UNIVERSITAS WIRARAJA

LEMBAGA PENELITIAN DAN PENGABDIAN KEPADA MASYARAKAT

Kampus : Jl. Raya Sumenep Pamekasan KM. 5 Patean, Sumenep, Madura 69451 Telp : (0328) 664272/673088

e-mail : lppm@wiraraja.ac.id Website : lppm.wiraraja.ac.id

SURAT PERNYATAAN

Nomor : 056/SP.HCP/LPPM/UNIJA/II/2021

Yang bertanda tangan di bawah ini :

Nama : Nailiy Huzaimah, S. Kep., Ns, M.Kep.
Jabatan : Sekretaris LPPM
Instansi : Universitas Wiraraja

Menyatakan bahwa :

1. Nama : Dr. Anik Anekawati, S.Si., S.Pd., M.Si.
   Jabatan : Staf Pengajar Fakultas Keguruan dan Ilmu Pendidikan

Telah melakukan cek plagiarisme ke LPPM menggunakan *software turnitin.com* untuk artikel dengan judul *"**GENERALIZED METHOD OF MOMENTS APPROACH TO SPATIAL STRUCTURAL EQUATION MODELING**"* dan mendapatkan hasil similarity sebesar 11%

Demikian surat pernyataan ini dibuat untuk digunakan dengan sebaik-baiknya.

Sumenep, 26 Februari 2021
a.n Sekretaris LPPM
Universitas Wiraraja,

Nailiy Huzaimah, S. Kep., Ns, M.Kep.
NIDN. 0727069003

GENERALIZED METHOD OF MOMENTS APPROACH TO SPATIAL STRUCTURAL EQUATION MODELING

by Anik Anekawati 260221

Submission date: 26-Feb-2021 11:15AM (UTC+0700)

Submission ID: 1518535862

File name: FJMS 2018.pdf (274.42K)

Word count: 5625

Character count: 27510

Far East Journal of Mathematical Sciences (FJMS)

http://www.pphmj.com
http://dx.doi.org/10.17654/MS103061057
Volume 103, Number 6, 2018, Pages 1057-1076 ISSN: 0972-0871

GENERALIZED METHOD OF MOMENTS APPROACH TO SPATIAL STRUCTURAL EQUATION MODELING

Anik Anekawati[1,2], Bambang Widjanarko Otok[2], Purhadi[2] and Sutikno[2]

[1]Faculty of Teacher Training and Education
Wiraraja University
Sumenep, Indonesia

[2]Departement of Statistics
Faculty of Mathematics, Computing and Data Science
Institut Teknologi Sepuluh Nopember
60111, Indonesia

Abstract

Currently, researchers are often faced with an analysis involving latent variables having spatial effects. Therefore, a statistical analysis technique called spatial structural equation modeling (spatial SEM) is required. In this work, it was applied to model the quality of education at the senior high school in Sumenep Regency in East Java, Indonesia. The latent variable in the structural model was estimated and became the score of a latent variable using the partial least square (PLS) method. The score of a latent variable in the form of regression equation was tested spatial effect using robust Lagrange multiplier test. This test results in a type of the model of spatial SEM, whether autoregressive spatial model (SAR in SEM) or spatial error model (SERM in SEM). The model of SAR in SEM was estimated using two-stage least square (2SLS), while the model of SERM in SEM was

Received: September 19, 2017; Accepted: December 23, 2017
2010 Mathematics Subject Classification: 62H11, 62G05, 62P25, 91D25.
Keywords and phrases: structural equation modeling (SEM), spatial SEM, SAR in SEM, SERM in SEM, generalized method of moments.

estimated using generalized method of moments (GMM). The result of spatial effect test leads to the model of SAR in SEM. The GMM method produced a better model than the maximum likelihood estimation (MLE) method.

1. Introduction

Researchers are often confronted with an analysis involving spatial effects. Studies related to the estimation procedures of the linear model with spatial dependence have been widely developed. Here are some researches related to the linear model estimation. Anselin [11] proposed an iterative two-stage procedure to maximize log-likelihood of the spatial error model. LeSage [8] has combined spatial information in regression relationships that show spatial autocorrelation. Kelejian and Prucha [6] and [7] described the practical difficulties of using quasi-maximum likelihood (QML) for the spatial autoregressive model (SAR). Estimation of the parameters of spatial autoregressive (ρ) on SAR required computational complexity for medium to large sample sizes. So they offered the method of generalized spatial two-stage least squares (GS2SLS). Lee [12] developed an estimation of spatial autoregressive processes using the generalized method of moments (GMM). The result of this method is characteristic of the estimator in a more consistent and asymptotic normal than QML. The estimation of GMM can be extended easily to estimate the spatial autoregressive process in high order processes without additional computational complexity. Lee [13] developed the estimation method of a mixed regression, the autoregressive spatial model (MRSAR) using the GMM and two-stage least square (2SLS) approaches. This method produces estimators which are asymptotically relatively more efficient than 2SLS estimators and even estimators from best two stage least square (B2SLS). This method can be easily extended to MRSAR model estimation with high order spatial lag. Arnold [14] in his paper showed an improvement of the GMM estimator for autoregressive parameters of the spatial autoregressive error model using the calculation that the unobserved regression disturbance differs from the observed regression residual. By using Monte Carlo simulation, the bias can be reduced by

65-80%. Drukker et al. [3] considered a spatial-autoregressive model with an autoregressive disturbance having possibility of an endogenous regress with an additional spatial lag of the dependent variable. They suggested using a two-step of GMM and extension of instrument variables (IV) from the previous estimation approach that has been done by Kelejian and Prucha [6] and [7].

The developed estimation procedures of the linear model that have spatial dependence were often limited to measurable variables only. However, researchers are often confronted with an analysis involving latent variables. One of the techniques of statistical analysis that has the ability to analyze patterns of relations among latent variables, and between the latent variable and its indicators is structural equation modeling (SEM). SEM is often illustrated using a path model where factors are seen as latent variables, namely variable that cannot be observed directly but measured through measurable indicators (Schumacker and Lomax [17]). SEM consists of two types of analysis, namely SEM of covariance-based and variance-based. The covariance-based SEM is based on large sample size theory and must fulfill the requirements of the normal multivariate distribution. As an alternative solution for this problem, SEM of variance-based or partial least square (PLS) was developed.

The estimation procedures of a parameter on models involving latent variables as well as having spatial effects have been widely studied. Wang and Wall [4] proposed a general spatial factor model and used the Bayesian approach. This study was applied in cases of cancer deaths. The research of Wang and Wall [4] continued by Liu et al. [20]. They proposed a generalized spatial structural equation model. They combined the common spatial factor models of Wang and Wall [4] to measure latent variables and modeled regression among latent variables using a Bayesian approach. Hogan and Tchernis [10] described the hierarchy model for factor analysis of spatially correlated multivariate data. They used a Bayesian approach for single spatial correlations among social indicators. Congdon [16] studied the spatial SEM with an effect of nonlinear construct using spline regression. Oud and Folmer

[9] studied spatial lag in SEM (SAR in SEM), spatial error models in SEM (SERM in SEM) and used the estimation method of full information maximum likelihood (FIML).

The estimation problems in the spatial SEM model are the adequacy of the amount of data, the fulfillment of normality assumptions, and computational complexity, so an easier estimation method is needed. Therefore, this work modeled the education quality of the high school in Sumenep Regency, which is a spatial SEM model and used the estimation method of GMM. GMM is one of the semi-parametric estimations since it is commonly used in data with a few information on distribution (Greene [19]). Since GMM does not impose restrictions on data distribution, it is a good alternative method to be used (Chausse [15]).

In this paper, the latent variable in the structural model was estimated to be the score of a latent variable. The score of a latent variable in the form of regression equation will be tested spatial effect and used Lagrange multiplier test. From the results of this test, a type of the model of spatial SEM can be determined, namely whether the model of SAR in SEM or the model of SERM in SEM. The model of SAR in SEM is estimated by 2SLS, while the model of SERM in SEM is estimated by GMM.

2. Spatial Model

Anselin [11] developed a spatial model using the spatial data of a cross section, which was generally written as follows:

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta}^* + \lambda^*\mathbf{W}\mathbf{y}^* + \mathbf{u}^*, \tag{1}$$

where $\mathbf{u}^* = \rho^*\mathbf{M}\mathbf{u}^* + \boldsymbol{\varepsilon}^*$, with $\boldsymbol{\varepsilon}^* \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. λ^* is the coefficient of autoregressive spatial on response variable with the value of $|\lambda^*| < 1$, ρ^* is the spatial coefficient in error with the value of $|\rho^*| < 1$, $\boldsymbol{\beta}^*$ is a vector of parameter in regression model, $\mathbf{X}^*$ is a matrix of predictor variable, $\mathbf{y}^*$ is a vector of response variable that has a dependence of spatial, $\mathbf{W}$ and $\mathbf{M}$ are the matrices of spatial weight with a diagonal element of 0, $\mathbf{u}^*$ is a vector of regression error that was autocorrelated, and $\boldsymbol{\varepsilon}^*$ is a vector of error.

The derived models of equation (1) are: (a) the model of OLS linear regression that has no spatial effect, it is obtained if $\rho^* = 0$ and $\lambda^* = 0$, then equation (1) changes to $\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta}^* + \boldsymbol{\varepsilon}^*$; (b) the spatial autoregressive model (SAR) which assumes autoregressive processes only at the response variable, it is obtained if $\rho^* = 0$ and $\lambda^* \neq 0$, then equation (1) changes to $\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta}^* + \lambda^*\mathbf{W}\mathbf{y}^* + \boldsymbol{\varepsilon}^*$; (c) the spatial error model (SERM) is obtained if $\rho^* \neq 0$ and $\lambda^* = 0$, then equation (1) changes to $\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta}^* + \mathbf{u}^*$ or $\mathbf{y}^* = \rho^*\mathbf{M}\mathbf{y}^* + \mathbf{X}^*\boldsymbol{\beta}^* - \rho^*\mathbf{M}\mathbf{X}^*\boldsymbol{\beta}^* + \boldsymbol{\varepsilon}^*$; (d) spatial autoregressive moving average (SARMA) is obtained if $\rho^* \neq 0$ and $\lambda^* \neq 0$, then obtained the model as equation (1).

3. Structural Equation Modeling (SEM) Based on Variance

SEM based-variance or PLS generally consists of two submodels, namely outer model and inner model. The inner model shows the strength of estimation among latent variables. The model equation can be written as follows (Trujillo [5]): $\boldsymbol{\eta} = \boldsymbol{\beta}_0^{\otimes} + \boldsymbol{\beta}^{\otimes}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}^{\otimes}$, $\boldsymbol{\eta}$ is the vector of an endogenous latent variable, $\boldsymbol{\xi}$ is the vector of an exogenous latent variable, $\boldsymbol{\beta}^{\otimes}$ is a coefficient matrix measuring the relationship of the influence of endogenous latent variable to other endogenous latent variables, $\boldsymbol{\Gamma}$ is a coefficient matrix measuring the relationship of the influence of exogenous latent variables to endogenous latent variable, and $\boldsymbol{\zeta}^{\otimes}$ is the vector of random error, with the value of expectation equal to zero.

The outer model specifies the relationships between the latent variable and its associated indicators. Trujillo [5] wrote the outer model equations for a reflective model such as the following: $x_{jk}^{\otimes} = \lambda_{jk}^{\otimes} + \varepsilon_{jk}^{\otimes}$, where $\lambda_{jk}^{\otimes}$ is the loading coefficient of the relationship between the latent variable and its kth indicator. The notation $\varepsilon_{jk}^{\otimes}$ is an error from each measurement variable and value of $E(\varepsilon_{jk}^{\otimes}) = E(\xi_j^{\otimes}\varepsilon_{jk}^{\otimes})$.

4. Estimation of Score of Latent Variable

The estimation of the parameter of SEM based-variance was categorized into three, namely: weight estimate, path estimate, and means and parameter location (regression constant value). Weight estimate is used to create the score of a latent variable.

Figure 1 shows the steps of an estimation of exogenous latent variables $(\boldsymbol{\xi})$ for obtaining the scores of the exogenous latent variable ($\boldsymbol{l}$) as written by Trujillo [5]. Estimation of the endogenous latent variable was obtained using the same the way as shown in Figure 1.

The first step is the outside approximation that is started by an initialization in each latent variable as a linear combination of the indicators and written as follows: $l_j = \sum_k \tilde{w}_{jk} x_{jk}^{\otimes}$, where $\tilde{w}_{jk}$ is an outer weight. For the first approach to the initial weights, in Trujillo report [5], which in Chin suggested using the same value of latent variables as a simple sum from their indicators.

The second step is the inside approximation that is used to get initialization on each latent variable in the inner model. The initialization process is done by calculating the weighted aggregate against adjacent latent variables, i.e., $Z_j = \sum_{i:\beta_{ij}\neq 0,\, \beta_{ji\neq 0}} e_{ij} l_i$, where e_{ij} is the inner weight which can be selected from three schemes, namely centroid, factor or path. These schemes were defined by Trujillo [5] as follows:

(i) Inner weight (e_{ij}) in the centroid scheme is obtained from the sign of the correlation between l_i and l_j.

(ii) Inner weight (e_{ij}) in the factor scheme is obtained from the sign and the value of the correlation coefficient between l_i and l_j.

(iii) Inner weight (e_{ij}) in the path scheme is the weighting of the latent variable of the neighbor depending on whether the neighbor variable is the antecedent or consequent of the latent variable that will be estimated. The path scheme was defined as follows: $e_{ij} = cor(l_j, l_i)$ if ξ_j is explained by ξ_i or $l_j = \sum_i e_{ji} l_i$, where e_{ij} is the coefficient of regression equation l_i on l_j.

The third step is updating of outer weights. Updating of outer weights is done because Z_j should consider their indicators. The weights w_{jk} on the outer model of reflective are as follows: $w_{jk} = (Z_j^T Z_j)^{-1} Z_j^T x_{jk}^{\otimes} = cor(x_{jk}^{\otimes}, Z_j)$.

The fourth step is the checking of convergence. In each iteration procedure, let $S = 1, 2, 3, ...,$ convergence is checked by comparing outer weight in step S to outer weight in step S-1. Trujillo report [5], which in Wold proposed $| \tilde{w}_{jk}^{s-1} - w_{jk}^{s} | < 10^{-5}$ as the convergence criterion.

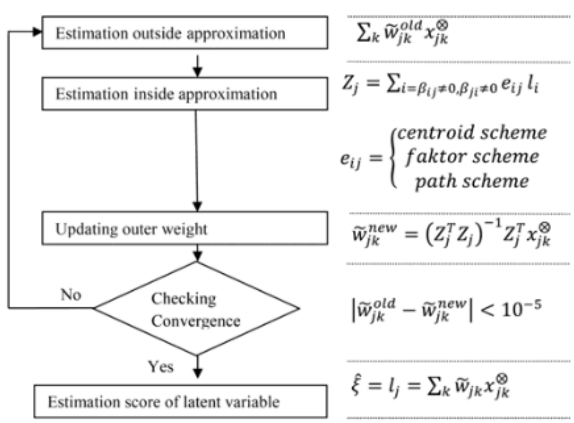

Figure 1. Procedure to estimate the score of latent variable.

5. Spatial Autoregressive Model (SAR) in SEM

The model of spatial SEM in this study is spatial dependence on the inner model, where the latent variables are replaced by the score of latent variable estimated as in the estimation procedure in Section 4 (Figure 1). The matrix of spatial weighted $\mathbf{W}$ shows spatial dependence among observations or locations. The matrix of the spatial weighted in this paper uses queen contiguity that is the common vertex and the common side.

SAR in SEM in the form of multiple indicators multiple causes (MIMIC) written by Oud and Folmer [9] is as follows:

$$\tilde{\boldsymbol{y}} = \rho^{\otimes}\mathbf{W}\tilde{\boldsymbol{y}} + \mathbf{X}^{\otimes}\boldsymbol{\gamma} + \tilde{\boldsymbol{\varepsilon}}, \quad (2)$$

where $\tilde{\boldsymbol{y}}$ is the vector of observations on the dependent variable y, $\mathbf{X}^{\otimes}$ is the matrix of observation of the explanatory variables, $\boldsymbol{\gamma}$ is a vector of a regression coefficient of explanatory variables, $\rho^{\otimes}$ is the coefficient of spatial autoregressive, $\tilde{\boldsymbol{\varepsilon}}$ is vector of an error, and $\mathbf{W}$ is the contiguity matrix.

In this study, the latent variables of SAR in SEM in equation (2) were replaced by the score of a latent variable as a unit sample. The model did not involve the MIMIC model because there were no exogenous or endogenous variables as observed variables. So SAR in SEM in equation (2) changes to the following:

$$\boldsymbol{l} = \lambda\mathbf{W}\boldsymbol{l} + \mathbf{K}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (3)$$

where $\boldsymbol{l}$ is the vector of the score of an endogenous latent variable of sized $n \times 1$, $\mathbf{K}$ is the matrix of the score of an exogenous latent variable of sized $n \times k$, λ is the coefficient of lag spatial from the score of an endogenous latent variable, $\boldsymbol{\beta}$ is the vector of regression parameter coefficient of sized $k \times 1$, $\boldsymbol{\varepsilon}$ is the error vector of sized $n \times 1$, and $\mathbf{W}$ is the weight matrix of sized $n \times n$ a diagonal element of zero.

6. Spatial Error Model (SERM) in SEM

SERM in SEM in the form of MIMIC written by Oud and Folmer [9] is as follows:

$$\tilde{y} = \mathbf{X}^{\otimes}\gamma + \tilde{\boldsymbol{\varepsilon}} \tag{4}$$

with $\tilde{\boldsymbol{\varepsilon}} = \lambda^{\otimes}\mathbf{W}\tilde{\boldsymbol{\varepsilon}} + \tilde{\boldsymbol{\zeta}}$ or was written as follows:

$$\tilde{y} = \lambda^{\otimes}\mathbf{W}\tilde{y} + \mathbf{X}^{\otimes}\gamma - \lambda^{\otimes}\mathbf{W}\mathbf{X}^{\otimes}\gamma + \tilde{\boldsymbol{\zeta}}. \tag{5}$$

In this study, the latent variables of SERM in SEM in equations (4) and (5) were replaced by the score of a latent variable as a unit sample. In this model, the MIMIC model was not used because there were no exogenous or endogenous variables as observed variables. So, SERM in SEM in equation (4) changes to equation (6):

$$\boldsymbol{l} = \mathbf{K}\boldsymbol{\beta} + \mathbf{u} \tag{6}$$

and equation (5) changes to equation (7):

$$\boldsymbol{l} = \rho\mathbf{W}\boldsymbol{l} + \mathbf{K}\boldsymbol{\beta} - \rho\mathbf{W}\mathbf{K}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{7}$$

$\boldsymbol{l}$ is the vector of the score of an endogenous latent variable of sized $n \times 1$, $\mathbf{K}$ is the matrix of the score of an exogenous latent variable of sized $n \times k$, $\boldsymbol{\beta}$ is the vector of regression parameter coefficient of sized $k \times 1$, $\boldsymbol{\varepsilon}$ is the error vector of sized $n \times 1$, ρ is the coefficient of on the spatial autoregressive structure in error $\boldsymbol{\varepsilon}$, $\mathbf{u}$ is the vector of a regression error that is assumed to have a random region effect and also the error of spatially autocorrelated, and $\mathbf{W}$ is the weight matrix of sized $n \times n$ a diagonal element of zero.

7. The Lagrange Multiplier Test for Model of Spatial SEM

There are several tests of spatial dependence of the model of standard spatial regression. Among them are Moran's I test, Lagrange multiplier test for a dependence of error spatial and Lagrange multiplier test for a dependence of autoregressive spatial. Moran's I test for spatial autocorrelation is a general test, while the Lagrange multiplier test is more specific.

Statistics test of Lagrange multiplier is defined by Breusch and Pagan [18], $LM = \widetilde{\mathbf{D}}^T \widetilde{\boldsymbol{\psi}}^{-1} \widetilde{\mathbf{D}}$. According to Anselin [11], a matrix of information $\widetilde{\boldsymbol{\psi}}$ for Lagrange multiplier test on the model of the SAR is $\widetilde{\boldsymbol{\psi}} = (\mathbf{A} + \mathbf{T})$, where $\mathbf{A} = \sigma^{-2} (\mathbf{W}\mathbf{X}^* \boldsymbol{\beta}^*)^T \mathbf{S} (\mathbf{M}\mathbf{X}^* \boldsymbol{\beta}^*)$, $T_{ij} = tr\{\mathbf{W}\mathbf{M} + \mathbf{W}^T \mathbf{M}\}$, $\mathbf{S} = \mathbf{I} - \mathbf{X}^* (\mathbf{X}^{*T} \mathbf{X}^*)^{-1} \mathbf{X}^{*T}$, $\mathbf{W}$ and $\mathbf{M}$ are the matrices of a spatial weight. If $\mathbf{W}$ and $\mathbf{M}$ are equal, then $T_{ij} = T = tr\{(\mathbf{W}^T + \mathbf{W})\mathbf{W}\}$.

Thus, statistics test of Lagrange multiplier for SAR model is obtained as follows:

$$LM_\lambda = \left[\frac{1}{\sigma^2} \mathbf{e}^{*T} \mathbf{W} \mathbf{y}^* \right]^2 (\mathbf{A} + \mathbf{T})^{-1}. \tag{8}$$

The Lagrange multiplier test for SAR in SEM is equal to equation (8), but independent variable $\mathbf{X}^*$ is replaced by the score of an exogenous latent variable $\mathbf{K}$, dependent variable $\mathbf{y}^*$ is replaced by the score of endogenous latent variable $\boldsymbol{l}$, so the Lagrange multiplier test for SAR in SEM is $LM_\lambda = \left[\frac{1}{\sigma^2} \mathbf{e}^{\odot T} \mathbf{W} \boldsymbol{l} \right]^2 (\mathbf{A} + \mathbf{T})^{-1}$, with $\mathbf{e}^{\odot} = (\boldsymbol{l} - \mathbf{K}\boldsymbol{\beta})$;

$$\mathbf{A} = \sigma^{-2} (\mathbf{W}\mathbf{k}\boldsymbol{\beta})^{\mathbf{T}} \mathbf{S} (\mathbf{W}\mathbf{K}\boldsymbol{\beta}); \quad T_{ij} = T = tr\{(\mathbf{W}^T + \mathbf{W})\mathbf{W}\}.$$

Statistics test of LM_λ follows χ^2 distribution. If $LM_\lambda > \chi^2$, then there is an effect of SAR in SEM.

According to Anselin [11], a matrix of information $\widetilde{\boldsymbol{\psi}}$ for Lagrange multiplier test of SERM is $\widetilde{\boldsymbol{\psi}} = T_{ij}$, where $T_{ij} = tr\{\mathbf{W}\mathbf{M} + \mathbf{W}^T \mathbf{M}\}$, $\mathbf{W}$ and $\mathbf{M}$ are the matrices of a spatial weight. If $\mathbf{W}$ and $\mathbf{M}$ are equal, then $T_{ij} = T = tr\{(\mathbf{W}^T + \mathbf{W})\mathbf{W}\}$.

Thus, statistics test of Lagrange multiplier for SERM is obtained as follows:

$$LM_{\rho} = \left[\frac{1}{\sigma^2}\mathbf{e}^{*T}\mathbf{W}\mathbf{e}^{*}\right]^2 \mathbf{T}^{-1}. \quad (9)$$

The Lagrange multiplier test for SERM in SEM is equal to equation (9), but independent variable $\mathbf{X}^*$ is replaced by the score of an exogenous latent variable $\mathbf{K}$, dependent variable $\mathbf{y}^*$ is replaced by the score of endogenous latent variable $\boldsymbol{l}$, so the Lagrange multiplier test for SERM in SEM is $LM_{\rho} = \left[\frac{1}{\sigma^2}\mathbf{e}^{*T}\mathbf{W}\mathbf{e}^{*}\right]^2 \mathbf{T}^{-1}$, with $\mathbf{e}^{\odot} = (\boldsymbol{l} - \mathbf{K}\boldsymbol{\beta})$ and $T_{ij} = T = tr\{(\mathbf{W}^T + \mathbf{W})\mathbf{W}\}$.

Statistics test of LM_{ρ} follows χ^2 distribution. If $LM_{\rho} > \chi^2$, then there is an effect of SERM in SEM.

8. Estimation of Spatial Autoregressive Model (SAR) in SEM

The parameters of SAR in SEM as equation (3) are estimated using the 2SLS method that has been reviewed by Kelejian and Prucha [6] and [7]. The 2SLS method requires an instrument variable. An instrument variable (**H**) is valid if it fulfills the following criteria: it does not correlate with **u** and contains only minimal variables of k or $p \geq k$ or such that **H** correlates with $\mathbf{W}\boldsymbol{l}$ regressor. Spatial SEM in equation (3) will use an instrument variable which is a combination of matrices **K** and **WK**, i.e., $\mathbf{H} = (\mathbf{K} \vdots \mathbf{WK})$.

Let $(\mathbf{WK})_t$ is the tth variable of **WK**. Then $\text{cov}[(\mathbf{WK})_t, \mathbf{u}] = E[(\mathbf{WK})_t, \mathbf{u}^T] - E[(\mathbf{WK})_t]E[\mathbf{u}^T] = \mathbf{W}\,\text{cov}[\boldsymbol{k}_t, \mathbf{u}]$, because $\text{cov}[\boldsymbol{k}_t, \mathbf{u}] = 0$. Thus $\text{cov}[(\mathbf{WK})_t, \mathbf{u}] = 0$. This represents that each score of the exogenous latent variables **K** and **WK** is not correlated with **u** or variables in matrix $\mathbf{H} = (\mathbf{K} \vdots \mathbf{WK})$ are not correlated with **u**. The first criterion as a valid instrument variable is fulfilled.

Furthermore, it is known that

$$\text{cov}[\boldsymbol{k}_t,\ \mathbf{W}\boldsymbol{l}] = E[\boldsymbol{k}_t(\mathbf{W}\boldsymbol{l})^T] - E[\boldsymbol{k}_t]E[(\mathbf{W}\boldsymbol{l})^T] = \text{cov}[\boldsymbol{k}_t,\ \boldsymbol{l}^T]\mathbf{W}^T,$$

because $\text{cov}[\boldsymbol{k}_t,\ \boldsymbol{l}^T] \neq 0$ and $\mathbf{W} \neq \mathbf{0}$, then $\text{cov}[\boldsymbol{k}_t,\ \mathbf{W}\boldsymbol{l}] \neq 0$, therefore the variables at **WK** correlate with $\mathbf{W}\boldsymbol{l}$.

Thus, the variables at **K** and **WK** correlate with $\mathbf{W}\boldsymbol{l}$ so that it is proven that the variables in the matrix $\mathbf{H} = (\mathbf{K} \vdots \mathbf{WK})$ are correlated with $\mathbf{W}\boldsymbol{l}$. The second criterion as a valid instrument variable is fulfilled. Then the matrix $\mathbf{H} = (\mathbf{K} \vdots \mathbf{WK})$ is the matrix of a valid instrument variable.

SAR in SEM as in equation (3) is simplified to:

$$\boldsymbol{l} = \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon}, \tag{10}$$

where $\mathbf{Z} = (\mathbf{X} \vdots \mathbf{W}\boldsymbol{l})$ and $\boldsymbol{\delta} = (\boldsymbol{\beta}^{\mathbf{T}} : \lambda)^{\mathbf{T}}$. $\boldsymbol{\delta}$ is estimated using 2SLS with the following steps: (a) set an instrument variable **H**, namely $\mathbf{H} = (\mathbf{K} \vdots \mathbf{WK})$; (b) set the value of $\mathbf{P_H}$, that is $\mathbf{P_H} = \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$; (c) the value of $\mathbf{P_H}$ is used to estimate $\mathbf{W}\boldsymbol{l}$, namely $\widehat{\mathbf{W}\boldsymbol{l}} = \mathbf{P}_H(\mathbf{W}\boldsymbol{l})$; (d) estimate **Z**, namely $\hat{\mathbf{Z}} = (\mathbf{K} \vdots \widehat{\mathbf{W}\boldsymbol{l}})$, (e) estimate $\boldsymbol{\delta}$, that is $\hat{\boldsymbol{\delta}} = (\hat{\mathbf{Z}}^T\hat{\mathbf{Z}})^{-1}\hat{\mathbf{Z}}^T\boldsymbol{l}$.

In this first stage, the estimated of SAR in SEM is generated, namely:

$$\hat{\boldsymbol{l}} = \hat{\boldsymbol{\beta}}\mathbf{K} + \hat{\lambda}\mathbf{W}\boldsymbol{l}, \tag{11}$$

where the residual results of each its observations are used to estimate the parameter of SERM in SEM at the second stage.

9. Estimation of Spatial Error Model (SERM) in SEM

The estimation of SERM in SEM uses GMM method that has been reviewed by Kelejian and Prucha [6] and [7]. The equation model (11) is an estimation result of SAR in SEM, while the deviation of $\boldsymbol{l}_i$ and $\hat{\boldsymbol{l}}_i$ is the residual value denoted by $\hat{\mathbf{u}}$. This value of $\hat{\mathbf{u}}$ will be used as an observation vector for the random variable $\mathbf{u}$ in SERM in SEM.

The general model of spatial SEM is written based on SAR in SEM in (3) and SERM in SEM in (6), is expressed as follows:

$$\boldsymbol{l} = \lambda \mathbf{W}\boldsymbol{l} + \boldsymbol{\beta}\mathbf{K} + \mathbf{u}, \tag{12}$$

$$\text{with } \mathbf{u} = \rho\mathbf{M}\mathbf{u} + \boldsymbol{\varepsilon}, \tag{13}$$

$$\text{or } \boldsymbol{\varepsilon} = \mathbf{u} - \rho\mathbf{M}\mathbf{u}. \tag{14}$$

Three of the moment equations are obtained based on equations (13) and (14) as follows:

$$2\rho n^{-1}E(\mathbf{u}^T\overline{\mathbf{u}}) - \rho^2 n^{-1}E(\overline{\mathbf{u}}^T\overline{\mathbf{u}}) + \sigma^2 - n^{-1}E(\mathbf{u}^T\mathbf{u}),$$

$$2\rho n^{-1}E(\overline{\mathbf{u}}^T\overline{\overline{\mathbf{u}}}) - \rho^2 n^{-1}E(\overline{\overline{\mathbf{u}}}^T\overline{\overline{\mathbf{u}}}) + \sigma^2 Tr(\mathbf{M}^T\mathbf{M}) - n^{-1}E(\overline{\mathbf{u}}^T\overline{\mathbf{u}}),$$

$$2\rho n^{-1}E(\mathbf{u}^T\overline{\overline{\mathbf{u}}} + \overline{\mathbf{u}}^T\overline{\mathbf{u}}) - \rho^2 n^{-1}E(\overline{\mathbf{u}}^T\overline{\overline{\mathbf{u}}}) - n^{-1}E(\mathbf{u}^T\overline{\mathbf{u}}).$$

Three of the moment equations can be presented in the following matrix form:

$$\begin{bmatrix} 2n^{-1}E(\mathbf{u}^T\overline{\mathbf{u}}) & -n^{-1}E(\overline{\mathbf{u}}^T\overline{\mathbf{u}}) & 1 \\ 2n^{-1}E(\overline{\mathbf{u}}^T\overline{\overline{\mathbf{u}}}) & -n^{-1}E(\overline{\overline{\mathbf{u}}}^T\overline{\overline{\mathbf{u}}}) & Tr(\mathbf{M}^T\mathbf{M}) \\ 2n^{-1}E(\mathbf{u}^T\overline{\overline{\mathbf{u}}} + \overline{\mathbf{u}}^T\overline{\mathbf{u}}) & -n^{-1}E(\overline{\mathbf{u}}^T\overline{\overline{\mathbf{u}}}) & 0 \end{bmatrix} \begin{bmatrix} \rho \\ \rho^2 \\ \sigma^2 \end{bmatrix} - \begin{bmatrix} n^{-1}E(\mathbf{u}^T\mathbf{u}) \\ n^{-1}E(\overline{\mathbf{u}}^T\overline{\mathbf{u}}) \\ n^{-1}E(\mathbf{u}^T\overline{\mathbf{u}}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

or it is summarized to be $\boldsymbol{\Gamma\alpha} - \boldsymbol{\gamma} = \mathbf{0}$. $\boldsymbol{\Gamma}$ can be estimated by $\mathbf{G}$, namely: $\hat{\boldsymbol{\Gamma}} = \mathbf{G} = \frac{1}{n}\begin{bmatrix} 2\hat{\mathbf{u}}^T\hat{\overline{\mathbf{u}}} & \hat{\overline{\mathbf{u}}}^T\hat{\overline{\mathbf{u}}} & 1 \\ 2\hat{\overline{\mathbf{u}}}^T\hat{\overline{\overline{\mathbf{u}}}} & \hat{\overline{\overline{\mathbf{u}}}}^T\hat{\overline{\overline{\mathbf{u}}}} & Tr(\mathbf{M}^T\mathbf{M}) \\ \hat{\mathbf{u}}^T\hat{\overline{\overline{\mathbf{u}}}} + \hat{\overline{\mathbf{u}}}^T\hat{\overline{\mathbf{u}}} & \hat{\overline{\mathbf{u}}}^T\hat{\overline{\overline{\mathbf{u}}}} & 0 \end{bmatrix}$, $\boldsymbol{\gamma}$ can be estimated by $\hat{\boldsymbol{\gamma}} = \mathbf{g} = \frac{1}{n}\begin{bmatrix} \hat{\mathbf{u}}^T\hat{\mathbf{u}} \\ \hat{\overline{\mathbf{u}}}^T\hat{\overline{\mathbf{u}}} \\ \hat{\mathbf{u}}^T\hat{\overline{\mathbf{u}}} \end{bmatrix}$, where $\hat{\mathbf{u}} = \boldsymbol{l} - \mathbf{Z}\hat{\boldsymbol{\delta}}$, $\hat{\overline{\mathbf{u}}} = \mathbf{M}\hat{\mathbf{u}}$ and $\hat{\overline{\overline{\mathbf{u}}}} = \mathbf{M}^2\hat{\mathbf{u}}$.

The empirical equations of the moment conditions can be rewritten as follows:

$$\mathbf{g} = \mathbf{G}\boldsymbol{\alpha} - \mathbf{v}. \tag{15}$$

Approximation using GMM is defined as the result of minimizing the sum squares of residual or $\mathbf{v}^T\mathbf{v}$. From equation (15), the value of $\mathbf{v} = \mathbf{G}\boldsymbol{\alpha} - \mathbf{g}$ is obtained so that the value squared of its residual is $\mathbf{v}^T\mathbf{v} = \boldsymbol{\alpha}^T\mathbf{G}^T\mathbf{G}\boldsymbol{\alpha} - 2(\mathbf{G}^T\mathbf{g})^T\boldsymbol{\alpha} + \mathbf{g}^T\mathbf{g}$.

The value of estimation $\boldsymbol{\alpha}$ is obtained by minimizing the sum squares of residual, namely: $\frac{\partial \mathbf{v}^T\mathbf{v}}{\partial \alpha} = 0$ and $\boldsymbol{\alpha} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{g}$, so the value of estimation $\boldsymbol{\alpha}$ is $\hat{\boldsymbol{\alpha}} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{g}$, where the estimation of $\boldsymbol{\alpha}$ consists ρ and σ^2.

10. Application

Spatial SEM was applied in the modeling of the education quality of high school, where there are three latent variables, namely the quality of education (four indicators) influenced by the school infrastructure (four indicators) and the socio-economic conditions (six indicators). The conceptual model was formed based on these dimensions as shown in Figure 2A.

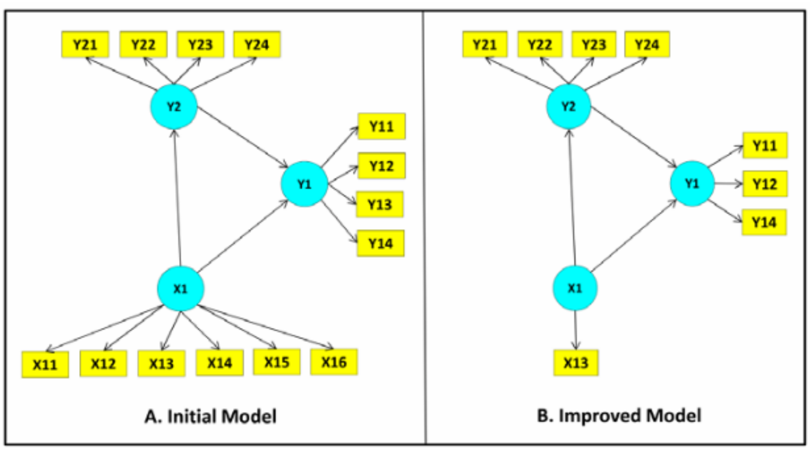

Figure 2. Inner model.

The latent variables to the model were estimated to obtain the score of a latent variable such as the procedure in Figure 1 through three schemes so there were three models, namely the model of centroid scheme, factor, and path. The three models were improved based on evaluations of the outer, inner model, and based on significance tests. Three new models were formed by maintaining valid indicators. The three improved models were re-evaluated based on the evaluation of the inner model, outer, and significance test. From the evaluation result, the best model was obtained, that is the model from the path scheme because all the indicators of the path scheme model were valid, and the value of R^2 had increased. The best conceptual model is as shown in Figure 2B. The process of getting this best model was discussed by Anekawati et al. [1].

The improvement model of the path scheme with the endogenous latent variable of educational quality $(\boldsymbol{\eta})$ has three valid indicators that were estimated to be $\boldsymbol{l}$, the exogenous latent variable of the school infrastructure $(\boldsymbol{\xi}_1)$ has four valid indicators, which were estimated to be $\mathbf{K}_1$, and the exogenous latent variable of socio-economic conditions $(\boldsymbol{\xi}_2)$ has one valid indicator that was estimated to be $\mathbf{K}_2$. So the model of spatial SEM becomes:

$$\hat{l}_i = \hat{\alpha} + \hat{\lambda}\sum\nolimits_{j=1,\, i \neq j}^{n}(W_{ij}l_i) + \hat{\beta}_1 k_1 + \hat{\beta}_2 k_2 + \hat{\rho}\sum\nolimits_{j=1,\, i \neq j}^{n}(W_{ij}u_i). \qquad (16)$$

The spatial effect test on the model (16) used robust Lagrange multiplier test and spatial weights used queen contiguity. The queen contiguity is defined as $W_{ij} = 1$ for the common side or common vertex met with the region of concern, $W_{ij} = 0$ for the other region (LeSage [8]). The result of the spatial effect test is summarized in Table 1. The test result shows that the model of the education quality leads to the model of SAR in SEM at the 5% significance level. So the model of equation (16) changes to be the SAR in SEM:

$$\hat{l}_i = \hat{\alpha} + \hat{\lambda}\sum_{j=1,\, i\neq j}^{n}(W_{ij}l_i) + \hat{\beta}_1 k_1 + \hat{\beta}_2 k_2. \quad (17)$$

Table 1. Identification of spatial effects for the model of education quality

Test of spatial dependence	Chi-sq.	*P*-value	Inference
LM (lag) Robust	3.84	0.05	Affected
LM (error) Robust	3.84	0.24	Not affected

The parameter of SAR in SEM at equation (17) was estimated using the 2STL method. The result of parameter estimation of the model is summarized in Table 2.

Table 2. The result of parameter estimation for model of SAR in SEM

Parameter	Coefficient	*T*-value	*P*-value
b_0	0.1206	0.3712	0.3566
b_1	0.5728	2.7980	0.0048
b_2	0.0210	0.0640	0.4747
The value of λ :	**–0.2471**		
***R*-squared :**	**0.8422**		

The estimation result of SAR in SEM for education quality of the high school in Sumenep Regency can be interpreted as follows: (a) the school infrastructure has a significant effect on the quality of education at a significance level of 5%. The value of coefficient shows that every 1 point increase in the school infrastructure will be followed by the education quality improvement by 0.57 point; (b) the *R*-square value (R^2) is interpreted that the data variation of education quality of the high school can be explained by school infrastructure and the socio-economic condition of 84.22%. The remaining 15.78% is influenced by the other factors than the school infrastructure and the socio-economic condition.

The model of SAR in SEM for the education quality of the high school is

$$\hat{l}_i = 0.121 - 0.247\sum_{j=1,\, i\neq j}^{n}(W_{ij}l_i) + 0.573k_1 + 0.021k_2.$$

The following are some examples of the quality education model of high school in some districts: Masalembo district with no neighboring district, Nonggunong district with one neighbor and Batuputih district with four neighbors:

$$\hat{l}_{masalembo} = 0.121 + 0.573k_1 + 0.021k_2,$$

$$\hat{l}_{nonggunong} = 0.121 - 0.2471_{gayam} + 0.573k_1 + 0.021k_2,$$

$$\hat{l}_{batuputih} = 0.121 - 0.0621_{dasuk} - 0.0621_{manding} - 0.0621_{gapura}$$
$$- 0.0621_{batangbatang} + 0.573k_1 + 0.021k_2.$$

Table 3 is a comparative summary of the adequacy of the model. The compared model is divided into two groups. The first group is the SEM model, i.e., the SEM of an initial model and SEM of an improved model. The second group is the model of a spatial SEM with different estimation methods, i.e., MLE and GMM. The spatial SEM model estimated using MLE method has been discussed by Anekawati et al. [2].

SEM of an improved model has an *R*-square value of 1.7% higher than SEM of an initial SEM one. It shows that the improved model is better than the initial model. This improved model is continued to be tested for spatial effect. The result of spatial effect test leads to SAR in SEM. SAR in SEM with estimation method of MLE has the *R*-square value of 13.16% higher than the model of improved SEM. It shows that the model of education quality of the high school in Sumenep Regency involving the spatial effect is better. The model of SAR in SEM with the estimation method of GMM has the *R*-square value of 78% higher than the model of SAR in SEM with estimation method of MLE. It shows that the GMM method produces a better model than the MLE method. From the four models, the model of SAR in SEM with the estimation method of GMM has the highest *R*-square value. The spatial SEM model estimated using MLE method has been discussed by Anekawati et al. [2].

Table 3. The comparison of model adequacy

Model of SEM	***R*-square**
Initial model	0.411
Improved model	0.418
Model of Spatial SEM	***R*-square**
Spatial SEM with method of MLE	0.473
Spatial SEM with method of GMM	0.842

It indicates that the model of education quality of the high school in Sumenep Regency has the spatial effect. This is supported by the real condition that coordination, development, and evaluation among districts within a regency are necessary in order to increase the education quality in one regency.

11. Summary and Discussion

Since the coefficient of lag spatial (the value of λ) is negative, further study on spatial weighting is required. The spatial weighting of queen contiguity was used in this study that notices the intersection of sides and angles. Sumenep Regency is an archipelago with 18 districts on the mainland and 9 districts on the islands. If there is a district that does not have side and angle intersection with the other districts, then the weighting value is zero. In terms of measuring the education quality, it is impossible not to relate one district to the others. It is due to the coordination, evaluation, development, and supervision by the central district as well as other aspects so that the district must be related to other districts, especially centre districts as the centre of control and coordination. For that, a more flexible weighting form is needed, which is customized weighting.

The value of R-square is 84.22%. It means, there are 15.78% other factors that affect the education quality besides the school infrastructure and socio-economic whereas the factor of socio-economic condition does not affect the education quality. Thus, it is necessary to study other factors that affect the education quality, such as the quality of teachers, the role of local government and others.

References

[1] A. Anekawati, B. W. Otok, Purhadi and Sutikno, Structural equation modeling with three schemes estimation of score factors on partial least square (case study: the quality of education level SMA/MA in Sumenep Regency), Journal of Physics: Conference Series 855 (2017), 012006.

[2] A. Anekawati, B. W. Otok, Purhadi and Sutikno, Modelling of the education quality of a high schools in Sumenep Regency using spatial structural equation modelling, Journal of Physics: Conference Series 890 (2017), 012094.

[3] D. M. Drukker, P. Egger and I. R. Prucha, On two-step estimation of a spatial autoregressive model with autoregressive disturbances and endogenous regressors, Econometric Rev. 32(5-6) (2013), 686-733.

[4] F. Wang and M. M. Wall, Generalized common spatial factor model, Biostatistics. 4 (2003), 569-582.

[5] G. S. Trujillo, PATHMOX approach: segmentation trees in partial least squares path modeling, Ph.D., Diss., Universitat Politècnica de Catalunya, 2009.

[6] H. H. 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(1) (1998), 99-21.

[7] H. H. Kelejian and I. R. Prucha, Generalized moment estimator for the autoregressive parameter in a spatial model, International Economic Review 40(2) (1999), 509-533.

[8] J. P. LeSage, Regression analysis of spatial data, The Journal of Regional Analysis and Policy 27(2) (1997), 83-94.

[9] J. Oud and H. Folmer, A structural equation approach to spatial dependence models, Geographical Analysis 40 (2008), 152-166.

[10] J. W. Hogan and R. Tchernis, Bayesian factor analysis for spatially correlated data, with application to summarizing area-level material deprivation from census data, J. Amer. Statist. Assoc. 99(466) (2004), 314-324.

[11] L. Anselin, Spatial Econometrics: Methods and Models, Kluwer Academic Publishers, Dordrecht, Boston and London, 1988.

[12] L. Lee, Generalized Method of Moment Estimation of Spatial Autoregressive Processes, Department of Economics, Ohio State University, Columbus, OH, 2001a.

[13] L. Lee, GMM and 2SLS Estimation of Mixed Regressive, Spatial Autoregressive Models, Department of Economics, Ohio State University, Columbus, OH, 2001b.

[14] M. Arnold, GMM estimation of the autoregressive parameter in a spatial autoregressive error model using regression residuals, Institut fur Wirtschafts und Sozialstatistik, Universität Dortmund, Dortmund, Germany, 2007.

[15] P. Chausse, Computing generalized method of moments and generalized empirical likelihood with R, Journal of Statistical Software 34 (2010), 1-35.

[16] P. Congdon, A spatial structural equation model for health outcomes, J. Statist. Plann. Inference 138 (2008), 2090-2105.

[17] Schumacker and Lomax, A Beginner's Guide to Structural Equation Modeling, Lawrence Erlbaum Associates Inc, New Jersey, 2004.

[18] T. S. Breusch and A. R. Pagan, The Lagrange multiplier test and its application to model specification in econometrics, Rev. Econom. Stud. 47 (1980), 239-254.

[19] W. H. Greene, Econometric Analysis, Pearson Education Inc, New Jersey, 2008.

[20] X. Liu, M. M. Wall and J. S. Hodges, Generalized spatial structural equation models, Biostatistics 6(4) (2005), 539-557.

Anik Anekawati: anekawa10@gmail.com

Bambang Widjanarko Otok: bambang_wo@statistika.its.ac.id

Purhadi: purhadi@statistika.its.ac.id

Sutikno: sutikno@statistika.its.ac.id

GENERALIZED METHOD OF MOMENTS APPROACH TO SPATIAL STRUCTURAL EQUATION MODELING

ORIGINALITY REPORT

11%	9%	7%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.mdpi.com Internet Source	2%
2	www.hindawi.com Internet Source	2%
3	jyx.jyu.fi Internet Source	1%
4	"Handbook of Regional Science", Springer Nature, 2014 Publication	1%
5	Submitted to Padjadjaran University Student Paper	1%
6	Yue Zhongwen, Han Pengfei, Zhao Ling. "Risk Analysis of Coal Mine Construction Project Based on Structural Equation Model", 2010 3rd International Conference on Information Management, Innovation Management and Industrial Engineering, 2010 Publication	1%

7	iopscience.iop.org Internet Source	1%
8	www.tandfonline.com Internet Source	<1%
9	academic.oup.com Internet Source	<1%
10	dx.doi.org Internet Source	<1%
11	Submitted to STIE Perbanas Surabaya Student Paper	<1%
12	www.coursehero.com Internet Source	<1%
13	"Advanced Methods for Modeling Markets", Springer Science and Business Media LLC, 2017 Publication	<1%
14	mafiadoc.com Internet Source	<1%
15	Jörg Henseler, Marko Sarstedt. "Goodness-of-fit indices for partial least squares path modeling", Computational Statistics, 2012 Publication	<1%
16	David W. Marcouiller, Kwang-Koo Kim, Steven C. Deller. "Natural amenities, tourism and	<1%

income distribution", Annals of Tourism Research, 2004

Publication

17	onlinelibrary.wiley.com Internet Source	<1%

Exclude quotes	On	Exclude matches	< 10 words
Exclude bibliography	On		