Application of the spatial empirical best linear unbiased prediction method for estimating per capita expenditure in Lampung province

View the article online for updates and enhancements.

Application of the spatial empirical best linear unbiased prediction method for estimating per capita expenditure in Lampung province

Widiarti[*], DD Oktafiani, M Usman and D Kurniasari

Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Lampung, Indonesia

*Corresponding author: widiarti08@gmail.com

Abstract. Spatial Empirical Best Linear Unbiased Prediction (SEBLUP) is one of the methods in small area estimation. This method is the development of EBLUP by observing the influence of random spatial correlated areas. One application of this method is the estimation of per capita expenditure in each district/city in Lampung Province. This estimation is based on the existence of additional information on the number of births of the population from each district/city. In this study, the SEBLUP method was applied to the two-level model Normal-Normal and weighting matrix determination based on the type of queen contiguity. Based on the monthly household expenditure data in the Province of Lampung in 2017, Bandarlampung and Metro are areas with high per capita income levels.

1. Introduction

BPS (Central Bureau of Statistics, Indonesia) has a role in providing data needed for development planning both sectoral and cross-sectoral. The National Socio-Economic Survey which is routinely carried out by BPS aims to provide data on public welfare in terms of education, health and people's purchasing power. The purchasing power of this community can reflect how much expenditure from a household. Unfortunately, this information has not been able to directly measure per capita expenditure because it must be seen in advance how many household members there are. The existence of limited sample information regarding the number of household members results in a direct estimate to suspect that this expenditure cannot be used.

Direct estimation on a small area in the sample will produce an unbiased estimator but has a large variety. Small areas are defined as sub-populations, where samples taken in these areas are not sufficient for direct estimation with accurate results. Therefore we need a method that can be used to get better results, namely small area estimation (SAE) [1]. Estimation of small areas has been widely applied in various fields. In the field of health research about SAE including published by Das et al., Islam et al., and Li et al. [2-4]. In the economic field, Jędrzejczak and Kubacki have applied SAE to estimate household expenditure, and Anjoy et al [5]. Applied SAE to map poverty in Odisha, India [6]. In fact, this method has also been applied to other statistical methods such as those conducted by Fabrizi et al, Frumento and Salvati, Innocent et al and Sam et al. [7-13].

Estimation of small areas is an effort to minimize the variation in small areas by utilizing information from the surrounding area related to the observed parameters. Various indirect estimation methods have been developed to obtain estimators for small areas. The methods that are often used are Empirical Best Linear Unbiased Prediction (EBLUP), Empirical Bayes (EB), and Hierarchical Bayes (HB). The EBLUP method is a technique of solving mixed effect models that minimize the Mean

Squared Error (MSE) generated with the assumption of unknown variant components. Yu et al. introduced a general estimation (GE) of BLUP to overcome the problem of negative or zero variance components, and compared the performance of the ML and REML methods for BLUP [14]. Molina et.al conduct preliminary testing procedures to determine the MSE of the small area estimator [15]. Simulation results show that for a small number of areas, the results of this test give better results especially when the variance of the area effect is small. Widiarti et al. used bootstrap method in estimation mean squared error of EB estimator and in other research, Widiarti et al. used the HB method to estimate the proportion of underprivileged families in 2015 in Bandarlampung. The results of this study indicate that the HB method provides a more accurate estimate than direct estimation [16-17].

The method contained in the SAE discussion is an indirect estimator, which is obtained by weighing based on the value of a random variable from a particular domain to increase the effectiveness of sample size and reduce variance. The variance of an area can be influenced by the surrounding area so that spatial effects can be placed into random effects. Spatial effects are things that occur between one area and another, which means that one area affects another area. Spatial Empirical Best Linear Unbiased Prediction (SEBLUP) is known as the EBLUP estimator, which also takes into account the random effect of spatially correlated areas. Some research on SEBLUP was conducted by Pusponegoro and Rachmawati and Buil-Gil et al. [18-19].

In this study, the SEBLUP method is applied to estimate monthly per capita expenditure in Lampung Province by taking district-level domain information. The quality of the estimator will be evaluated based on the MSE value. The SEBLUP parameter estimation method used is the Restricted Maximum Likelihood Estimation (REML) method.

2. Method

2.1. Data Sources and Research Variables

The data used in this study are household expenditure data in each district/city in Lampung Province in 2017 obtained from BPS publication in Lampung Province in 2018. The auxiliary variable used in this study is the number of births of residents in Lampung Province in each district/city in 2017.

2.2. Small Area Based Spatial EBLUP Model

Area-based model is a model based on the availability of supporting data that only exists for a certain area level, $\boldsymbol{x_i}^T = (x_{1i}, x_{2i}, \ldots, x_{pi})^T$ with parameters to be estimated are θ_i and assumed to have an association with $\boldsymbol{x_i^T}$ following the model:

$$\theta_i = \boldsymbol{x_i^T\beta} + b_i v_i \quad , i = 1, 2, \ldots, m \tag{1}$$

Where $v_i \sim N(0, A)$ as a random effect that is assumed to be normally distributed, b_i is a known positive constant, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is a regression coefficient vector measuring $p \, x \, 1$. Conclusion about θ_i can be known directly if y_i is available, that is:

$$y_i = \theta_i + e_i, \quad i = 1, 2, \ldots, m \tag{2}$$

Where $e_i \sim N(0, \, D_i)$ and D_i known. From equation (1) and (2) the combined model is obtained:

$$y_i = \boldsymbol{x_i^T\beta} + b_i v_i + e_i, \quad i = 1, 2, \ldots, m \tag{3}$$

with assumption v_i and e_i independent [1].

In this study the two level model used to estimate per capita expenditure in small areas is as follows:

i. $y_i|\theta_i \sim N \, (\theta_i, D_i)$
ii. $\theta_i \sim N(x_i^T\beta, A) \qquad , i = 1, 2, \ldots, m$

where:
y_i = direct estimator
θ_i = average monthly per capita expenditure in area i
D_i = variance in the area
$\boldsymbol{x_i^T}$ = auxiliary variable

$\boldsymbol{\beta}$ = regression coefficient

A = variance of random effect

The SAE model includes spatial correlations between areas that were first introduced by Cressie by assuming spatial dependencies following the Conditional Autoregressive (CAR) process [20]. In the CAR model, the random effect vector $v = v_i$ satisfy:

$$y = X\boldsymbol{\beta} + v$$
$$v \sim N(0, \sigma_u^2(I - \rho W)^{-1})$$
$$v = \rho W v + u \tag{4}$$

Coefficient ρ in equation (4) is a spatial autoregressive coefficient which shows the strength of the spatial relationship between random effect, W denoted as a spatial weighting matrix that describes the neighborhood structure of a small area in the form of row standardization (the sum of each row in the matrix W is 1). In this study, the weighting matrix used is the type of queen contiguity, v is random effect in the area, and u is an error vector of an area random variable with a mean equal to zero and variance σ_u^2. In the CAR model, the covariance matrix is $\sigma_u^2(I - \rho W)^{-1}$. In the CAR model, W is a symmetric matrix, and $(I - \rho W)$ is a positive definite matrix. Let $Var(\theta_i) = A = \sigma_u^2$ where σ_u^2 is a variance from random effect.

2.3. Parameter Estimation

The normality assumption of random effect is used to estimate parameters σ_u^2 and ρ by using the REML procedure with a log-likelihood function

$$l(\boldsymbol{\beta}, \sigma_u^2, \rho) = -\frac{1}{2} n \log 2\pi - \frac{1}{2}\log|V| - \frac{1}{2}(y - x_i^T\boldsymbol{\beta})^T V^{-1}(y - x_i^T\boldsymbol{\beta})$$

where $V = diag(D_i) + \sigma_u^2(I - \rho W)^{-1}$. Partial derivative of $l(\sigma_u^2, \rho)$ by using the REML method is:

$$s\sigma_u^2(\sigma_u^2, \rho) = \frac{\partial l}{\partial \sigma_u^2} = -\frac{1}{2}tr\{PZR^{-1}Z^T\} + \frac{1}{2}y^T PZR^{-1}Z^T Py$$

and

$$s\rho(\sigma_u^2, \rho) = \frac{\partial l}{\partial \rho} = -\frac{1}{2}tr\{PZ\sigma_u^2[R^{-1}WR^{-1}]Z^T\} + \frac{1}{2}y^T PZ\sigma_u^2[R^{-1}WR^{-1}]Z^T Py$$

where $R = (I - \rho W)$ dan $P = V^{-1} - V^{-1} X (X^T V^{-1} X)^{-1} X^T V^{-1}$. The matrix containing the second derivative of the log-likelihood function is

$$\mathfrak{I}(\sigma_u^2, \rho) = \begin{bmatrix} \frac{1}{2}tr\{PZR^{-1}Z^T PZR^{-1}Z^T\} & \frac{1}{2}tr\{PZR^{-1}Z^T PZ\sigma_u^2 R^{-1}WR^{-1}Z^T\} \\ \frac{1}{2}tr\{PZ\sigma_u^2 R^{-1}WR^{-1}Z^T PZR^{-1}Z^T\} & \frac{1}{2}tr\{PZ\sigma_u^2 R^{-1}WR^{-1}Z^T PZ\sigma_u^2 R^{-1}WR^{-1}Z^T\} \end{bmatrix}$$

REML estimator $\hat{\sigma}_{u_{REML}}^2$ and $\hat{\rho}_{REML}$ obtained iteratively using the scoring algorithm:

$$\begin{bmatrix} \sigma_u^2 \\ \rho \end{bmatrix}^{(a+1)} = \begin{bmatrix} \sigma_u^2 \\ \rho \end{bmatrix}^{(a)} + \left[\mathfrak{I}\left(\sigma_u^{2(a)}, \rho^a\right)\right]^{-1} s\left(\sigma_u^{2(a)}, \rho^a\right)$$

So the SEBLUP estimator is obtained as follows:

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

where $\hat{\boldsymbol{\beta}} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y$.

2.4. MSE of EBLUP Estimator

Saei and Chambers [21] states that MSE of EBLUP estimator (σ_u^2, ρ) is:

$$MSE\left(\hat{\theta}_i^{EBLUP}\right) = g1i(\sigma_u^2, \rho) + g2i(\sigma_u^2, \rho)$$

where:

$$g1i(\sigma_u^2, \rho) = b_i^T\{\sigma_u^2(I - \rho W)^{-1} - \sigma_u^2(I - \rho W)^{-1}Z^T \times \{diag(D_i) + Z\sigma_u^2(I - \rho W)^{-1}Z^{-1}\}^{-1}$$
$$Z\sigma_u^2(I - \rho W)^{-1}\}b_i$$

and

$$g2i(\sigma_u^2, \rho) = (x_i - b_i^T \sigma_u^2 (I - \rho W)^{-1} Z^T \{diag(D_i) + Z\sigma_u^2(I - \rho W)^{-1}Z^{-1}\}^{-1} \times (X^T\{diag(D_i)$$
$$+Z\sigma_u^2(I - \rho W)^{-1}Z^{-1}\}^{-1} X)^{-1} \times (x_i - b_i^T \sigma_u^2 (I - \rho W)^{-1}$$
$$Z^T\{diag(D_i) + Z\sigma_u^2(I - \rho W)^{-1}Z^{-1}\}^{-1}X)^T$$

The MSE was obtained by substituting $\hat{\sigma}_u^2$ and $\hat{\rho}$, that is:

$$\text{MSE}\left(\hat{\theta}_i^{SEBLUP}\right) = \text{MSE}\left(\hat{\theta}_i^{EBLUP}\right) + \text{E}\left[\left(\hat{\theta}_i^{SEBLUP} - \hat{\theta}_i^{SBLUP}\right)\right]^2$$
$$= g1i(\hat{\sigma}_u^2, \hat{\rho}) + g2i(\hat{\sigma}_u^2, \hat{\rho}) + g3i(\hat{\sigma}_u^2, \hat{\rho})$$

where $g3i(\hat{\sigma}_u^2, \hat{\rho})$ obtained by approach

$$tr\left\{\begin{bmatrix} b_i^T \left(R^{-1}Z^TV^{-1} + \sigma_u^2 R^{-1}Z^T\left(-V^{-1}ZR^{-1}Z^TV^{-1}\right)\right) \\ b_i^T \left(\sigma_u^2 R^{-1}WR^{-1}Z^TV^{-1} + \sigma_u^2 R^{-1}Z^T\left(-V^{-1}Z\sigma_u^2 R^{-1}WR^{-1}Z^TV^{-1}\right)\right) \end{bmatrix} V \right.$$
$$\times \left. \begin{bmatrix} b_i^T \left(R^{-1}Z^TV^{-1} + \sigma_u^2 R^{-1}Z^T\left(-V^{-1}ZR^{-1}Z^TV^{-1}\right)\right) \\ b_i^T \left(\sigma_u^2 R^{-1}WR^{-1}Z^TV^{-1} + \sigma_u^2 R^{-1}Z^T\left(-V^{-1}Z\sigma_u^2 R^{-1}WR^{-1}Z^TV^{-1}\right)\right) \end{bmatrix}^T \bar{V}(\hat{\sigma}_u^2, \hat{\rho}) \right\}$$

3. Result and Discussion

Expenditures for consumption needs can reflect the level of the economic capacity and the purchasing power of the community, which gives an idea of the community welfare level. The higher the purchasing power of the community shows how increasing the ability to meet their needs and will have an impact on increasing the welfare of the community. Changes in a person's income will affect the shift in spending patterns. The higher the income figure, the higher the expenditure figure will be [22]. The estimated value of district/city per capita expenditure in Lampung Province in 2017 and MSE using the SEBLUP method presented in Table 1.

With the SEBLUP method, the highest expenditure per capita is in Bandarlampung City of Rp.336,620.30, and the lowest expenditure per capita is in the Pringsewu of Rp176,278.20. The MSE value generated by this method is also small, so the resulting estimator is accurate in estimating the actual value.

Table 1. Estimated Value and MSE of District/City Expenditures Per Capita in Lampung Province in 2017 (multiplied by Rp.10,000)

District/City	$\left(\hat{\theta}_i^{SEBLUP}\right)$	$\left(MSE^{SEBLUP}\right)$
Lampung Barat	23.51517	0.9670401
Tanggamus	18.00390	0.9692262
Lampung Selatan	19.70642	0.9489172
Lampung Timur	18.99756	0.9565741
Lampung Tengah	19.81492	0.9626178
Lampung Utara	18.81710	0.9525283
Way Kanan	18.85890	0.9582983
Tulang Bawang	21.00187	0.9542108
Pesawaran	17.92321	0.9566865
Pringsewu	17.62782	0.9487739
Mesuji	20.58719	0.9493190
Tulang Bawang Barat	17.81861	0.9619209
Pesisir Barat	18.84199	0.9559596
Bandar Lampung	33.53578	1.0036238
Metro	32.70883	0.9771448

Figure 1. Map of Lampung Province

Figure 1 shows the distribution of monthly per capita expenditure in 15 districts/cities in Lampung Province using the GIS archview application. In this study, population welfare was divided into three categories, that is high (yellow area), medium (green area), and low (blue area). The low category that is with per capita expenditure which is at an interval of 176278.2 - 198149.2 found in ten districts. The medium category, which is in the interval 198149.2 - 235151.7 in three districts, and region is an area with a high level of population welfare in per capita expenditure, which is at an interval of 235151.7 - 335357.8 that is Bandarlampung and Metro.

4. Conclusion

The results of this study indicate that the SEBLUP estimator gives an accurate estimator with a reasonably small error value. Based on district/city per capita expenditure data in Lampung Province in 2017, areas with per capita expenditure are in the high category, which is Bandarlampung City and Metro City.

Acknowledgments

References

[1]   Rao J N K and Molina I 2015 *Small Area Estimation second edition* (John Wiley and Sons, Hoboken New Jersey)
[2]   Das S, Chandra H and Saha U 2019 *PLoS ONE* **14**(2) e0211062
[3]   Islam A, Das S, Kumar B and Kawsar L 2018 *Proc.5th Int, Conf, Nat. Sci. Technol.* (Chittagong, Bangladesh)
[4]   Li Z, Hsiao Y, Godwin J, Martin B, Wakefield J and Clark S 2019 *PLoS ONE* 14(1):e0210645
[5]   Jędrzejczak A and Kubacki J 2017 *Stat. Transit. new ser.* **18**(4) 725
[6]   Anjoy P, Chandra H and Basak P 2019 *Soc. Indic. Res.* **144**(4)
[7]   Fabrizi E, Salvati N, Pratesi M and Tzavidis N 2014 *Biom. J.* **56** (1)
[8]   Fabrizi E, Salvati N and Tzavidis N 2018 *Int. Stat. Rev.*
[9]   Fabrizi E, Salvati N and Trivisano C 2020 *Comput. Stat. Data Anal.* **45** 106900
[10] Frumento P and Salvati N 2019 *J. R. Stat. Soc. (Stat. Soc.)* DOI: 10.1111/rssa/12495
[11] Innocent N, Nzabanita J, van Rosen D and Singull M 2016 *Commun. Stat.-Theory Methods*

46(21)

[12] Innocent N, van Rosen D and Singull M 2018 *Commun. Stat.-Theory Methods* **48**(8)

[13] Sam W, Peijin X, Ching R Y and Kelly H Z 2018 *Biostat. Biom. Open Access J.* **4**(4) 555641

[14] Yu C R, Zou K, Carlsson M and Weerahandi S 2015 *Commun. Stat.: Simul. Comput.* **44**(3)

[15] Molina I, Rao J N K and Datta G S 2015 *Surv. methodol.* **41**(1) 19

[16] Widiarti, Adityawati N and Nusyirwan 2019 *J. Phys.: Conf, Ser.* 1338 012039.

[17] Widiarti, Malinda S T, Suharsono S, Warsono and Usman M 2018 *Sci.Int.*(*Lahore*) **30**(4) 523

[18] Pusponegoro N H and Rachmawati N R 2018 *Procedia Comput. Sci.* **135** 712

[19] Buil-Gil D, Moretti A, Shlomo N and Medina J 2020 *Appl. Spat. Anal. Policy* https://doi.org/10.1007/s12061-020-09333-8

[20] Cressie N 1991 *Proc. Stat. Symp. 90: Meas. Improv. Data Qual.* (Ottawa Canada: Statistics Canada) p 93

[21] Saei A and Chambers R 2003 *Small Area Estimation: A Review of Methods based on the application of mixed models* (Southampton Statistical Sciences Research Institue WP MO3/16 Southampton)

[22] Badan Pusat Statistika 2018 *Provinsi Lampung Dalam Angka 2018* (BPS Provinsi Lampung: Lampung)