Journal of Theoretical and Applied Information Technology
15th March 2017. Vol.95. No 5

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# THE ArcVIEW AND GeoDa APPLICATION IN OPTIMIZATION OF SPATIAL REGRESSION ESTIMATE

**[1]SCOLASTIKA MARIANI,[2]WARDONO, [3]MASRUKAN , [4]FATKHUROKHMAN FAUZI**

[1]Doctor, Department of Mathematics, Semarang State University, Indonesia
[2]Doctor, Department of Mathematics, Semarang State University, Indonesia
[3]Doctor, Department of Mathematics, Semarang State University, Indonesia
[4]Student, Department of Mathematics, Semarang State University, Indonesia
E-mail: [1] scmariani.unnes@gmail.com, [2] wardono@mail.unnes.ac.id
,[4]fwaratkhurokhmanfauzi@gmail.com

**ABSTRACT**

ArcView is a geographic information system software product produced by and GeoDa is a software tool developed to implement techniques for exploratory spatial data to the analysis (ESDA) on a lattice of data (points and polygons). There were spatial diversity between the regions on the human development index (HDI) variable. This research would look for the best spatial regression model for the HDI in Central Java, Indonesia. Testing spatial effects were done by looking at the value Moran`s I, there were five variables with positive autocorrelation, and a negative autocorrelation variable is unemployment. The next test with the test Lagrange Multiplier (LM), only Lagrange Multiplier (error) which is included in the admission criteria, then the regression model used is the Spatial Error Model (SEM). In this study used two kinds of weighting matrix, i.e. rock contiguity and queen contiguity. The next step research was comparing the value of R-Square and the value of Akaike Info Criterion (AIC) and then obtained best regression model i.e.Spatial Error Model (SEM) with weighting rock contiguity with the value of R-Square is 99.8119 % and the smallest Akaike Info Criterion (AIC) was 4.6362. The spatial regression equation was : $HDI = 4.5565 + 0.4176\ LE + 0.8445\ HPS + 1.4262\ ALS + 0.0011\ PCS$ with $HDI$ : human development index, $LE$ : life expectancy, $HPS$ : hope and period school, $ALS$ : average length of school, $PCS$ : per capita spending..

**Keywords :** *Arcview, Geoda, Optimization, Spatial Regression Estimate*

## 1. INTRODUCTION

Softwares to perform spatial analysis have been extended over the years to include geostatistical techniques [1]. Spatial statistics comprises a set of techniques for describing and modeling spatial data. In many ways they extend what the mind and eyes do, intuitively, to assess spatial patterns, distributions, trends, processes and relationships. Unlike traditional (non-spatial) statistical techniques, *spatial* statistical techniques actually use space – area, length, proximity, orientation, or spatial relationships – directly in their mathematics. [2]

The development of software for spatial data analysis has seen rapid growth since the lack of such tools was lamented in the late 1980s by Haining (1995) [3] and cited as a major impediment to the adoption and use of spatial statistics by GIS researchers. How to integrate spatial statistical methods and a GIS environment and which techniques would be most fruitfully included in such a framework. Familiar reviews of these issues are represented in, among others, [4], [5], [6], [7], [8], [9]. Today, the situation is quite different, and a fairly substantial collection of spatial data analysis software is readily available, ranging from niche programs, customized scripts and extensions for commercial statistical and GIS packages, to a burgeoning open source effort using software environments such as R, Java and Python. *GeoDa* is the outcome of this effort. It is envisaged as an 'introduction to spatial data analysis' where the latter is taken to consist of visualization, exploration and explanation of *interesting* patterns in geographic data. The main objective of the software is to provide the user with a natural path through an empirical spatial data analysis exercise,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

starting with simple mapping and geovisualization, moving on to exploration, spatial autocorrelation analysis, and ending up with spatial regression.

ArcView is a geographic information system software produced by Esri, it is able to view and edit GIS data held in a flat file database or, through ArcSDE, ST-Links PgMap view data held in a relational database management system. The ArcView software is split between ArcMap & ArcCatalog. ArcMap is used for map composition and geographic analysis. ArcCatalog is used for geographic data management [10]. GeoDa is a free software program that serves as an introduction to spatial data to the analysis. GeoDa is a software tool developed by Dr. Prof. Luc Anselin [11]. It is designed to implement techniques for exploratory spatial data to the analysis (ESDA) on a lattice of data (points and polygons). The free program provides a user friendly and graphical interface to methods of descriptive spatial analysis of data, such as spatial autocorrelation statistics, as well as basic spatial regression functionality. Geographic information systems are conducting four elementary functions on space data : input, storage, analysis and output. Spatial analysis has a wide range of different techniques, from basic description all the way up to complex modelling based on inferential statistical methods [4].

Human Development Index (HDI) was determined from three variables, namely the length of life, knowledge / level of education and decent living standard. To measure the dimension of health, used life expectancy on birth, knowledge dimension used expectation of the length of formal study and the average length of the school, for the dimension of decent living used the ability of purchasing power of some staples indicator as seen from the average amount of per capita expenditure.

HDI also influenced by the variable of unemployment. There is also a Gini factor. The Gini coefficient is a measure of the unbalanced distribution. Gini coefficient is in the interval of 0 to 1, that is if the Gini ratio is approaching 1 that means the most high unbalance is occurred, one person controlling everything (income), but if the Gini value close to 0 then there is equity in terms of revenue. Regarding income if the income of a society more equitable or not going unbalance then the quality or the gap between the communities are not much different. Indirectly also affect human development.

Spatial approach is an approach that examines a series of the equations of differences geosphere phenomena in space. In this spatial approach was necessary to note the widespread using of space and the provision of space that will be utilized. Data that requires including regional planning, econometrics, climate and environmental studies, the spread of diseases and the human development index (HDI).

To overcome this problem used the spatial approach method that enables measurement of HDI and it's factor, displayed in visualization form to provide information which easily understood in the analysis, especially in comparing. Visualization in map form expected to describe the tendency for a better spatial analysis in view of the spatial pattern of HDI and it's factors. Spatial method is a method to obtain information observation that influenced the effects of space or location. Spatial effects often occur between one region to another. In spatial data, observations in one location often depends on observations of nearby locations.

Spatial regression was a result of the development of the classical linear regression. This development was based on the influence of the place or spatial data analysis. In this case the HDI was not only influenced by independent variables, but spatial effects therein. Spatial regression modeling can used to produce a better estimate than the classical regression. Some studies using spatial methods are now widely used including the following, Kekez [12] in his Geoinformatics thesis implement concept of spatial autocorrelation as a center question of investigation of the processes of spatial clustering and formation of specific spatial clusters in Helsinki Metropolitan Area among immigration population. This thesis represents comparative study of computing capabilities of ESDA methods (global and local Moran's Index) performed in two GIS software packages (ArcGIS and GeoDa). Quality and accuracy of the results (maps, statistical values, etc.) are going to be tested and presented. Spatial Regression Model For Children in School Age Less 15 Years in Medan [13], Modeling Spatial Error Model (SEM) for the HDI in Java Central [14], Spatial Regression for Determining Factors Poverty in East Java [15], Application of Spatial Regression Model for modeling enrollment Study in high school in Central Java [16], Testing for Spatial Lag and Spatial Dependence Using Doble Length Error Artifical regressions [17], Usage Spatial Weighting Matrix Rock and Queen Contiguity in Spatial Tobit Regression [18]. In studies Spatial Regression Model For Children in School Age Less 15 Years in Medan [13], the results of this study showed a spatial effects should be considered, and proved

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

that the spatial methods more appropriate than the usual regression.

This study takes 6 parameters that affect the HDI were life expectancy, the average of the length school, period school expectation, adjusted per capita spending, unemployment, gini ratio by taking into account location factors (spatial). Researchers wanted to test more about the best spatial regression model to HDI. In this study, the problem was restricted to the data for the region of Central Java in 2014 and uses a weighted rock and queen contiguity, spatial regression models were used including Spatial Autoregressive Model (SAR) and Spatial Error Model (SEM). Stages used in this paper to perform spatial modeling are multiple linear regression, residual assumption test, multicollinearity test, spatial model, Spatial Autoregressive Regression (SAR), Spatial Error Model (SEM), and Test Lagrange Multiplier (LM).

## 2. HUMAN DEVELOPMENT INDEX (HDI)

The Human Development Index (HDI) is a summary measure of average achievement in key dimensions of human development: a long and healthy life, being knowledgeable and have a decent standard of living. The HDI is the geometric mean of normalized indices for each of the three dimensions.The health dimension is assessed by life expectancy at birth, the education dimension is measured by mean of years of schooling for adults aged 25 years and more and expected years of schooling for children of school entering age. The standard of living dimension is measured by gross national income per capita. The HDI uses the logarithm of income, to reflect the diminishing importance of income with increasing GNI. The scores for the three HDI dimension indices are then aggregated into a composite index using geometric mean. (United Nations Development Programme, [19] ). Indonesia in the category country with medium human development , ranked 110th in the world with an HDI value = 0.684 and average annual HDI growth = 1.06 in 2014.

## 3. SPATIAL STATISTICS

Spatial data analysis is a statistical study of certain phenomenon manifested in space [20]. Special techniques and methods are developed for classification of objects which have topological, geometric and geographic properties. All together these techniques are called spatial analysis or spatial statistics techniques and mostly they are used in the analysis of different geographic data and its spatial dispersal. With advanced development of computers many automatic spatial techniques algorithms have been created or re-introduced from field of statistics for measuring different sorts of spatial dispersal including Mantels test, Pearson's correlation test, Moran's I, Geary's C, Getis-Ord General G, etc. Probably the best term describing this process is *geocomputation*. Geocomputation represents quantitative analysis conducted by computer in which computer is having a key role [21]. Spatial statistics comprises a set of techniques for describing and modeling spatial data. In many ways they extend what the mind and eyes do, intuitively, to assess spatial patterns, distributions, trends, processes and relationships. Unlike traditional (non-spatial) statistical techniques, spatial statistical techniques actually use space – area, length, proximity, orientation, or spatial relationships directly in their mathematics [2], ; [22]. There are many different types of spatial statistics : descriptive, inferential, exploratory, geostatistical and econometric statistics are just some of the most widely used [23]. Inferential statistic is trying to reach conclusions that extend beyond immediate data alone and it's opposite to descriptive statistics, which is organizing and describing already existing data [24]. Methods used in this paper belong to inferential statistic. Inferential statistical techniques are using statistical tests, which are gathering accurate probabilistic inferences from data set [25].

## 4. SPATIAL MODELING

Spatial modeling is a modeling that related to point and area approach. Stages to perform spatial modeling is multiple linear regression, residual assumption test, multicollinearity test, spatial modeling, Spatial Autoregressive Model (SAR), Spatial Error Model (SEM), and Test Lagrange Multiplier (LM).

### a. Regression

Regression is an mathematics equation to explain the relationship between the response variable and predictor variables [26]. In general, multiple linear regression model as follows:

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i$$

Information :

$y_i$: Variable responses of the observation$-i$ $(i = 1,2,\dots,n)$

$\beta_0$ : constant

$x_{ij}$ : $j$ −th regression parameter $(j = 1,2,\dots,k)$

$\beta_j$ : $j$ −th predictor variable at $i$ −th observation

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$\varepsilon_i$ : Residual that assumed identical, independent, and normal distribution with zero mean and variance $\sigma^2$

$n$ : the number of observations

In the form of a matrix can be described as follows:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

with :

$$\boldsymbol{y} = [\boldsymbol{y_1}, \boldsymbol{y_2}, \dots, \boldsymbol{y_n}]^T \quad ; \quad \boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon_1}, \boldsymbol{\varepsilon_2}, \dots, \boldsymbol{\varepsilon_n}]^T$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & & x_{2k} \\ & \vdots & & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \quad ; \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

**b. Residual Assumption Test**

Residual Assumptions in the regression model must have characteristic : identical, independent, normally distributed [26]. Classical regression modeling with Ordinary Least Square (OLS) is very strict on several assumptions. If there is an assumption that is not met, then there is an indication of the influence of spatial [26].

To perform regression analysis required residual assumptions that must be occur include :

1) Identical is one important residual assumption of the regression model. Some of the tests that can be used to test identical assumption are Glejser test, park test, plots of residuals and fit.

Statistical Test :

$$F_{counting} = \frac{MSR}{MSE}$$

with :

$$MSR = \frac{\sum_{i=1}^{n}(|\hat{e}_i| - |\bar{e}|)^2}{k} \quad ; \quad MSE = \frac{\sum_{i=1}^{n}(|e_i| - |\hat{e}_i|)^2}{n-k-1}$$

2) Independent Assumption or residual autocorrelation test, which was worked to determine whether there is a correlation between residuals. Some tests can be done to test these assumptions are independent Durbin-Watson test and Autocorrelation Function (ACF) plot. [26]

Statistical test:

$$d_{counting} = \frac{\sum_{i=1}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} {e_i}^2}$$

**3) The normal assumption**

The normal assumption used to determine whether residual has normal distribution. If the assumption of normality is not accured, the OLS estimates can not be used. Some tests can be done for normal distribution assumption is Anderson Darling test, Kolmogorov-Smirnov test, Jarque-Bera test, and Skewnes-kurtosis.

Statistical test:

$$D = maks\,|F_0(x) - S_n(x)|$$

$$S_n(x) = \frac{i}{n}$$

$F_0(x)$ : is the cumulative distribution function of theoretical (is a function of the cumulative of a random sample observation with $i$ is the observation and $n$ is the number of observations.

**4) Multicollinearity Test**

Multicollinearity means that there is a strong correlation between some or all of the predictor variables. This test aims to test whether the regression model found a correlation between the predictor variables. How to detect the presence of multicollinearity is to look at the value tolerance dan variance inflation factor (VIF) of the analysis results. If the VIF value is smaller than 10 it can be concluded not happen multicollinearity [26].

## 5. SPATIAL MODELS

Based on data types, spatial modeling can be divided into point and area modeling approach. The point of approach i.e. : Geographically Weighted Regression (GWR), Geographically Weighted Poisson Regression (GWPR), Logistic Geographically Weighted Regression (GWLR), Space-Time Autoregressive (STAR), and Generalized Space Time Autoregressive (GSTAR). According LeSage [27], the area of approach including : Mixed Regressive-Autoregressive or Spatial Autoregressive Models (SAR), Spatial Error Models (SEM), Spatial Durbin Model (SDM), Conditional Autoregressive Models (CAR), Spatial Autoregressive Moving Average (SARMA) and panel data. Spatial modeling very closely with autoregressive process, indicated by the dependency relationship between a set of observations or location. The relationship can also be expressed by the value of a location depends on the value of other locations adjacent or neighboring. For example there are two locations for the adjacent $i = 1$ and $j = 2$, then the shape of the model is expressed as follows [27] :

$$y_i = \alpha_i y_j + X_i\beta + \varepsilon_i$$

$$y_j = \alpha_j y_i + X_j\beta + \varepsilon_j$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$\varepsilon_j \sim N(0, \sigma^2)$$

Autoregressive process is analogous to the general model of spatial autoregressive as in the following equation :

$$\boldsymbol{y = \rho W_1 y + X\beta + u}$$

with :

$$\boldsymbol{u = \lambda W_2 u + \varepsilon} \quad ; \quad \boldsymbol{\varepsilon \sim N(0, \sigma^2 I)}$$

with :

$\mathbf{y}$ : vector of response variables $(n \times 1)$
$\boldsymbol{X}$ : matrix of predictor variables $(n \times (k + 1))$
$\boldsymbol{u}$ : error vector in equation $(n \times 1)$

Model $u$ have error normally distributed with mean zero and variance $\sigma^2 \boldsymbol{I}$ . Parameters in the estimation are $\beta, \rho$ and $\lambda$. $\rho$ these parameters are the coefficient of the spatial lag of dependent variable and parameter λ is the coefficient of the spatial lag on error, $n$ is the number of observations or location $(i = 1, 2, 3, \ldots, n)$ and $k$ is the number of predictor variables $(k = 1, 2, 3, \ldots, l)$. The influence of the spatial between locations in the model established in the weighting matrix $\boldsymbol{W_1, W_2}$ $(n \times n)$.

In matrix form as follows :

$$\boldsymbol{y} = [y_1 \ y_2 \ \ldots \ y_n]^T \quad ; \quad \boldsymbol{u} = [u_1 \ u_2 \ \ldots \ u_n]^T \quad ; \quad \boldsymbol{\varepsilon} = [\varepsilon_1 \ \varepsilon_2 \ \ldots \ \varepsilon_n]^T$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & & x_{2k} \\ & \vdots & & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \quad ; \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

### a. Spatial Autoregressive Model (SAR)

According to Anselin [28], Spatial Autoregresive Model is a model that combines a regression model with a spatial lag on dependent variables using cross section data. Autoregressive spatial model was formed when $\boldsymbol{W_2 = 0}$ and $\lambda = 0$, so that this model assumes that the process autoregressive only on the response variable [29]. The general model SAR is shown by the following equation :

$$\boldsymbol{y = \rho W_1 y + X\beta + \varepsilon}$$

$$\boldsymbol{\varepsilon \sim N(0, \sigma^2 I)}$$

This model is the development of the first order autoregressive model, where the response variable in addition affected by the lag of response variable itself is also influenced by the predictor variables. Autoregressive has similarity with the analysis of the time series as the first order autoregressive spatial models. The development of the SAR model is SAC and SARMA [27].

### b. Spatial Error Model (SEM)

Spatial Error Model is a model which the error has spatial correlation [26]. The error spatial model formed when $W_1 = 0$ and $\rho = 0$, so that this model assumes that the autoregressive process only on error model. The general model of SEM is shown by the equation :

$$\boldsymbol{y = X\beta + \lambda W_2 u + \varepsilon}$$

$$\boldsymbol{\varepsilon \sim N(0, \sigma^2 I)}$$

Where $\boldsymbol{\lambda W_2 u}$ shows the spatial structure of $\boldsymbol{\lambda W_2}$ on spatially dependent error $(\boldsymbol{\varepsilon})$. This model can be developed into another model, a model example of the result of the development of SEM is spatial Durbin error model (SDEM). Development of SEM models can be applied in the economic field. One of the studies in the field of economics to model SEM is a Spatial Regression Model to Detect Factors Poverty in East Java Province [16].

The maximum likelihood parameter estimates SEM model has the following formula :

$$\boldsymbol{\beta(\lambda) = \left(X(\lambda)' X(\lambda)\right)^{-1} X(\lambda)' y(\lambda)}$$

To estimate the parameter $\lambda$ required a numerical iterations to get estimate by maximizes the log likelihood function [26].

## 6. LAGRANGE MULTIPLIER TEST (LM)

Lagrange Multiplier Test (LM) is used to select the appropriate spatial regression model and to identify the existence of a spatial model [27]. The first step in this test by a simple regression model with Ordinary Least Square (OLS). If LM error significant then the appropriate model is SEM, and if significant LM lag the appropriate model is SAR. If both are significant then the appropriate model is SARMA. Robust Lagrange Multiplier test also done when both are significant. LM lag used to identify SAR models, but it can also model SDM.

The hypothesis used in LM lag

Statistical test:

$$LM_{lag} = \frac{\left(\frac{\boldsymbol{e^T W_1 y}}{s^2}\right)^2}{\frac{(\boldsymbol{W_1 X\beta})^T \boldsymbol{M} (\boldsymbol{W_1 X\beta}) + T s^2)}{s^2}}$$

with :

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$M = I - X(X^T X)^{-1} X^T$$

$$s^2 = \frac{e^T e}{n}$$

$$LM_{error} = \frac{\left(\frac{e^T W_2 y}{\sigma^2}\right)^2}{T}$$

$$T = tr\left((W_2^T + W_2)W_2\right)$$

## 7. SPATIAL PATTERNS

The spatial pattern is something related to the placement or arrangement of objects on the Earth's surface [30]. Every spatial patterns changes will describe spatial process which shown by environmental or cultural factors. According McGarigal and Marks [31], the spatial pattern is a quantitative parameterization of the composition and configure these spatial objects. The spatial pattern describes how geographic phenomena is distributed and how it compares with other phenomena. In this case, spatial statistics is a tool that is widely used to describe and analyze the spatial patterns, ie how the geographic objects happen and change at a given location. It also can compare patterns of objects found in other locations. The spatial pattern can be shown with spatial autocorrelation.

Spatial autocorrelation is an assessment of the correlation between observations on a variable. If the observations $X_1, X_2, \ldots, X_n$, shows the interdependence of space, then the data is said to be spatially autocorrelated. So that spatial autocorrelation is used to analyze the spatial pattern of the spread of the dots to distinguish the location and specific attributes or variables. Some testing in spatial spatial autocorrelation are Moran's I, Geary's ratio, and Local Indicator of Spatial Autocorrelation (LISA).

## 8. SPATIAL AUTOCORRELATION

Spatial autocorrelation is trying to understand the degree of similarity between objects or activities on one spot of Earth's surface and location nearby. *First law of geography* defined by Tobler [32] : "*Everything is related to everything else, but near things are more related than distant things"* has described spatial autocorrelation in the most precise manner [33]. If we have certain variable $Z$, which we are observing on certain spatial location $s$ which is determined by certain coordinates $x$ and $y$ then we can explain spatial autocorrelation as a correlation between $Z(s_i)$ and $Z(s_j)$. Autocorrelation is the correlation of variable with itself, but spatial autocorrelation is correlation of variable with itself on different spatial locations [34].

Spatial autocorrelation modeling started to develop more at the end of 1940's and throughout 1950's. At the end of that decade Moran [35] revealed Moran's Index. Some year afterwards Geary [36] has implemented same but slightly different concept, by presenting Geary's C. The work of Whittle [37] was important additional contribution to the field. Based on these works following example of older colleagues, Cliff & Ord [38] , [39] are employing revolutionary concept of spatial autocorrelation. Further development, especially visual representations of the gained results of the inferential statistics were developed by John Tukey [40]. His work was extremely important for the development of what today we know as Exploratory Spatial Data Analysis [20], [41], [42] with his concept of Exploratory Data Analysis (EDA). It marked a huge discovery at that time and it opened up new horizons and possibilities, for further development. In following years, spatial autocorrelation analysis has been used increasingly for making inferences concerning the factors that underlie observed patterns of spatial variation in processes like human and animal migration [43]. Contemporary analysis is marked with Exploratory Spatial Data Analysis (ESDA) conceptualized by Anselin [20], following up the path and tradition of Tukey.

Statistical equations used for the calculations by these methods are the same in ArcGIS and GeoDa [44] . Final outcome of their results is interesting for comparison and further analysis. Theoretical and methodological approach of global and local methods of spatial autocorrelation in ArcGIS and GeoDa is almost the same, but visual representation of the gained results is slightly different. Therefore,certain prerequisites are needed to be taken into account before global and local methods of spatial autocorrelation are implemented.

## 9. MORAN'S INDEX

First measure of spatial autocorrelation was presented by Moran [35] , [45] . He was studying random or nonrandom distribution of certain phenomena in space in one or two dimensions. It is used to calculate the strength of correlation between observations as a function of the distance separating them [46]. Moran's Index is calculating spatial autocorrelation, similarity between certain features, which is based on a feature location and values for that certain feature simultaneously and at

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the same time multi-directionally. It compares neighboring areal units over complete study area, and informs us about positive spatial autocorrelation (clustering) if the neighboring units have similar values. If the values of the neighboring units are dissimilar it indicates negative spatial autocorrelation (dispersal) [47]. Dispersion with geographic data is less common than clustering, but might be seen with some kind of competitive or territorial spatial process, where similar features try to be as far away from each other as possible.

Global Moran's Index is defined as [48]. The Moran's I statistic for spatial autocorrelation is defined as:

$$I = \frac{n}{S_0}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} z_i z_j}{\sum_{i=1}^{n} z_i^2}$$

$n$ : is total number of spatial units indexed by $i$ and $j$
$i$ and : are spatial units
$z_i$ : deviation of an attribute for feature $i$ from its mean $(x_i - \bar{X})$
$x_i$ : variable of interest
$\bar{X}$ : mean of $x_i$
$w_{ij}$ : the spatial weight between feature $i$ and $j$
$S_0$ : the aggregate of all the spatial weight

$$S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}$$

The score for the statistic is computed as :

$$z_I = \frac{I - E[I]}{\sqrt{V - [I]}}$$

which is based on :

$$E[I] = \frac{-1}{n-1}$$

$$V[I] = E[I^2] - E[I]^2$$

Additional calculations for Moran's Index:

$$E[I^2] = \frac{A - B}{C}$$

$$A = n\{(n^2 - 3n + 3)\, S_1 - nS_2 + 3S_0^2\}$$

$$B = D\{(n^2 - n)\, S_1 - 2nS_2 + 6S_0^2\}$$

$$C = (n-1)(n-2)(n-3)S_0^2$$

$$D = \frac{\sum_{i=1}^{n} z_i^4}{(\sum_{i=1}^{n} z_i^2)^2}$$

$$S_1 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(w_{ij} + w_{ji}\right)^2$$

$$S_2 = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} w_{ij} + \sum_{j=1}^{n} w_{ji}\right)^2$$

Deviation of the mean is calculated, by calculation of difference between each value in data set and mean. After that deviation values for all the neighboring features (neighboring grid cells in this case) are multiplied together to form a cross-product. The cross-products of the deviations from the mean are then summed for all pairs of areal units as long as they are neighbors. Cross-product's results can vary, dependent on the feature values, value of mean and deviations in data. Because of this summed cross-product is always used in this version of Global Moran's Index equation. If both neighboring values are above the mean, the product is a positive number. Product is negative, if both neighboring values are below the mean (product of two negative numbers). So the bigger value of deviation from the mean is, the higher cross-product result is.

When values in dataset have intention to cluster spatially (high value clusters close to other high value clusters and low value clusters close to other low value clusters) Global Moran's Index is positive, which reflects the presence of positive spatial autocorrelation, where similar values are next to each other. But if the value of one areal unit is above the mean and the value of the neighboring unit is below the mean, which are at the same time neighboring units, the product of the two mean deviations will be negative, indicating the presence of negative spatial autocorrelation and a negative value of Global Moran's Index The final result which can occur is that positive and negative cross-product values are in balance, which would lead to that Global Moran's Index value would be zero. Global Moran's Index values are ranging between $-1$ and $+1$. The denominator of Moran's I is essentially the sum of the squared deviations scaled by the total weight of the matrix.

## 10. MAPE

The mean absolute percentage error (MAPE),also known as mean absolute percentage deviation (MAPD), is a measure of prediction accuracy of a forecasting method in statistics, for example in trend estimation. It usually expresses accuracy as a percentage, and is defined by the formula:

Journal of Theoretical and Applied Information Technology
15th March 2017. Vol.95. No 5

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$M = \frac{100}{n}\sum_{i=1}^{n}\frac{(A_i - F_i)}{A_i}$$

where $A_i$ is the actual value and $F_i$ is the forecast value. The difference between $A_i$ and $F_i$ is divided by the Actual value $A_i$ again. The absolute value in this calculation is summed for every forecasted point in time and divided by the number of fitted points $n$. Multiplying by 100 makes it a percentage error. Although the concept of MAPE sounds very simple and convincing, it has major drawbacks in practical application. [49].

## 11. RESEARCH METHOD

### a. Data Sources and Research Variables

Data in this research were : the value of HDI, life expectancy, the average of the length school, period school expectations, adjusted per capita spending, unemployment, and the Gini ratio in Central Java province covering 35 cities / districts. The variables used in this study were 7 variables consisting of one dependent variable and six independent variables.

*Table 1. The Structure Data*

| **Dependent variables** | **Independent Variables** | | | | | |
|---|---|---|---|---|---|---|
| **HDI** | **LE** | **HPS** | **ALS** | **CPS** | **UNM** | **Gini Rasio** |
| $Y$ | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ |
| $y_1$ | $x_{11}$ | $x_{21}$ | $x_{31}$ | $x_{41}$ | $x_{51}$ | $x_{61}$ |
| $y_2$ | $x_{12}$ | $x_{22}$ | $x_{32}$ | $x_{42}$ | $x_{52}$ | $x_{62}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| $y_n$ | $x_{1n}$ | $x_{2n}$ | $x_{3n}$ | $x_{4n}$ | $x_{5n}$ | $x_{6n}$ |

### b. Analysis Method

In this study, the software that used were ArcView and Geoda. The steps of the analysis conducted role in this research are: (1) exploration of thematic maps to determine the distribution patterns and dependencies on each variable and scatterplot to determine the pattern of relationships variables $X$ and $Y$, (2) modeling regression by Ordinary least Square (OLS), which includes estimating the parameters and the significance of the model, (3) test dependencies or correlation, (4) Identificating the existence of spatial effects by Lagrange Multiplier (LM) test and Moran's I Statistics, (5) the process of modeling, the data was modeled with Autoregressive Spatial Model (SAR) and Spatial Error Model (SEM).

## 12. RESULTS AND DISCUSSION

The pattern of spread of the HDI and the variables that influenced using ArcView application obtained the following output :

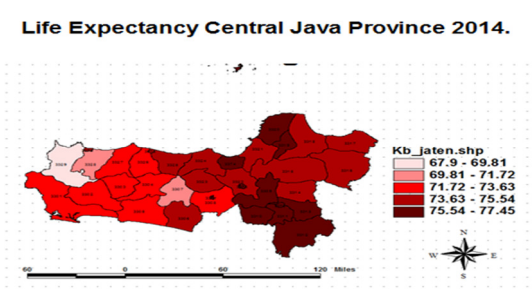


*Figure 1. Life Expectancy*

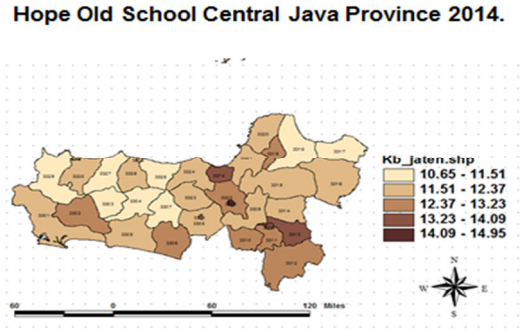


*Figure 2. Hope Period School*

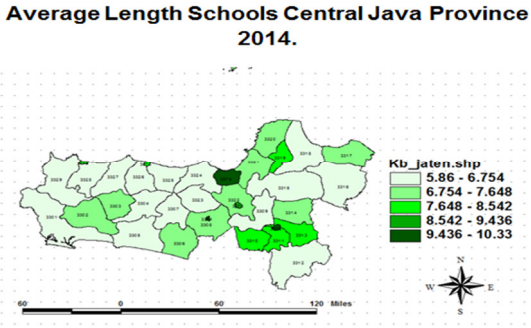


*Figure 3. Average of Period school*

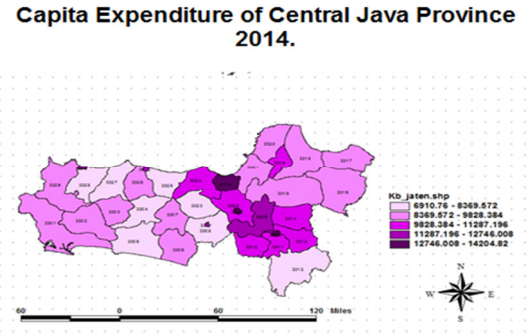


*Figure 4. Per Capita Expenditure*

Journal of Theoretical and Applied Information Technology
15th March 2017. Vol.95. No 5

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

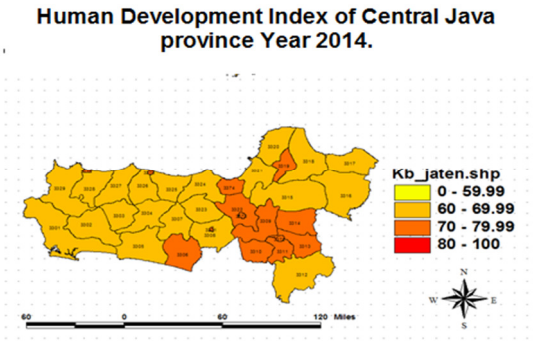


*Figure 5. Human Development Index*

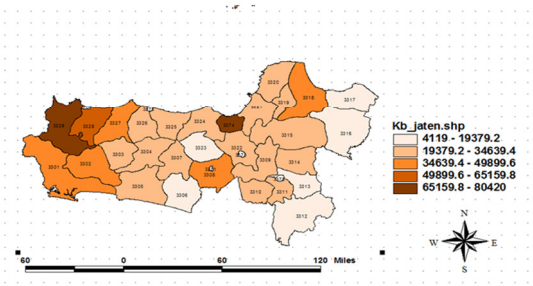


*Figure 6. Unemployment*

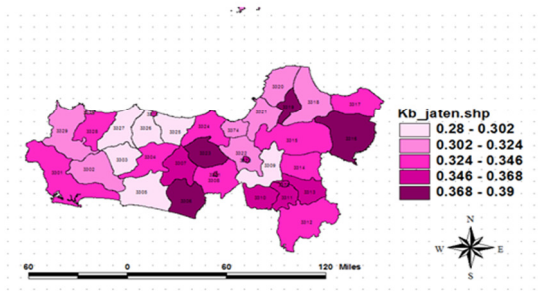


*Figure 7. Gini Ratio*

### a. OLS Regression Model

In regression modeling, parameter estimation was done using Ordinary Last Square (OLS). In testing the OLS model obtained the normality of data, the parameters were significant or not having an effect on the HDI, as well as heteroscedastisity / spatial diversity.

*Table 2. OLS Regression Testing Parameters HDI Central Java Province Based on Output Geoda*

| Variable | Coefficient | Std.Error | t-Statistic | Probability |
|---|---|---|---|---|
| **Constant** | 4.556509 | 2.519293 | 1.808646 | 0.08126 |
| **LE** | 0.4176043 | 2.519293 | 13.46038 | 0.00000 |
| **HPS** | 0.8444755 | 0.09066972 | 9.313754 | 0.00000 |
| **ALS** | 1.426227 | 0.07275583 | 19.60292 | 0.00000 |
| **PCS** | 0.00107774 | 4.77e-005 | 22.59221 | 0.00000 |
| **UNM** | 0.1401131 | 0.0872824 | 1.605286 | 0.11965 |
| **Rasio** | 0.1401131 | 1.828649 | 1.944986 | 0.06188 |
| | | | | |
| ***Probability test of normality (Jarque-Bera)*** | | | | 0.1200 |
| ***Probability of Breusch-Pagan test*** | | | | 0.0270 |
| ***R-square*** | | | | 99.7178% |
| ***Akike Info Criterion*** | | | | 14.0269 |
| ***Moran's I (error)*** | | | | 0.0370 |

To test the normality was expressed in Probability test of normality (Jarque-Bera) amounted to 0.1200 > 0.05 then the data was said to be normal. As for the test parameters see Table 1. the value of probability for the variable LE (= life expectancy = **LE**), HPS (hope and period school = **HPS**), ALS (average length of school = **ALS**), per capita (=per capita spending= **PCS**), obtained 0.0000 < 0.05 then the parameters significantly, while for the variable unemployment (PENGANGGUR = **UNM**) and ratios obtained probability value respectively 0.1197 and 0.0619 <0.05 then the parameter was not significant that the two variables omitted from the model. So obtained the following equation :

$$HDI = 4.5565 + 0.4176\ LE + 0.8445\ HPS + 1.4262\ ALS + 0.0011\ PCS$$

***HDI*** : human development index
***LE*** : life expectancy
***HPS*** : hope and period school
***ALS*** : average length of school
***PCS*** : per capita spending

In an OLS regression tested spatial diversity or heteroskesdastistas also between regions with a view to the test Breusch Probability value-pagan test of 0.0270 <0.05 then we can conclude there is heteroskesdastisitas between regions. Rated R-Square on OLS regression is 99.7178% and the value Akike Info Criterion (AIC) is 14.0269.

In the autocorrelation test by test Moran's I (error) of 0.0370 obtained probab value of <0.05. It berbarti that there spatial autocorrelation it is necessary to further test the spatial regression models.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

**b. Spatial Effects**

Spatial Effects testing done to see if the data of each variable has influenced spatial location, Spatial dependence testing using statistical Moran`s I. The value of each Moran`s I are as follows:

*Table 3. Index Moran`s I*

| Variable | Moran's Indexs I |
|---|---|
| HDI | 0.301571 |
| LE | 0.570926 |
| HPS | 0.322061 |
| ALS | 0.17193 |
| PCS | 0.210319 |
| UNM | -0.0462018 |
| Rasio | 0.0256364 |

$$I_0 = -\frac{1}{n-1}$$
$$I_0 = -\frac{1}{35-1}$$
$$I_0 = -\frac{1}{34}$$
$$I_0 = -0.0294$$

Based on Table 3 and the value seen that the value Moran`s I value is greater than $I_0$ , which means that all independent variables and the dependent variable has a positive autocorrelation exception unemployment smaller this means that unemployment has a negative autocorrelation means more unemployment smaller development index human.

**c. Lagrange Multiplier (LM) Test**

Selection of a spatial model is done by using Lagrange Multiplier (LM) as the initial identification. Lagrange Multiplier is used to identify more specific spatial dependencies are dependencies lag, error or both (lag and error). LM Testing Results are presented in Table 4 as follows:

*Table 4. Analysis of Spatial Dependencies*

| Spatial Test Dependencies | Probability |
|---|---|
| Moran's I | 0.03694 |
| Lagrange Multiplier (lag) | 0.08667 |
| Lagrange Multiplier (error) | 0.01489 |
| Significance | 0.05 |

Based on Table 4 is known that the probability Moran`s I equal to 0.0370 and less than 0.05 (level exhibited significantly). So it means that there H0 spatial dependencies in regression error.

Lagrange Multiplier test (lag) aims to identify the relationship between cities / districts. Based on Table 4 it can be seen that the value of the Lagrange Multiplier (lag) is greater than 0.05 (significance level) then H0 accepted then there are no dependencies spatial lag so no need to proceed to the modeling of Spatial Autoregressive Model (SAR). However, the value of the Lagrange Multiplier (error) is worth 0.0149, then reject H0 means there is a spatial dependencies in an error that needs to be continued in the modeling Spatial Error Model (SEM).

**d. Spatial Error Model (SEM) with a weighted Rook Contiguity**

Furthermore, regression modeling using SEM models with weighting rook contiguity use applications obtained output geoda each parameter as follows :

*Table 5. Estimation Parameters Rook SEM weighting Contiguity HDI Central Java Province Based Output Geoda*

| Variable | Coefficient | Std.Error | z-value | Probability |
|---|---|---|---|---|
| Constant | 3.307369 | 1.450686 | 2.279865 | 0.02262 |
| LE | 0.4302332 | 0.017048 | 25.23684 | 0.00000 |
| HPS | 0.8604433 | 0.060525 | 14.21629 | 0.00000 |
| ALS | 1.377209 | 0.046663 | 29.51378 | 0.00000 |
| PCS | 0.001106 | 3.046e-005 | 36.31416 | 0.00000 |
| UNM | 0.1600864 | 0.062214 | 2.573155 | 0.01008 |
| Rasio | 3.578425 | 1.293102 | 2.767319 | 0.00565 |
| LAMBDA | -0.7631367 | 0.201368 | -3.789765 | 0.00015 |
| | | | | |
| *Probability of Breusch-Pagan test* | | | | 0.0197 |
| *R-square* | | | | 99.8119% |
| *Akike Info Criterion* | | | | 4.6362 |

Based on the output table 5 the results of SEM with weighting rock contiguity that showed a dependency spatial error it appears from LE, HPS, ALS, per-capita (PCS), unemployment (UNM), Gini ratio, and lambda is significant at the 5% level, it can be concluded that there are linkages between the regions with other areas and these variables can be inserted into the equation model SEM. In addition to be indicated that neighboring region has the same characteristics. Model SEM weighted contiguity rock formed is as follows:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$HDI = 3.3073 + 0.4302\,LE + 0.8604\,HPS + 1.3772\,ALS + 0.0011\,PCS + 0.1601\,UNM + 3.5784\,Rasio + \mu_i$$

With

$$\mu_i = -0.7631 \sum_{i=1, i \neq j}^{n} w_{ij} y_i + \varepsilon_i$$

Probability values obtained on test-pagan breusch test of 0.0197 <0.05 then it can be concluded that under the SEM weighted contiguity rock there that there are variations between regions. For the R-Suquare and Akaike Info Criterion (AIC) SEM models is 99.8119% and 4.6362.

**e. Spatial Error Model (SEM) with a weighted Queen Contiguity**

Further models of SEM with the queen contiguity weighted output obtained each parameter based on the output geoda as follows:

*Table 6. SEM Parameter Estimation weighting Queen Contiguity HDI Central Java Province Based Output Geoda*

| Variable | Coefficient | Std.Error | z-value | Probability |
|---|---|---|---|---|
| **Constant** | 3.446014 | 1.469044 | 2.345752 | 0.01899 |
| **LE** | 0.4287813 | 0.0174184 | 24.61659 | 0.00000 |
| **HPS** | 0.8624558 | 0.061761 | 13.9644 | 0.00000 |
| **ALS** | 1.37696 | 0.0475365 | 28.96639 | 0.00000 |
| **PCS** | 0.001108 | 3.109e-005 | 35.63629 | 0.00000 |
| **UNM** | 0.1579015 | 0.063806 | 2.474701 | 0.01333 |
| **Rasio** | 3.433249 | 1.322961 | 2.595125 | 0.00946 |
| **LAMBDA** | -0.7629005 | 0.2123974 | -3.591854 | 0.00033 |
| | | | | |
| ***Probability of Breusch-Pagan test*** | | | | 0.0301 |
| ***R-square*** | | | | 99.8048 % |
| ***Akike Info Criterion*** | | | | 5.6108 |

Based on the output table 6 the results of SEM by weighting queen contiguity that showed a dependency spatial error this can be seen from LE, HPS, ALS, per capita, unemployment, ratio, and lambda is significant at the 5% level, it can be concluded that there are linkages between the region with other areas and these variables can be inserted into the equation model SEM. It also indicated that nearby regions have the same characteristics. SEM models with weighting queen contiguity are formed as follows:

$$HDI = 3.4460 + 0.4288\,LE + 0.8625\,HPS + 1.3770\,ALS + 0.0011\,PCS + 0.1579\,UNM + 3.4332\,Rasio + \mu_i$$

With

$$\mu_i = -0.7629 \sum_{i=1, i \neq j}^{n} w_{ij} y_i + \varepsilon_i$$

Probability values obtained on test-pagan breusch test on SEM with queen contiguity weighting of 0.0301 <0.05 then it can be concluded that under the SEM with queen contiguity weighting are that there are variations between regions. For the R-Suquare and Akaike Info Criterion (AIC) SEM models is 99.8048% and 5.6108.

**f. Determining the Best Model**

To determine the best spatial regression model by seeing the value of R-Square and the Akaike Info Criterion (AIC), best spatial regression models with the criteria of the largest R-Square and the value of Akaike Info Criterion (AIC), the smallest. Comparable values are as follows:

*Table 7. Determination of the Best Model*

| Model | R-square | AIC |
|---|---|---|
| **SEM with weighting rock contiguity** | 99.8119% | 4.6362 |
| **SEM with weighting queen contiguity** | 99.8048% | 5.6108 |
| **OLS** | 99.7178% | 14.6108 |

Based on the table 7 showed that the biggest of R-Square value and the smallest AIC value is Spatial Error Model (SEM) with weighting rock contiguity. So we can conclude that the best spatial regression model is a model Spatial Error Model (SEM) with weighting rock contiguity.

## CONCLUSION

The pattern of the spread of HDI in Central Java province seems patterned clustered between regions adjacent to each other. Based on the relationship between HDI with LE, HPS, ALS, PCS and UNM can be interpreted that the similarities and differences in the characteristics in each municipality / town adjacent may cause an increase or decrease HDI in Central Java.

SEM with weighted regression model rook contiguity is a regression model of the Human Development Index (HDI) by the following equation:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$HDI = 3.3073 + 0.4302\ LE + 0.8604\ HPS + 1.3772\ ALS + 0.0011\ PCS + 0.1601\ UNM + 3.5784\ Rasio + \mu_i$$

With

$$\mu_i = -0.7631 \sum_{i=1, i \neq j}^{n} w_{ij} y_i + \varepsilon_i$$

with $HDI$ : human development index, $LE$ : life expectancy, $HPS$ : hope and period school, $ALS$ : average length of school, $PCS$ : per capita spending.

For further research on applied spatial regression estimate can be developed al. analyzing data outliers, Bayesian spatial regression, spatial regression panel data, etc., with development in the use of software for example : spatial statistical analysis in the open-source R language environment, CrimeStat spatial statistics program, a GIS-based toolbox for network data, SANET (Spatial Analysis on a NETwork), the package STARS (Space-Time Analysis of Regional Systems), a three - dimensional visualization extension cto the SAND (Spatial and Nonspatial Data) spatial database system, package ChoroWare (which adopts a multiobjective approach to the construction of choropleth map class intervals), dll. CrimeStat is a stand-alone Windows program for the analysis of the spatial pattern of crimes and is designed to interface with most desktop GIS programs. The interface is carried out through using graphical objects produced byCrimeStat, which are linked to packages such asArcView, ArcGis, and MapInfo.

## REFFERENCES

[1] M.J. Smith, Goodchild MF, Longley PA (2006) Geospatial analysis. Troubador, Leicester

[2] Scott L, and Getis A (2008) Spatial statistics. InKemp K (ed) Encyclopedia of geographic informations. Sage, Thousand Oaks, CA

[3] Haining, R. (1995). Data problems in spatial econometric modelling. In L. Anselin, & R. Florax (Eds.), New directions in spatial econometrics(pp. 156–171). Berlín: Springer August

[4] Anselin, L. and Getis, A. (1992). Spatial data analysis with GIS: an introduction to application in the social sciences. Technical report 92–10 National Center for Geographic Information and Analysis, University of California, Santa Barbara, USA, 1–54

[5] Goodchild, M.F.1992. Geographical Data Modeling. Comput Geosci 18:401-408

[6] Fisher M. Manfred & Peter Nijkamp.1993. Handbook of Regional Scince.New York:Springer.

[7] Fotheringham,A.S. and Rogerson, P. 1994. Spatial Analysis and GIS. London : Taylor & Francis

[8] Fisher,M.M.,Scholten,H., and Unwin,D. 1996. Spatial Analytical Perspectives in GISin Environmental and Socio-economic Science. London : Taylor & Francis

[9] Fisher,M.M., and Getis,A . 1977. Recent Developments in Spatial Analysis. Berlin : Springer Verlag

[10] Budiyanto, E. 2010. Sistem Informasi Geografis ArcView GIS. Yohyakarta : Andi Offset.

[11] Anselin, Luc, Syabri, Ibnu, and Kho, Youngihn (2006). Geoda, An Introduction To Spatial Data Analysis. Geographical Analysis. Geographical Analysis 38 (2006) 5–22

[12] Kekez, F. 2015. Clustering Of Immigration Population In Helsinki Metropolitan Area, Finland: A Comparative Study Of Exploratory Spatial Data Analysis Methods. Master's Thesis. Geography and Geoinformatics. University Of Helsinki. Department Of Geosciences And Geography. Division Of Geography

[13] Rati, M.. 2013. Model regresi spasial untuk anak tidak bersekolah usia kurang 15 tahun di kota medan. Skripsi. Medan: FMIPA Universitas Sumatra Utara.

[14] Diana,W.S., Moh Yamin Darsyah, Tiani Wahyu Utami. 2014. Modeling Spatial Error Model (SEM) for the HDI in Java Central. Statistika, Vol. 2, No. 2, November 2014

[15] Djuraidah A, Aji Hamim W. 2012. Regresi Spasial untuk Menentukan Faktor-faktor Kemiskinan di Provinsi Jawa Tengah.Jurnal Statistika 12(1),1-8

[16] Arisanti, R. 2011. Performance Spatial Regression Models for detecting factors of poverty in East Java Province. Sekolah Pascasarjana. Institut Pertanian Bogor. Bogor

[17] Baltagi H. Badi & Long Liu. 2012. Testing for Spatial Lag and Spatial Error Dependence Using Doble Length Artifical Regressions.Stat Papers(2014)55: 477-486

[18] Sholikhah,M. 2014. Penggunaan Matriks Pembobot Spasial Tipe Queen Contiguity Dan Rook Contiguity Pada Regresi Tobit Spasial.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Jurnal Mahasiswa Statistik. Vol 2, No 3 (2014). Jurusan Matematika, F.MIPA, Universitas Brawijaya

[19]http://hdr.undp.org/en/content/human-development-index-hdi

[20] Anselin, L. (1996). The Moran Scatterplot As An ESDA Tool To Assess Local Instability In Spatial Association. In Spatial Analytical Perspectives On GIS, Edited By Fischer, M. et al., Taylor & Francis, London, 111–125

[21] Fotheringham, A.S. (1998). Trends In Quantitative Methods II: Stressing The Computational, Progress In Human Geography, Volume 22, Sagepub, 283–292

[22] Scott, L. M. and Mark V. Janikas. Spatial Statistics in ArcGIS. 2001

[23] ESRI (2013a). ArcGIS Support: GIS dictionary, Spatial Statistics. Environmental Systems Research Institute, Redlands, California. http://support.esri.com/en/knowledgebase/GISDictionary/term/spatial%20statistics

[24] Rice, P.G. and A.J. Venables (2003) 'Equilibrium regional disparities; theory and British evidence', Regional Studies, 37, 675-686

[25] Taylor, L. R. (1977). Migration and the spatial dynamics of an aphid, Myzus persicae. Journal of Animal Ecology, 46,411-23

[26] Anselin L., (1988), Spatial Econometrics: Methods and Models, Academic Piblishers, Dordrecht

[27] LeSage, J. 2009. Intoduction to Spatial Econometrics. CRC Press, Taylor and Francis Group.

[28] Anselin, L. (1989). What Is Special About Spatial Data?: Alternative Perspectives On Spatial Data Analysis. Symposium On Spatial Statistics, Past, Present and Future National Center for Geographic Information and Analysis, University of California, Santa Barbara, USA, 63–77

[29] Lee, L.F. and J. Yu, 2010. Estimation of spatial autoregressive panel data models with Öxed e§ects. Journal of Econometrics 154, 165-185

[30] Lee, J. dan Wong, D. W. S. (2001). Statistical Analysis with Arcview GIS. New York: John Wiley and Sons

[31] McGarigal, K. & Marks, B.J. (1995). FRAGSTATS: spatial pattern analysis program for quantifying landscape structure. USDA For. Serv. Gen. Tech. Rep. PNW-351. 141 p

[32] Tobler W., (1970) "A computer movie simulating urban growth in the Detroit region". Economic Geography, 46(2): 234-240

[33] Goodchild, M.F. 1987. A Spatial Analytical Perspective on Geographical Information Systems. Int J Geogr Inf Sci I : 327-334

[34] Schabenberger, O., & Gotway, C. A. (2005).Statistical methods for spatial data analysis. Boca Raton,FL: Chapman & Hall/CRC Press

[35] Moran,P.A.P. 1948. The Interpretation of Statistical Maps. Journal of The Royal Statistical Society **B10** 243-251

[36] Geary,R.C. 1954. The Contiguity Rasio and Statistical Mapping. The Incorporated Statistician **5** 115-145

[37] Whittle,P. 1954. On Stationary Processes in The Plane. **41** 434-449

[38] Cliff A.D. & J.K. Ord (1969). The problem of spatial autocorrelation. In Studies in Regional Science, edited by A.J.Scott, Pion Press, London, 25–55

[39] Cliff A.D. & J.K. Ord (1970). Spatial Autocorrelation: A Review Of Existing And New Measures With Applications. Economic Geography, Volume 46, International Geographical Union. Commission on Quantitative Methods, Clark University, 269–292

[40] Tukey, J. W. (1977). Exploratory Data Analysis, Addison-Wesley, Reading

[41] Anselin, L. (1999). Interactive techniques and exploratory spatial data analysis. In Geographical Information Systems: Principles, Techniques, Management and Applications, edited by P. Longley et al., Geoinformation Int., Cambridge, 253–266

[42] Messner,S.F., Luc Anselin, Robert D. Baller, Darnell F. Hawkins, Glenn Deane, and Stewart E. Tolnay. 1999. The Spatial Patterning of County Homicide Rates: An Application of Exploratory Spatial Data Analysis. Journal of Quantitative Criminology, Vol. 15, No. 4, 1999

[45] Moran,P.A.P. 1950. Notes on Continuous Stochastic Phenomena. Biometrika **37** 17-23

[43] Sokal, R.R., Jacquez, G.M., Wooten, M.C. 1989. Spatial autocorrelation analysis of migration and selection. Genetics 121:845-855

[44] Anselin, L and Rey. 2010. Prespective on Spatial Data Analysis.New York:Springer.

[45] Moran,P.A.P. 1950. Notes on Continuous Stochastic Phenomena. Biometrika **37** 17-23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[46] Oliveau, S., Christophe Guilmoto. Spatial correlation and demography.: Exploring India's demographic patterns.. XXVe Congr`es International de la Population, Jul 2005, Tours, France.

[47] ESRI, (2013b). ArcGIS Desktop Help: Spatial Statistics toolbox, Annalyzing Patterns toolset, How Spatial Autocorrelation (Global Moran's I) works. Environmental Systems Research Institute, Redlands, California. 4.18.2013

[48] Getis, A. Ord JK. 1992. The Analysis of Spatial Assosiation by Use of Distance Statistics. Geogr Anal 24 : 189-206

[49]Tofallis, C. 2015. A Better Measure of Relative Prediction Accuracy for Model Selection and Model Estimation. J Oper Res Soc 66 : 524