International Journal of Advanced Science and Technology
Vol. 29, No. 6, (2020), pp. 3296 – 3309

# On Taylor Series and Chebyshev Polynomial Approximation in Parameter Estimation for the Spatial Autoregressive Model

Jajang[*1] , Budi Pratikno,[2]

[1,2] Department of Mathematics, Faculty of Mathematics and Natural Science, Jenderal Soedirman University, 53122 - Central Java, Indonesia.

[1*]rzajang@yahoo.com, [2] pratikto@gmail.com,

***Abstract***

*Parameter estimation for the spatial autoregressive (SAR) model by using maximum likelihood (MLE) method involve log determinant of spatial weight matrix where its dimension is large. Therefore, to solve log determinant of this matrix often used approximation. The paper studied performances of Taylor series and Chebyshev polynomial methods in parameter estimation of SAR model. In this paper, we also studied performance of two types of spatial matrices, W-AMOEBA and W-contiguity, to choose the best of spatial weighted matrix in SAR model. Evaluation of approximation methods to solve log determinant and to choose the best performance model, we used data simulation and root mean square error (RMSE) criteria. The data simulation is generated by Monte Carlo simulation methods. Furthermore, the best model (the best performances of approximation method and spatial weight matrix) is implemented to model human development index (HDI) and its factors in Central Java Province. The HDI factors are population, gross enrolment rate, district minimum wage, the number of poor people and poverty line. The results showed that RMSE of models used to Chebyshev polynomial is smaller than Taylor series. Therefore, Chebyshev polynomial approximation is more accurate than Taylor series approximation. Furthermore, the Chebyshev polynomial is used to analysis the human development index (HDI) and its factors by using SAR model. The result showed that the gross enrollment rate, district minimum wage, and poverty line then the HDI have positive impact. It means that increasing of theirs factors can improve HDI.*

***Keywords:*** *W-AMOEBA, SAR, HDI, Taylor series , Chebyshev polynomial, poverty line.*

## 1. Introduction

In recent years, spatial regression model have been developed to take spatial dependence. The models that involve statistical dependence are often more realistic [7], [8]. A fundamental concern of spatial analysts is to find patterns in spatial data that lead to the identification of spatial autocorrelation or association [16]. Taking spatial dependences into account when dealing with spatial data is very important, and neglecting them can cause problems. For example, ignoring spatial lag structures causes ordinary least squares (OLS) estimators to become bias and inconsistent. The spatial weights matrix is one of the most convenient ways to summarize spatial relationship in the data. Spatial weights matrix is a nonnegative matrix that specifies the neighborhood set for each observation. Here, the data are collected from different spatial locations. Spatial weights characterize cross-section dependence in useful ways their measurement

ISSN: 2005-4238 IJAST

has an important effect on the estimation of a spatial dependence model [1], [9], [15]. The prediction result becomes accurate if we found a representative spatial weight matrix and parameter estimation method. There are many to create spatial weight matrix [10]. However the most commonly use spatial weight matrix is a binary matrix based on geographic distance and contiguity. Furthermore, the spatial weight matrix is also found on Aldstadt and Getis [1], namely a multidirectional optimum ecotope-based algorithm (AMOEBA). Here, elements of AMOEBA matrix depend on both neighborhood among spatial units and variable [1], [13], [14]. Moreover, Aldstadt and Getis [2] used local Getis statistic to create the matrix. Jajang et al [13],[14] shown that the in spatial dynamic model performance of AMOEBA matrix is better than Contiguity matrix (W.Contiguity).

In the spatial model, we also found endogenous problem in the model. Therefore, classic method such as ordinary least square (OLS) is not relevant to solve it problem. The OLS estimator will be biased as well as inconsistent for the parameters of the spatial model [2]. The inappropriateness of the least squares estimator for models that incorporate spatial dependence has focused attention on the maximum likelihood (MLE), generalized method of moment (GMM), and Two-Stage Least Square (TSLS) methods approach as alternative [11], [12]. In this paper, we use maximum likelihood method to estimate the parameters of SAR model.

Parameter estimation of spatial autoregressive model (SAR) use maximum likelihood method involve log determinant of large matrix, so it is computationally expensive. Therefore, approximation method to find solution in this problem is needed. Two approximation methods, Taylor series and Chebyshev polynomial approximation, are often used to solve this problem.

Based on the description above, this study aims to determine the best performances of two approximation method and two spatial weighted matrix in modeling HDI and its factor in Central Java Province.

## 2. Material and Method

In this section, the materials and methods used are described as follows.

### 2.1. Material

In this paper, we use simulation data and real data. The simulation data is generated by using Monte Carlo simulation method. In this simulation, we only use one predictor variable, spatial weight matrix, and take values of intercept is 1 and slope is 2. In addition, the real data used in SAR model is human development index (HDI and its factors in Central Java Province 2017. The HDI factors are population, gross enrolment rate, district minimum wage, the number of poor people and poverty line.

### 2.2 Method

In this section, we discuss the methods used for modeling HDI data in central Java Province. Discussions include MLE estimation in SAR model, approximation method to calculate logarithm of determinant spatial weight matrix, and implementation the model to real data.

## 3. Statistical models

This section discusses the related material that will be used in the parameter estimation of SAR model and its implementation to HDI data.

### 3.1 Maximum likelihood estimation

Spatial lag dependence or spatial autoregressive model in a regression model is similar to the inclusion of serially autoregressive term for dependent variable in a time-series context. Spatial autoregressive model (SAR) is specified as [2] ,[3]

$$\boldsymbol{y} = \rho \boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

where $y$ is the n x 1 of the response variable, $\boldsymbol{X}$ is the n x k matrix of the non-stochastic explanatory variables, $W$ is the n x n non-stochastic weights matrix, $\rho$ is a spatial autoregressive parameter, $\boldsymbol{\beta}$ is a parameter vector, and $\boldsymbol{\varepsilon}$ is an n x 1 vector of innovations. To estimate parameters of the model use maximum likelihood method, we assume vector of innovations are normally distributed.

Maximum likelihood estimation (MLE) of the SAR models described involves maximizing the log likelihood function with respect to the parameters. The MLE selects the set of values of the model parameters that maximizes the likelihood function. The MLE for estimate SAR parameters was first outlined by Ord [12]. The model (1) represent as equilibrium, so $(I - \rho W)$ is assumed invertible. To avoid calculating the determinant of the $(I - \rho \boldsymbol{W})$ matrix, Ward and Kristiani [21] proposed that $ln|I - \rho W| = \sum_i \ln(1 - \rho\omega_i)$, where $\omega_i$, i=1,2,….,n are eigenvalues of the matrix **W**.

The equilibrium vector y is given by $y = (I - \rho W)^{-1}(X\beta + \varepsilon)$. We assumed that errors are normally distributed, $\varepsilon \sim N(0, I\sigma^2)$. For this errors are normally distributed, we can be expressed by

$$f(\varepsilon) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left[-\frac{\varepsilon'\varepsilon}{2\sigma^2}\right]. \tag{2}$$

Based on the equation in (1) and invertible condition of matrix $(I - \rho W)$, then the equation in (1) can be rewritten by

$$y = (I - \rho W)^{-1}(X\beta + \varepsilon). \tag{3}$$

Furthermore, the probability density function (pdf) in equation in (2) can be transformed into **Y** to obtain likelihood function of **Y.** From the equation (1) we have $\varepsilon = (I - \rho W)y - X\beta$, and from here we ca use Jacobi transformation method to

ISSN: 2005-4238 IJAST

find pdf of ***Y***. The Jacobi transformation in this here is $= \left|\frac{d\varepsilon}{dy}\right| = |I - \rho W|$ . Based on this Jacobi, we can get pdf y as follows [4,6]

$$f(y) = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} \exp\left[-\frac{(y(I-\rho W)-X\beta)'(y(I-\rho W)-X\beta)}{2\sigma^2}\right].|I-\rho W|. \quad (4)$$

Here, the maximum likelihood Estimation method (MLE) is the method maximizes likelihood function. The likelihood function in this as follow

$$L(\rho,\sigma^2,\beta) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left[-\frac{(y(I-\rho W)-X\beta)'(y(I-\rho W)-X\beta)}{2\sigma^2}\right]. \quad (5)$$

The expression in (5) is actually quite a pain to differentiate, so it is almost always simplified by taking the natural logarithm of the expression. This is absolutely fine because the natural logarithm is a monotonically increasing function. This is important because it ensures that the maximum value of the log of the probability occurs at the same point as the original probability function. Therefore, we can work with the simpler log-likelihood instead of the original likelihood. The logarithm of likelihood function of (5) can be rewrite as

$$ln\big(L(\rho,\sigma^2,\beta;y)\big) = ln|I-\rho W| - \frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{(\boldsymbol{y}(\boldsymbol{I}-\rho\boldsymbol{W})-\boldsymbol{X\beta})'(\boldsymbol{y}(\boldsymbol{I}-\rho\boldsymbol{W})-\boldsymbol{X\beta})}{2\sigma^2}. \quad (6)$$

There are the following requirements, the existence of the log likelihood function for the parameter values under consideration, continuous differentiability of the log likelihood, boundness of various partial derivatives; the existence, non-singularity of covariance matrices; and the finiteness of various quadratic forms. Here, there are conditions to ensure that these assumptios are hold. These conditions are all diagonal elements of **W** are zero, the matrices $(I-\rho W)$ is nonsingular for $0 < |\omega_i| < 1$, i=1,2,…,n. The innovations are independent identically distribution, $E(\varepsilon_i) = 0, E(\varepsilon_i{}^2) = \sigma^2$>0, and $E(|\varepsilon|^{4+\eta} < \infty$, for some $\eta$.

We can see that in the equation (3) involve determinant of large matrix, so analytical solution to solve this problem is not easy. Therefore, to avoid analytical solution $ln|I-\rho\boldsymbol{W}|$ , we then use methods Chebyshev polynomial and Taylor series approximation and then compare their accuracy to choose the best performance.

### 3.2 Chebyshev's Polynomial and Taylor series approximation

Chebyshev polynomials are a sequence of orthogonal polynomials which are related to de Moivre's formula and which can be defined recursively. The Chebyshev polynomial approach uses the symmetric equivalent of the neighborhood matrix $\boldsymbol{W}$. The eigenvalues of symmetric matrix W are the same of those of the neighborhood. The Chebyshev solution tries to approximate the log determinant of $(I-\rho W)$ involving a symmetric neighborhood matrix W which is the relationship of the Chebyshev polynomial to the log determinant of $(I-\rho W)$ matrix. Approximation of $\left|I-\rho\widetilde{\boldsymbol{W}}\right|$ :

$$ln\left|I-\rho\widetilde{\boldsymbol{W}}\right| = \sum_{j=1}^{q+1} c_j(\rho)tr\left(T_{j-1}(\widetilde{W})\right) - \frac{1}{2}c_1(\rho), \tag{7}$$

where, $T_0 = diag(n)$, $T_1 = \widetilde{\boldsymbol{W}}$, $T_2 = 2\widetilde{\boldsymbol{W}} - T_0$ , $T_{k+1}(\widetilde{\boldsymbol{W}}) = 2T_k(\widetilde{\boldsymbol{W}}) - T_{k-1}(\widetilde{\boldsymbol{W}})$, and $c_j(\rho) = \frac{2}{q+1}\sum_{k=1}^{q+1} ln\left(1-\rho\cos\left(\frac{\pi\left(k-\frac{1}{2}\right)}{q+1}\right)\right)\cos\left(\frac{\pi\left(j-1\right)\left(k-\frac{1}{2}\right)}{q+1}\right)$. Substitution $ln\left|I-\rho\widetilde{\boldsymbol{W}}\right|$ in (7) to the equation in (6), we obtain

$$ln\left(L(\rho,\sigma^2,\boldsymbol{\beta};\boldsymbol{y})\right) = \sum_{j=1}^{q+1} c_j(\rho)tr\left(T_{j-1}(\widetilde{W})\right) - \frac{1}{2}c_1(\rho) - \frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{(\boldsymbol{y}(\boldsymbol{I}-\rho\boldsymbol{W})-\boldsymbol{X\beta})'(\boldsymbol{y}(\boldsymbol{I}-\rho\boldsymbol{W})-\boldsymbol{X\beta})}{2\sigma^2}. \tag{8}$$

The Taylor's series method is an approximation method for a function that is represented by a series of powers. There are two conditions must be fulfilled, (1) it must have (n+1)$^{th}$ derivative at a point exists and (2) the n$^{th}$ derivative for n to infinite (n→∞) , then residual value close to zero [21]. In the approximation by Taylors's series of log determinant of $(I-\rho W)$ matrix use the powers of the neighborhood matrix ($W$). The matrix $W$ is a stochastic matrix which have main diagonal is 0, nonsingular matrix having a spectral radius of less than 1, $r(I-\rho W) < 1$. Furthermore, the matrix $(I-\rho W)$ can be written by $\exp\left(ln(I-\rho W)\right)$, and $|I-\rho W| = \left|\exp\left(ln(I-\rho W)\right)\right|$. Because of $|I-\rho W| = \left|\exp\left(ln(I-\rho W)\right)\right|$, then $ln|I-\rho W| = tr(\ln(I-\rho W))$, and finally, we have $ln|I-\rho\boldsymbol{W}| = \sum_{k=0}^{\infty}\frac{\rho^k tr(\boldsymbol{W}^k)}{k}$. Therefore, the log likelihood function in (6) can be expressed by

$$ln\left(L(\rho,\sigma^2,\boldsymbol{\beta};\boldsymbol{y})\right) = \sum_{k=0}^{\infty}\frac{\rho^k tr(\boldsymbol{W}^k)}{k} - \frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{(\boldsymbol{y}(\boldsymbol{I}-\rho\boldsymbol{W})-\boldsymbol{X\beta})'(\boldsymbol{y}(\boldsymbol{I}-\rho\boldsymbol{W})-\boldsymbol{X\beta})}{2\sigma^2}. \tag{9}$$

### 3.3 Golden search section algorithm

The *golden section search* is one of algorithm that use to search optimum a function.

(1). Given initial interval $[a_1, b_1]$ and precision e. Set φ=0.618.

Calculate $x_1{}^1 = b_1 - 0.168(b_1 - a_1)$ and $x_2{}^1 = a_1 + 0.168(b_1 - a_1)$, Set i=1,

(2). If $f\left(x_2{}^i\right) > f(x_1{}^i)$,

$a_{i+1} = a_i$
$b_{i+1} = x_2{}^i$
$x_2{}^{i+1} = x_1{}^i$
$x_1{}^{i+1} = b_{i+1} - 0.168(b_{i+1} - a_{i+1})$

if $f\left(x_2{}^i\right) \le f(x_1{}^i)$,

$a_{i+1} = x_1{}^i$
$b_{i+1} = b_i$
$x_1{}^{i+1} = x_2{}^i$
$x_2{}^{i+1} = a_{i+1} + 0.168(b_{i+1} - a_{i+1})$

ISSN: 2005-4238 IJAST

(3) $if\ |b_{i+1} - a_{i+1}| < \varepsilon$ stop, otherwise, set i=i+1, and go to (2)

### 3.4 W-AMOEBA matrix

Following Aldstadt and Getis [1] ,[2], the spatial structure can be considered in two part framework, namely separated spatially and associated data non spatially [1], [2]. Furthermore, the spatial units are clustered using local spatial statistic. Moreover, Aldstadt and Getis used local Getis statistics for clustering spatial units. Local Getis statistics less than d, $G_i(d)$ (hereinafter, abbreviated $G_i$) is denoted by [1], [2], [11], [12], [19] ,[20]

$$G_i = \frac{\sum_{j=1}^{n} w_{ij} x_j}{\sum_{j=1}^{n} x_j},\ i \neq j\ , \tag{10}$$

where $w_{ij} = 1$ when the spatial unit $j$ is within the distance $d$ from unit $i$, and $w_{ij}$=0 for others. Expectation and variation of $G_i$ are, respectively, $E(G_i) = \frac{w_{i.}}{n-1}$, $Var(G_i) = \frac{w_{i.}(n-1-w_{i.})}{(n-1)(n-2)} \cdot \left[\frac{s_{(i)}}{\bar{x}(i)}\right]^2$, where $\bar{x}(i) = \frac{\sum_{j=1}^{n} x_j}{n-1}$ and $s_{(i)} = \frac{\sum_{j=1}^{n} x_j^2}{n-1} - (\bar{x}(i))^2$ [8] ,[9]. Based on Expectation and variation, the standardized of local Getis statistic is given by

$$\boldsymbol{G_i^* = \frac{G_i - E(G_i)}{Var(G_i)}} \tag{11}$$

AMOEBA is an algorithm that can be used to create spatial weighted matrix. In this algorithm, each spatial units are clustered by local Getis statistic. The outlines of the AMOEBA procedure of Aldstadt and Getis [2] are as follows:

(1). compute $G_i^*(0)$ is the value for spatial unit $i$ itself). A $G_i^*(0)$ value greater than zero is indicated that the value at location $i$ is larger than mean of all units, correspondingly, a value less than zero indicates that the value at location $i$ is smaller than the mean of all units.
(2). compute $G_i^*(1)$ $G_i^*(1)$ is the value for each region that contains $i$ and all combinations of its contiguous neighbors. At each succeeding step, contiguous units that are not in the ecotope, they are eliminated from further consideration. Likewise, units include in the ecotope remain in the ecotope, and
(3). These process continue for $k$ number of links,k=2,3,…,$k_{max}$ where $k_{max}$ is determined by the absolute the $G_i^*$. Here, $k_{max}$ is chosen if some addition contiguous units into ecotope cannot improve the absolute $G_i^*(0)$.

If $k_{max}$ is obtained, then we create AMOEBA matrix as follows :

(a) If $k_{max} > 1$, then

$$w_{ij} = \begin{cases} \dfrac{\{P[z \leq G_i^*(k_{max})] - P[z \leq G_i^*(k_j)]\}}{\{P[z \leq G_i^*(k_{max})] - P[z \leq G_i^*(0)]\}}, & 0 < k_j \leq k_{max} \\ 0 & ,\ others \end{cases}$$

(b) If $k_{max} = 1$, then

$$w_{ij} = \begin{cases} 1, for\ k_j = 1 \\ 0,\ \text{others} \end{cases}$$

(c) If $k_{max} = 0$, then

$$w_{ij} = 0\,, \qquad for\ all\ k_j$$

## 4. Numerical simulation and implementation to HDI data

In this section discusses methods used in performance of approximation methods in SAR model and its implementation. First, we generated data to compare approximation methods between Taylor series and Chebyshev polynomial by using Monte Carlo simulation. Here, we use one predictor variable, spatial weight matrix, and take parameters intercept is 1 and slope is 2 for variation ρ and n. Second, we use the best aproximation method to estimate parameter of SAR model on HDI data.

### 4.1. Monte Carlo Simulation

Based on the information, response variable is obtained. The procedure to obtain pair predictor and response variables are below:

(1). $W = \{wij\}$ is given as spatial weights matrix
(2). Here, the parameters $\beta_0 = 1$ , $\beta_1 = 2, \rho = 0.1, 0.2, 0.3, 0.4, 0.5$
(3). Generate predictor variable $X$ and error, $X \sim U(20{,}60)$ and $\varepsilon \sim N(0{,}1)$
(4). Determine y , $y = (I - \rho W)^{-1}(X\beta + \varepsilon)$
(5). Estimate parameter ( $\beta_0$ , $\beta_1$, $\rho$ ) by using MLE method. One is use of Chebyshev polynomial approximation and another I use using Taylor's series approximation.
(6). Estimate RMSE for both models and compare their RMSE's.

For each lag coefficient and sample size, the step (1) – (6) are repeated 100 times and we then compute RMSEs. The process is repeated for different lag coefficient (ρ) and sample size (n). In this, we use sample size n = 20, 40, ..., 250, and $\rho = 0.1, 0.2, 0.3, 0.4, 0.5$ . Hereinafter, we use Chebyshev RMSE's term to describe RMSE of the model and Chebyshev polynomial approximation. Similarly, we use Taylors RMSE's term to describe RMSE of the model and Taylor series approximation. The Chebyshev and Taylor RMSE's of simulation results are presented in Table 1.

**Table 1**. Chebyshev and Taylor RMSE's for different n and ρ

| | n=20 | | n=40 | | n=60 | | n=80 | |
|---|---|---|---|---|---|---|---|---|
| Rho | Chebyshev | Taylor | Chebyshev | Taylor | Chebyshev | Taylor | Chebyshev | Taylor |
| 0.1 | 0.8348 | 0.7408 | 0.7775 | 0.8032 | 1.0047 | 0.8840 | 1.0744 | 0.9040 |
| 0.2 | 1.3281 | 0.8170 | 0.8469 | 0.9785 | 1.0152 | 0.9822 | 1.0192 | 0.9975 |
| 0.3 | 0.7494 | 0.8429 | 0.8565 | 0.9243 | 1.1077 | 1.0001 | 1.0134 | 0.9151 |
| 0.4 | 0.7670 | 1.0631 | 1.0033 | 0.9904 | 1.0212 | 0.9666 | 1.0495 | 0.9984 |
| 0.5 | 0.7119 | 0.9749 | 1.1161 | 0.8148 | 1.0364 | 1.0498 | 0.9467 | 1.0347 |

| | n=100 | | n=150 | | n=200 | | n=250 | |
|---|---|---|---|---|---|---|---|---|
| Rho | Chebyshev | Taylor | Chebyshev | Taylor | Chebyshev | Taylor | Chebyshev | Taylor |
| 0.1 | 0.9216 | 0.9541 | 1.0273 | 0.9799 | 1.0711 | 0.9469 | 0.9873 | 0.9631 |
| 0.2 | 0.9832 | 1.0449 | 1.0493 | 1.0455 | 0.9910 | 1.0422 | 1.0497 | 1.0209 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 0.3 | 1.0223 | 1.0575 | 0.9874 | 0.9261 | 0.9395 | 1.0291 | 0.9717 | 0.9674 |
| 0.4 | 1.0261 | 0.9758 | 0.9659 | 0.9630 | 0.9422 | 1.0182 | 1.0047 | 0.9280 |
| 0.5 | 1.0025 | 1.0411 | 1.0033 | 0.9284 | 1.0343 | 0.9963 | 0.9929 | 0.9697 |

The Chebyshev and Taylor RMSE's were fluctuated for variation n. Furthermore, based on Table 1, we can see that differences between Chebyshev RMSE's and Taylor RMSE's are small. From this simulation results, we can't conclude the best solution. Therefore, for next analysis we use Chebyshev polynomial and Taylor series approximation to approximate ln $|I - \rho W|$in MLE.

### 4.2. Description statistics of HDI and its factors

The data used in this study were taken from BPS statistics (central Bureau of Statistic) of Central Java province. Central Java Province consists of 35 districts. HDI for a collection of geographic areas are commonly display on maps. The Map of Central Java Province is presented in Figure 1. In this study, we use the human development index (HDI) as response variable, population, number of poor people, gross enrollment rate, district minimum wage and poverty line as predictors variables (Table 2). Summary statistics for these variables are shown in Table 3.


**Figure 1** Map of Central Java Province

**Table 2**. HDI data in Central Java province data 2017

| ID | District Name | HDI | Population | Gross enrolment rate | District minimum wage | Poor people | Poverty line | Neighbors |
|---|---|---|---|---|---|---|---|---|
| 1 | Cilacap | 68.9 | 1711627 | 87.28 | 1693690 | 13.94 | 307041 | 2,5,29 |
| 2 | Banyumas | 70.75 | 1665025 | 85.43 | 1461400 | 17.05 | 357748 | 1,3,4,5,27,28,29 |
| 3 | Purbalingga | 67.72 | 916427 | 72.83 | 1522500 | 18.8 | 313343 | 2,4,26,27 |
| 4 | Banjarnegara | 65.86 | 912917 | 66.77 | 1370000 | 17.21 | 264387 | 2,3,5,7,25,26 |
| 5 | Kebumen | 68.29 | 1192007 | 104.89 | 1433900 | 19.6 | 325819 | 1,2,4,6,7 |
| 6 | Purworejo | 71.31 | 714574 | 102.81 | 1445000 | 13.81 | 325871 | 5,7,8 |
| 7 | Wonosobo | 66.89 | 784207 | 52.98 | 1457100 | 20.32 | 308553 | 4,5,6,8,23,24,25 |

ISSN: 2005-4238 IJAST

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 8 | Magelang | 68.39 | 1268396 | 75.56 | 1570000 | 12.42 | 281237 | 6,7,9,22,23, 30 |
| 9 | Boyolali | 72.64 | 974579 | 77.45 | 1519290 | 11.96 | 293405 | 8,10,11,13, 14,15,22,31 |
| 10 | Klaten | 74.25 | 1167401 | 100.58 | 1528500 | 14.15 | 376305 | 9,11 |
| 11 | Sukoharjo | 75.56 | 878374 | 96.11 | 1513000 | 8.75 | 337037 | 9,10,12,13, 31 |
| 12 | Wonogiri | 68.66 | 954706 | 86.58 | 1401000 | 12.9 | 284710 | 11,13 |
| 13 | Karanganyar | 75.22 | 871596 | 83.11 | 1560000 | 12.28 | 340538 | 9,11,12,14, 31 |
| 14 | Sragen | 72.4 | 885122 | 106.49 | 1422590 | 14.02 | 292544 | 9,13,15, |
| 15 | Grobogan | 68.87 | 1365207 | 81.28 | 1435000 | 13.27 | 345379 | 9,14,16,18, 19,21,22 |
| 16 | Blora | 67.52 | 858865 | 84.82 | 1438100 | 13.04 | 291114 | 15,17,18, |
| 17 | Rembang | 68.95 | 628922 | 72.05 | 1408000 | 18.35 | 354440 | 16,18, |
| 18 | Pati | 70.12 | 1246691 | 91.14 | 1420500 | 11.38 | 393817 | 15,16,17,19 ,20, |
| 19 | Kudus | 73.84 | 851478 | 93.35 | 1740900 | 7.59 | 373224 | 15,18,20,21 , |
| 20 | Jepara | 70.79 | 1223198 | 87.05 | 1600000 | 8.12 | 355607 | 18,19,21 |
| 21 | Demak | 70.41 | 1140675 | 91.7 | 1900000 | 13.41 | 371525 | 15,19,20,22 ,33 |
| 22 | Semarang | 73.2 | 1027489 | 78.21 | 1745000 | 7.78 | 317935 | 8,9,15,21,2 3,24,32,33 |
| 23 | Temanggung | 68.34 | 759128 | 70.09 | 1431500 | 11.46 | 277707 | 7,8,22,24 |
| 24 | Kendal | 70.62 | 957024 | 87.1 | 1774800 | 11.1 | 335497 | 7,22,23,25, 33 |
| 25 | Batang | 67.35 | 756079 | 73.93 | 1603000 | 10.8 | 249292 | 4,7,24,26,3 4 |
| 26 | Pekalongan | 68.4 | 886197 | 55.13 | 1583700 | 12.61 | 354435 | 3,4,25,27,3 4 |
| 27 | Pemalang | 65.04 | 1296281 | 71.38 | 1460000 | 17.37 | 331584 | 2,3,26,28 |
| 28 | Tegal | 66.44 | 1433515 | 75.44 | 1487000 | 9.9 | 319758 | 2,29,27,35 |
| 29 | Brebes | 64.86 | 1796004 | 76.51 | 1418100 | 19.14 | 382125 | 2,1,28,35 |
| 30 | Magelang city | 77.84 | 121474 | 107.24 | 1453000 | 8.75 | 450908 | 8 |
| 31 | Surakarta city | 80.85 | 516102 | 103.55 | 1534990 | 10.65 | 448062 | 9,11,13 |
| 32 | Salatiga city | 81.68 | 188928 | 109.61 | 1596850 | 5.07 | 359944 | 22, |
| 33 | Semarang city | 82.01 | 1757686 | 107.82 | 2125000 | 4.62 | 402297 | 21,22,24 |
| 34 | Pekalongan city | 73.77 | 301870 | 92.04 | 1623750 | 7.47 | 390555 | 25,26 |
| 35 | Tegal city | 73.95 | 248094 | 87.08 | 1499500 | 8.11 | 418845 | 29,28 |

**Table 3**. Summary statistics for response (HDI) and predictor variables

| variables | N | Mean | StDev | Min | Q1 | Median | Q3 | Max |
|---|---|---|---|---|---|---|---|---|
| HDI | 35 | 71.19 | 4.48 | 64.86 | 68.29 | 70.41 | 73.84 | 82.01 |
| Population | 35 | 978796 | 421350 | 121474 | 759128 | 916427 | 1246691 | 1796004 |
| district | 35 | 85.58 | 14.54 | 52.98 | 75.44 | 86.58 | 96.11 | 109.61 |

ISSN: 2005-4238 IJAST

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| minimum wage | | | | | | | | |
| district minimum wage | 35 | 1547905 | 157827 | 1370000 | 1435000 | 1513000 | 1600000 | 2125000 |
| The number of Poor People | 35 | 12.49 | 4.13 | 4.62 | 8.75 | 12.42 | 14.15 | 20.32 |
| Poverty line | 35 | 340931 | 48936 | 249292 | 307041 | 337037 | 373224 | 450908 |

### 4.3. Implementation model for the HDI data

For implementation the model to HDI data, we use the SAR model, MLE method, and Chebyshev polynomial approximation with refer to simulation results. In addition, we also use AMOEBA matrix as spatial weighted matrix in SAR model. In above simulation, this matrix isn't involve because it is created by algorithm, so it's time consuming. However, from our previous research [6], the performance of this matrix is better than contiguity matrix.

Quantile map of HDI data in Central Java province are shown in Figure 2. Figure 2 show that dark blue color describe the district which is lowest HDI (64.9), turquoise color describe the district which is second lowest HDI (64.9-66.4), and the dark brown color is describe the district which is highest HDI.

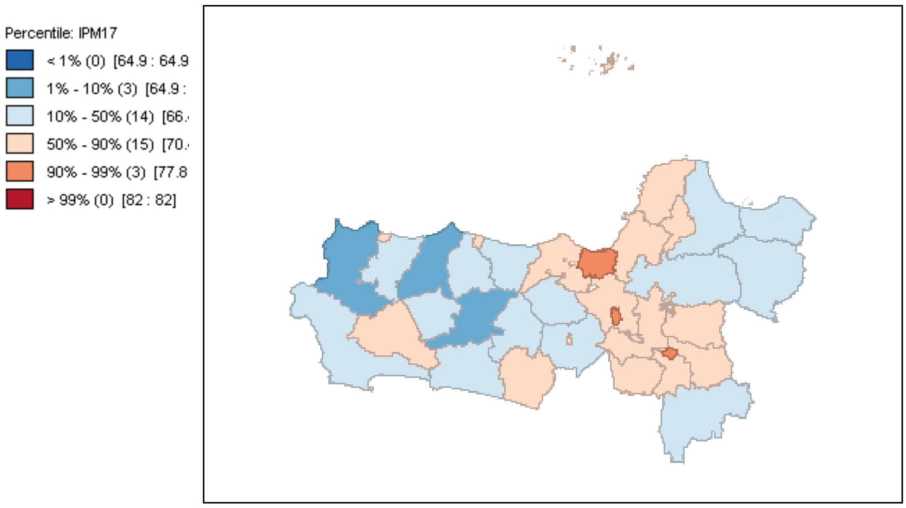


**Figure 2**. Quantile map of HDI data in Central Java Province

From Figure 2, we can see that not all of adjacent districts have similar HDI value. Therefore, this information is a clue to create spatial weight matrix such as W-AMOEBA matrix, because the W-AMOEBA matrix can accommodate both spatial and nonspatial component in the spatial modeling. Furthermore, for modeling HDI and it factors, we use two spatial , W-Contiguity and W-AMOEBA matrices. In addition, to find logarithm determinant of matrix I-ρW, we use Chebyshev polynomial approximation as the best performance as previous analysis (in section 3.1).

Before the W-AMOEBA is created, we explore about significance of HDI in Central Java province. Figure 3 shown that there are three cluster, high, medium and low clusters. The high cluster describes districts which have highest HDI value around them. Meanwhile, the low cluster describes districts which have lowest HDI value around them.

The ANOVA Table of SAR models with W-contiguity and W-AMOEBA are presented in Table 3 and Table 4, respectively. Based on Table 3 and Table 4, we can see that all of predictor variables are significant, so we can say that they are important factors to explain HDI. The coefficients of population and poor people are negative. This means that to improve HDI the population and poor people must be reduced. The coefficients of district minimum wage, district minimum wage and Poverty line are positive impacts. This means that if their factors (district minimum wage, district minimum wage and poverty line) are increase, then HDI are also increase.

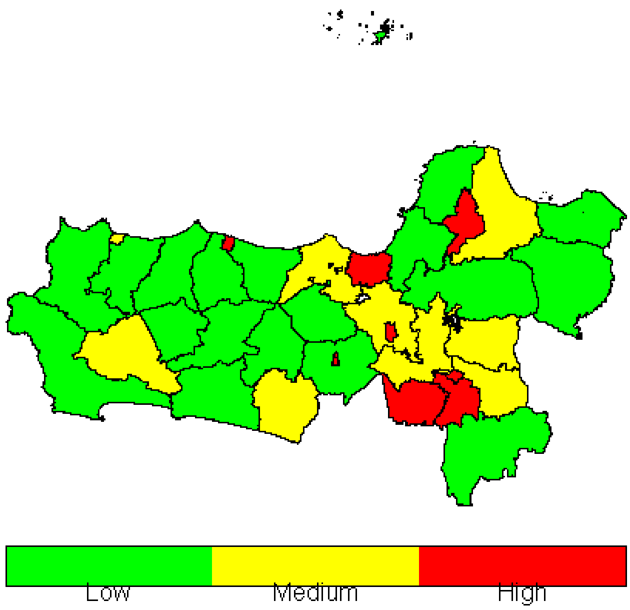


**Figure 3.** Significance of Clustering HDI use local Getis statistic

**Table 3.** ANOVA of SAR Model with W.contiguity

| Coefficients: (asymptotic standard errors) | Estimate | Std. Error | z-value | Pr(>\|z\|) |
|---|---|---|---|---|
| Intercept | 2.8002e+01 | 1.0140e+01 | 2.7614 | 0.0057547 |
| Population | -2.7181e-06 | 9.0475e-07 | -3.0042 | 0.0026625 |
| gross enrollment rate. | 1.0511e-01 | 2.6621e-02 | 3.9484 | 7.868e-05 |
| District minimum wage | 5.5992e-06 | 2.6923e-06 | 2.0797 | 0.0375500 |
| Poor people | -2.3157e-01 | 1.1488e-01 | -2.0157 | 0.0438268 |
| Poverty line | 2.5350e-05 | 7.5799e-06 | 3.3444 | 0.0008247 |

Rho: 0.31856, LR test value: 5.7208, p-value: 0.016765
RMSE : 1.8521
AIC: 159.36, (AIC for lm: 163.08)

**Table 4**. ANOVA of SAR Model with W.AMOEBA

| | Coefficients: Estimate | (asymptotic standard errors) Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | 7.6212e-01 | 8.3596e+00 | 0.0912 | 0.9273597 |
| Population | -2.3122e-06 | 7.8855e-07 | -2.9322 | 0.0033654 |
| gross enrollment rate | 7.8151e-02 | 2.3704e-02 | 3.2970 | 0.0009771 |
| District minimum wage | 6.1336e-06 | 2.3495e-06 | 2.6105 | 0.0090401 |
| Poor people | -1.6678e-01 | 9.7846e-02 | -1.7045 | 0.0882884 |
| Poverty line | 1.6648e-05 | 6.6573e-06 | 2.5008 | 0.0123927 |

Rho: 0.75119, LR test value: 13.354, p-value: 0.0002578
RMSE: 1.6251
AIC: 151.72, (AIC for lm: 163.08)

Evaluating of the performance of SAR model with W-contiguity and W-AMOEBA, then we use criteria RMSE from that models. From the Table 3 and Table 4, we can see that RMSE's of SAR model with W-AMOEBA is smaller than RMSE's of SAR model with W-Contiguity. Therefore, we can say that accuracy of SAR model with W-AMOEBA is better than of SAR model with W-Contiguity.

## 5. Conclusion

The difference in accuracy between Chebyshev polynomial and Taylor series approximations are not significant the same. However, Based on the all simulation for variation lag coefficient and sample size, Chebyshev polynomial slightly better. Therefore, we used Chebyschev polynomial approximation to solve log determinant matrix (I-ρW) in the SAR model for modeling HDI data in Central Java Province.

Mapping plot of HDI values in Central Java Province shown that not all among adjacent districts are similar. Based on this condition, to analysis HDI and their factors in Central Java Province, we used two spatial matrix in the SAR models, W.contiguity and W-AMOEBA matrix. To evaluate the performances of W-contiguity and W-AMOEBA in SAR model, we used Chebyschev polynomial as approximation method in MLE. The RMSE of SAR Model with W.Contiguity and W.AMOEBA are 1.8521 and 1.6251, respectively. Here, we conclude that the best model is SAR model with W-AMOEBA.

Implementation of the selected model showed that to improve or increase HDI in Central Java Province, the population and poor people and increasing must be decreased, meanwhile of gross enrollment rate, district minimum wage and Poverty line of poor people must be increased.

## Acknowledgments

This research was supported by Jenderal Soedirman University (UNSOED). The Authors thank for financial support UNSOED through research fund of research competence. The Authors is greathful for the help Jenderal Soedirman University at funding.

**References**

[1]. J Aldstadt and A Getis, Constructing the Spatial Weights matrices Using Local Statistic. Geographical Analysis: 36 (2004), 90-104.

[2]. J Aldstadt and A Getis, Using AMOEBA to create a spatial weights matrices and identify spatial clusters. Geographical Analysis 8 (2006),327-343.

[3]. L. Anselin, Spatial Econometrics: Methods and Models. Dordecht: Kluwer Academic Publishers, Netherlands, 1988.

[4]. L. Anselin, Local indicators of spatial association-LISA.Geographical Analysis 27 1995, 93-115.

[5]. LJ Bain and M Engelhardt, Introduction to Probability and Mathematical Statistics. Brooks/Cole, 1992.

[6]. BPS-Statistics of Jawa Tengah Province, Jawa Tengah Province in Figures, 2016.

[7]. G Casella and R Berger, Statistical inference. Wadsworth, Belmont. California, (1990)

[8]. NAC Cressie, Statistics for Spatial Data, John Wiley and Sons, New York, (1993)

[9]. KM Cubukcu, The spatial distribution of economic base multipliers: A GIS and spatial statistics-based cluster analysis. ITU A Z. 2 (K M. 2011), 49-62.

[10]. H Folmer H and JHL Oud, How to get rid of W: a Latent variables approach to modeling spatially lagged variables, Environment and Planning A 40 (2008), 2526–2538

[11]. C Gaetan and X Guyon, Spatial Statistics and Modelling, John Wiley & Sons, New York, (2010)

[12]. A Getis and JK Ord, The Analysis of Spatial Association by Use of Distance Statistics. Geographical Analysis 24 (1992), 189-206.

[13]. Jajang, A. Saefuddin, I. W. Mangku and H. Siregar, Normal asymptotic of modified local Getis statistic, Far East J. Math. Sci. (FJMS) 80 (2013), 155-167.

[14]. Jajang, A Saefuddin A, IW Mangku IW and H Siregar, Comparing Performances of WG, WGnew and WC on Dynamic Spatial Panel Model By Monte Carlo Simulation. Far East Journal Of Mathematical Sciences, 2 (2014), 155-167.

[15]. HH Kelejian and I.R.Prucha, A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances, Journal of Real Estate Finance and Economics, 17(1998): 99-121.

[16]. A Liu, H Folmer and JHL Oud, W-Based vs Latent Variables Spatial Autoregressive Models: Evidence from Monte Carlo Simulation, Ann Reg Sci. 47(2011a), 619–639.

[17]. A Liu, H Folmer and JHL Oud, Estimating regression coefficients by W-based and latent variables spatial autoregressive models in the presence of spillovers from hotspots : evidence from Monte Carlo simulations. Lett Spat Resour Sei. 4 (2011b), 71-80.

[18]. K Ord, Estimation Methods for Models of Spatial Interaction, Journal of the American Statistical Association 70(1975) :120-126.

[19]. JK Ord and A Getis, Local spatial autocorrelation statistics: distributional issues and an application .Journal Geographical Analysis 27 (1995), 286-306.

[20]. JK Ord and A Getis, Testing for local autocorrelation in the presence of global autocorrelation, Journal of Regional Science 41(2001): 411-43.

[21]. DE Vargerg and EJ Purcell. Calculus With Analytic Geometry. 6th Edition, Prentice Hall, (1991).

[22]. MD Ward and SG Kristiani, Spatial Regression Models Sereies: Quantitative Application in the Social Science, California, Sage Publications Inc,( 2008).

ISSN: 2005-4238 IJAST