Media Statistika

p-ISSN 1979 – 3693 e-ISSN 2477 – 0647
MEDIA STATISTIKA 16(1) 2023: 100-111
http://ejournal.undip.ac.id/index.php/media_statistika

COMPARISON OF SPATIAL WEIGHTED MATRIX BETWEEN POWER AND QUEEN ON THE SPATIAL EMPIRICAL BEST LINEAR UNBIASED PREDICTION MODEL
(Study on Per Capita Expenditure in East Java Province in 2019)

Luthfatul Amaliana, Andi Prasetya
Department of Statistics, Brawijaya University, Indonesia

e-mail: luthfatul@ub.ac.id

DOI: 10.14710/medstat.16.1.100-111

Article Info:
Received: 28 June 2023
Accepted: 22 September 2023
Available Online: 12 December 2023

Keywords:
Spatial Analysis, Indirect Estimation, Queen, Power, Expenditure Per Capita, Small Area Estimation
.

Abstract: This study aims to make a comparison related to the spatial weighted matrix of power and queen in the SEBLUP model to estimate per capita expenditure in East Java in 2019. The data used is secondary data then the data were analyzed by the Spatial Empirical Best Linear Unbiased Prediction (SEBLUP). The results of this study indicate that the best spatial weighted matrix for estimating per capita expenditure in East Java using the SEBLUP model is the spatial weighted matrix of Queen, because it produces the smallest MSE value. In this study, the factors that significantly affect East Java's per capita expenditure are population density (X1), number of health facilities (X2), number of public elementary schools (X3), and the percentage of residents who have BPJS as the Fund Assistance Recipients (X5). The novelty of this study are combining multiple determinant factors that have demonstrated their substantial/significant effect on the average per capita expenditure and focusing on the regions characters in intermediate size ($16<n<64$).

1. INTRODUCTION

Small Area Estimation (SAE) is an approach used when direct estimation of a small area by sampling cannot produce sufficient precision. It is often necessary to use indirect estimators to increase the effective sample size (Rao, 2003). The problem of poverty in Indonesia has become a priority that must be resolved in each leadership period. The government always makes improvements related to poverty problems and always collects data related to this poverty data. The government's efforts to control poverty is the holding of the National Economic Survey (SUSENAS) conducted by the Indonesian Bureau of Statistics (*Badan Pusat Statistik*) (BPS, 2016).

To produce more precise information from the data, a data analysis method that has good accuracy is needed, so that in making policy it can achieve the target. Most surveys mostly used sampling to obtain data, because the use of samples is more cost effective and can obtain information on various interesting topics at frequent intervals from time to time. The use of samples in surveys is also widely used to provide estimates not only for the total population of interest, but also for various sub-populations/domains (Rao, 2003).

Rao (2003) states that estimates made with variable values that are of concern only to the time period and unit sample area are referred to as direct estimates. Direct estimates can be used if all areas in the population are used as samples and this estimator is based on a sampling design (Chandra et al., 2007). Insufficient sample size for small area level makes poverty measurement with direct estimation result in large standard error, so that analysis based on these conditions becomes unreliable. To overcome this problem, an estimation method is needed that can provide a better level of accuracy, namely by combining survey data with other supporting data, such as previous census data containing variables with the same characteristics as survey data (Rao, 2003). One method that is often used is Small Area Estimation (SAE). Darsyah (2013) states that SAE is one of the statistical techniques used to estimate the sub-population/domain parameters that have a small sample size. In this estimation, a model is obtained that connects related areas using "strength assistance" from supporting variables. In the case where the observed small area has a random influence that is spatially correlated with each other, the suitable method for prediction is the Spatial Empirical Best Linear Unbiased Prediction (SEBLUP) method.

This method is a development of the Empirical Best Linear Unbiased Prediction (EBLUP) method because the EBLUP model is only suitable for use on data that does not have a random effect that is spatially correlated with each other (not spatial data). Based on research Mutualage (2012), the SEBLUP method is better used than the direct estimation method and EBLUP in estimating the average expenditure per capita in a village or sub-district in Jember Regency. Nusrang et al. (2017) also found similar results, namely the SEBLUP method is better than the EBLUP method in generation data that violates the heteroscedasticity assumption in the error and there is a strong autocorrelation between areas.

According to Getis & Alstads (2004), the selection of a spatial weighted matrix is an important thing in forming a spatial model. Research on determining the spatial weighted matrix in the SEBLUP model has been studied by Asfar (2016). In his findings, it was found that in estimating the average monthly expenditure per household at the sub-district level of Kota and Kabupaten Bogor in 2010, the number of areas observed had an effect on the selection of the spatial weighted matrix used and each case study has a different spatial weighted matrix as well. Asfar's research (2016) states that for a small number of areas (m=16), a good spatial weighted matrix to use is queen, double power (double rank), and a combination of radial and queen. For a medium number of areas (m=64), a good spatial weighted matrix to use is KNN, power (rank), and a combination of radial and queen.

Hence, this study incorporates recommendations drawn from the extensive research conducted by Asfar (2016), particularly concerning the utilization of spatial weighted matrices. Two types of spatial weighting matrices are employed, as suggested by Asfar; firstly, the distance-based spatial weighted matrix is adopted in the form of a power/rank-type matrix. Secondly, the contiguity-based spatial weighted matrix is queen-type matrix. The spatial weighted matrix of power (rank) assumes that the weights are functions of the negative power (rank) of distance (Smith, 2014; Jajang et al., 2017). The power (rank) degree (α) is a positive number and d_{ij} is the Euclidean distance from location i to location j. On the other hand, in the spatial weighted matrix of queen's rule, a space unit can be said to be neighboring another space unit if the two regions share a corner or edge (Suryowati et al., 2018).

Distinguishing itself from Asfar's (2016) and previous research, this study goes a step further by combining multiple determinant factors that have demonstrated their substantial/significant effect on the average per capita expenditure. By integrating these diverse factors, this research seeks to provide a more comprehensive and nuanced understanding of the dynamics affecting per capita expenditure levels within the studied regions. Moreover, this study deliberately focuses on regions characterized by intermediate size, where 16<n<64, thereby introducing a novel perspective into the Small Area Estimation (SAE) methods. This approach not only extends the applicability of SAE but also contributes to the ongoing discourse surrounding economic disparities at the regional level, shedding light on previously unexplored territories within this domain.

2. LITERATURE REVIEW

2.1. Small Area Estimation

Small Area Estimation (SAE) is defined by Saei & Chambers (2003) as a statistical technique that is used when a study has a small sample size or even zero in the desired area; where if direct estimation is carried out, the resulting variance is large or even direct estimation cannot be used if there are no samples in the desired small area. In SAE it works by using indirect estimation or model-based. Indirect estimation is an estimation that is done by borrowing strength from the auxiliary variable that has a relationship with the desired variable (Y) in the same area. According to Rao (2003), indirect estimation in small area estimation has several advantages, namely: 1) The optimum estimator can be formed based on the model that has been formed; 2) The size of the variance of each area can be associated with each of the resulting estimators; and 3) The model can be validated from sample data.

Rao (2003) also stated that small area models can be divided into 2 broad categories, namely aggregate/area level models and unit level models. In the aggregate/area level model, accompanying variables are only available in specific areas, so the use of a unit level model is not possible. Meanwhile, in the unit level model, accompanying variables are available in specific units.

2.2. Spatial Empirical Unbiased Linear Prediction (SEBLUP)

Spatial Empirical Unbiased Linear Prediction (SEBLUP) is a model resulting from the development of the Empirical Unbiased Linear Prediction (EBLUP) model by including random effects between areas. Molina et al. (2007) states that models that have random effects of areas that are independent of each other are as follows:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zv} + \boldsymbol{e} \qquad (1)$$

with $\boldsymbol{y}$: $(y_1, \dots, y_m)^T$, vector that predicts the response variable; $\boldsymbol{X}$: $(x_1^T, \dots, x_m^T)^T$, matrix with ordo $m \times p$ which has the full rank of the accompanying variable whose elements are known; $\boldsymbol{\beta}$: regression parameter vector with ordo $p \times 1$ is fixed and its value is unknown; $\boldsymbol{Z}$: $diag(z_1, \dots, z_m)$, matrix with ordo $m \times m$ where each element is a constant positive value; $\boldsymbol{v}$: $(v_1, \dots, v_m)^T$, vector of random area effect; and $\boldsymbol{e}$: $(e_1, \dots, e_m)^T$, vector of error sample.

The model in equation (1) in fact cannot be used if there is a significant spatial correlation between adjacent areas. So that the previous random area effect needs to be changed to the following equation (2).

$$\boldsymbol{v} = \rho \boldsymbol{Wv} + \boldsymbol{u} = (\boldsymbol{I} - \boldsymbol{\rho W})^{-\boldsymbol{1}}\boldsymbol{u} \qquad (2)$$

with $\boldsymbol{v}$: $(v_1, \dots, v_m)^T$, vector of random area effect; $\boldsymbol{I}$: matrix identity with ordo $m \times m$; ρ: spatial autoregressive coefficient which has an value between -1 and 1 $(-1 < \rho < 1)$; $\boldsymbol{W}$: spatial weighted matrix with ordo $m \times m$; $\boldsymbol{u}$: vector of *error* with ordo $m \times 1$ and independent (μ = 0 dan variance = σ_u^2).

By substituting equation (2) into equation (1), we can obtain a model with an area effect that has a spatial correlation shown in equation (3).

$$y = X\beta + Z\left((I - \rho W)^{-1} u\right) + e \tag{3}$$

In estimating the parameter y_i with ρ, σ_e^2, and σ_u^2; the Spatial Best Linear Unbiased Predictor (SBLUP) model can use the following equation (4)

$$\hat{y}_i^{SBLUP} = x_i\hat{\beta} + b_i^T\{GZ^T\} \times [V]^{-1}(\tilde{\theta} - X\hat{\beta}) \tag{4}$$

Because the results of estimating SBLUP produce estimators whose values are dependent on the values of ρ and σ_u^2 2 whose values are unknown, then if the parameter estimators are replaced with $\hat{\rho}$ and $\hat{\sigma}_u^2$, then estimators for y_i can be obtained in the SEBLUP model. So that the estimation of parameter y_i with $\hat{\rho}$ and $\hat{\sigma}_u^2$ can use the following equation (5)

$$\hat{y}_i^{SEBLUP} = x_i\hat{\beta} + b_i^T\{\hat{\sigma}_u^2[(I - \hat{\rho}W)(I - \hat{\rho}W)^T]^{-1}\}Z^T \tag{5}$$

$$\times \left\{diag(\hat{\sigma}_e^2) + Z\hat{\sigma}_u^2[(I - \hat{\rho}W)(I - \hat{\rho}W)^T]^{-1}Z^T\right\}^{-1}(\tilde{\theta} - X\hat{\beta})$$

with: $\hat{\beta}$**:** matrix that obtained from $(X^TV^{-1}X)^{-1}X^TV^{-1}\hat{\theta}$); b_i^T: $(0, 0, \dots, 0, 1, 0, \dots, 0)$, vector with ordo $1 \times n$ (1 show that the area *i*); $E = \sigma^2 = diag(\sigma_e^2)$; e for error; $G = \sigma_u^2[(I - \rho W)(I - \rho W)^T]^{-1}$

2.3. Parameter Estimation for SEBLUP

From the formation of the existing model, the next stage is to estimate the parameters. Because equation (1) is a liner mixed model (LMM) equation, one parameter estimation method that can be used is Generalized Least Square (GLS). Henderson (1984) said that the working principle of GLS in estimating parameters in a linear mixed model is to minimize $e^TV^{-1}e$. The equation of $e^TV^{-1}e$ can be written in equation (6).

$$\begin{aligned} e^TV^{-1}e &= (y - X\beta)^TV^{-1}(y - X\beta) \\ &= [y^TV^{-1} - (X\beta)^TV^{-1}](y - X\beta) \\ &= y^TV^{-1}y - (X\beta)^TV^{-1}y - y^TV^{-1}X\beta + (X\beta)^TV^{-1}X\beta \\ &= y^TV^{-1}y - 2(X\beta)^TV^{-1}y + (X\beta)^TV^{-1}X\beta \end{aligned} \tag{6}$$

$\hat{\beta}$ obtained by minimizing $e^TV^{-1}e$**;** differentiating $e^TV^{-1}e$ on β and equated to 0, which then results in equation (7):

$$\frac{\partial e^TV^{-1}e}{\partial \beta} = 0 \tag{7}$$

$$\frac{\partial y^TV^{-1}y - 2(X\beta)^TV^{-1}y + (X\beta)^TV^{-1}X\beta}{\partial \beta} = 0$$

$$-X^TV^{-1}y + X^TV^{-1}X\hat{\beta} = 0$$

$$X^TV^{-1}X\hat{\beta} = X^TV^{-1}y$$

$$\hat{\beta} = (X^TV^{-1}X)X^TV^{-1}y$$

2.4. Spatial Weighted Matrix

Spatial weighted matrix is an important element in conducting spatial analysis. Selection of the appropriate spatial weighted matrix will produce a relatively small error and can estimate the desired parameters with good accuracy. According to Smith (2014), spatial weighted matrices are generally divided into three types, namely 1) spatial weighted matrices based on distance; 2) spatial weight matrix based on boundaries; and 3) spatial weighted matrix based on a combination of distance and boundaries (combined distance-boundary).

The spatial weighted matrix of power (rank) assumes that the weights are functions of the negative power (rank) of distance. The equation of the weighting matrix elements is presented in equation (8) (Smith, 2014; Jajang et al., 2017):

$$w_{ij} = d_{ij}^{-\alpha} \tag{8}$$

The power (rank) degree (α) is a positive number and d_{ij} is the Euclidean distance from location i to location j. According to Ma et al. 2018) and Behrens et al. (2018), euclidean distance is the geometric distance between two data objects. The Euclidean distance between location i and location j can be calculated using equation (9).

$$d(x_i, x_j) = \sqrt{\sum_{a=1}^{n} (x_i - x_j)^2} \tag{9}$$

where: $d\ (x_i, x_j)$: distance between location i to location j; x_i: centroid point of location i; x_j: centroid point of location j.

On the other hand, in the queen weighting matrix; a spatial unit can be said to be a neighbor of another spatial unit if the two areas share a corner or edge (Suryowati et al., 2018). So that the elements of the weighting matrix can be defined as in equation (10):

$$w_{ij} = \begin{cases} 1 & \text{if i and j intersect} \\ 0 & \text{else} \end{cases} \tag{10}$$

where: i: location i; j: location j

2.5. Assumptions in SEBLUP Modeling

To model data using the SEBLUP method it is necessary to fulfill the following assumptions:

(1) Spatial Autocorrelation

According to Fitriani and Efendi (2019), spatial autocorrelation is the correlation between the same variables in other locations. There are several types of spatial autocorrelation, namely:

a) Spatial autocorrelation which is positive if the observed values of variables at adjacent locations are similar.
b) Spatial autocorrelation which is negative if the variable observation values at adjacent locations are very different.
c) There is no spatial autocorrelation if the observed values of the variables have a random effect.

One of the tests used to detect spatial autocorrelation is the Moran-I test. The hypothesis used in the Moran-I test:

H_0: $\forall$ corr(X_i, X_j) = 0, $i \neq j$ (no spatial autocorrelation on variable X)

H_1: $\exists$ corr(X_i, X_j) = 0, $i \neq j$ (there is spatial autocorrelation on variable X)

The Moran-I test statistics can be written as follows (Fitriani and Efendi, 2019):

$$I = \frac{\sum_{j=1}^{n}\sum_{i=1}^{n} w_{ij}(X_i - \bar{X})(X_J - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}, i, j = 1,2, \dots, n \tag{11}$$

where: X_i: the variable that observed at location i; X_i: the variable that observed at location i; $\bar{X}$: mean of variable X; w_{ij}: ij^{th} element of the spatial weighted matrix **W**.

Test statistics I are then standardized and the p-value of these statistics is calculated (Anselin, 2013):

$$Z = \frac{I-E(I)}{Var(I)}, i,j = 1,2,\dots,n \quad (12)$$

where: $E(I)$: expected value of test statistic $I\left(\frac{-1}{n-1}\right)$; $Var(I)$: variance of test statistic I. According to Fitriani and Efendi (2019), $Var(I)$ has the following equation:

$$Var(I) = \frac{nS_4 - S_3S_1(1-2n)}{(n-1)(n-2)(n-3)\sum_{i=1}^{n} w_{ij}{}^2} \quad (13)$$

(2) Random Effect Normality Test

The random effects used in the SEBLUP model must meet the normality assumption. Normality testing of random effects was carried out using the Anderson-Darling test. The following are the hypotheses in the Anderson-Darling Test:

H_0 : Random effects are normally distributed

H_1 : Random effects aren't normally distributed

The Anderson Darling test statistic defined by Anderson and Darling (1954) can be written in equation (14).

$$W_n^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[\ln u_i + \{1 - \ln u_{n-i+1}\}] \quad (14)$$

where $u_i = F(X_i)$ is the cumulative distribution function for a particular distribution and n is the number of samples to be tested. If the statistical value $W_n{}^2$ is smaller than $W_n{}^2$-table or the p-value is greater than the α (5%), then it can be interpreted that the random effect has a normal distribution.

(3) Residuals Normality Test

The next assumption that needs to be met is that the resulting error is required to spread normally. To test the normality of errors, the Anderson Darling test can be used. This test focuses on measuring the distance between points and a fitted line which is based on the existing distribution (in this case the normal distribution, which has a fitted line in the form of a linear diagonal line). The following are the hypotheses in the Anderson-Darling Test:

H_0 : Residuals are normally distributed

H_1 : Residuals aren't normally distributed

The Anderson Darling test statistic are shown in equation (14), if the statistical value $W_n{}^2$ is smaller than $W_n{}^2$-table or the p-value is greater than the α (5%), then it can be interpreted that the residuals have a normal distribution.

2.6. Parameter Significance Test

In the significance test, the parameters will be tested partially for the influence of the predictor variable on the response variable using the t test. In the t-test, the test statistic will be compared with a previously known critical point or can also be done by comparing the probability value (p-value) and the α.

3. MATERIAL AND METHOD

Based on Tobler's Law, it is that things that are closer are more correlated with each other than things that are far away (Cressie, 1993). Therefore, in making estimates, geographical or spatial aspects are worth considering. In this study, spatially per capita expenditure is strongly influenced by geographical/spatial aspects. For example, someone who lives in a city is generally more consumptive than someone who lives in a village. Therefore, the expenses incurred by someone who lives in a certain city will be higher than someone who lives in the countryside.

In estimating the per capita expenditure in an area, the EBLUP (Empirical Best Linear Unbiased Prediction) method can be used. To include spatial aspects in the EBLUP method, Rao (2003) with reference to Cressie (1989) introduced spatial-based EBLUP whose model formation follows the conditional autoregressive (CAR) process. The EBLUP method with spatial aspects is then referred to as the Spatial Empirical Best Linear Unbiased Prediction (Spatial EBLUP) method. Salvati (2004) brought the Spatial EBLUP method from the CAR process to the simultaneously autoregressive (SAR) process in building the model. Mutualage (2012) compared the direct estimation method, the EBLUP method and the Spatial EBLUP method using only one accompanying variable and using a certain weighting matrix. Mutualage (2012) concluded that the MSE and RRMSE values for Spatial EBLUP were much smaller, so it was concluded that the Spatial EBLUP method could improve parameter estimates which were carried out either directly or with the EBLUP method.

In this study, the data that used is secondary data sourced from BPS in 2019. The population in this study is the regencies/cities in East Java Province, as many as 38 districts/cities. The predictor variables used are five variables and one response variable, namely:

i. Population density (X_1)
ii. The number of health facilities (X_2)
iii. The number of public elementary schools (X_3)
iv. The average of household members (X_4)
v. The percentage of the population who have BPJS as Fund Assistance Recipients (X_5).
vi. The average per capita expenditure (Y)

The steps taken in this research are as follows:

1) Collecting data from BPS related to the research variables that exist in each district/city in East Java.
2) Testing the assumption of non-spatial autocorrelation with the Moran-I test. If there is a spatial autocorrelation, it can be continued with SEBLUP modeling.
3) Calculate the centroid coordinates of each district/city in East Java and form a Euclidean distance matrix.
4) Forming the spatial weighted matrix of power (rank) based on the Euclidean distance matrix in step 3.
5) Forming the spatial weighted matrix of queen.
6) Estimating the average expenditure per capita by applying the SEBLUP model for districts/cities in East Java using the two spatial weighted matrices that were formed in the previous step.
7) Estimating the regression coefficient using GLS for each SEBLUP model.
8) Estimating the random effect (v_i) using GLS on each SEBLUP model.
9) Conducting a normality test on the random effect (v_i) using the Anderson-Darling test in each SEBLUP model.
10) Estimating the MSE value of each SEBLUP model.
11) Performing a normality test on the residuals using the Anderson-Darling test in each SEBLUP model.
12) Comparing MSE values between SEBLUP models. The SEBLUP model with the smallest MSE value is the best model.
13) Performing a significance test on each predictor variable based on the best SEBLUP model.

4. RESULTS AND DISCUSSION

In this study, the response variable will be modeled on the predictor variables using multiple linear regression using the Ordinary Least Square (OLS) method. From this model, the residual model will be obtained which will be the data representation and then tested using the Moran-I test. In testing assumptions, the residuals in the multiple linear regression model have resulted in a decision that the model has passed the assumption test. Furthermore, the residuals were tested against the existing spatial autocorrelation. By using Moran-I test, the spatial weighted matrix of power (rank) and queen yields p-values of 0.01512 and 0.01004 respectively. Therefore, it can be concluded that by using the Power (Rank) and Queen, the data have spatial autocorrelation with each other so that the SEBLUP model can be used. To find out in full/complete the spatial weighting matrix of power/rank and queen is attached in Appendix 1.

Then the SEBLUP model equation can be formed after estimation of the regression ($\hat{\beta}_i$) and random effect ($\hat{v}_i$). The equation of the SEBLUP model with the spatial weighted matrix of queen is given in the following equation.

$$\begin{aligned}\hat{y}_i^{Queen} &= \boldsymbol{x_i}\widehat{\boldsymbol{\beta}} + \hat{v}_i + e_i \\ &= x_{1i}\hat{\beta}_1 + x_{2i}\hat{\beta}_2 + x_{3i}\hat{\beta}_3 + x_{4i}\hat{\beta}_4 + x_{5i}\hat{\beta}_5 + \hat{v}_i + e_i \\ &= 3.476{,}962x_{1i} + 1.973{,}7x_{2i} + 333{,}0717\, x_{3i} + 3{,}630\, x_{4i} \\ &\quad +28{,}896x_{5i} + \hat{v}_i + e_i\end{aligned} \tag{15}$$

The equation of the SEBLUP model with the spatial weighted matrix of power (rank) is given in the following equation.

$$\begin{aligned}\hat{y}_i^{Power} &= \boldsymbol{x_i}\widehat{\boldsymbol{\beta}} + \hat{v}_i + e_i \\ &= x_{1i}\hat{\beta}_1 + x_{2i}\hat{\beta}_2 + x_{3i}\hat{\beta}_3 + x_{4i}\hat{\beta}_4 + x_{5i}\hat{\beta}_5 + \hat{v}_i + e_i \\ &= 3.391{,}923x_{1i} + 1.951{,}873\, x_{2i} + 348{,}763\, x_{3i} + 3{,}635\, x_{4i} \\ &\quad +29{,}003x_{5i} + \hat{v}_i + e_i\end{aligned} \tag{16}$$

where: $\hat{y}_i$: The estimated value of average expenditure; $\boldsymbol{x_i}$**:** Predictor variable- i; $\widehat{\boldsymbol{\beta}}$: The coefficient and intercept of the model; $\hat{v}_i$: The estimated value of random effects; e_i: The errors of the model.

The comparison of the estimated value of the Average Per Capita Expenditure (SEBLUP model with spatial weighted matrix of Power (Rank) and Queen) based on the model that obtained in the equation (15) and (16), so the results can be seen in Table 1.

Table 1. The Result of Estimated Value of Average Expenditure Per Capita (SEBLUP Model with Spatial Weighted Matrix of Power and Queen)

Kab/Kota	SEBLUP Model with Spatial Weighted Matrix of Power	SEBLUP Model with Spatial Weighted Matrix of *Queen*
Kab. Pacitan	852,665.10	868,784.60
Kab. Ponorogo	820,024.40	827,252.70
Kab. Trenggalek	802,554.10	834,922.00
Kab. Tulungagung	817,581.60	850,694.40
Kab. Blitar	916,158.10	867,123.10
Kab. Kediri	918,230.00	962,671.90
Kab. Malang	951,210.00	902,807.00
Kab. Lumajang	847,768.60	870,563.00

Kab. Jember	912,694.80	895,979.50
Kab. Banyuwangi	959,937.80	935,428.50
Kab. Bondowoso	812,849.50	799,416.10
Kab. Situbondo	922,606.80	909,952.90
Kab. Probolinggo	758,483.80	776,438.70
Kab. Pasuruan	939,056.70	971,826.70
Kab. Sidoarjo	1,346,032.00	1,358,058.00
Kab. Mojokerto	1,094,893.00	1,071,806.00
Kab. Jombang	976,137.40	977,572.70
Kab. Nganjuk	762,440.30	789,993.90
Kab. Madiun	899,080.30	863,839.70
Kab. Magetan	870,658.60	851,584.20
Kab. Ngawi	842,900.30	826,591.90
Kab. Bojonegoro	809,944.30	810,079.70
Kab. Tuban	837,935.90	834,035.70
Kab. Lamongan	917,298.50	905,731.60
Kab. Gresik	1,048,230.00	1,045,491.00
Kab. Bangkalan	806,573.40	768,922.30
Kab. Sampang	792,783.10	751,549.90
Kab. Pamekasan	882,375.40	849,959.00
Kab. Sumenep	915,152.50	896,601.60
Kota Kediri	1,180,012.00	1,282,197.00
Kota Blitar	1,318,557.00	1,319,330.00
Kota Malang	1,533,494.00	1,475,018.00
Kota Probolinggo	1,181,675.00	1,256,233.00
Kota Pasuruan	1,276,273.00	1,355,056.00
Kota Mojokerto	1,513,378.00	1,527,945.00
Kota Madiun	1,378,312.00	1,369,396.00
Kota Surabaya	1,891,631.00	1,928,338.00
Kota Batu	1,171,957.00	1,096,726.00

Note: Kab. is regency, Kota is city.

From Table 1 it can be seen that the two spatial weighted matrices used in several estimates of the average expenditure per capita of districts/cities in East Java produce values that are not much different. This can be seen from the two lines at some points almost coincide with each other.

The MSE value between the SEBLUP model with the power (rank) matrix and the SEBLUP model with the queen matrix will be compared through the average value of the MSE generated from each region, this is because it will be difficult to determine the best model between the two when viewed from one region to another. another region. Therefore, the average MSE calculated for each SEBLUP model is as follows (see Table 2).

Table 2. Average MSE of each SEBLUP Model

SEBLUP Model	Average MSE
Power	0.15570
Queen	0.14729

Based on the results in the table, it shows that the SEBLUP model formed from the Queen matrix has a better MSE value than the power (rank) matrix. This is shown by the smaller MSE value, which indicates that the error value produced by the model tends to be smaller.

The parameter significance test is used to determine whether the parameters used have a significant effect. Significance testing can be done using the t-test. The following are the hypotheses in the t-test:

$H_0 : \beta_i = 0$ (variable has no significant effect)

H_1: $\beta_i \neq 0$ (variable has a significant effect)

The results of parameter significance testing in the SEBLUP model are presented in Table 3.

Table 3. The results of parameter significance testing for each SEBLUP Model

Variable	SEBLUP Model with Spatial Weighted Matrix of Power		SEBLUP Model with Spatial Weighted Matrix of Queen	
	Coefficient $\hat{\beta}$	P-value	Coefficient $\hat{\beta}$	P-value
Population density (X1)	0.678	<0.001	0.717	<0.001
Number of health facilities (X2)	0.624	0.0124	0.651	0.0028
Number of public elementary schools (X3)	-0.611	0.0296	-0.671	0.0064
Average of household members (X4)	-0.039	0.3525	-0.059	0.2800
Percentage of the population who have BPJS as Fund Assistance Recipients (X5)	-0.140	0.0884	-0.152	0.0372

Based on the table above, it can be concluded that there are significant differences in research variables in modeling the two SEBLUP models. In the SEBLUP model with a spatial weighted matrix of power (rank). population density (X_1), the number of health facilities (X_2), and the number of public elementary schools (X_3) are significant. Meanwhile, in the SEBLUP model with the spatial weighted matrix of queen. population density (X_1), number of health facilities (X_2), number of public elementary schools (X_3), and the percentage of the population who have BPJS as Fund Assistance Recipients (X_5) are significant.

This result supports the research of Ningtyas et al. (2015) which states that the number of health facilities (public health center, village health polyclinic, medical center, special hospital, maternity hospital) has a significant relationship to per capita expenditure. In addition, this research's results also support research by Dewi (2020) which states that population size and the number of public elementary schools have a significant relationship to per capita expenditure.

The population density variable (X_1) is one of the variables that affect per capita expenditure in terms of population, the variable number of health facilities (X_2) and the percentage of the population who have BPJS as Fund Assistance Recipients (X_5) is one of the variables that affect per capita expenditure in terms of health, the number of public elementary schools (X_3) is one of the variables that affect per capita expenditure in terms of education. Meanwhile, the average of household members (X_4) in both SEBLUP models does not have a significant effect on the per capita expenditure variable because it does not represent per capita expenditure. This is probably caused by the data used in the study that is not able to explain the relationship between the average number of household members and the expenditure variable per capita.

5. CONCLUSION

This research concludes that the SEBLUP model with the spatial weighted matrix of queen has a smaller Mean Square Error (MSE) value than the SEBLUP model with the spatial weighted matrix of power (rank). It can be concluded that the SEBLUP model with the spatial weighted matrix of queen has a better model accuracy than the SEBLUP model with the spatial weighted matrix of power (rank). Thus, it can be said that the estimation of the average per capita expenditure of districts/cities in East Java Province using the SEBLUP model with the spatial weighted matrix of queen is better than the SEBLUP model with the spatial weighted matrix of power (rank).

To effectively address poverty concerns in various regions, it is crucial for regional governments to prioritize districts where the average per capita expenditure hovers around the poverty line. Monitoring demographic indicators like population density is vital, alongside non-demographic factors such as the presence of health facilities, the availability of public elementary schools, and the proportion of the population benefiting from BPJS (*Badan Penyelenggara Jaminan Sosial* or Social Security Organizing Agency) as Fund Assistance Recipients. Research has demonstrated that these factors are proven to have a significant influence on the average average per capita expenditure . This data-driven approach ensures targeted interventions to uplift economically disadvantaged regions and improve the overall economic well-being of the population.

REFERENCES

Asfar (2016). Studi Penentuan Matriks Pembobot Spasial Optimum dalam Pendugaan Area Kecil. *Dissertation*. Bogor Agricultral University (IPB).

Behrens, T., Schmidt, K., Viscarra Rossel, R. A., Gries, P., Scholten, T., & MacMillan, R. A. (2018). Spatial Modelling with Euclidean Distance Fields and Machine Learning. *European Journal of Soil Science*, *69*(5), 757-770.

BPS (2016). Indonesia - *Survei Sosial Ekonomi Nasional 2016* Maret (KOR). https://mikrodata.bps.go.id/mikrodata/index.php/catalog/769 accessed in September 1st, 2020.

Cressie, N.A. (1989). Spatial Data Analysis of Regional Counts. *Biometrical Journal*, *31*(6), 699-719.

Cressie, N.A. (1993), Statistics for Spatial Data. New York: John Wiley & Sons, Inc.

Darsyah, M.Y. (2013). Small Area Estimation terhadap Pengeluaran per Kapita di Kabupaten Sumenep dengan Pendekatan Nonparametrik. *Jurnal Statistika Universitas Muhammadiyah Semarang*, *1*(2), 28-36.

Dewi, M.L.S. (2020). Metode Prediksi Tak Bias Linier Terbaik Empiris Pada Area Kecil Untuk Pengeluaran Per Kapita Per Kecamatan Di Provinsi Bali. Universitas Brawijaya, Malang.

Fitriani, R. & Efendi, A. (2019). *Ekonometrika Spasial Terapan dengan R*. Malang: Universitas Brawijaya Press.

Getis, A. & Aldstadt, J. (2004). Constructing the Spatial Weights Matrix using a Local Statistic. *Geographical Analysis*, *36*(2), 90-104.

Jajang, J., Pratikno, B., & Mashuri, M. (2017). Kajian Matriks W-Amoeba dan W-Contiguity dalam Spatial Lag Model dengan Metode Estimasi Maximum Likelihood. *Prosiding Seminar Nasional LPPM Unsoed*, 7(1).

Ma, X., Zhang, J., Ding, C., & Wang, Y. (2018). A Geographically and Temporally Weighted Regression Model to Explore The Spatiotemporal Influence of Built Environment on Transit Ridership. *Computers, Environment and Urban Systems*, *70*, 113-124.

Molina, I., Salvati, N., & Pratesi M. (2007). Bootstrap for Estimating the MSE of the Spatial EBLUP. *Working Paper 07-34*, *Statistic and Econometric Series 08*.

Mutualage, D. (2012). *Metode Prediksi Tak Bias Linear Terbaik Empiris Spasial Pada Area Kecil Untuk Pendugaan Pengeluaran Per Kapita*. Bogor: Sekolah Pasca Sarjana, IPB.

Ningtyas, R., Rahmawati, R., Wilandari, Y. (2015). Penerapan Metode Empirical Best Linear Unbiased Prediction (EBLUP) pada Model Penduga Area Kecil dalam Pendugaan Pengeluaran Per Kapita Di Kabupaten Brebes. *Jurnal Gaussian*, *4*(4), 977-986

Nusrang, M., Annas, S., Asfar, A., Hastuty, H., & Jajang, J. (2017). Spatial EBLUP dalam Pendugaan Area Kecil. *Sainsmat*, *6*(1), 59-66.

Rao, J. N. K. (2003). *Small Area Estimation*. New Jersey (US): John Wiley and Sons.

Saei, A. & Chambers, R. (2003). Small Area Estimation: A Review of Methods Based on The Application of Mixed Models. *S3RI Methodology Working Paper*, M03/16

Salvati, N. (2004). Small Area Estimation by Spatial Models: The Spatial Empirical Best Linear Unbiased Prediction (Spatial EBLUP), Dipartimento di Statistica "Giuseppe Parenti" viale Morgagni, *Working Paper 2004/03*, University of Florence, Italy.

Smith T. E. (2014). *Areal Data Analysis (Part III)-Spatial Weights Matrices. Notebook for Spatial Data Analysis.* https://annas-archive.org/md5/c478e6cebea1d10930ad0bf4e4238cf3 Accessed 13rd June 2020

Suryowati, K., Bekti, R. D., & Faradila, A. (2018). A Comparison of Weights Matrices on Computation of Dengue Spatial Autocorrelation. *IOP Conference Series: Materials Science and Engineering*, *335*(1), p. 012052.