VOL. 11, NO. 24, DECEMBER 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

THE COMPARISON OF SPATIAL ECONOMETRIC MODELS TO ESTIMATE SPILLOVER EFFECT BY MEANS OF MONTE CARLO SIMULATION

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

ABSTRACT

The spatial econometrics models has been developed covering from estimation method, selection of appropriate weight matrix (W) and the issue of spatial spillover. In regional science, spatial spillovers are the main interest. They can be defined as the impact of changes to explanatory variables in a particular unit i on the dependent variable values in other units j (i≠j). The spatial spillover appears because there are endogenous interaction effects among the dependent variable and exogenous interaction effects among the explanatory variables. There are several spatial econometrics model that have been used to cover those interaction effects i.e.: General Nesting Spatial model (GNS), Spatial Lag Combined model (SAC), Spatial Durbin Error Model (SDEM), Spatial Lag Model (SEM) and Spatial Lag of X model (SLX). Compared to the other models, the last model is the simplest model proposed to estimate spatial spillover effect Vega *et al* (2015) have compared this model to estimate the spillover effect based on empirical study. This paper tries to make a general conclusion of a better model to estimate the spillover effect based on minimum bias and mean square error of regression and spillover effect. The comparison has been done by Monte Carlo Simulation using several conditions. The SDM model is the simplest model which can explain the spillover effect easier and results in minimum bias.

Keyword: spatial econometric, SLX, monte carlo, spillover.

1. INTRODUCTION

The spatial econometrics models have been developed covering from estimation method, selection of appropriate weight matrix (**W**) and the issue of spatial spillover. The focus of spatial econometrics models is specifying, estimating and testing for the presence of spatial interaction [1]. In the specifying model, often a wide gap between the theoreticians and practitioners. Theoreticians are mainly interested in the model containing endogenous interaction effects among the dependent variable, exogenous interaction effects among the independent variables, and interaction effect among the error terms or a combination of all interaction. The model has been developed are General Nesting Spatial model (GNS), Spatial Durbin Error Model (SDEM), Spatial Lag Combined model (SAC), Spatial Durbin Model (SDM), Spatial Lag Model (SEM), Spatial Error Model (SEM) [2]. Otherwise, practitioners are interested in a simple model that ignores some interaction components. Their focus is choosing a model that easily computed and interpreted. The problem arises when there is no valid statistical testing to select the better model.

In regional science, spatial spillovers are the main interest. They can be defined as the impact of changes to explanatory variables in a particular unit i on the dependent variable values in other units j ($i \neq j$). The spatial spillover appears because there are endogenous interaction effects among the dependent variable and exogenous interaction effects among the explanatory variables.

This paper tries to make a general conclusion of a better model to estimate the spillover effect based on minimum bias and mean square error of regression and spillover effect. The comparison has been done by Monte Carlo Simulation using several conditions.

2. SPATIAL ECONOMETRICS MODELS

Specification model

Spatial econometrics models usually use to modeling spatial dependence in economic data. Its models aim to take account the spatial dependence component to the model, therefore, the value of the response variable in one region does not only depends on own covariates but also the value of response variable in neighboring regions [3].

By the contagious characteristics of the spatial data, the spatial econometrics model is appropriate to apply in econometrics analysis for modeling the spatial pattern of spatial data.

Commonly the spatial econometrics models are easy to apply for continuous data. The spatial dependence is modeled explicitly; the value of response variable at an observed region is centered at a weighted average of the observed values at its neighbors and also the effect of other covariates at the observation. Several models have been developed to modeling spatial dependence. Spatial Lag Model (SLM), Spatial Error Model (SEM) and Spatial Durbin Model (SDM) are the three popular models [4]

The response variable **y** in SLM is modeled as depending on a weighted sum of the responses at their neighbours, plus a linear term on covariates and an error term:

$$\mathbf{y} = \rho_{\mathrm{Lag}}\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}; \qquad \boldsymbol{\varepsilon} \sim \mathrm{MVN}(0, \sigma^2\mathbf{I}_n) \tag{1}$$

VOL. 11, NO. 24, DECEMBER 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

where $\mathbf{y} = (y_1, \dots, y_n)$ is the vector of response variable, X is a design matrix of $k+1$ covariates include constant term, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)$ are the coefficient of the covariates include intercept, $\mathbf{I}_n$ is the identity matrix of dimension nxn, and $\mathbf{W}$ is a row-standardized adjacency matrix which, ρ_{Lag} is parameter of spatial lag that measure spatial autocorrelation. The error term $\boldsymbol{\varepsilon}$ is follows multivariate normal (MVN) distribution with zero mean and with constants variance $\sigma^2$for each region.
The SLM model can be written in reducing form as:

$$\mathbf{y} = (\mathbf{I}_\mathrm{n} - \rho_\mathrm{Lag}\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad (2)$$

with:

$$\mathbf{e} \sim \mathrm{MVN}(0, \sigma^2(\mathbf{I}_\mathrm{n} - \rho_\mathrm{Lag}\mathbf{W})^{-1}(\mathbf{I}_\mathrm{n} - \rho_\mathrm{Lag}\mathbf{W}')^{-1})$$

The second model in SEM, in this model, the error term is modeled as depending on a weighted sum of the error term (δ) at their neighbours plus some random noise $\boldsymbol{\varepsilon}$:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\delta} (3)$$

with $\boldsymbol{\delta} = \rho_{Err}\boldsymbol{W} + \boldsymbol{\varepsilon}$; $\quad \boldsymbol{\varepsilon} \sim MVN(0, \sigma^2\mathbf{I}_n)$

Where the noise ε is follows multivariate normal (MVN) distribution with zero mean and with constants variance $\sigma^2$for each region.

The SEM model can be written in reduce form as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad (4)$$

with

$$\mathbf{e} \sim \mathrm{MVN}(0, \sigma^2(\mathbf{I}_\mathrm{n} - \rho_\mathrm{Err}\mathbf{W})^{-1}(\mathbf{I}_\mathrm{n} - \rho_\mathrm{Err}\mathbf{W}')^{-1})$$

The third model that is usually used in Spatial Econometrics is SDM. In this model, the response variable is not only modeled as depending on a weighted sum of the response at neighbors and covariates, but also on the weight values of the covariates at neighbors. The Spatial Durbin model can be expressed as a Spatial Lag model as follows:

$$\mathbf{y} = \rho_\mathrm{Lag}\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \quad (5)$$

with: $\boldsymbol{\varepsilon} \sim MVN(0, \sigma^2\mathbf{I}_\mathrm{n})$

Where $\boldsymbol{\gamma}$ is a vector of coefficients for the spatially lagged covariates $\mathbf{WX}$,

$$\mathbf{y} = \boldsymbol{\rho}_\mathbf{Lag}\mathbf{W}\mathbf{y} + \mathbf{X}^*\boldsymbol{\beta}^* + \boldsymbol{\varepsilon} \quad (6)$$

with $\mathbf{X}^* = [\mathbf{X}, \mathbf{WX}]$; $\quad \boldsymbol{\beta}^* = [\boldsymbol{\beta}, \boldsymbol{\gamma}]$and

$$\mathbf{y} = \left(\mathbf{I}_\mathbf{n} - \boldsymbol{\rho}_\mathbf{Lag}\mathbf{W}\right)\mathbf{X}^*\boldsymbol{\beta}^* + \mathbf{e} \quad (7)$$

Those model can be developed become eight models based on deletion or added the parameters spatial lag dependencies, spatial lag error, and spatially lagged covariates.

Estimation of the parameters model

Several methods are usually used to estimate the parameter model, i.e., maximum likelihood method (ML), Instrumental Variable with Two Stage Least Square (2LS) and Bayesian Method [5]

For ML method we obtain the parameter estimates for SLM based on maximizing the likelihood function bellow:

$$\max_\rho \ln \mathrm{L_c}(\rho) = \mathrm{C} - \frac{\mathrm{n}}{2}\ln\left[\frac{1}{\mathrm{n}}(\mathbf{e_0} - \rho\mathbf{e_L})^\mathrm{T}(\mathbf{e_0} - \rho\mathbf{e_L})\right] + \ln|\mathbf{I} - \rho\mathbf{W}| \quad (8)$$

and for the SEM model we use

$$\ln \mathrm{L}(\lambda) = \mathrm{C} - \frac{\mathrm{n}}{2}\ln\frac{1}{\mathrm{n}}\mathbf{e}^\mathrm{T}(\mathbf{I} - \lambda\mathbf{W})^\mathrm{T}(\mathbf{I} - \lambda\mathbf{W})\mathbf{e} + \ln|\boldsymbol{I} - \lambda\boldsymbol{W}| \ (9)$$

We solve those likelihood functions to obtain the parameter estimates of the other models

Specification test

The presence of spatial error autocorrelation can be assessed by the Moran Test applied on the residuals of the standard regression model. However, this test does not inform the better model should be chosen. No specific spatial test is available for the spatial cross-regressive model. F-test for linear restrictions on the regression coefficient is suggested to identify spatial autocorrelation due to omitted lagged exogenous variable [6]. For the SEM and SEM models, Lagrange Multiplier (LM) test is generally used [5].

a) The classical and robust LM lag test
b) The classical and robust LM error test

Those tests are very limited. They only use for SEM and SEM specification. There are no general tests for all the models. Therefore we use simulation method to obtain the robust model for several conditions.

Direct and indirect effect

The effect of exogenous variables to the endogenous variable can be divided become two components, i.e., direct and indirect effect (spillover effect). Spillover effect is computed based endogenous interaction effects among the dependent variable, exogenous interaction effects among the independent variables.

ARPN Journal of Engineering and Applied Sciences

Table-1. Direct and spillover effects of different model specifications (Elhorst, 2014).

	Direct Effects (γ)	**Indirect Effect** (δ)
OLS/ SEM	β_k	0
SEM/ SAC	Diagonal elements of $(\mathbf{I_n} - \boldsymbol{\rho}\mathbf{W})^{-1}\boldsymbol{\beta}_\mathbf{k}$	Off-diagonal elements $(\mathbf{I_n} - \boldsymbol{\rho}\mathbf{W})^{-1}\boldsymbol{\beta}_\mathbf{k}$
SLX/ SDEM	β_k	θ_k
SDM/ GNS	Diagonal elements of $(\mathbf{I_n} - \boldsymbol{\rho}\mathbf{W})^{-1} \cdot (\beta_k + \mathbf{W}\boldsymbol{\theta}_\mathbf{k})$	Off-diagonal elements $(\mathbf{I_n} - \boldsymbol{\rho}\mathbf{W})^{-1} \cdot (\beta_k + \mathbf{W}\boldsymbol{\theta}_\mathbf{k})$

3. SIMULATION STUDY

The Monte Carlo design experiment is conducted using the standard procedure typically used to evaluate spatial econometric estimators. First, four square maps are generated with $N = \{25,\ 64,\ 100,\ 400\}$ number region. The four different numbers of N represent four categories, i.e.: small, medium, large, and very large observe numbers. Each area is indexed $i = (X, Y)$, for $X = 1,2,..,m$ and $Y = 1,2,...,m$ with m as a square root of N. We use two different types of matrix, i.e.: contiguity based and distance based. The continuity weight matrix is generated by calculating the Euclidean distance (d_{ij}) among the regions. The element matrix of $\mathbf{W}$ is equal to one if the distance among regions is one; and zero for the others. While for the spatial weight matrix based on distance, we used directly the Euclidean distance. The data generating process is defined by

$$y = \rho Wy + X\beta + WX\theta + u;\ \ u = \lambda Wu + \varepsilon \tag{10}$$

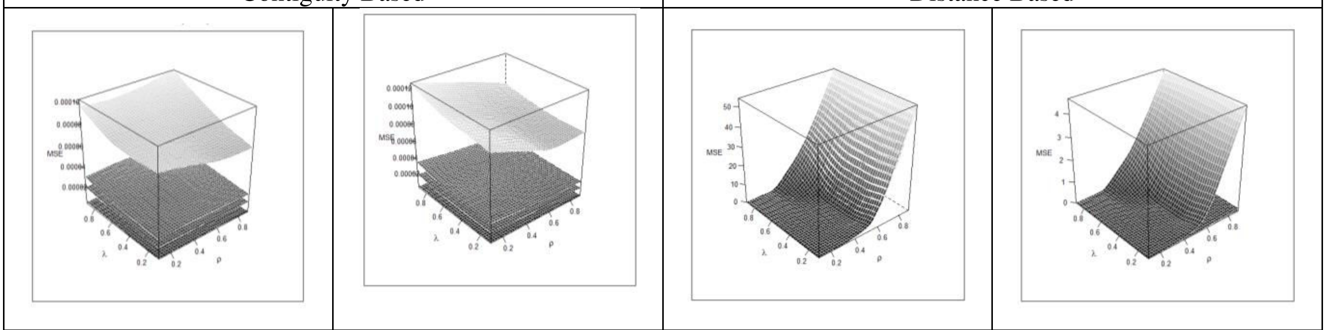
Where the $\mathbf{X}$ is aNx 2 matrix of explanatory variables without a constant term, i.e.: the model without an intercept. The explanatory variables are drawn from the normal distribution. The first explanatory variable is $x_1 \sim N(30,3)$ and second is $x_2 \sim N(50,5)$. The error term ε is drawn from the standard normal distribution with $\sigma = 0.1$. The parameter regressions are $\beta_1 = \beta_2 = 1$ and the parameter of spillover effects are $\theta_1 = \theta_2 = 1$. We take value $0 \leq (\rho,\lambda) \leq 1$. The best model is a model that has a minimum value of Mean Square error for each parameter estimate.

We used Lasso smoothing regression to present the output in three-dimension graph. The graphs show the means square error of the parameter estimates related to the value of spatial lag dependencies, spatial lag error, and sample size.

The MSE of β_1 estimate

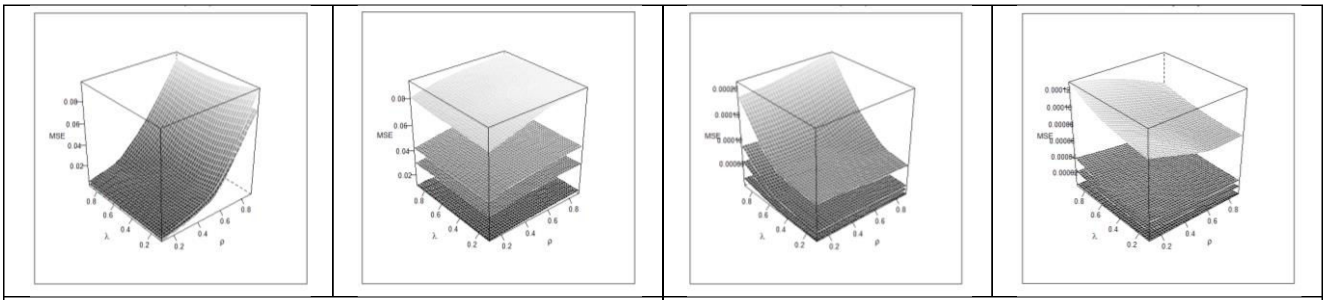
The spillover effect is a function of regression parameters. The MSE of the regression parameter will influence the MSE of the spillover effect. The first analysis of the quality estimate of the slop regression β. For the preliminary result, we only present the regression coefficient of β_1.

Contiguity Based | Distance Based

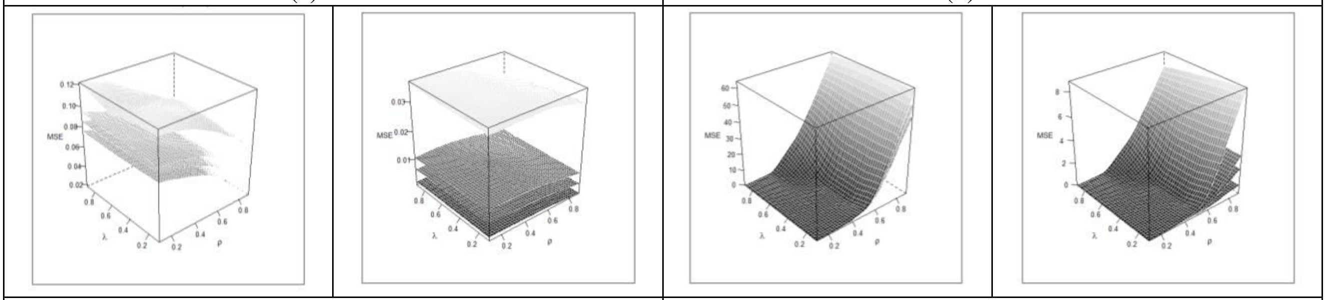


(a) GNS | (b) SDEM

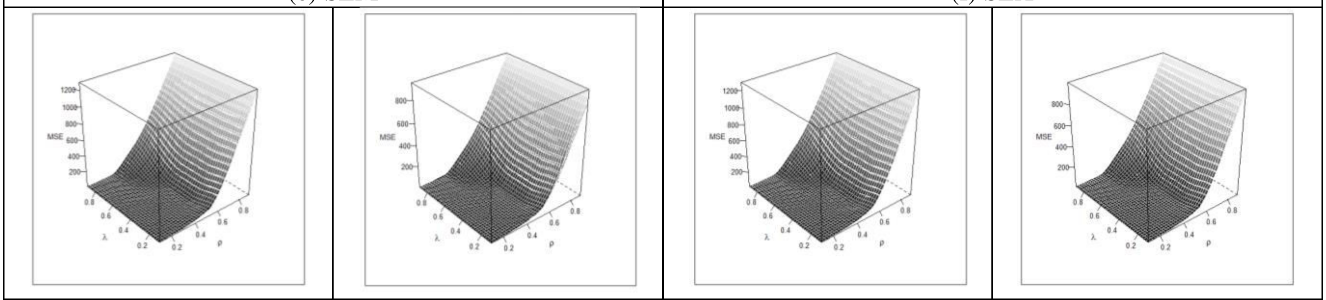
VOL. 11, NO. 24, DECEMBER 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences



(c) SAC (d) SDM



(e) SEM (f) SLX



(g) SEM (h) LM

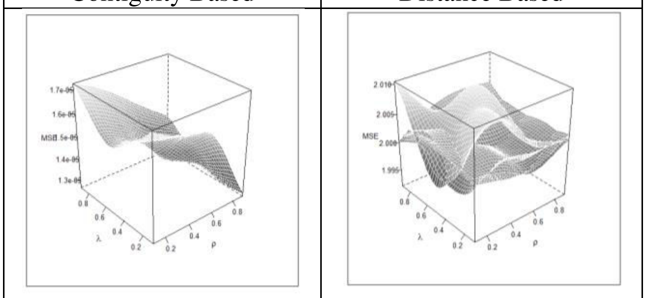
Figure-1. The MSE of β_1 estimate.

Figure-1 shows the MSE of the parameter estimate β a function of spatial lag dependence (ρ), spatial lag error (λ), and sample size (N). Generally, the GNS model has smallest MSE compare than other models. This model also has the same pattern and the same value of MSE for spatial weight based on contiguity and distance. For a large number of samples, GNS model is strictly robust. It is shown from the pattern of the parameter estimate β that almost flat. For small sample size, the MSE will increase for a large value of λ. It means the GNS model is more robust for spatial lag dependence compare than spatial lag error. The SDM model has the same result with the GNS Model. This is evident from the pattern and same value of MSE. The SAC model is the very sensitive to the sample size. The MSE of this model looks different for small and large sample size. The SLX model has a large MSE however still smaller than SEM and LM Model. For all the models, the spatial weight matrix based on distance produces smaller MSE.

The MSE of θ_1 estimate

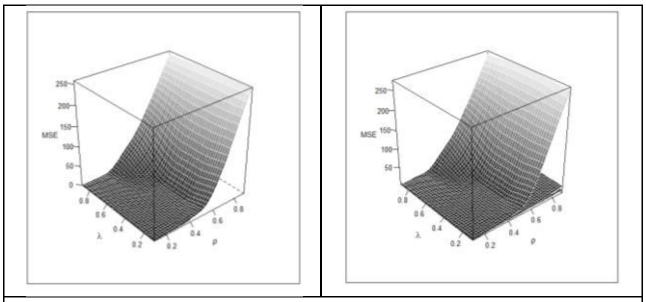
The lambda (θ) is a parameter of spatial lag X model. Substantive spatial dependence can be captured by spatial lags in the exploratory variables X or the endogenous variable Y. In the former case, the spatial lag variables Wx will be incorporated into the standard regression model as additional regressors. We term the regression model with spatially lagged exogenous regressors spatial cross-regressive model. Substantive spatial interaction can occur in different applications. Output growth of a region may not only depend on own region's initial income but as well on income in adjacent regions. In this case, spillover effects are restricted to neighborhood regions.

Contiguity Based | Distance Based

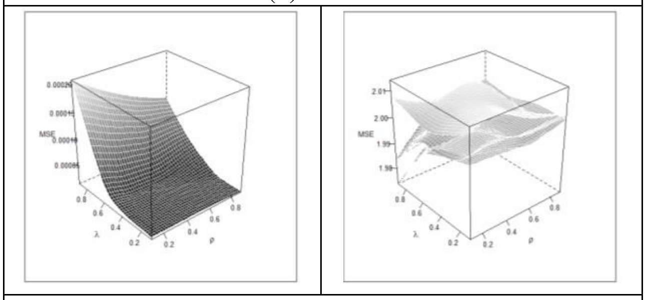


(a) GNS

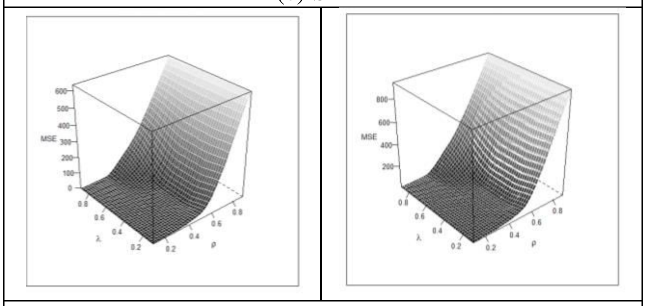
VOL. 11, NO. 24, DECEMBER 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences



(b) SDEM



(c) SDM



(d) SLX

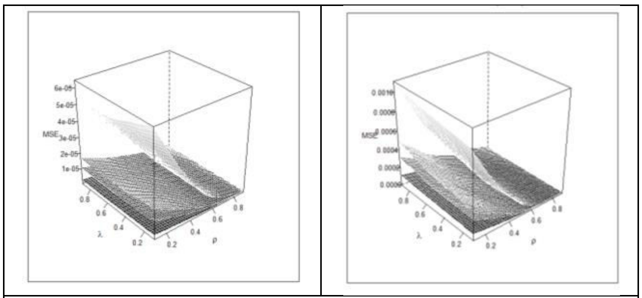
Figure-2. The MSE of θ_1 estimate.

Figure-2 shows the MSE for spatial weight matrix based on distance larger than contiguity matrix. The GNS model has smaller MSE compare than the other model. For contiguity matrix, GNS model has smaller MSE for large ρ and small λ. For the ρ and λ are not equal to zero, the SLX model is the worst model.

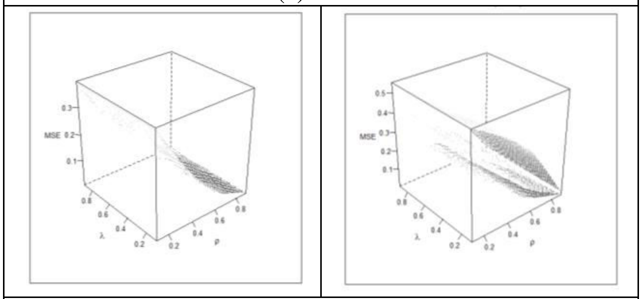
The MSE of ρ_1 estimate

The rho (ρ) is a parameter of spatial lag dependence model. The spatial lag model captures as well substantial spatial dependencies like external effects or spatial interactions. It assumes that such dependencies manifests in the spatial lag Wy of the dependent variable Y. In this case, spillover effects are not restricted to adjacent regions but propagated over the entire regional system.

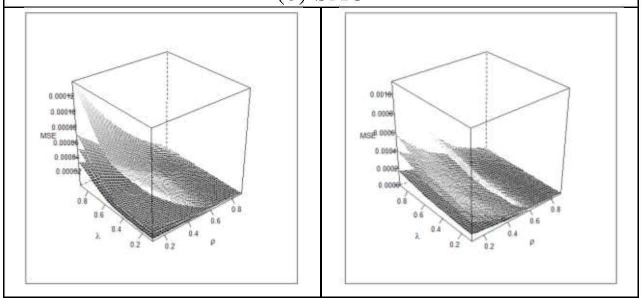
Contiguity Based	Distance Based



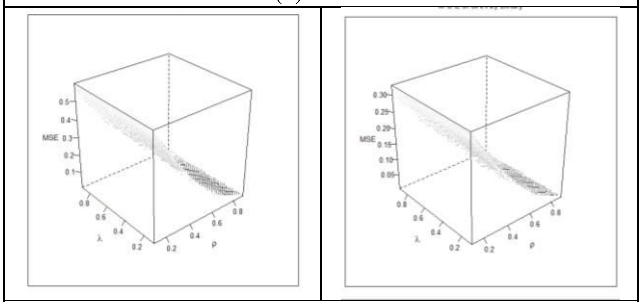
(a) GNS



(b) SAC



(c) SDM



(d) SEM

Figure-3. The MSE of ρ estimate.

Figure-3 shows the MSE of the $\hat{\rho}$ estimate for spatial weight matrix based on distance smaller than contiguity matrix. The GNS model has smaller MSE for estimating spatial lag dependence compare than the other model. For contiguity matrix, GNS model has smaller MSE for large of ρ. The SAC and SEM model have same pattern and value of MSE. Those model has large MSE for large λ and small ρ.

The MSE of λ estimate

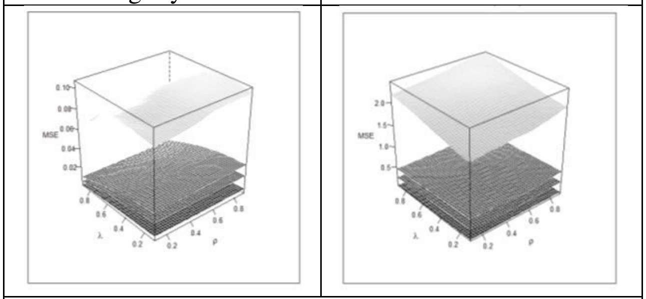
The lambda (λ) is a parameter of spatial lag error model. The spatial error model is applicable when spatial autocorrelation occurs as nuisance resulting from misspecification or inadequate delineation of spatial units.

VOL. 11, NO. 24, DECEMBER 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

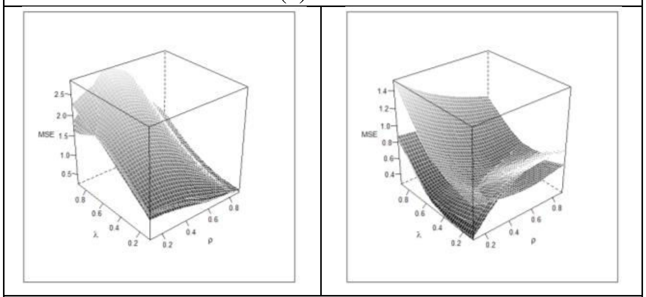
www.arpnjournals.com

Unmodelled interactions among regions are restricted to the error terms. In convergence studies, the convergence rate will be properly assessed by standard estimation methods. However, a random shock occurring in a specific region is not restricted to that region and its neighborhood but will diffuse across the entire regional system.

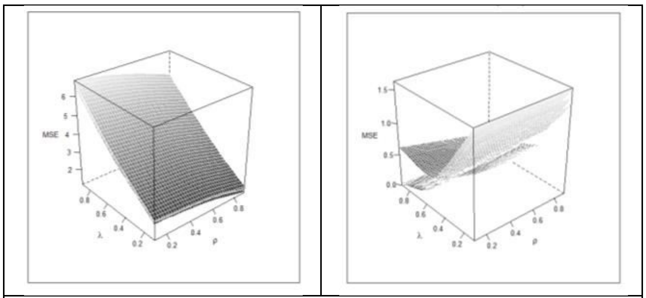
Contiguity Based | Distance Based



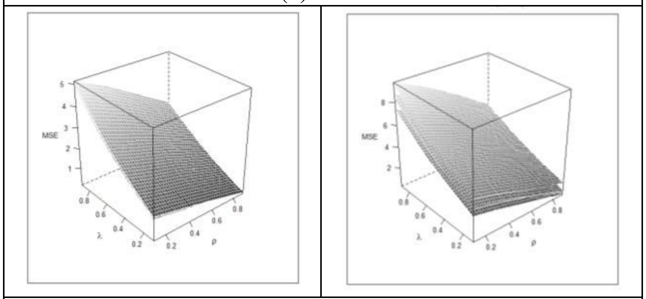
(a) GNS



(b) SDEM



(c) SAC



(d) SEM

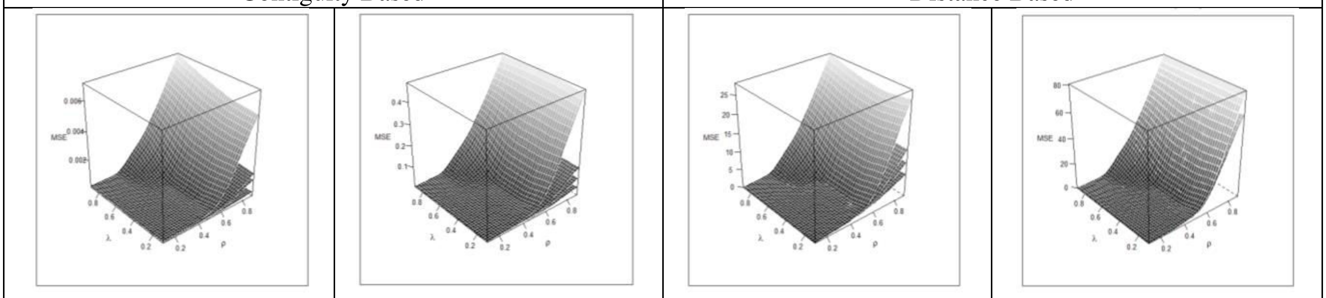
Figure-4. The MSE of λ estimate.

Figure-4 shows the strange results. The SEM model has small MSE for estimating spatial lag error when the spatial lag dependence is large. It shows dependence in error term can be solved by spatial lag dependence.

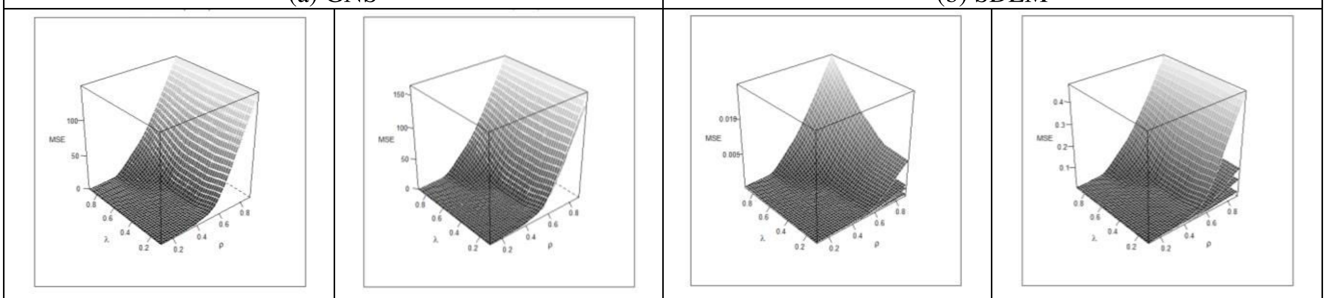
The MSE of γ estimate

The gamma (γ) is the average of direct effect – averaged over all n regions providing a summary measure of the impact arising from changes in the i-th observation of variable k

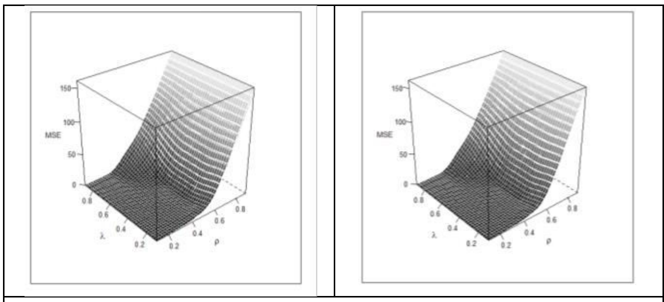
Contiguity Based | Distance Based

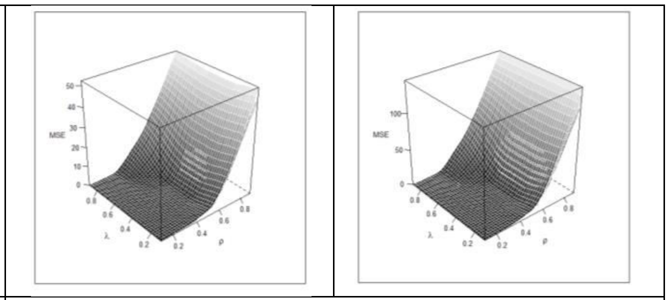


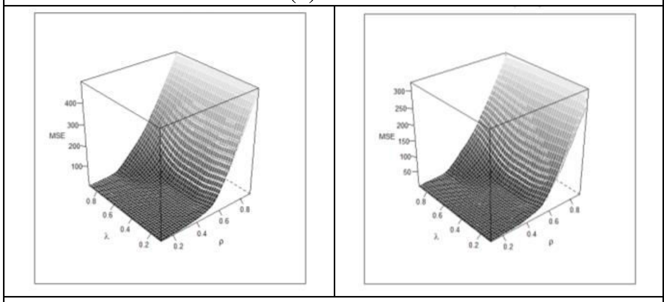
(a) GNS | (b) SDEM



(c) SAC | (d) SDM


(e) SEM


(f) SLX


(g) SEM

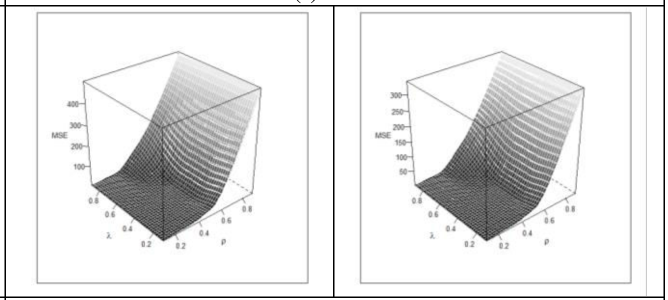

(h) LM

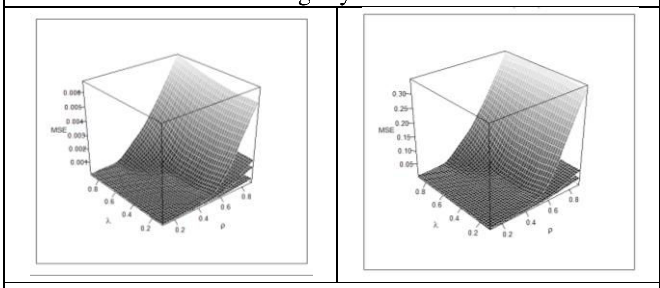
Figure-5. The MSE of γ estimate.

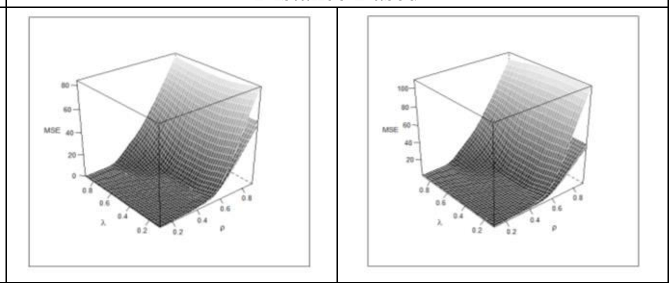
All figures show almost the same pattern of MSE for $\hat{\boldsymbol{\gamma}}$. The MSE of direct effect will increase for strong spatial lag dependence in the data. The GNS Model is a better model to estimate the direct effect.

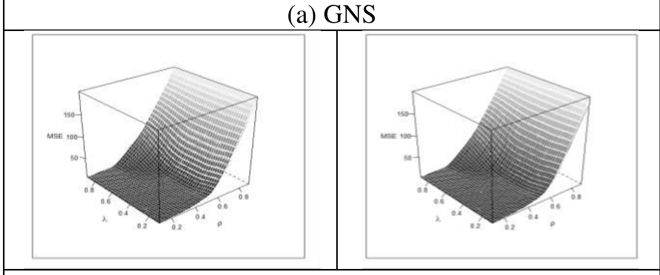
The MSE of δ estimate

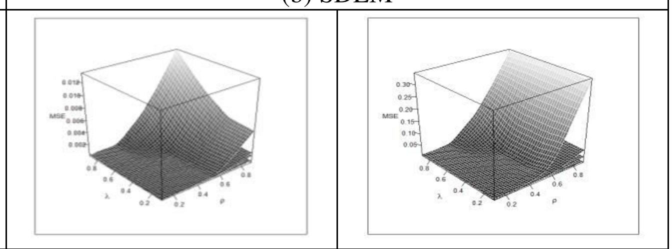
The delta (δ) is the average of indirect effect - averaged over all *n* regions providing a summary measure of the impact arising from changes in the neighboring observation of variable

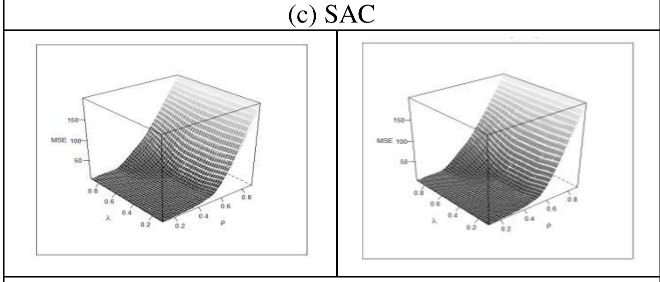
Contiguity Based	Distance Based


(a) GNS


(b) SDEM


(c) SAC


(d) SDM


(e) SEM

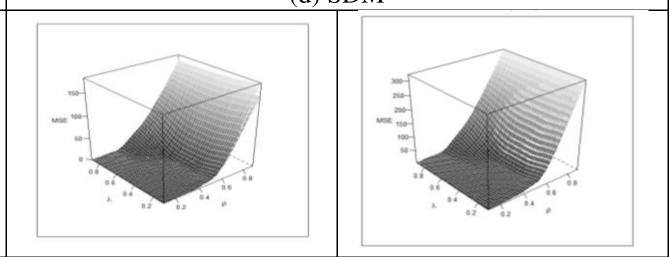

(f) SLX

Figure-6. The MSE of δ estimate.

All figures show almost the same pattern of MSE for $\hat{\boldsymbol{\delta}}$. The MSE of direct effect will increase for strong spatial lag dependence in the data. The GNS Model is a better model to estimate the direct effect.

4. CONCLUSIONS

The GNS and SDM are two models with small MSE for all conditions. The SDM is a best model to estimate the direct and spillover effect with minimum MSE. An increasing number of the sample cannot improve MSE for SDEM, SLX, SEM, LM model. This simulation has not accommodated the violation of the Gauss-Markov assumption. For the next paper, we will develop simulation which accommodates the violation of Gauss-Markov assumption.

ACKNOWLEDGEMENTS

The work reported in this paper is supported by the Directorate General of Higher Education, Ministry of Research-Technology and Higher Education, Indonesia, through Research Cooperation of Foreign Affairs and International Publication 2016

REFERENCES

[1] Anselin L. 1988. Spatial Econometrics: Methods and Models, Kluwer, Dordrecht.

[2] Ehlhorst P. 2014. Spatial Econometrics-From Cross-Sectional Data to Spatial Panels, Springer, Heidelberg, New York.

[3] Lawson AB. 2014. Hierarchical modeling in spatial WIREs. ComputStat. doi:10.1002/wics.1315.

[4] Bivand RS, Gómez-Rubio V, Rue H. 2014. Approximate bayesian inference for spatial econometrics models. Spatial Statistics. 9:146-165.

[5] LeSage J., Pace R.K. 2009. Introduction to Spatial Econometrics, Chapman & Hall, Boca Raton, FL.

[6] Florax, Raymond and Henk Folmer. 1992. Specification and Estimation of Spatial Linear Regression Models: Monte Carlo Evaluation of Pre-test Estimators, Regional Science and Urban Economics. 22, 405-432.