

# Comparison of Uniform and Kernel Gaussian Weight Matrix in Generalized Spatial Panel Data Model

Tuti Purwaningsih, Erfiani, Anik Djuraidah

Departement of Statistics, Graduate School of Bogor Agricultural University, Bogor, Indonesia  
Email: [purwaningsiht@yahoo.com](mailto:purwaningsiht@yahoo.com)

Received 7 February 2015; accepted 25 February 2015; published 28 February 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc.  
This work is licensed under the Creative Commons Attribution International License (CC BY).  
<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

Panel data combine cross-section data and time series data. If the cross-section is locations, there is a need to check the correlation among locations.  $\rho$  and  $\lambda$  are parameters in generalized spatial model to cover effect of correlation between locations. Value of  $\rho$  or  $\lambda$  will influence the goodness of fit model, so it is important to make parameter estimation. The effect of another location is covered by making contiguity matrix until it gets spatial weighted matrix ( $W$ ). There are some types of  $W$ —uniform  $W$ , binary  $W$ , kernel Gaussian  $W$  and some  $W$  from real case of economics condition or transportation condition from locations. This study is aimed to compare uniform  $W$  and kernel Gaussian  $W$  in spatial panel data model using RMSE value. The result of analysis showed that uniform weight had RMSE value less than kernel Gaussian model. Uniform  $W$  had stabil value for all the combinations.

## Keywords

Component, Uniform Weight, Kernel Gaussian Weight, Generalized Spatial Panel Data Model

---

## 1. Introduction

Panel data analysis combines cross-section data and time series data, in sampling when the data are taken from different locations. It's commonly found that the observation value at one location depends on observation value in another location. In the other name, there is spatial correlation between the observations, which is spatial dependence. Spatial dependence in this study is covered by generalized spatial model which is focussed on dependence between locations and errors [1]. If there is spatial influence but not involved in model so error assumption that between observations must be independent will not fulfilled. So the model will be in bad condi-

tion, for that need, a model that involves spatial influence in the analysis panel data will be mentioned as Spatial Panel Data Model.

Some recent literature of spatial cross-section data is Spatial Ordinal Logistic Regression by Aidi and Purwaningsih [2], and Geographically Weighted Regression [3]. Some of the recent literature of Spatial Panel Data is forecasting with spatial panel data [3] and spatial panel models [4]. For accomodating spatial dependence in the model, there is spatial weighted matrix ( $W$ ) that is an important component to calculate the spatial correlation between locations. Spatial parameter in generalized spatial panel data model, is known as  $\rho$  or  $\lambda$ . There are some types of  $W$ —uniform  $W$ , binary  $W$ , inverse distance  $W$  and some  $W$  from real cases of economics condition or transportation condition from the area. This research is aimed to compare uniform  $W$  and kernel Gaussian  $W$  in generalized spatial panel data model using RMSE value which is obtained from simulation.

## 2. Literature Review

### 2.1. Data Panel Analysis

Data used in the panel data model is a combination of cross section and time-series data. Crosssection data is data collected at one time of many units of observation, then time-series data is data collected over time to an observation. If each unit has a number of observations a cross individuals in the same period of time series, it is called a balanced panel data. Conversely, if each individual unit has a number of observations a cross different period of time series, it is called an unbalanced panel data (unbalanced panel data).

In general, panel data regression model is expressed as follows:

$$y_{it} = \alpha + \mathbf{x}'_{it}\boldsymbol{\beta} + u_{it} \quad i = 1, 2, \dots, N ; t = 1, 2, \dots, T \quad (1)$$

with  $i$  is an index for crossection data and  $t$  is index of time series.  $\alpha$  is a constant value,  $\boldsymbol{\beta}$  is a vector of size  $K \times 1$ , with  $K$  specifies the number of explanatory variables. Then  $y_{it}$  is the response to the individual cross- $i$  for all time period stand  $\mathbf{x}_{it}$  are sized  $K \times 1$  vector for observation  $i$ -th individual cross and all time periods  $t$  and  $u_{it}$  is the residual/error [5].

Residual components of the direction of the regression model in Equation (1) can be defined as follows:

$$u_{it} = \mu_i + \varepsilon_{it} \quad (2)$$

where  $\mu_i$  is an individual-specific effect that is not observed, and  $\varepsilon_{it}$  is a remnant of crossection- $i$  and time series- $t$  [5].

### 2.2. Spatial Weighted Matrix ( $W$ )

Spatial weighted matrix is basically a matrix that describes the relationship between regions and obtained by distance or neighbourhood information. Diagonal of the matrix is generally filled with zero value. Since the weighting matrix shows the relationship between the overall observation, the dimension of this matrix is  $N \times N$  [6]. There are several approaches that can be done to show the spatial relationship between the location, including the concept of intersection (contiguity). There are three types of intersection, namely Rook Contiguity, Bishop Contiguity and Queen Contiguity [6].

After determining the spatial weighting matrix to be used, further normalization in the spatial weighting matrix. In general, the matrix used for normalization row (row-normalize). This means that the matrix is transformed so that the sum of each row of the matrix becomes equal to one. There are other alternatives in the normalization of this matrix is to normalize the columns of the matrix so that the sum of each column in the weighting matrix be equal to one. Also, it can also perform normalization by dividing the elements of the weighting matrix with the largest characteristic root of the matrix ([6] [7]).

There are several types of Spatial Weight ( $W$ ): binary  $W$ , uniform  $W$ , inverse distance  $W$  (non uniform weight) and some  $W$  from real case of economics condition or transportation condition from the area. Binary weight matrix has values 0 and 1 in off-diagonal entries; uniform weight is determined by the number of sites surrounding a certain site in  $\ell$ -th spatial order; and non-uniform weight gives unequal weight for different sites. The element of the uniform weight matrix is formulated as,

$$W_{ij} = \begin{cases} \frac{1}{n_i^{(l)}}, & j \text{ is neighbor of } i \text{ in } l\text{-th order} \\ 0, & \text{others} \end{cases} \quad (3)$$

$n_i^{(l)}$  is the number of neighbor locations with site- $i$  in  $l$ -th order. The non-uniform weight may become uniform weight when some conditions are met. One method in building non-uniform weight is based on inverse distance. The weight matrix of spatial lag  $k$  is based on the inverse weights  $1/(1+d_{ij})$  for sites  $i$  and  $j$  whose Euclidean distance  $d_{ij}$  lies within a fixed distance range, and otherwise is weight zero. Kernel Gaussian Weight follow this formula:

$$w_j(i) = \exp\left[-1/2(d_{ij}/b)^2\right] \quad (4)$$

with  $d$  is distance between location  $i$  and  $j$ , then  $b$  is *bandwidth* which is a parameter for smoothing function.

### 2.3. Generalized Spatial Panel Data Model

Generalized spatial model expressed in the following equation:

$$y_{it} = \rho \sum_{j=1}^N w_{ij} y_{jt} + \mathbf{x}'_i \boldsymbol{\beta} + \mu_i + \phi_{it} \quad \text{dengan} \quad \phi_{it} = \lambda \sum_{j=1}^N w_{ij} \phi_{jt} + \varepsilon_{it} \quad (5)$$

where  $\rho$  is spatial autoregressive coefficient,  $w_{ij}$  is elements of the spatial weighted matrix which has been normalized ( $W$ ) and  $\lambda$  is spatial autocorrelation between error [7].

## 3. Methodology

Data used in this study was gotten from simulation using generalized spatial panel data model as Equation (5) with initiation of some parameter. Simulation was done use R program. The following step is used to generate the spatial data panel which is consist of index  $n$  and  $t$ . In dexn indicates the number of locations and indext indicates the number of period in each locations. Here is the process:

- 1) Determining the number of locations to be simulated is  $N = 3$ ,  $N = 9$  and  $N = 25$ .
- 2) Makes 3 types of map location on step 1.
- 3) Creating a binary spatial weighted matrix based on the concept of queen contiguity of each type of map locations. In this step, to map the 3 locations it will form a  $3 \times 3$  matrix, 9 locations will form a  $9 \times 9$  matrix and 25 locations form a  $25 \times 25$  matrix.
- 4) Creating spatial uniform weighted matrix based on the concept of queen contiguity of each type of map locations.
- 5) Making weighted matrix kernel Gaussian based on the concept of distance. To make this matrix, previously researchers randomize the centroid points of each location. After setting centroid points, then measure the distance between centroids and used it as a reference to build kernel Gaussian  $W$ . Gaussian kernel  $W$  as follows:

$$w_j(i) = \exp\left[-\frac{1}{2}\left(d_{ij}/b\right)^2\right] \quad [3].$$

- 6) Specifies the number of time periods to be simulated is  $T = 3$ ,  $T = 6$ ,  $T = 12$  and  $T = 24$ .
- 7) Generating the data  $Y$  and  $X$  based on generalized spatial panel data models follows Equation (5).
- 8) Cronecker multiplication between matrix identity of time periods and  $W$ , then get new matrix named  $IW$ .
- 9) Multiply matrix  $IW$  and  $Y$  to obtain vector  $WY$ .
- 10) Build a spatial panel data models and get the value of RMSE.
- 11) Repeat steps 7)-9) until 1000 replications for each combination on types of  $W$ ,  $N$ ,  $T$ ,  $\rho$  and  $\lambda$ .

Description:

Types of  $W$ :  $W$  binary,  $W$  uniform and Gaussian kernel  $W$ ;

Types of  $N$ : 3, 9 and 25 locations;

Types of  $T$ : 3, 6, 12 and 36 series;

Types of  $\rho = 0.3, 0.5, 0.8$  and  $\lambda = 0.3, 0.5, 0.8$ .

12) Get the RMSE value for all of 1000 replicationsoh each combination between  $W$ ,  $N$ ,  $\rho$  and  $\lambda$ .

13) Determine the best  $W$  based on the smallest RMSE for all combinations.

#### 4. Results and Discussions

Simulation generate data for vector  $Y$  as dependent variable and  $X$  matrix as independent variable.  $Y$  and  $X$  is generate with parameter initiation. After doing simulation, we can get RMSE for each combinations and processing it, then we can calculate RMSE for each  $W$ ,  $N$ ,  $T$ ,  $\rho$  and  $\lambda$ . Here is the result. With the result in **Table 1** then continued to figure it into graphs in order to look the comparison easily.

**Table 1.** Value of RMSE resulted from simulation for all the combinations ( $W$ ,  $N$ ,  $T$ ,  $\rho$  and  $\lambda$ ).

W types	Location types	Periods types	Generalized spatial panel data model			Average RMSE	Average RMSE
			$\rho = 0.3,$ $\lambda = 0.3$	$\rho = 0.5,$ $\lambda = 0.5$	$\rho = 0.8,$ $\lambda = 0.8$		
Uniform $W$	$N = 3$	$T = 3$	1.076	1.23	2.06	1.771	1.634
		$T = 6$	1.223	1.387	2.684		
		$T = 12$	1.251	1.464	2.957		
		$T = 36$	1.296	1.524	3.099		
		Average	1.211	1.401	2.7		
	$N = 9$	$T = 3$	1.293	1.365	1.775	1.578	
		$T = 6$	1.341	1.401	1.976		
		$T = 12$	1.357	1.429	2.054		
		$T = 36$	1.362	1.448	2.139		
		Average	1.338	1.411	1.986		
	$N = 25$	$T = 3$	1.383	1.433	1.755	1.553	
		$T = 6$	1.397	1.446	1.812		
$T = 12$		1.403	1.467	1.843			
$T = 36$		1.407	1.409	1.877			
Average		1.398	1.439	1.822			
Kernel Gaussian $W$	$N = 3$	$T = 3$	1.137	1.137	1.137	1.748	
		$T = 6$	1.352	1.352	1.806		
		$T = 12$	1.405	2.971	2.014		
		$T = 36$	1.461	3.098	2.11		
		Average	1.339	2.14	1.767		
	$N = 9$	$T = 3$	2.101	1.115	1.056	1.243	
		$T = 6$	1.353	1.138	1.097		
		$T = 12$	1.255	1.15	1.106		
		$T = 36$	1.261	1.161	1.119		
		Average	1.493	1.141	1.095		
$N = 25$	$T = 3$	1.49	1.282	1.168	2.436		
	$T = 6$	5.705	1.286	1.169			
	$T = 12$	6.004	1.293	1.177			
	$T = 36$	6.19	1.294	1.179			
	Average	4.847	1.289	1.173			

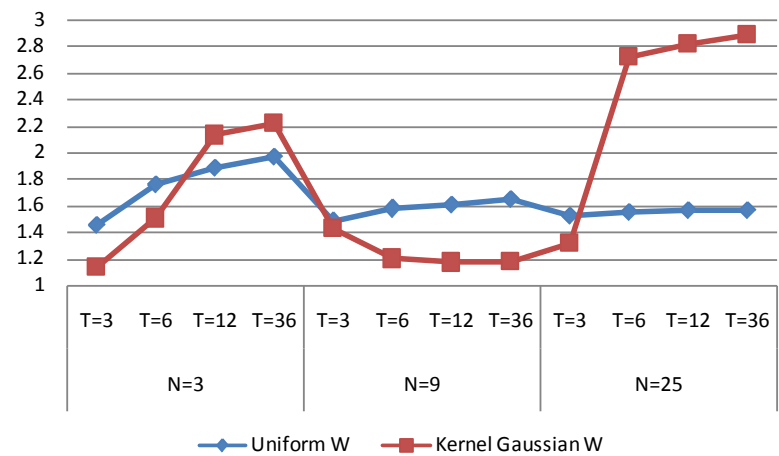


Figure 1. Comparison of RMSE between uniform W and kernel Gaussian W for all combinations.

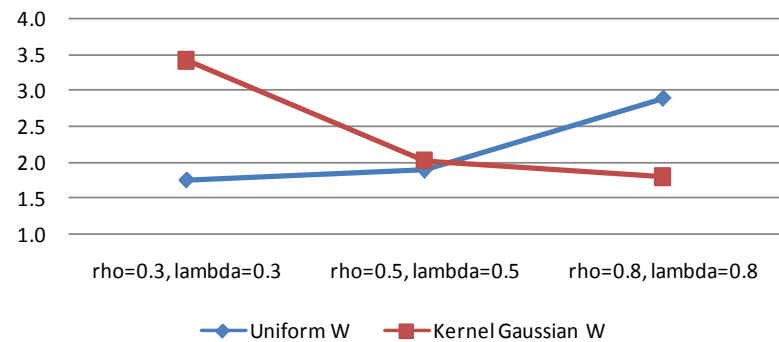


Figure 2. Comparison RMSE each W for each parameter.

Based on Figure 1 can be said that uniform W has smaller RMSE than kernel Gaussian W for  $T = 12, T = 36$  on location  $N = 3$ , then for  $T = 6, 12, 36$  on location  $N = 25$  and the remaining combinations, kernel Gaussian is higher. If we look the level of stabilization, uniform W is better than kernel Gaussian W. We can look at the graph in blue line as uniform W, it has value only in range 1, 4 until 2 then kernel Gaussian W has range from 1 - 3. So can be concluded that uniform W is better than kernel Gaussian W.

Based on Figure 2, we can look that average RMSE of uniform W is smaller in  $\rho = 0.3, \lambda = 0.3$  and  $\rho = 0.5, \lambda = 0.5$  while kernel Gaussian W is smaller only in  $\rho = 0.8, \lambda = 0.8$ .

### 5. Conclusion

After looking at the result, it can be concluded that uniform W is better than kernel Gaussian W almost for all combinations of  $N$  and  $T$ . Then uniform W is better in  $\rho$  and  $\lambda$  in small value until medium (less than 0.5).

### Acknowledgements

The first, authors would like to thankful to Allah SWT, my parents, lecturer and all of friends. This research was supported by private funds.

### References

- [1] Anselin, L., Gallo, J. and Jayet, H. (2008) The Econometrics of Panel Data. Springer, Berlin.
- [2] Aidi, M.N. and Purwaningsih, T. (2012) Modelling Spatial Ordinal Logistic Regression and the Principal Component to Predict Poverty Status of Districts in Java Island. *International Journal of Statistics and Application*, 3, 1-8.
- [3] Fotheringham, A.S., Brunson, C. and Charlton, M. (2002) Geographically Weighted Regression, the Analysis of Spa-

tially Varying Relationships. John Wiley and Sons, Ltd., Hoboken.

- [4] Elhorst, J.P. (2011) Spatial Panel Models. Regional Science and Urban Econometric.
- [5] Baltagi, B.H. (2005) Econometrics Analysis of Panel Data. 3rd Edition, John Wiley and Sons, Ltd., England.
- [6] Dubin, R. (2009) Spatial Weights. In: Fotheringham, A.S. and Rogerson, P.A., Eds., *Handbook of Spatial Analysis*, Sage Publications, London. <http://dx.doi.org/10.4135/9780857020130.n8>
- [7] Elhorst, J.P. (2010) Spatial Panel Data Models. In: Fischer, M.M. and Getis, A., Eds., *Handbook of Applied Spatial Analysis*, Springer, New York. [http://dx.doi.org/10.1007/978-3-642-03647-7\\_19](http://dx.doi.org/10.1007/978-3-642-03647-7_19)

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either [submit@scirp.org](mailto:submit@scirp.org) or [Online Submission Portal](#).

