ISSN 2442-6571

Computation of spatial error model with matrix exponential spatial specification approach

Marsono [a,1], Setiawan [b,2,*], Heri Kuswanto [b,3]

[a] BPS-Statistics of Sulawesi Barat Province, Mamuju, 91511, Indonesia
[b] Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia
[1] marsono@bps.go.id; [2] setiawan@statistika.its.ac.id; [3] heri_k@statistika.its.ac.id
* corresponding author

ARTICLE INFO

Article history
Received February 09, 2024
Revised April 14, 2024
Accepted April 26, 2024
Available online August 31, 2024

Keywords
Spatial error model
MESS(0,1)
Maximum likelihood estimation
Gross regional domestic product

ABSTRACT

In spatial regression analysis, we not only consider a location's internal factors but also consider the spatial factors that may affect the relationship. The model of spatial dependence between regions caused by unknown factors or errors is known as the Spatial Error Model (SEM). In its application to large datasets, SEM suffers from several problems in parameter estimation and computational time. One of the methods to solve this problem is to use Matrix Exponential Spatial Specification (MESS). This research aims to find another alternative to modeling data containing spatial autocorrelation errors as a substitute for SEM. MESS(0,1) is named as an alternative model to SEM. With the advantage of MESS features, the MESS(0,1) model is expected to be faster in analytics and computation compared to SEM when using Maximum Likelihood Estimation (MLE). This study aimed to evaluate the effectiveness of the MESS (0,1) model as an alternative to SEM using MLE based on simulation studies and real data analysis. Simulation studies were conducted by generating data from small samples to large samples and then estimating parameters with the MESS(0,1) and SEM models. Then we compared the performance of the two models with the time used during estimation and the root mean square error (RMSE). In addition, it is applied to real data, namely Gross Regional Domestic Product (GRDP) data. The real data used is the GRDP of the construction category on Java Island in 2021. This is in line with the massive infrastructure development as a government program. The independent variables used and considered influential on the GRDP of the construction sector are domestic investment, foreign investment, labor, and wages. Based on the simulation study results, the computation time for estimating the parameters of MESS(0,1) is faster than the SEM model. In addition, in terms of accuracy, the RMSE indicator shows MESS(0,1) is more accurate than the SEM. In addition, the MESS(0,1) and SEM models were applied to the real data. The modeling real data results show that all variables significantly positively affect GRDP in the construction category.

1. Introduction

The availability of georeferenced data sets and the exponential rise of geographic information system (GIS) technology have made it possible for researchers to do spatial statistical analyses on thousands, even millions, of observations. Large spatial data is one of the parts of huge spatial data [1]. In daily life, statistical analysis of large spatial data sets is necessary to draw thorough conclusions, particularly when examining the link across regions (spatial dependency).

Spatial regression is a type of data analysis that examines the relationship between independent and dependent variables while considering spatial dependency [2]. The field of spatial regression has advanced quickly in the past ten years. The model that has been constructed comprises specification, estimation of parameters, and interpretation. A common spatial linear regression specification is spatial autocorrelation in the dependent variable, or spatial autoregressive (SAR) [3], and spatial autocorrelation in the errors or spatial error model (SEM) [4]. In order to choose either the SAR model or the SEM model for modeling, the spatial dependency test should be conducted first [5]–[8]. When the spatial dependency test results show the existence of error correlation, the model used is the SEM model [9], [10]. Error autocorrelation in SEM can be written in the form of $\lambda\mathbf{Wu}$, where λ is the spatial autocorrelation error parameter, $\mathbf{W}$ is a spatial weight matrix, and u is a vector of unobservable error terms. MLE is utilized since conventional techniques for estimating SEM models, including simple least squares, are biased and ineffective [5]. The existence of the Jacobian matrix component in the form of $|\mathbf{I} - \lambda\mathbf{W}|$ in MLE distinguishes it from least squares estimation [11].

Researchers find it challenging to estimate the parameters of SEM models when working with large data sets. The difficulties encountered are the length of estimation and computation time due to numerical iteration to obtain parameters and the computational problem in computing the Jacobian matrix [9], [10]. Computationally, the larger the data or spatial unit used, the more likely it is to form a singular matrix in parameter estimation so that the estimator obtained will be biased. This computational issue has multiple solutions available, beginning with the eigenvalue computation in the spatial weight matrix, sparsity and Cholesky decomposition applied to the Jacobian matrix, determinant approximation, and a polynomial series approach to spatial weights [12]. In general, some of the solutions developed still perform decomposition on the Jacobian matrix [13]. The researcher offers another solution to overcome computational problems: using an exponential matrix approach called MESS [12], [13]. MESS was chosen because it has several properties that make MESS more advantageous from an analytical and computational perspective than SEM when using MLE. MESS does not require the computation of the Jacobian matrix in its computation and has a unique solution in parameter estimation [13].

Lesage and Pace originally introduced the MESS model in 2007 as a substitute for the SAR model. Some researchers name MESS lag, or MESS(1,0), as a substitute for the SAR model, while for the SEM, the substitute is MESS error, or MESS(0,1) [9]. We chose the term MESS(0,1) in this study based on recent research. The $(\mathbf{I} - \lambda\mathbf{W})$ matrix in SEM is substituted with a matrix $e^{\tau\mathbf{W}}$ so that it becomes the MESS(0,1) model with .τ .is a spatial scalar parameter in the error. The use of $e^{\tau\mathbf{W}}$ has many advantages, one of its properties is $|e^{\tau\mathbf{W}}| = 1$ so that it does not calculate the Jacobian matrix as in the SEM. MESS(0.1) with MLE estimation has not been developed by researchers, so in this study, we want to estimate MESS(0.1) with MLE and see the performance of MESS(0,1) with a simulation study and real data modeling.

Real data modeling uses GRDP data sets by category. One of the important data sets that the government focuses on and is large in size is the district's GRDP. The manufacturing, construction, mining, agricultural, and service categories are among the classifications or business sectors that make up the GRDP. The government, whether at the federal or municipal levels, is eager to raise GRDP since it gauges how well a region has developed. A community will be more prosperous if its GRDP is higher. For modeling using MESS(0,1), the construction category was chosen because the Indonesian government is very intensive in infrastructure development so as to see the influence of investment, employee, and wage variables on GRDP in the construction category. In addition, the results of the spatial dependency test before modeling show that the GRDP construction category shows a spatial error model.

This study aimed to evaluate the effectiveness of the MESS (0,1) model as an alternative to SEM using MLE based on simulation studies and real data analysis. Simulation studies were conducted by generating data from small samples to large samples and then estimating parameters with the MESS(0,1) and SEM models. In addition, it is applied to real data, namely Gross Regional Domestic Product (GRDP) data. The real data used is the GRDP of the construction category on Java Island in 2021. This research is very useful for the government when it wants to analyze spatially large data more quickly, such as GRDP data.

2. Related Work

This section discusses related approaches such as the spatial error model, matrix exponential spatial, spatial weights, the GRDP, and the concept of the independent variable.

2.1. Spatial Error Model

The concept of spatiality stems from Tobler's Law, which reads: "Everything is related to everything else, but near things are more related than distant things." This law is a pillar of regional science studies. Spatial effects are common between one region and another. Several models have been built with the addition of spatial interaction effects in either the dependent variable or the error. One of them is the spatial error model (SEM) [5]. SEM is produced when geographic units and error factors are combined in an ordinary least squares model. The SEM shown in (1), expressed in matrix notation, contains a dependent variable that depends on the autocorrelation of the error term

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

where $\mathbf{y}$ is the dependent variable, $\mathbf{X}$ is the independent variable, λ is the spatial autocorrelation coefficient of the error, $\mathbf{W}$ is the spatial weighting matrix on the error with size (N x N), $\boldsymbol{\beta}$ is the model parameter, and ε is a normally distributed, identical, and independent with a mean of zero and variance σ^2 [5]. Although it is often suggested in the literature, the value λ for the interval (-1, +1). One way of making the λ values stationary is rows or columns can normalize the matrix W. For row normalization, $\mathbf{W}$ is normalized so that the elements of each column sum to one [10].

To make it easier to estimate parameters, the SEM in (1) is changed to (2):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{B}^{-1}\boldsymbol{\varepsilon} \tag{2}$$

with $\mathbf{B} = \mathbf{I} - \lambda\mathbf{W}$ and assuming $\varepsilon \sim N(0, \sigma^2)$ with the density function for ε_i is:

$$f(\varepsilon_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\varepsilon_i^2}{2\sigma^2}\right) \; ; -\infty < \varepsilon_i < \infty \; ; i = 1, 2, \dots, N \tag{3}$$

There are two main approaches in estimating models containing spatial interaction effects: one based on the MLE principle and the other on instrumental variables or the generalized method of moment (IV/GMM). IV/GMM estimators differ from MLE estimators in that they do not rely on the assumption of normality of the errors. One of the drawbacks of the IV/GMM estimator is the possibility that the final estimated coefficient λ will be out of range. This is in contrast to MLE, as it is restricted to the parameter space by the Jacobian log-likelihood function [14]. Parameter estimation is done using the MLE by maximizing the likelihood function. The likelihood function SEM with $\theta'_{SEM} = \left[\boldsymbol{\beta}', \lambda, \sigma^2\right]$ becomes (4):

$$L(\boldsymbol{\theta}_{SEM}) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}|\mathbf{B}|\exp\left(-\frac{1}{2\sigma^2}(\mathbf{B}(\mathbf{y}-\mathbf{X\beta}))'(\mathbf{B}(\mathbf{y}-\mathbf{X\beta})\right) \quad (4)$$

The log-likelihood function of the SEM model looks like in (5)

$$In\ L(\boldsymbol{\theta}_{SEM}) = -\frac{N}{2}In(2\pi\sigma^2) + \text{In}|\mathbf{B}| - \frac{(\mathbf{B}(\mathbf{y}-\mathbf{X\beta}))'\mathbf{B}(\mathbf{y}-\mathbf{X\beta}))}{2\sigma^2} \quad (5)$$

By equalizing the first derivative of the likelihood function and equating it to zero, the MLE estimator $\boldsymbol{\theta}^{\wedge}$ of the parameter can be obtained. This is the result obtained in (7) -(9).

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = \frac{1}{\sigma^2}[(\mathbf{X}'\mathbf{B}'^{\mathbf{B}_\mathbf{y}}) - (\mathbf{X}'\mathbf{B}'\mathbf{BX})\boldsymbol{\beta})] \quad (6)$$

$$\frac{\partial L}{\partial \lambda} = -(tr(\mathbf{B}^{-1})\mathbf{W}) + \frac{1}{\sigma^{\wedge 2}}(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}})'\mathbf{B}'\mathbf{W}(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}) \quad (7)$$

$$\frac{\partial L}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}(\mathbf{y}-\mathbf{X\beta})'\mathbf{B}'\mathbf{B}(\mathbf{y}-\mathbf{X\beta}) \quad (8)$$

By equating (7) and (9) to zero, the estimators $\widehat{\boldsymbol{\beta}}$ and $\sigma^{\wedge 2}$ can be obtained as follows;

$$\boldsymbol{\beta}^{\wedge} = (\mathbf{X}'\mathbf{B}'\mathbf{BX})^{-1}\,\mathbf{X}'\mathbf{B}'\mathbf{By} \quad (9)$$

and

$$\sigma^{\wedge 2} = \frac{1}{N}(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}})'\mathbf{B}'\mathbf{B}(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}) \quad (10)$$

Computationally, (10) and (11) cannot be solved because they still contain other unknown parameters, namely λ . To get the estimator $\lambda^{\wedge}$, use numerical iteration to obtain the parameters [9]. In addition, calculating the determinant of the Jacobian matrix will make the estimation time longer when using large data. This has led researchers to look for such solutions. This computational issue has multiple solutions available, beginning with the eigenvalue computation in the spatial weight matrix, sparsity and Cholesky decomposition applied to the Jacobian matrix, determinant approximation, and a polynomial series approach to spatial weights [12]. The researcher offers another solution to overcome computational problems: using an exponential matrix approach called MESS [12], [13].

2.2. Matrix Exponential Spatial

The MESS model is a spatial model that uses an exponential matrix in its decay. The exponential matrix is defined as,

$$e^{\tau\mathbf{W}} = \sum_{\text{i}=0}^{\infty}\frac{\tau^{\text{i}}\mathbf{W}^{\text{i}}}{\text{i}!} = \text{I} + \tau\mathbf{W} + \frac{\tau^2\mathbf{W}^2}{2} + \cdots + \frac{\tau^{\text{i}}\mathbf{W}^{\text{i}}}{\text{i}!} + \cdots \quad (11)$$

where τ is the scalar spatial coefficient, i is the spatial unit, and W is the spatial weight matrix. The MESS model offers many benefits as a substitute to overcome the difficulties that SEM models face in parameter estimation. Due to the features of MESS, exponential matrices are particularly advantageous, such as [15], [16]:

$\text{e}^{\tau\mathbf{W}}$ is a non – singular matrix

$$(e^{\tau \mathbf{W}})^{-1} = e^{-\tau \mathbf{W}}$$

$$|e^{\tau \mathbf{W}}| = e^{trace(\tau \mathbf{W})} = 1$$

The first property will result in a covariance matrix that is always positive. The second property will make it easier to analyze mathematically. The third property will speed up computation because there is no need to calculate the Jacobian matrix. With the properties that MESS has, it is overall very effective when analyzing large amounts of data [9], [10], [13], [17]. MESS can be applied to replace autoregressive type models, including SEM; for the MESS(0,1) substitute the SEM [9].

2.3. Spatial Weights

One of the key components of spatial regression that demonstrates spatial dependency is spatial weight. The particular relationship between regions of spatial observation units is described by the spatial weight, which is a non-negative $N \times N$ matrix and is represented by $\mathbf{W}$ in the model [5]. The spatial weight can be the weak side of the spatial model because the researcher can subjectively determine it into customized weights. The spatial weight (**W**) can be obtained based on distance, neighbors, and contiguity. Spatial contiguity relationships consist of linear, rook, bhisop, and queen [18].

The spatial weighting **W** can be determined by specifying the economic science underlying the spatial theory of econometrics [10]. The impact of misspecifying the spatial weight can lead to differences in the parameters obtained from the estimation results. Therefore, it is necessary to choose weights that can explain the relationship between regions in real terms. In this study, queen contiguity weights were chosen because they represent GRDP data for the construction category in Java. In general, economic growth is closely related to regional contiguity.

2.4. GRDP and Independent Variables Concept

GRDP is a crucial metric that can be employed to ascertain a region's economic state over an extended period of time. Three perspectives—the production, income, and expenditure sides—can be used to analyze the concept of GRDP itself. The amount that the three methods should yield is the same in theory: the spending must match the quantity of finished goods and services produced as well as the quantity of revenue from the factors of production. Therefore, the gross domestic product (GDP) is essentially the amount of added value produced by all business units in a given area or the total worth of final goods and services provided by all economic units (typically for one year) [19].

These production units are grouped into 17 categories. Among the 17 categories of GRDP in Indonesia, there are four main categories that show the largest percentage contribution compared to other categories. With the total contribution of the four categories reaching 61 percent in 2021, the main sectors include the Manufacturing Industry, Agriculture, Forestry, and Fisheries category, Wholesale and Retail Trade category, and Construction category [20]. Of the four categories, the construction category is the main focus of the current government, namely infrastructure development. The GRDP of the construction category includes three activities: building construction, civil building construction, and special construction [19].

The independent variables used in this study are domestic investment, foreign investment, employees, and wages in the construction sector. Investment, which is also commonly referred to as capital investment or capital formation, can be defined as the expenditure or expenditure of investors or companies to purchase capital goods and production equipment to increase the ability to produce goods and services available in the economy. The regional investment pattern, which functions as a capital builder for regional development to achieve various development goals, can be grouped into two categories: domestic and foreign investment patterns. A net wage or salary is the reward received during a month by laborers/employees in the form of money or goods paid by the company, office, or employer [21]. An employee is everyone who can do work to produce goods and/or services to meet their own needs and for the community [21].

According to previous research, labor, domestic investment, and foreign investment have a positive relationship with GRDP, which means that the more labor used, the higher the GRDP. The same thing happens when investment gets bigger, and GRDP will increase [22], [23]. Other research suggests that human capital positively influences economic growth, as calculated by GRDP [24]. Some infrastructure built in the construction sector, such as roads and bridges, can increase economic activity, which in turn increases economic growth [25] . The relationship between wages and economic growth has a positive correlation [26]. Of the four dependent variables used in real data analysis, they are strongly suspected to influence the GRDP construction sector positively.

3. Method

This section contains an explanation of the creation of datasets for simulation studies and real data analysis, the proposed model and model specification for real data analysis, and the parameter estimation of MESS(0,1).

3.1. Creating Simulated Data and Real Data

Simulation studies and real-data analysis were conducted to obtain the research objectives. The purpose of the simulation study here is to see the performance during the parameter estimation process so that the computation time can be known. In addition, the accuracy of the prediction values obtained must also be considered. This simulation uses the Matlab toolbox built by Lesage and has been modified [18]. The simulation uses Monte Carlo simulation with one thousand iterations, four independent variables, τ set at four, and the generated spatial weights. The simulation compares the estimation results of the SEM and MESS(0,1) models using the same computer. Experiments were started using small samples (N is less than 1,000), medium samples (N between 1,000 and 10,000), and large samples (N is larger than 10,000).

For real data analysis, it is better to use large data sets. However, due to the limited availability of data owned by researchers, it uses cross-sectional data on GRDP in the construction category in 2021 on Java Island, so the data used is 119 units. Such a large sample is still classified as a small sample (N less than 1,000), so the difference in computing time between the SEM and MESS models is not significant. However, because it aims to model and analyze data to see the factors that affect the GRDP of the construction category, the GRDP data used in this research comes from BPS-Statistics Indonesia. The data used is GRDP data at current prices; the independent variables used consist of employees, wages, domestic investment, and foreign investment in the construction category. Investment data comes from the Ministry of Investment and the Investment Coordinating Board of the Republic of Indonesia. The variables used in this study are presented in Table 1.

The GRDP of the construction category is one of the main contributing sectors to the national GDP. This is in line with the rapid development of infrastructure throughout Indonesia. Infrastructure development is part of the GRDP of the construction category, so it needs to be seen as a spatial dependency. The construction category was chosen because the Indonesian government has been very aggressive in infrastructure development in recent years, such as roads, bridges, airports, and others.

Researchers want to see the effect of labor, wages, and investment on GRDP in the construction category. In addition, the results of the spatial dependency test before modeling show that the construction category GRDP shows a spatial error model.

Table 1. Information Research Variables

Variables	Description	Unit
y	GRDP by contruction category	Billion Rupiah
X1	Foreign investment by contruction category	US Dollar
X 2	Domestic investment by contruction category	US Dollar
X3	Employee by contruction category	Million Residents
X 4	Wage by contruction category	Million Residents

3.2. Proposed Model

By utilizing the exponential spatial matrix, the proposed model is MESS(0,1). The model is obtained by replacing the $\mathbf{B} = \mathbf{I} - \lambda\mathbf{W}$ matrix in SEM with $e^{\tau\mathbf{W}}$ so that (2) becomes:

$$y = \mathbf{X}\boldsymbol{\beta} + \mathrm{e}^{-\tau\mathbf{W}}\boldsymbol{\varepsilon} \tag{12}$$

where ε is a properly distributed, identically independent error with zero mean and variance σ^2 .

There are several assumptions used in model estimation:

- Assumption 1. $\mathbf{W}$ is a constant spatial weight matrix or nonstochastic with zero diagonals, $\mathbf{W}$ is uniformly bounded in both row and column sums in absolute value.
- Assumption 2. The disturbances ε are identic and independent with mean 0 and variance σ^2
- Assumption 3. $\tau \in A$ where A is the compact interval. Moreover, τ is in the interior of A .
- Assumption 4. The predictor matrix $\mathbf{X}$ has full rank. Furthermore, the elements of $\mathbf{X}$ are constantly bounded for all n such that $\lim_{N\to\infty} \frac{\mathbf{X}'\mathbf{X}}{N}$
- Assumption 5. The number of sample observations (N) is large.

The following are steps in parameter estimation with MLE:

- Assuming the MESS(0,1) model has a normal distribution, find its error distribution function.
- Form the MESS(0,1) model using normally distributed errors as the previous premise.
- Create the MESS(0,1) model equation's likelihood function.
- Create a log-likelihood function by utilizing the likelihood function generated in the preceding step.
- Using the first derivative of the log-likelihood function with regard to the parameter, find the formula for parameter estimation. The first derivative of the log-likelihood function is compared with the estimated parameter, $\boldsymbol{\theta}'_{MESS(0,1)} = (\boldsymbol{\beta}', \tau, \sigma^2)$, by equating it to zero.
- Calculating $\tau^{\wedge}$ parameters with the method suggested by [13].
- Calculate the $\beta^{\wedge}$ parameters and $\sigma^{\wedge 2}$

After estimating the parameters, a hypothesis test is performed on the parameters. However, hypothesis testing is not the focus of this study because we want to see the performance during parameter estimation. The indicator used is the estimation speed, so the total time required for estimation is calculated. In addition, to measure the accuracy of a model's estimated results using RMSE, it is calculated by squaring the error divided by the amount of data then rooted. The RMSE formula is in (14).

$$RMSE = \sqrt{\frac{\Sigma(y-y^{\wedge})^2}{N}} \tag{13}$$

A low RMSE value indicates that the variation in the value produced by a prediction model is close to the variation in the observed value. The smaller the RMSE value, the closer the predicted value is to the observed value.

To see the performance of the proposed model and apply it to real data, a comparison is made between MESS(0,1) and SEM. Model specifications for the MESS(0,1) model in (15) and the SEM model in (16) are as follows:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u,\ e^{\tau \mathbf{W}} u = \boldsymbol{\varepsilon} \tag{14}$$

and

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u,\ u = \lambda \mathbf{W}\mathrm{u} + \boldsymbol{\varepsilon} \tag{15}$$

The specification of the model built refers to the variables used by the Indonesian Ministry of National Development Planning (Bappenas) in 2006 in regional and cross-sector planning, where the dependent variable used is GRDP data and the independent variables are the number of workers, labor wages, and domestic and foreign investment. The spatial weights **W** used for parameter estimation use queen contiguity weights with the consideration that construction buildings, namely highways, connect adjacent areas

3.3. Parameter Estimation

The step in MLE is to build a likelihood function MESS(0,1) with parameters.

$\boldsymbol{\theta}'_{MESS(0,1)} = (\boldsymbol{\beta}', \tau, \sigma^2)$ is:

$$\boldsymbol{L}(\boldsymbol{\theta}'_{MESS(0,1)}) = \frac{1}{(2\pi\sigma^2)^{N/2}} |e^{\tau W}| \exp\left(-\frac{1}{2\sigma^2}(e^{\tau W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' e^{\tau W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right) \tag{16}$$

As explained earlier, MESS has the properties $|e^{\tau W}| = 1$, the log-likelihood function becomes;

$$\text{In } L(\boldsymbol{\theta}'_{MESS(0,1)}) = -\frac{\mathrm{N}}{2}\text{in}(2\pi\sigma^2) - \frac{(e^{\tau\mathbf{W}}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'e^{\tau\mathbf{W}}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}))}{2\sigma^2} \tag{17}$$

To obtain the $\boldsymbol{\theta}'_{MESS(0,1)} = (\boldsymbol{\beta}', \tau, \sigma^2)$ parameter with MLE, we maximize the function of (18). One common way is to derive the likelihood function for each parameter and equate it to zero. The result is (19)-(21);

$$\frac{\partial \mathrm{L}}{\partial \boldsymbol{\beta}} = \frac{1}{\sigma^2}\left[\left(\mathbf{X}'{e^{\tau\mathbf{W}}}'e^{\tau\mathbf{W}}\mathbf{y}\right) - (\mathbf{X}'{e^{\tau\mathbf{W}}}'e^{\tau\mathbf{W}}\mathbf{X})\boldsymbol{\beta})\right] \tag{18}$$

$$\frac{\partial \mathrm{L}}{\partial \boldsymbol{\beta}} = \frac{1}{\sigma^{\wedge 2}}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' e^{\tau\mathbf{W}}{}' \mathbf{W}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \tag{19}$$

$$\frac{\partial L}{\partial \sigma^2} = \frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' e^{\tau\mathbf{W}'} e^{\tau\mathbf{W}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \tag{20}$$

By equating (18) and (20) to zero, the estimators $\boldsymbol{\beta}^{\wedge}$ and $\sigma^{\wedge 2}$ can be obtained as follows:

$$\boldsymbol{\beta}^{\wedge} = (\boldsymbol{X}'e^{\tau W}{}'e^{\tau W}\boldsymbol{X})^{-1}\boldsymbol{X}'e^{\tau W}{}'e^{\tau W}{}'\boldsymbol{y} \quad (21)$$

and

$$\sigma^{\wedge 2} = \frac{1}{N}(\mathrm{e}^{\tau\mathbf{W}}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'\mathrm{e}^{\tau\mathbf{W}}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})) \quad (22)$$

The MESS(0,1) model allows closed-form solutions for its parameters, in contrast to the SEM model, which requires numerical iteration to determine the spatial error coefficients [13]. The steps to obtain the τ parameter in various ways are carried out, one of which is the matrix-vector product approach [9]. After obtaining $\tau^{\wedge}$, the spatial error coefficient can be calculated using the approach in (23):

$$\lambda^{\wedge} \approx 1 - \exp(\tau) \quad (23)$$

4. Results and Discussion

This section contains an explanation of the results of the simulation study and the results of real data analysis.

4.1. Simulation Study

The purpose of the simulation study here is to see the performance during the parameter estimation process so that the computation time can be known. In addition, the accuracy of the prediction values obtained should also be considered. The simulation compares the estimation results of the SEM and MESS(0,1) models using the same computer. Experiments were started using small samples (N is less than 1,000), medium samples (N between 1,000 and 10,000), and large samples (N is larger than 10,000). The results obtained can be seen in Table 2.

Table 2. Comparison of simulation results of SEM and MESS(0,1) models

Number of observations	Model Evaluation	SEM	MESS(0,1)
N=100	Time Estimation(s)	0.1450	0.0730
	RMSE	6.0637	5.9576
N=500	Time Estimation(s)	0.7140	0.0760
	RMSE	6.8492	6.5223
N=1000	Time Estimation(s)	0.8520	0.0790
	RMSE	5.5323	6.0245
N=5000	Time Estimation(s)	7.9200	0.0960
	RMSE	6.2762	6.0888
N=10000	Time Estimation(s)	16.7230	0.1210
	RMSE	6.3396	6.1561
N=25000	Time Estimation(s)	66.5300	0.1440
	RMSE	6.1340	6.1716
N=50000	Time Estimation(s)	207.0730	0.3760
	RMSE	6.1804	6.1084
N=75000	Time Estimation(s)	488.1320	0.4400
	RMSE	6.5733	6.1863
N=100000	Time Estimation(s)	703.8330	0.7390
	RMSE	6.2352	6.1216

Based on Table 2, when using a small sample (less than 1,000), there is no significant difference in estimation time between the SEM and MESS(0,1) models, which is still below one second. The difference in estimation time begins to appear when using a medium sample (N between 1,000 and 10,000). The SEM model takes longer than MESS(0.1). The difference is very clear when using a large sample (N is larger than 10,000), and the time required by the SEM model is longer than the MESS(0.1) model. The simulation results prove that MESS(0.1) is faster in terms of computation. As shown in Fig. 1, the parameter estimation time with a large sample (N = 100,000) required by the MESS(0,1) model is very fast, which is less than one second compared to the SEM model, which takes a long time.

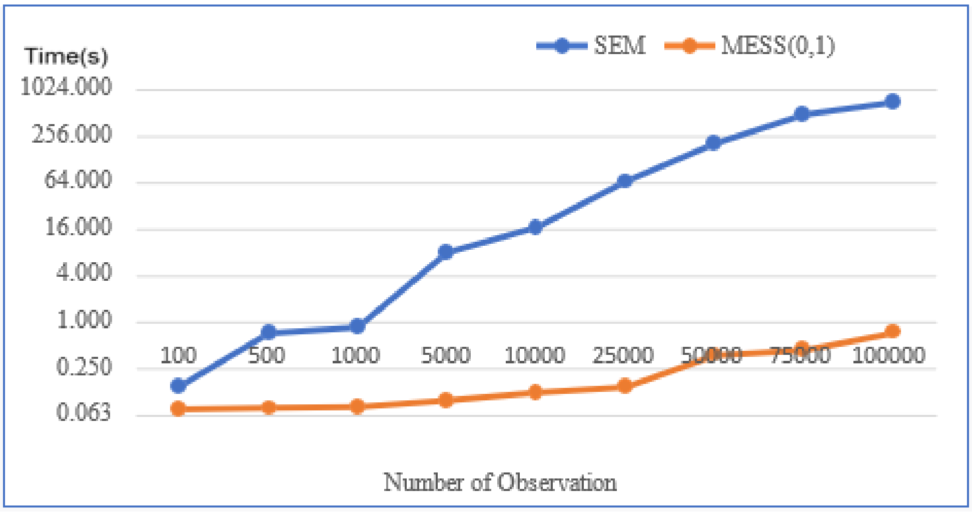

Fig. 1. The difference in estimation time between SEM and MESS(0,1)

Based on Table 2, when comparing the RMSE values of the SEM and MESS(0.1) models, it can be seen that the RMSE value of MESS(0.1) is lower using both small and large samples. Although not significantly different, it shows that the MESS(0.1) model is more accurate than the SEM model. When using N = 100, the RMSE value of the SEM model is 6.0637, while the MESS (0.1) model is 5.9576. With increasing samples, when using N = 100000, the RMSE value of the SEM model is 6.2352, while the MESS (0.1) model is 6.1216.

4.2. Real Data Analysis

In accordance with the simulation study results, the MESS(0,1) model is very effective in using large data. However, due to the limited availability of data owned by researchers, it uses cross-sectional data on GRDP in the construction category 2021 on the Java Island, so the data used is 119 units. Such a large sample is still classified as a small sample (N less than 1,000), so the difference in computing time between the SEM and MESS models is insignificant. However, because it aims at modeling and analyzing data to see the factors that affect the GRDP of the construction category. Before estimating the parameters, it is necessary to visually see the relationship between the independent variables and GRDP in the construction category from the scatterplot in Fig. 2.

Based on the scatterplot in Fig. 2, The vertical axis is the dependent variable, GRDP construction category, and the horizontal axis is the independent variable. It shows that all independent variables, including foreign investment, domestic investment, employee, and wage, by construction category have a positive relationship pattern with the GDP of the construction category. The greater the value of the independent variable, the greater the value of GRDP in the construction category.

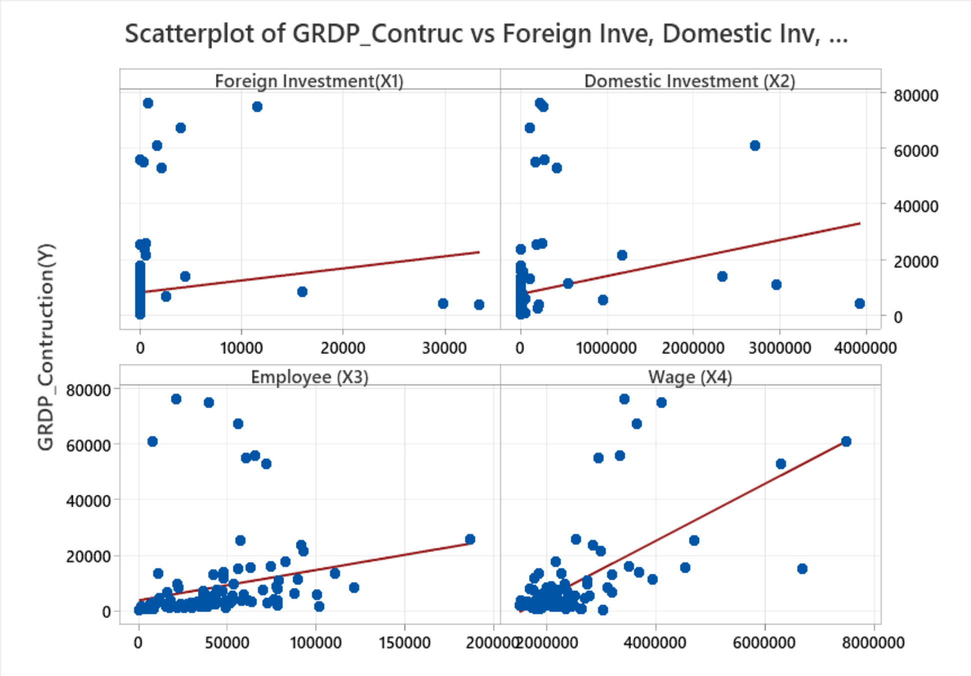

Fig. 2. Scatterplot of Independent Variables against GRDP in the Construction Sector

The distribution of GDRP in the construction sector in Java Island is shown in Fig. 3.

Fig. 3. Map of Distribution of GRDP Value in Construction category in Java Island in 2021

Based on Fig.3, GRDP values are divided into three classes according to the color gradation: low, medium, and high GRDP. GRDP below 25531.157 (billion rupiah) is 111 districts/cities, GRDP between 25531.157 -50910.993 (billion rupiah) is 1 district/cities, and GRDP above 50910.993 (billion rupiah) is 7 districts/cities., so to see the magnitude of the effect, estimation is carried out with the MESS(0,1) and SEM with with Queen Contiguity weights. Parameter estimation using the MESS(0,1) and SEM is shown in Table 3. Of all the independent variables used, all significantly influence the GRDP of the construction sector in Java Island using either the MESS(0,1) or SEM model. The parameter coefficient values estimated by both models show almost the same value and sign, indicating their influence on GRDP in the construction sector.

Table 3. Comparison of estimation results and performance of MESS(0,1) and SEM with GRDP data of Construction Sector in Java Island

Parameter	MESS(0,1)		SEM	
	Coefficient	*P-Value*	*Coefficient*	*P-Value*
β_0	-15107.40770	0.00001	-15063.02100	0.00001
β_1	0.12948	0.00001	0.12912	0.00001
β_2	0.02895	0.00000	0.02893	0.00000
β_3	0.03301	0.00303	0.03315	0.00208
β_4	0.00889	0.00000	0.00887	0.00000
τ	-0.24895	0.09631		
λ	0.220*		0.22700	0.06038
Estimation Time	0.10500		0.14000	
R-Square	0.84020		0.84170	
RMSE	9965.14859		9964.83459	

[a] Remarks*: Calculated using (24)

Based on Table 3, the parameters for the MESS(0,1) model intercept (β_0) is -15107.40770, foreign investment (β_1) is 0.12948, domestic investment (β_2) is 0.02895, employee (β_3) is 0.03301, wage (β_4) is 0.00889 with a scalar spatial coefficient (τ) of -0.24895. Parameter for SEM, intercept (β_0) is-15063.02100, foreign investment (β_1) is 0.12912, domestic investment (β_2) is 0.02893, employee (β_3) is 0.03315, wage (β_4) is 0.00887 with a spatial autoregressive coefficient of 0.22700. The parameter estimation results of the MESS(0,1) model have similarities with the SEM model, namely that all independent variables have a significant effect (α = 5%). In addition, the R square and RMSE values are also not different between MESS(0,1) and SEM.

In addition to the independent variables, the influence of other unknown variables (the error term) of neighbors that intersect also has a significant effect (α = 10%) on the GRDP of a region. This is reflected in the positive spatial autocorrelation error coefficient (λ = 0.227). The magnitude of the autocorrelation coefficient in the MESS(0,1) model can be calculated with the approach in (24). So the value of the autocorrelation error coefficient in the MESS(0,1) model is 0.220.

A notable difference between the MESS(0,1) and SEM models lies in the time required for parameter estimation. Since the data used is N = 119, the observation time required to estimate the MESS(0,1) parameters is 0.10 seconds, while the SEM is (0.14). Regarding estimation time in computation, the MESS(0,1) model is better than the SEM model because the time required is faster even though it is below one second because the data used is a small sample. When using a large data set, the difference in time required for estimation will be more obvious. Fast estimation is needed for policymakers in important situations and conditions. Regarding accuracy using RMSE and R square, the MESS(0,1) and SEM models give similar results. In other words, MESS(0,1) can replace the SEM model well for small and large data.

Based on the analysis of real data with MESS(0,1) and SEM in this study, it can be utilized for government policy in increasing GRDP or economic growth in the construction sector. Both foreign and domestic investment variables have a positive and significant effect on GRDP, so the government must try to include foreign and domestic investors in investing in Indonesia. In addition, employment in the construction sector must be increased to absorb unemployment, and wages given to workers must also be increased for the welfare of workers, which ultimately increases GRDP or economic growth.

5. Conclusion

Based on this research, it can be concluded that MESS(0,1) is an excellent substitute for the SEM model. This can be proven from simulation studies that show that the MESS(0,1) model is faster in computation and more accurate in the parameter estimation process compared to the SEM model. In addition, the results of real data analysis by modeling construction category GRDP data with queen contiguity weights show the same performance because the data used is still a small sample. Independent variables have the same results between MESS (0,1) and SEM in terms of value, sign, and significance. Future work should be able to use large data sets so that the performance efficiency of MESS(0,1) can be seen. In addition, the spatial weights used can be customized weights related to the research. In research that still uses cross-sectional data, it is recommended to use panel data to make the analysis more comprehensive by considering the effect of time.

Acknowledgment

The authors are grateful to Research Program Funding and Community Service Program Fiscal Year 2023-2024 from Ministry of Education, Culture, Research and Technology under the PPS-PDD grant and Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia.

Declarations

Author contribution. The novelty of our research is about quick computation in spatial error model estimation with matrix exponential spatial specification approach. and applied to model real data namely gross regional domestic product (GRDP). all authors have read the final manuscript, have approved the submission to the journal, and accept full responsibility regarding the submission and content of the manuscript

Funding statement. Research Program Funding and Community Service Program Fiscal Year 2023 from Ministry of Education, Culture, Research and Technology under the PPS-PDD grant and Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia.

Conflict of interest. The authors declare that they have no conflicts of interest to report regarding the present study.

Additional information. No additional information is available for this paper.

Data and Software Availability Statements

We are willing to share the data and software used in this research. The real data and source code used in this research can be downloaded in https://drive.bps.go.id/f/27997594

References

[1] I. Sabek and M. F. Mokbel, "Machine Learning Meets Big Spatial Data," in *2020 IEEE 36th International Conference on Data Engineering (ICDE)*, Apr. 2020, vol. 2020-April, pp. 1782–1785, doi: 10.1109/ICDE48307.2020.00169.

[2] S. J. Rey and R. S. Franklin, *Handbook of Spatial Analysis in the Social Sciences*. Edward Elgar Publishing, pp. 1-588, 2022, doi: 10.4337/9781789903942.

[3] M. Koley and A. K. Bera, "Testing for spatial dependence in a spatial autoregressive (SAR) model in the presence of endogenous regressors," *J. Spat. Econom.*, vol. 3, no. 1, p. 11, Dec. 2022, doi: 10.1007/s43071-022-00026-7.

[4] R. Chakir and J. Le Gallo, "Spatial Autocorrelation in Econometric Land Use Models: An Overview," in *Advances in Contemporary Statistics and Econometrics*, Cham: Springer International Publishing, 2021, pp. 339–362, doi: 10.1007/978-3-030-73249-3_18.

[5] L. Anselin, *Spatial Econometrics: Methods and Models*, vol. 4. Dordrecht: Springer Netherlands, p. 284, 1988, doi: 10.1007/978-94-015-7799-1.

[6] T. Rüttenauer, "Spatial Regression Models: A Systematic Comparison of Different Model Specifications Using Monte Carlo Experiments," *Sociol. Methods Res.*, vol. 51, no. 2, pp. 728–759, May 2022, doi: 10.1177/0049124119882467.

[7] R. Franzese and L. Curini, *The SAGE Handbook of Research Methods in Political Science and International Relations*. SAGE Publications Ltd, pp. 1-1332, 2020. [Online]. Available at: https://www.torrossa.com/en/resources/an/5018785#page=775.

[8] R. De Siano, V. Leone Sciabolazza, and A. Sapio, "Spatial Econometric Models: Theory," in *SpringerBriefs in Regional Science*, Springer, Cham, 2020, pp. 31–43, doi: 10.1007/978-3-030-54588-8_3.

[9] Y. Yamagata and H. Seya, *Spatial Analysis Using Big Data : Econometrical Methods and Applications.* Elsevier Academic Press, p. 316, 2019. [Online]. Available at: http://www.sciencedirect.com:5070/book/9780128131275/spatial-analysis-using-big-data?via=ihub=.

[10] J. P. Elhorst, *Spatial Econometrics*. Berlin, Heidelberg: Springer Berlin Heidelberg, p. 119, 2014. [Online]. Available at: https://link.springer.com/10.1007/978-3-642-40340-8.

[11] L.-F. Lee, "Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models," *Econometrica*, vol. 72, no. 6, pp. 1899–1925, Nov. 2004, doi: 10.1111/j.1468-0262.2004.00558.x.

[12] Marsono, Setiawan, and H. Kuswanto, "Fast Maximum Likelihood Estimation Of Big Data Spatial Autoregressive Model: Matrix Exponential Spatial Specification Approach," in *2023 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, Oct. 2023, pp. 376–381, doi: 10.1109/IC3INA60834.2023.10285776.

[13] Y. Yang, O. Doğan, and S. Taşpınar, "Fast estimation of matrix exponential spatial models," *J. Spat. Econom.*, vol. 2, no. 1, p. 9, Dec. 2021, doi: 10.1007/s43071-021-00015-2.

[14] J. P. Elhorst, "Spatial Panel Data Models," in *Handbook of Applied Spatial Analysis*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 377–407, doi: 10.1007/978-3-642-03647-7_19.

[15] O. Doğan, Y. Yang, and S. Taşpınar, "Information criteria for matrix exponential spatial specifications," *Spat. Stat.*, vol. 57, p. 100776, Oct. 2023, doi: 10.1016/j.spasta.2023.100776.

[16] Y. Yang, O. Doğan, and S. T. Inar, "Estimation of Matrix Exponential Unbalanced Panel Data Models with Fixed Effects: An Application to US Outward FDI Stock," *J. Bus. Econ. Stat.*, vol. 42, no. 2, pp. 469–484, Apr. 2024, doi: 10.1080/07350015.2023.2200486.

[17] N. Debarsy, F. Jin, and L. Lee, "Large sample properties of the matrix exponential spatial specification with an application to FDI," *J. Econom.*, vol. 188, no. 1, pp. 1–21, Sep. 2015, doi: 10.1016/j.jeconom.2015.02.046.

[18] J. LeSage and R. K. Pace, *Introduction to Spatial Econometrics*. Chapman and Hall/CRC, p. 340, 2009, doi: 10.1201/9781420064254.

[19] B. P. Statistik, "Gross Regional Domestic Product of Regency/City in Indonesia 2017-2021," p. 202, 2022. [Online]. Available at: https://www.bps.go.id/id/publication/2022/06/07/59cf6a8a96b61d09c6d65260.

[20] B. P. Statistik, "Gross Regional Domestic Product of Provinces in Indonesia by Industry 2017-2021 - Central Bureau of Statistics of Indonesia," p. 200, 2022. [Online]. Available at: https://www.bps.go.id/id/publication/2022/04/05/aeec064ce0205363edd1d58c/produk-domestik-regional-bruto-provinsi-provinsi-di-indonesia-menurut-lapangan-usaha-2017-2021.html.

[21] B. P. S. Indonesia, "Analysis of Labor Mobility Results from Sakernas 2022," p. 242, 2023. [Online]. Available at: https://www.bps.go.id/id/publication/2023/10/27/83d4aaefa732e79efa75107b/analisis-mobilitas-tenaga-kerja-hasil-sakernas-2022.html.

[22] Y. Lee, A. M. B. Azlan, and W. Lim, "Dataset on macroeconomic indicators and fiscal decentralisation indices' variables in Central Java Province, Indonesia," *Data Br.*, vol. 50, p. 109554, Oct. 2023, doi: 10.1016/j.dib.2023.109554.

[23] R. N. Wilantari, S. Latifah, W. Wibowo, and H. Al Azies, "Additive mixed modeling of impact of investment, labor, education and information technology on regional income disparity: An empirical analysis using the statistics Indonesia dataset," *Data Br.*, vol. 45, p. 108619, Dec. 2022, doi: 10.1016/j.dib.2022.108619.

[24] H. Saleh, B. Surya, D. N. Annisa Ahmad, and D. Manda, "The Role of Natural and Human Resources on Economic Growth and Regional Development: With Discussion of Open Innovation Dynamics," *J. Open Innov. Technol. Mark. Complex.*, vol. 6, no. 4, p. 103, Dec. 2020, doi: 10.3390/joitmc6040103.

[25] A. F. Aritenang, "The crucial role of motorcycle-based ride-hailing among commuters: The case of Jakarta and Bandung metropolitan areas," *J. Public Transp.*, vol. 26, p. 100082, Jan. 2024, doi: 10.1016/j.jpubtr.2024.100082.

[26] R. T. Siregar, H. P. Silitonga, K. Lubis, and A. Sudirman, "The Impact of GRDP and RWP on Regional Minimum Wage," *JEJAK*, vol. 13, no. 2, pp. 292–306, Oct. 2020, doi: 10.15294/jejak.v13i2.23398.