THE 1st INTERNATIONAL CONFERENCE ON DATA SCIENCE AND OFFICIAL STATISTICS
ICDSOS
"Strengthening the Contribution of Data Science and Official Statistics to the Society in the Distruption Era"
2021

Application of Spatial Empirical Best Linear Unbiased Prediction (SEBLUP) of Open Unemployment Rate on Sub-District Level Estimation in Banten Province

Apriliansyah, I Y Wulansari

Polytechnic of Statistics STIS, DKI Jakarta, Indonesia

*Corresponding author's e-mail: ikayuni@stis.ac.id

Abstract. The open unemployement rate is an indicator for measuring unemployment. Banten Province recorded as the highest on open unemployment rate number in Indonesia on 2018. A high open unemployment rates indicate serious problems in society. This problem must be resolved synergistically from the national level to the level of small areas such as sub-districts. However, data for the small area level has not been fulfilled due to the insufficient number of samples. We apply spatial EBLUP to estimate the open unemployment rates in the districs of Banten. Such a method of small area estimation is essential because some districts have small labor forces and direct estimation for them is not reliable. SEBLUP takes advantage of the correlation of the neighboring districts. Data that used for direct estimation is from National Labor Survey (Sakernas) and Village Potential (Podes) 2018. This research showed that SEBLUP model can increased the precision from direct estimation method or EBLUP. There are two districts that have highest category of open unemployment rate which are Curugbitung, and Koroncong.

1. Introduction

Based on the results of population projections in 2015, Indonesia has a population of around 250 million people with a population growth rate until 2015 reaching 1.38 percent per year. This number makes Indonesia the fourth most populous country in the world [1]. John Stuart Mill, a philosopher and economist in the 19th century, argued that the physical capital of a country comes from the quantity of the population in a country. It happened because if the optimal population quantity will result in high per capita production. However, another opposite theory regarding population size was proposed by Malthus. Malthus stated that the available resources, such as land, were even more limited due to the population growth. This causes an excess of labor supply while the available demand is limited. [2] explains that the difference between demand and supply is a labor surplus economy.

One of the employment indicators in measuring unemployment is the open unemployment rate. [3] explains that open unemployment is a condition of really not having a job. Unemployment is a complex problem that affects and is influenced by many factors that interact following a pattern and are not easy to understand [4]. One of the problems faced if the open unemployment rate is high is poverty. [5] has shown that there is a positive relationship between poverty and unemployment so that open unemployment indicators are always needed to inform labor and social policies that are interrelated with one another. This linkage is not only related to the interrelationships between policy indicators, but also the interrelationships between one area and another. This is due to the relationship between the open unemployment rate and the wages paid. The wages tend to have similar characteristics between regions. This is in accordance with the statement of [6], the first law of geography, that something close will have a stronger effect than something far away.

According to [7] more than 50 percent of the unemployed are in Java compared to outside Java. This fact is further exacerbated by evidence that Banten Province is always in the top five in the number of open unemployment rates in Indonesia from 2016 to 2018 and it is often in the first position. Furthermore, the population is always increasing every year because Banten Province has a growth rate of 1.94%. If population growth in this province is not accompanied by the provision of jobs, resulting in excess supply which causes unemployment to increase. Therefore, it is necessary to have a synergistic response from the national level to small levels, such as the sub-district level, to achieve the targets listed in the eighth goal of the Global Sustainability Goals (SDG's). This goal is to achieve full and productive employment for all people and reduce the proportion of young people who are unemployed. However, the availability of this data is still a problem because the sample available by the National Employment Survey (Sakernas) is only available at the provincial and district/city levels. This causes the need for additional samples if you want to use the survey results to estimate the smallest level such as sub-districts. According to [8], the relationship between sample size and statistical error is inversely proportional so that the smaller sample that used in the estimation, the greater statistical error will obtained. If the estimation results are used, or called direct estimation, the statistical error will be even greater.

Those problems can be solved by using a model-based estimation method or often referred as indirect estimation. Small Area Estimation (SAE) method is often used for this purpose. The SAE method is an estimation using information obtained from data collection that does not have errors such as census or regional registration. [9] explain that the auxiliary variables account for much of the variation in the unemployment rate. However, phenomena in a region can sometimes be influenced by other regions, so it is necessary to include regional influences into the model.[10] introduced an estimation model by incorporating regional elements represented by spatial weights. This model is called the Spatial Empirical Best Linear Unbiased Predictor (SEBLUP) estimation model. It is said that including spatial dependence of each area can increased the precision of result of estimation. In this study, we estimate the open unemployment rate at the Banten Province sub-district level in 2018 using the SEBLUP estimation model. The results of the estimates obtained will be compared using the relative values of the root mean square error (RRMSE).

2. Methods

Small Area Model with Spatial Dependence

[11] explained that SAE is an estimation method based on modeling. Two models can be used in the SAE model, namely the unit model and the area model. The fundamental difference between these two models is the supporting data used. Area level model proposed by Fay-Herriot. This a linear mixed model form which include random effect of the area. The model used can be written as:

$$\hat{\boldsymbol{\theta}} = \boldsymbol{X\beta} + \boldsymbol{Zv} + \boldsymbol{e} \tag{1}$$

where $\hat{\boldsymbol{\theta}}$ be the $m \times 1$ vector of parameter that we interested in, $\boldsymbol{e}$ be the $m \times 1$ vector of sampling errors that follow normal distribution with mean 0 and variance ψ_i, β be the $p \times 1$ vector of fixed effect parameters, $\mathbf{Z}$ be the $m \times m$ matrix of positive constant, and $\boldsymbol{v}$ be the $m \times 1$ vector of random effects that follow normal distribution with mean 0 and variance $\sigma_v^2 \boldsymbol{I}$. The Fay-Herriot model assumes independencies of the effect between areas. However, this assumption sometimes will be violated because of relation of neighborhood areas. It remains on Tobler's first law of geography which is stated that everything close will have an effect than something far away. Based on this theory, vector $\boldsymbol{v}$ can be manipulated by including value of area correlation which is ρ and proximity matrix of $\boldsymbol{W}$. Let say there is an autoregressive between $\boldsymbol{v}, \boldsymbol{\rho}, \boldsymbol{u}$, and $\boldsymbol{W}$ [12]

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

where $\boldsymbol{u}$ be the $m \times 1$ vector of independent error. Therefore, component $\boldsymbol{v}$ on equation (1) can be changed with equation (2) into

$$\widehat{\boldsymbol{\theta}} = \boldsymbol{X\beta} + \boldsymbol{Z}(\boldsymbol{I} - \boldsymbol{\rho W})^{-1}\boldsymbol{u} + \boldsymbol{e} \tag{3}$$

Vector $\boldsymbol{v}$ will be following normal distribution with mean 0 and variance **G** which is [13]

$$\begin{aligned} Var(\boldsymbol{v}) &= E\left[(\boldsymbol{v} - E(\boldsymbol{v}))(\boldsymbol{v} - E(\boldsymbol{v}))^T\right] \\ &= E((\boldsymbol{vv})^T \\ &= E\left[((\boldsymbol{I} - \rho\boldsymbol{W})^{-1}\boldsymbol{u})((\boldsymbol{I} - \rho\boldsymbol{W})^{-1}\boldsymbol{u})^{\boldsymbol{T}}\right] \\ &= E\left[((\boldsymbol{I} - \rho\boldsymbol{W})^{-1})(\boldsymbol{uu}^T)(\boldsymbol{I} - \rho\boldsymbol{W}^{\boldsymbol{T}})^{-1}\right] \\ &= \sigma_u^2[(\boldsymbol{I} - \rho\boldsymbol{W})(\boldsymbol{I} - \rho\boldsymbol{W})^T]^{-1} = \mathbf{G} \end{aligned}$$

From that model, it can be obtained the Spatial Best Linear Unbiased Prediction (SBLUP) estimator of $\widehat{\boldsymbol{\theta}}$ is

$$\widehat{\theta_\iota}^{SBLUP} = \boldsymbol{x}_{\boldsymbol{i}}^{\boldsymbol{T}}\widehat{\boldsymbol{\beta}} + \boldsymbol{b}_i^T\{\sigma_u^2[(\boldsymbol{I} - \rho\boldsymbol{W})(\boldsymbol{I} - \rho\boldsymbol{W})^T]^{-1}\}\boldsymbol{Z}^T \times \{diag(\psi_i) + \boldsymbol{Z}[\sigma_u^2[(\boldsymbol{I} - \rho\boldsymbol{W})(\boldsymbol{I} - \rho\boldsymbol{W})^T]^{-1}]\boldsymbol{Z}^T\}^{-1}(\widehat{\boldsymbol{\theta}} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \tag{5}$$

where $\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X^T V^{-1} X}\right)^{-1}\boldsymbol{X^T V^{-1}}\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{b}_i^T = (0,0, \ldots ,0,0,1, \ldots ,0)$.However, σ_u^2 and ρ is often unknown. Therefore, replacing those components with $\widehat{\sigma_u^2}$ and $\hat{\rho}$ respectively will obtain $\widehat{\boldsymbol{\theta}}_{\boldsymbol{I}}$ which is

$$\widehat{\theta_\iota}^{SEBLUP} = \boldsymbol{x}_{\boldsymbol{i}}^{\boldsymbol{T}}\widehat{\boldsymbol{\beta}} + \boldsymbol{b}_i^T\{\widehat{\sigma_u}[(\boldsymbol{I} - \hat{\rho}\boldsymbol{W})(\boldsymbol{I} - \hat{\rho}\boldsymbol{W})^T]^{-1}\}\boldsymbol{Z}^T \times \{diag(\psi_i) + \boldsymbol{Z}[\widehat{\sigma_v}[(\boldsymbol{I} - \hat{\rho}\boldsymbol{W})(\boldsymbol{I} - \hat{\rho}\boldsymbol{W})^T]^{-1}]\boldsymbol{Z}^T\}^{-1}(\widehat{\boldsymbol{\theta}} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \tag{6}$$

Estimation of Parameters

Suppose a sample $x_1, x_2, \ldots , x_n$ where $N \in n$ is drawn form a population. Maximum likelihood (ML) is one of the methods for point estimator. According to [14], let say $x_1, x_2, \ldots , x_n$ is independent random variables with one parameter θ. Then, joint probability density function for this sample is $f(x_1, x_2, \ldots , x_n \,|\theta)$. This function can be said as likelihood function and notated as L. The process of ML by derivate the log-likelihood function for maximizing the parameter. On the other hand, [15] state that ML process produce bias. This bias can be happened because have not considered loss in degrees of freedom. So that, REML is modification of ML that eliminate this bias. [16] explained that process of REML is begin where the observation of $\boldsymbol{Y}$ is changed by transformation into $\boldsymbol{Y}^*$. Transformation of **Y** is

$$\boldsymbol{Y}^* = \boldsymbol{A}^T\boldsymbol{Y} \tag{7}$$

where the matrix **A** is random orthogonal matrix sized $n \times (n - p)$ with p is rank of **X** Process of derivate the log-likelihood function is same as ML. The different between these two procedures is the likelihood function that used. Previously, explained that σ_u^2 and ρ is often unknown. Those components will be estimated by REML procedure.

Mean Square Error (MSE)

[17] explained that goodness of value of estimator can be rated from mean square error (MSE) number. Criteria that expected is the best estimator have a smallest. MSE of SBLUP can be expressed as [12]

$$MSE\left(\hat{\theta}_i^{SBLUP}(\sigma_u^2, \rho)\right) = g_{1i}(\sigma_u^2, \rho) + g_{2i}(\sigma_u^2, \rho) \tag{8}$$

where $g_{1i}(\sigma_u^2,\rho) = \boldsymbol{b}_i^T\{\sigma_u^2[(\boldsymbol{I}-\rho\boldsymbol{W})(\boldsymbol{I}-\rho\boldsymbol{W})^T]^{-1}\}\boldsymbol{Z}^T \times \{diag(\psi_i)+\boldsymbol{Z}[\sigma_u^2[(\boldsymbol{I}-\rho\boldsymbol{W})(\boldsymbol{I}-\rho\boldsymbol{W})^T]^{-1}]\boldsymbol{Z}^T\}^{-1}\boldsymbol{Z}\sigma_u^2 \times [(\boldsymbol{I}-\rho\boldsymbol{W})(\boldsymbol{I}-\rho\boldsymbol{W})^T]^{-1}\boldsymbol{b}_i$ and $g_{2i}(\sigma_u^2,\rho) = (\boldsymbol{x_i} - \boldsymbol{b}_i^T\{\sigma_u^2[(\boldsymbol{I}-\rho\boldsymbol{W})(\boldsymbol{I}-\rho\boldsymbol{W})^T]^{-1}\}\boldsymbol{Z}^T \times \{diag(\psi_i)+\boldsymbol{Z}[\sigma_u^2[(\boldsymbol{I}-\rho\boldsymbol{W})(\boldsymbol{I}-\rho\boldsymbol{W})^T]^{-1}]\boldsymbol{Z}^T\}^{-1}\boldsymbol{X}) \times (\boldsymbol{X}^T\{diag(\psi_i)+\boldsymbol{Z}[\sigma_u^2[(\boldsymbol{I}-\rho\boldsymbol{W})(\boldsymbol{I}-\rho\boldsymbol{W})^T]^{-1}]\boldsymbol{Z}^T\}^{-1}\boldsymbol{X})^{-\mathbf{1}} \times (\boldsymbol{x_i} - \boldsymbol{b}_i^T\{\sigma_u^2[(\boldsymbol{I}-\rho\boldsymbol{W})(\boldsymbol{I}-\rho\boldsymbol{W})^T]^{-1}\}\boldsymbol{Z}^T \times (\{diag(\psi_i)+\boldsymbol{Z}[\sigma_u^2[(\boldsymbol{I}-\rho\boldsymbol{W})(\boldsymbol{I}-\rho\boldsymbol{W})^T]^{-1}]\boldsymbol{Z}^T\}^{-1}\boldsymbol{X})^T$.

[18] expressed the value of general MSE is

$$MSE(\hat{\theta}) = E\left[(\hat{\theta}-\theta)^2\right] \tag{9}$$

if it applies to obtain value of $MSE(\hat{\theta}^{SEBLUP})$ will follow

$$\begin{aligned}
MSE\left(\hat{\theta}_i^{SEBLUP}(\widehat{\sigma_v^2},\hat{\rho})\right) &= E\left[\hat{\theta}_i^{SEBLUP}(\widehat{\sigma_v^2},\hat{\rho}) - \theta_i^{SBLUP}\right]^2 \\
&= E\left[\hat{\theta}_i^{SEBLUP}(\widehat{\sigma_v^2},\hat{\rho}) - \theta_i^{SBLUP} + \hat{\theta}_i^{SBLUP}(\sigma_u^2,\rho) - \hat{\theta}_i^{SBLUP}(\sigma_u^2,\rho)\right]^2 \\
&= E\left[\left(\hat{\theta}_i^{SBLUP}(\sigma_u^2,\rho) - \theta_i^{SBLUP}\right) + \left(\hat{\theta}_i^{SEBLUP}(\widehat{\sigma_v^2},\hat{\rho}) - \hat{\theta}_i^{SBLUP}(\sigma_u^2,\rho)\right)\right]^2 \\
MSE\left(\hat{\theta}_i^{SEBLUP}(\widehat{\sigma_v^2},\hat{\rho})\right) &= E\left(\hat{\theta}_i^{SBLUP}(\sigma_u^2,\rho) - \theta_i^{SBLUP}\right)^2 + E\left(\hat{\theta}_i^{SEBLUP}(\widehat{\sigma_v^2},\hat{\rho}) - \hat{\theta}_i^{SBLUP}(\sigma_u^2,\rho)\right)^2 \\
&\quad + 2cov\left[\left(\hat{\theta}_i^{SBLUP}(\sigma_u^2,\rho)\right) - \theta_i^{SBLUP}\right), \left(\hat{\theta}_i^{SEBLUP}(\widehat{\sigma_v^2},\hat{\rho}) - \hat{\theta}_i^{SBLUP}(\sigma_u^2,\rho)\right)\right]
\end{aligned} \tag{10}$$

It is known that form of $E\left(\hat{\theta}_i^{SBLUP}(\sigma_u^2,\rho) - \theta_i^{SBLUP}\right)^2$ is MSE for $\hat{\theta}_i^{SBLUP}$ estimator. Then, correlation between $\left(\hat{\theta}_i^{SBLUP}(\sigma_u^2,\rho)\right) - \theta_i^{SBLUP}\right), \left(\hat{\theta}_i^{SEBLUP}(\widehat{\sigma_v^2},\hat{\rho}) - \hat{\theta}_i^{SBLUP}(\sigma_u^2,\rho)\right)$ is zero [19]. Value of $2cov\left[\left(\hat{\theta}_i^{SBLUP}(\sigma_u^2,\rho)\right) - \theta_i^{SBLUP}\right), \left(\hat{\theta}_i^{SEBLUP}(\widehat{\sigma_v^2},\hat{\rho}) - \hat{\theta}_i^{SBLUP}(\sigma_u^2,\rho)\right)\right]$ equal to zero. [19] on [12] obtained a heuristic approximation for the last term which is $E\left(\hat{\theta}_i^{SEBLUP}(\widehat{\sigma_v^2},\hat{\rho}) - \hat{\theta}_i^{SBLUP}(\sigma_u^2,\rho)\right)^2$. This approximation called $g_{3i}\left(\widehat{\sigma_v^2},\hat{\rho}\right)$. Form of this part is

$$g_{3i}\left(\widehat{\sigma_v^2},\hat{\rho}\right) = tr\left\{\begin{bmatrix} \boldsymbol{b}_i^T\left(\boldsymbol{C}^{-1}\boldsymbol{Z}^T\boldsymbol{V}^{-1} + \sigma_u^2\boldsymbol{C}^{-1}\boldsymbol{Z}^T(-\boldsymbol{V}^{-1}\boldsymbol{Z}\boldsymbol{C}^{-1}\boldsymbol{Z}^T\boldsymbol{V}^{-1})\right) \\ \boldsymbol{b}_i^T\boldsymbol{A}\boldsymbol{Z}^T\boldsymbol{V}^{-1} + \sigma_u^2\boldsymbol{C}^{-1}\boldsymbol{Z}^T(-\boldsymbol{V}^{-1}\boldsymbol{Z}\boldsymbol{A}\boldsymbol{Z}^T\boldsymbol{V}^{-1}) \end{bmatrix}\boldsymbol{V} \times \begin{bmatrix} \boldsymbol{b}_i^T\left(\boldsymbol{C}^{-1}\boldsymbol{Z}^T\boldsymbol{V}^{-1} + \sigma_u^2\boldsymbol{C}^{-1}\boldsymbol{Z}^T(-\boldsymbol{V}^{-1}\boldsymbol{Z}\boldsymbol{C}^{-1}\boldsymbol{Z}^T\boldsymbol{V}^{-1})\right) \\ \boldsymbol{b}_i^T\boldsymbol{A}\boldsymbol{Z}^T\boldsymbol{V}^{-1} + \sigma_u^2\boldsymbol{C}^{-1}\boldsymbol{Z}^T(-\boldsymbol{V}^{-1}\boldsymbol{Z}\boldsymbol{A}\boldsymbol{Z}^T\boldsymbol{V}^{-1}) \end{bmatrix}^T \bar{V}(\widehat{\sigma_v^2},\hat{\rho})\right\} \tag{11}$$

Therefore, form of MSE($\hat{\theta}_i^{SEBLUP}$) is

$$MSE\left(\hat{\theta}_i^{SEBLUP}(\widehat{\sigma_v^2},\hat{\rho})\right) \approx g_{1i}\left(\widehat{\sigma_v^2},\hat{\rho}\right) + g_{2i}\left(\widehat{\sigma_v^2},\hat{\rho}\right) + 2g_{3i}\left(\widehat{\sigma_v^2},\hat{\rho}\right) \tag{12}$$

We compare the relative standard error given by the expression

$$RSE(\hat{\theta}) = \frac{\sqrt{MSE(\hat{\theta})}}{\hat{\theta}} \times 100\% \tag{13}$$

Better model-based estimator than direct estimator will be used. This estimator should have smaller value of RSE.

Data and Procedures

The focus of this research is the set of the 155 sub-districts of Banten Province. The data used is secondary data from the National Labor Survey (Sakernas) Banten Province in August 2018 and Village Potential Data (Podes) Banten Province in 2018. The software used are R-Studio and Microsoft Excel. The interest variable is the Open Unemployment Rate (TPT) calculated from Sakernas, while 43 candidate auxiliary variables are taken from Podes. All 43 auxiliary variables were obtained based on a literature review. Some of literature that are used as a reference for auxiliary variables such as [20],[21], and [22] The following procedure is an analytical step of this research.

1. Preparation
    1. Prepare open unemployment rate (TPT) data from raw dataset come from National Labor Survey (Sakernas) August 2018.
    2. Direct estimation of TPT that used is percentage and also estimation of sampling error variance for each sub-district. Those two data will be used for indirect estimation Spatial EBLUP.
    3. Count a value of MSE of direct estimation for each sub-district. Value of direct estimator will be unbiased so the value of MSE will equal to the variance. Then, evaluate the RSE.
2. Building the model
    1. Pearson correlation test with purpose to obtain auxiliary variables that have a significant correlation.
    2. Non-multicollinearity assumption test is done with purpose to guarantee there is no high correlation between each auxiliary variables.
    3. Build proximity matrix inverse distance weighting type.
    4. Spatial autocorrelation test of direct estimator number
    5. Fit a model of SAE Spatial EBLUP by REML method.
    6. Count a value of MSE of Spatial EBLUP, then evaluate RSE.

3. Result and Discussion

3.1 Direct Estimation

Banten Province has an average of 2018 open unemployment rate at the sub-district level of 9.575%. The meaning of this figure is that on average there are around 9 to 10 unemployed individuals out of 100 members of the labor force groups. The lowest value of the open unemployment rate resulting from a direct estimate is zero or there is one or several sub-districts that have a zero unemployment rate. Then, the maximum value generated is 71.4% or there are one or several sub-districts with unemployed individuals of 71 to 72 out of 100 individuals in the labor force group.

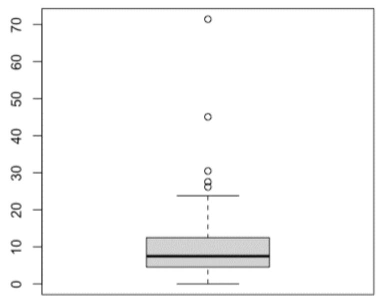

Figure 1. Direct Estimation of TPT

However, the interpretation of the lowest (minimum) value for this variable is incorrect because there are eight sub-districts that have an estimated TPT value equal to zero, also have a variance equal to zero.

Furthermore, after processing the estimation results, there are also thirty sub-districts that have a variance equal to zero even though the estimated TPT value is not equal to zero. This is due to the lack of samples used in the direct estimation process. So the maximum and minimum values generated are invalid. Therefore, 38 sub-districts were not included in the indirect estimation analysis but were estimated using synthetic estimates.

3.2 Spatial EBLUP

Next process is doing estimation by SAE Spatial EBLUP. Before building model, it is important to obtain a strong correlation between auxiliary variables and direct estimator. From forty-three candidate of auxiliary variables there are ten variables which is significant correlated or have a probability value less than five percent. These variables are listed in Table 1.

Table 1. Auxiliary variables identity

Name of Variables	Notation	Source
Ratio of electricity user family per 1000 citizens	x_1	
Ratio of total elementary school per 100 citizens	x_2	
Total of individual business	x_{11}	
Total of Base Transceiver Station tower	x_{16}	
Ratio of total small business of wood production per 100 citizens.	x_{19}	
Ratio of total small business of food and beverages production per 100 citizens	x_{24}	*Podes 2018*
Ratio of technology developing villages	x_{28}	
Ratio of sanitation developing villages	x_{29}	
Ratio of education, health and culture developing villages	x_{34}	
Ratio of transportation empowerment village	x_{35}	

It is also important to proof that observation data is having spatial autocorrelation. By conducting Moran's I spatial autocorrelation, it obtain result of probability value 0.021 which is having conclusion that observation data is having spatial autocorrelation. This result is required for developing spatial EBLUP model. On the first stage of model development, all variables enter model. It is known that there is one variable that significant in significance level of five percent which is x_{24}. Next procedure is to reduce variables by dropping insignificant variable or significant value near one. By looking Akaike Information Criteria (AIC) number, it will obtain most parsimony model. The following table is summary of AIC score and dropped variables on each process.

Table 2. Process of choosing best Spatial EBLUP model

Stage	Variables	AIC
1	$x_1, x_2, x_{11}, x_{16}, x_{19}, x_{24}, x_{28}, x_{29}, x_{34}, x_{35}$	670.0626
2	$x_2, x_{19}, x_{24}, x_{28}, x_{29}, x_{35}$	663.9182
3	$x_{19}, x_{24}, x_{28}, x_{35}$	659.4046
4	x_{19}, x_{24}, x_{28}	658.1535
5	x_{24}, x_{28}	658.9199

From the result, minimum AIC score will be obtained in the fourth iteration. This model will be used as indirect estimator of Spatial EBLUP. The following table is summary of regression coefficient, random effect coefficient, and autoregressive coefficient on minimum AIC score Spatial EBLUP model.

Table 3. Summary of regression coefficient, random effect coefficient, and autoregressive coefficient on minimum AIC score Spatial EBLUP model

Variable	$\widehat{\beta}_{Spatial\ EBLUP}$	p-value
Intercept	9.3293	0.0000
x_{19}	-4.0391	0.0830
x_{24}	0.6105	0.0218
x_{28}	-3.9549	0.0375
$\widehat{\sigma_v^2}$	18.4330	-
$\widehat{\rho}$	0.7878	-

This table shows that, it is know that x_{24} and x_{28} are significant on significance level of five percent whereas x_{19} is significant on significance level of ten percent. Then, value of autoregressive coefficient is 0.7878 which is showed that there is a strong area correlation on sub-district open unemployment rate. Descriptive statistics on open unemployment rate based on Spatial EBLUP is presented in Table 4.

Table 4. Summary of estimation result of Banten Province sub-district open unemployment rate based on Spatial EBLUP model

Statistics	TPT of Spatial EBLUP
Mean	8.5720
Minimum	2.2920
Median	8.0610
Maximum	25.6190
Total of Observation	106 Sub-district

3.3 RSE Comparison

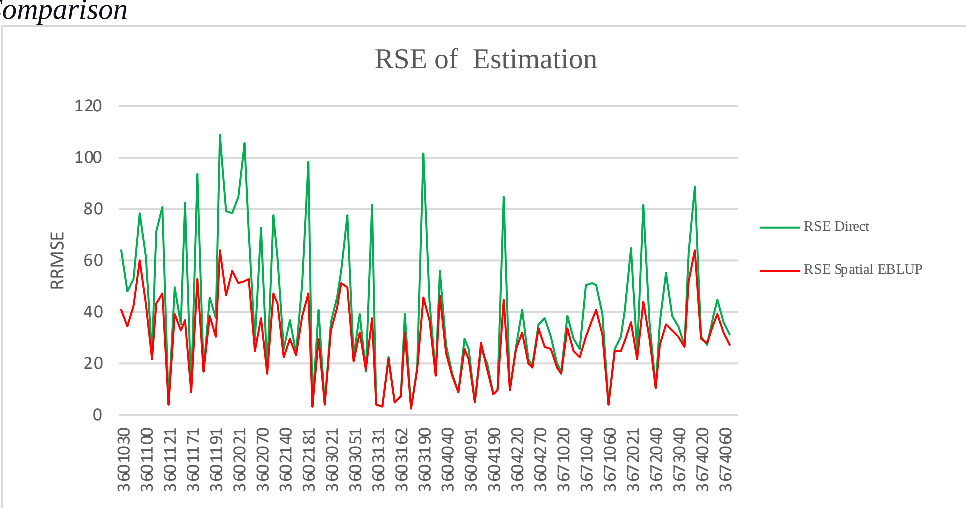

Figure 2. RSE Comparison Line Graph

RSE of direct estimator, shown in green, has mean 40.2040, minimum 2.0530, and maximum 108.4330. Then, RSE of spatial EBLUP, shown in red, is having mean 29.2770, minimum score 2.0540, and maximum score 29.1290. By fact, RSE score of spatial EBLUP is lower than direct estimator. So that, from Figure 2 and previous explanation it can be concluded that spatial EBLUP is more precise than direct estimator. So, spatial EBLUP model will be used for next analysis which is estimate number of Banten Province sub-district level open unemployment rate.

3.4 Implementation of Whole Area Estimation by Selected Model

After selecting better model, next step is doing estimation for whole area estimation by selected model which is spatial EBLUP. This whole area, 155 sub-districts, includes forty-nine area that not included on analysis. Each sub-district has a mean of 8.7367% or eight until nine people is being unemployed. It has a maximum value 25.6187 which is Koroncong sub-district. Then, it has a minimum value 2.2925% which is Kasemen sub-district. There are two sub-district which has a highest category, shown in purple, Koroncong and Curugbitung sub-district. Figure 3 presents a thematic map of result of spatial EBLUP estimation for all sub-district

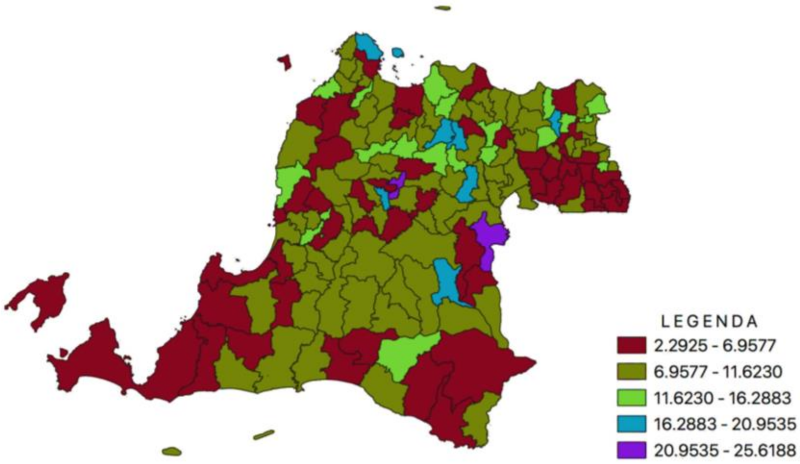

Figure 3. Result of indirect estimation of open unemployment rate for sub-district in Banten Province 2018

Based from Figure 3, it is known that there are fifty-two sub-district on first category which is shown in brown. On this first category, there are twenty-eight sub-district which is on under the value of national open unemployment rate as big as 5.34%. Therefore, the rest of total from this category is already on above the value of national open unemployment rate. Then, on the second category, which is shown in dark-green, there are seventy-six sub-district. Next, on the third category, shown in light-green, counted as eighteen sub-district. On the next category, shown in blue, there are seven sub-district. There are two sub-district on last category, shown in purple, which is the highest category. Overall, from 155 sub-district, there are 127 sub-district which is on above the value of national open unemployment rate.

4. Conclusion

This study produces an estimate of the open unemployment rate at the sub-district level in Banten Province in 2018 using the indirect approach (SEBLUP). We have shown that SEBLUP yields estimates with lower RSE than other alternatives, including direct estimation. Therefore, Spatial EBLUP is a better method for doing estimation for sub-district level of open unemployment rate in Banten Province. The result of the analysis can be used by the provincial government of Banten to set the priorities for reducing highes TPT, such as Koroncong and Curugbitung. The phenomenon of high and low TPT in each sub-district also shows that there is a spatial correlation between sub-districts. The Banten Provincial Government can focus on the sub-districts with the highest TPT which will ultimately have a positive impact on decreasing TPT in other nearby sub-districts.

References

[1] Population ranking. (2011, July 1). World Bank Data Catalog. https://datacatalog.worldbank.org/dataset/population-ranking

[2] Fbi, J. C. H., & Ranis, G. (1964). Development of the labor surplus economy: theory and policy. Development of the labor surplus economy: theory and policy.

[3] Sukirno, S. (2010). Teori Pengantar Makro Ekonomi Edisi Ketiga. Raja Grafindo. Jakarta.

[4] Muslim, M. R. (2014). Pengangguran terbuka dan determinannya. Jurnal Ekonomi & Studi Pembangunan, 15(2), 171-181.
[5] Seran, S. (2017). Hubungan Antara Pendidikan, Pengangguran, Dan Pertumbuhan Ekonomi Dengan Kemiskinan. Jurnal Ekonomi Kuantitatif Terapan, 10(1), 59-71.
[6] Tobler, W. R. (1976). The geometry of mental maps. Spatial choice and spatial behavior, 69-81.
[7] Andriati, R. D. (2010). Pengangguran dan Kemiskinan. Leuser Cita Pustaka.
[8] Kerlinger, F. N. (1973). Foundations of Behavioral Research. Holt. Rinehart and Winston Inc.
[9] Notodipuro, K. A., Kurnia, A., & Sadik, K. (2008). Statistical Models for Small Area Estimation. Proceedings The 3rd International Conference on Mathematics and Statistics (ICoMS-3), 677–684.
[10] Cressie, N. (1993). Statistics for spatial data. John Wiley & Sons.
[11] Dedianto, D. & Wulansari, I. Y. (2018). Aplikasi Small Area Estimation (SAE) Metode Pseudo-EBLUP Pada Official Statistics di Indonesia Studi Kasus: Estimasi Pengeluaran Rumah Tangga di Provinsi Jawa Timur Tahun 2016. Jurnal Aplikasi Statistika dan Komputasi Statistik, 10(2), 33–38.
[12] Pratesi, M., & Salvati, N. (2008). Small area estimation: the EBLUP estimator based on spatially correlated random area effects. Statistical methods and applications, 17(1), 113-141.
[13] Setyawan, A. (2016). Small Area Estimation Metode Spatial Empirical Best Linear Unbiased Prediction Untuk Estimasi Persentase Wanita Usia Subur Dengan Fertilitas Tinggi Di Kabupaten Mamuju Dan Mamuju Tengah [Thesis]. Surabaya: Institut Teknologi Sepuluh Nopember.
[14] Hogg, R. V., McKean, J., & Craig, A. T. (1995). Introduction to mathematical statistics. Pearson Education.
[15] Harville, D. A. (1974). Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American statistical association, 72(358), 320-338.
[16] Callanan, T. P. (1985). Restricted maximum likelihood estimation of variance components: computational aspects. (Doctoral dissertation). Iowa State University: Iowa.
[17] Rao JNK. (2003). Small area estimation. London: Wiley.
[18] Bain, L. J. Engelhardt. M.(1987). Introduction to Probability and Mathematical Statistics. California : Duxbury Press
[19] Kackar, R. N., & Harville, D. A. (1984). Approximations for Standard Errors of Estimators of Fixed and Random Effect in Mixed Linear Models. Journal of the American Statistical Association, 79(388), 853–862. https://doi.org/10.2307/2288715
[20] Afandi, W. N. (2010). Identifikasi Karakteristik Rumah Tangga Miskin di Kabupaten Padang Pariaman (Studi Kasus Nagari Malai V Suku). Jurnal Kemiskinan.
[21] Ikbal, M., Mustafa, S. W., & Bustami, L. (2018). Peran Usaha Mikro, Kecil Dan Menengah Dalam Mengurangi Pengangguran Di Kota Palopo. Jurnal Ekonomi Pembangunan STIE Muhammadiyah Palopo, 4(1).
[22] Arifin, S., & Firmansyah, F. (2017). Pengaruh Tingkat Pendidikan Dan Kesempatan Kerja Terhadap Pengangguran Di Provinsi Banten. Jurnal Ekonomi-Qu, 7(2).