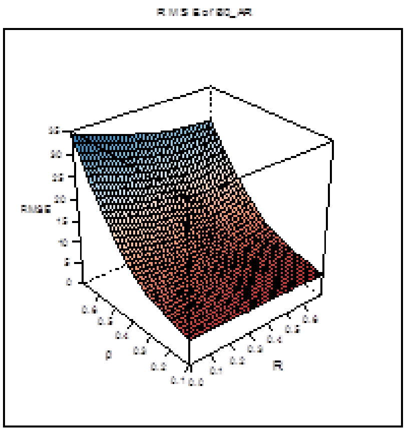
(a) RMSE β_0 from GWR



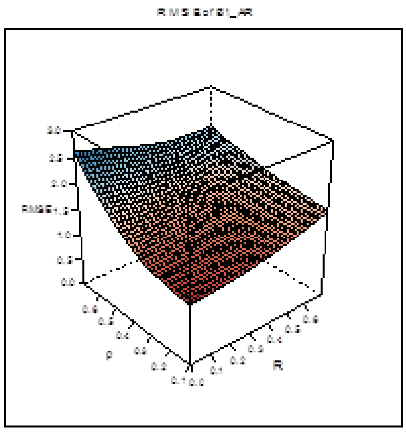
(b) RMSE β_1 from GWR



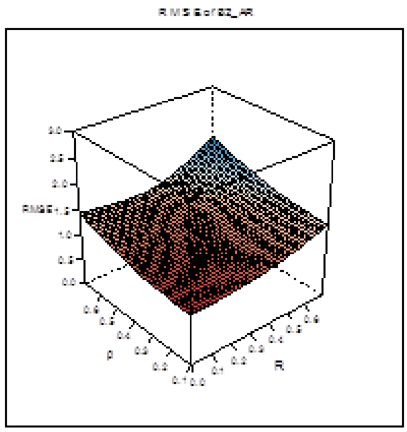
(c) RMSE β_2 from GWR



(d) RMSE β_0 from SAR-GWR



(e) RMSE β_1 from SAR-GWR



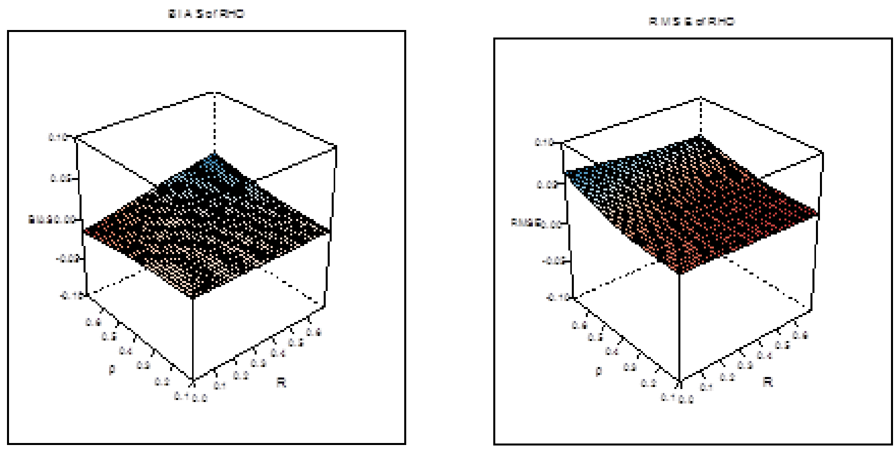
(f) RMSE β_2 from SAR-GWR

Figure-3. The Comparison of the RMSE of GWR and SAR-GWR.

The second criteria used to compare GWR and SAR-GWR is RMSEA. The consistent results can be seen from Figure-3 (a-f). The smallest RMSE come from the smallest spatial autoregressive and collinearity. The biggest effect to the RMSE is given by the spatial autoregressive. The collinearity does not have significant effect to increase RMSEA. The SAR-GWR model has smallest RMSE compared than GWR.

If we compare the Bias and RMSE we can see that the RMSE is more influenced by the variance estimate ratherthan the BIAS itself. It means that the existence of the spatial autoregressive more increases the inefficiency compare than the Bias.

VOL. 13, NO. 1, JANUARY 2018 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences
©2006-2018 Asian Research Publishing Network (ARPN). All rights reserved.
www.arpnjournals.com



(a) Bias of Spatial Autoregressive (ρ) (b) RMSE of Spatial Autoregressive (ρ)

Figure-4. The Bias and RMSE of the Spatial Autoregressive (ρ).

The SAR-GWR model is estimated using IV approach with two-stageleast square estimation. Figure-4(a-b) shows that the estimation of the spatial autoregressive coefficient has small Bias and RMSE. It means that the IV with TSLS is a good estimator for SAR-GWR. The resume of the Bias and RMSE GWR and SAR GWR can be seen in Table-1 below:

Table-1. The Comparison Bias Estimate of GWR and SAR-GWR.

Condition	GWR			SAR-GWR			
	β_0	β_1	β_2	β_0	β_1	β_2	ρ
$R = 0, \rho = 0.1$	18.266	-1.079	-0.036	-0.137	0.064	0.000	0.000
$R = 0, \rho = 0.3$	18.320	-0.956	-0.272	-0.161	0.069	0.005	0.000
$R = 0, \rho = 0.5$	18.248	-0.691	-0.516	-0.160	0.065	0.009	0.000
$R = 0, \rho = 0.7$	18.060	-0.134	-0.950	-0.146	0.055	0.017	0.000
$R = 0.3, \rho = 0.1$	68.746	-2.977	0.428	-0.040	0.064	0.022	0.000
$R = 0.3, \rho = 0.3$	68.720	-2.367	-0.371	-0.254	0.067	0.026	0.001
$R = 0.3, \rho = 0.5$	68.412	-1.398	-1.228	-0.388	0.062	0.034	0.001
$R = 0.3, \rho = 0.7$	67.740	0.509	-2.717	-0.484	0.038	0.053	0.002
$R = 0.5, \rho = 0.1$	158.105	-5.109	1.505	0.877	0.092	0.069	-0.004
$R = 0.5, \rho = 0.3$	157.180	-3.424	0.279	-0.593	0.104	0.071	0.001
$R = 0.5, \rho = 0.5$	156.408	-1.443	-1.386	-1.478	0.087	0.101	0.004
$R = 0.5, \rho = 0.7$	155.043	2.174	-4.235	-2.161	-0.007	0.180	0.006
$R = 0.7, \rho = 0.1$	365.578	-7.929	2.961	8.054	0.221	0.257	-0.017
$R = 0.7, \rho = 0.3$	363.289	-5.347	1.516	-0.777	0.238	0.231	0.000
$R = 0.7, \rho = 0.5$	360.126	-0.903	-0.770	-6.392	0.182	0.321	0.011
$R = 0.7, \rho = 0.7$	357.377	5.352	-5.696	-11.245	-0.179	0.671	0.021

GWR has a poor performance in the case of spatial autoregressive. The Bias and RMSE of the parameters estimate of GWR model increase due to strongly spatial autoregressive

4. APPLICATION

We applied our model to the Tuberculosis (TB^+) cases in Bandung city. Bandung is a capital city of WestJava. The Tuberculosis still a major health problem in Bandung city. To control or manage the increase and spread of this disease, we have to identify the potential risk factor of TB^+. In 2013, Bandung recorded as many as 1507 TB^+ cases, increased significantly to 1872 cases in 2014 and back down in 2015 to 1584 cases (Bandung Health Office). Two factors identified to influence the prevalence rate of TB^+ are Healthy House index and

©2006-2018 Asian Research Publishing Network (ARPN). All rights reserved.

Cleaning Water. We use the SAR-GWR to model the relationship between prevalence on the healthy house and cleaning water. The statistics of the parameter model can be seen in Table-3 below:

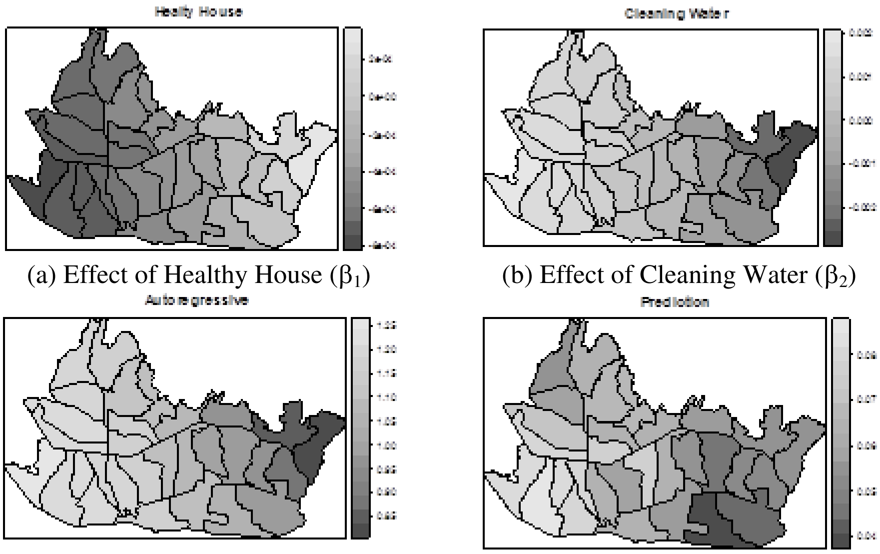
Table-2. Parameter estimate of modeling TB.

	Min.	1st Qu.	Median	3rd Qu.	Max.	Global
Intercept	-0.1386	-0.1010	-0.0543	0.0497	0.2417	-0.0026
Healthy House*)	-0.0007	-0.0006	-0.0005	-0.0002	0.0003	-0.0003
Cleaning Water	-0.0026	-0.0004	0.0008	0.0013	0.0018	0.0002
Spatial Auto regressive (Wy)*)	0.8320	1.0380	1.1500	1.2000	1.2390	1.0479

*) significant in level α=0.05

For global parameters, the Healthy House has a negative and significant effect on the prevalence rate of TB^+. The increase of the healthy house index will decrease the prevalence rate of TB^+. The cleaning water does not a significant effect. However, we can see from Figure-5(b) that the cleaning water gives a negative effect on the prevalence rate of TB^+. It means, the water quality still a factor to be considered in manage and control the spread of TB^+. The spatial autoregressive coefficient has a big and significant effect on the prevalence rate. It means that the location with highest TB^+has abig effect in increasing the prevalence rate in neighboring locations.

From the SAR-GWR result, the Government as policymakershas to campaign and take an action to increase the index of the healthy house and maintaining the water quality. Both of these variables are able to decrease the number of TB + cases.



(a) Effect of Healthy House (β_1) (b) Effect of Cleaning Water (β_2)

(c) Spatial Autoregressive (ρ) (e) Predictive of prevalence rate

Figure-5. Distribution of Parameter estimate of SAR-GWR model.

The maps show that the effect of the healthy house, cleaning water and autoregressive variable make clusters. This is because of the high impact of the spatial dependence. The negative big effect of Healthy House index on the prevalence rate occurs at the west of Bandung City. The higher prediction of prevalence rate also found at West of Bandung city. This result informs the house index in the west of Bandung city still need to be improved to decrease the number of cases of TB^+.

The SAR-GWR provides the valuable information to control and manage the spread of the TB+ disease. If we did not include the spatial dependence in the model, the result may wrong and the information will be misleading.

5. CONCLUSION AND DISCUSSION

We developed SAR-GWR model to model spatial dependence and heterogeneity simultaneously. We use Instrumental Variable approach and estimate the parameters by mean two-stage least square. Monte Carlo simulation has been done to prove the Bias and Efficiency of our method. The parameters estimate of GRW based on

©2006-2018 Asian Research Publishing Network (ARPN). All rights reserved.

WLS estimation lead to bias and inefficient when the data has spatial autoregressive structure. Furthermore, SAR-GWR with TSLS has minimum Bias and RMSE. It means that the TSLS can improve the quality estimate of the WLS for spatial autoregressive structure. The SAR-GWR also provides better results in case multicollinearity. The SAR-GWR model also success in modeling the Tuberculosis (TB^+) in Bandung City. We found that the spatial dependence influence the result which showed in the maps the regression coefficient model for every variable makes clusters. It means there is a positive spatial autocorrelation and has to be accommodated in the model via spatial autoregressive. If the model avoids the spatial structure the conclusion being misleading with high Bias and inefficiency [13]. For the next research, we are going to develop the Bayesian approaches for SAR-GWR to give a better result for inference of the parameters model [13].

ACKNOWLEDGEMENT

This research is fully supported by HIU grand. The authors fully acknowledged Rector of Universitas Padjadjaran for the approved fund which makes this important research viable and effective.

REFERENCES

[1] Sen A. & Srivastava M. 1990. Regression Analysis: Theory, Methods, and Applications. New York: Springer.

[2] Brunsdon C., Fotheringham A. & Charlton M. 1988. Spatial nonstationarity and autoregressive models. Environment and Planning A. 30: 957-973.

[3] Fotheringham A., Brunsdon C. & Charlton M. 2002. Geographically weighted regression: the analysis of spatially. New York: John Wiley and Sons.

[4] Jaya, I. M., Abdullah A. S., Hermawan E., & Ruchjana B. N. 2016. Bayesian Spatial Modeling and Mapping of Dengue Fever: A Case Study of Dengue Fever in the City of Bandung, Indonesia. IJAMAS. 94-103.

[5] Anselin L. 1988. Spatial Econometrics: Methods and Models. California: Springer.

[6] Jaya, I. M., Ruchjana B.N. 2016. The Comparison of Spatial Econometric Models to Estimate Spillover Effect By Means of Monte Carlo Simulation, ARPN Journal of Engineering and Applied Sciences. 11(24): 14168- 14174.

[7] Vega H. S. & Elhorst J. P. 2015. The SLX Model. Journal of Regional Science. 55(3): 339-363.

[8] LeSage J. & Pace R. K. 2009. Introduction to Spatial Econometrics. USA: Chapman & Hall/CRC.

[9] Hayes A. F. & Cai L. 2007. Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. Behavior Research Methods. 39: 4.

[10] Wheeler D. & Tiefelsdorf M. 2005. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. J. Geograph Syst. 7: 161-187.

[11] 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

[12] Klotz S. 2004. Cross-Sectional Dependence in Spatial Econometrics Models with an Application to German Start-up Activity Data, London: Transaction Publisher.

[13] Jaya, I. M., Folmer, H., Ruchjana, B. N., Kristiani, F., & Andriyana, Y. 2017. Modeling of Infectious Disease: A Core Research Topic for The Next Hundred Year. In J. Randall, & S. Peter, Regional Research Frontiers (pp. 681-701). US: Springer