InPrime: Indonesian Journal of Pure and Applied Mathematics
Vol. 6, No. 1 (2024), pp. 14 – 26, doi: 10.15408/inprime.v6i1.33749
p-ISSN 2686-5335, e-ISSN: 2716-2478

Small Area Estimation for Per Capita Expenditure in Sulawesi Selatan Using Empirical Best Linear Unbiased Prediction

Alya Safira Irtiqa Miolo, Nurtiti Sunusi*, Siswanto
Department of Statistics, Hasanuddin University, Makassar 90245, Indonesia
*Email: nurtitisunusi@unhas.ac.id

Abstract

Small area estimation (SAE) is an important technique for estimating parameters in regions or sub-populations with limited sample sizes, particularly when direct estimators are inadequate in capturing area-specific information. The Empirical Best Linear Unbiased Prediction (EBLUP) method is one of the SAE parameter estimation approaches, aiming to minimize Mean Square Error (MSE) by incorporating unknown variations of components. In this research, we derive an SAE model parameter estimator and compare its outcomes with both the direct estimator and EBLUP-SAE. The dataset used in this study consists of per capita expenditure data obtained from the March 2019 National Socioeconomic Survey (*Susenas*) conducted in South Sulawesi, providing a benchmark for assessing household purchasing power. The estimation of SAE parameters was performed using the maximum likelihood method. The results using the EBLUP method reveals that Makassar City recording the highest per capita expenditure at Rp.1,206,352.79 and Jeneponto Regency with the lowest at Rp.1,000,887.29, reflecting significant disparities. Furthermore, the estimated variance of random influence was determined to be 0.010. The study's findings indicate that the EBLUP method outperforms the direct estimation method in estimating per capita expenditure. This is evidenced by the significantly lower MSE value of the EBLUP method, averaging 0.001, compared to the direct estimator's average MSE value of 0.002. The finding not only emphasizes the reliability of the EBLUP method but also enhances the robustness of socioeconomic analyses and contributes to the advancement of small area estimation techniques. This provides a novelty in understanding regional disparities and informing policy decisions.

Keywords: small area estimation, direct estimation, EBLUP, per capita expenditure.

Abstrak

Small area estimation (SAE) *merupakan metode yang digunakan untuk menduga parameter yang berasal dari area atau sub populasi dengan ukuran sampel yang kecil, ketika estimasi menggunakan penduga langsung tidak mampu menyampaikan informasi area terkait.* Metode Empirical Best Linear Unbiased Prediction (EBLUP) *merupakan salah satu metode estimasi parameter* SAE *yang meminimumkan* Mean Square Error (MSE) *yang dihasilkan dengan asumsi komponen ragam yang tidak diketahui. Penelitian ini bertujuan untuk memperoleh estimator parameter model* SAE *dan memperoleh perbandingan hasil penduga langsung dan* EBLUP-SAE. *Data yang digunakan dalam penelitian ini yaitu data pengeluaran per kapita berdasarkan hasil Survei Sosial Ekonomi Nasional (Susenas) Maret 2019 di Sulawesi Selatan, yang berfungsi sebagai tolak ukur untuk menilai kekuatan beli rumah tangga. Estimasi parameter* SAE *dilakukan menggunakan metode* maximum likelihood. *Berdasarkan metode* EBLUP, *diperoleh bahwa nilai pengeluaran per kapita terbesar terjadi di Kota Makassar, yaitu sebesar Rp.1,206,352.79, sedangkan nilai pengeluaran per kapita terkecil terjadi di Kabupaten Jeneponto, yaitu sebesar Rp.1,000,887.29, mencerminkan disparitas yang signifikan. Sementara itu, diperoleh nilai estimasi varians dari pengaruh acak sebesar 0.010. Hasil estimasi dari penelitian ini menunjukkan bahwa metode* EBLUP *lebih baik dalam melakukan estimasi pengeluaran per kapita dibandingkan metode penduga langsung. Hal ini ditunjukkan dengan nilai* MSE *dari metode* EBLUP *menghasilkan rata-rata nilai* MSE *yang lebih kecil, yaitu sebesar 0.001 dibandingkan dengan rata-rata nilai* MSE *penduga langsung, yaitu sebesar 0.002. Hal ini tidak hanya menekankan reliabilitas*

* Corresponding author
Submitted July 12th, 2023, Revised May 6th, 2024,
Accepted for publication May 7th, 2024, Published Online May 31st, 2024

metode EBLUP tetapi juga meningkatkan ketangguhan analisis sosial ekonomi dan berkontribusi pada kemajuan teknik estimasi area kecil. Hal ini memberikan kebaruan dalam pemahaman disparitas regional dan pengambilan keputusan kebijakan.
Kata Kunci: small area estimation*, penduga langsung,* EBLUP*, pengeluaran per kapita.*

2020MSC: 62J05

1. INTRODUCTION

Per capita expenditure represents an adjustment to the consumer price index and a decrease in marginal utility. Based on these two things, per capita expenditure can be described by the level of people's purchasing power and the progress of human development in a region [1]. The calculation of per capita expenditure data in Indonesia has been accommodated through survey results and a direct estimator method called the National Socioeconomic Survey (Susenas). Susenas are one source of information to get an overview of the socio-economic conditions of the community [2]. The results of such surveys are obtained using direct sampling techniques in specific areas [3].

Pusponegoro and Rachmawati [4] explain that the direct estimation method in surveys estimates an area according to sample data from that area. However, direct estimation can be ineffective in estimating parameters due to limited sample adequacy in delivering them [5]. This limitation means that direct estimators cannot produce accurate estimates [6]. Therefore, to produce better predictions, indirect estimation methods can be used in small areas [6].

The SAE refers to a statistical approach employed in regions with limited sample sizes to enhance the accuracy of direct estimation [7]. By incorporating additional information or correlated variables, SAE utilizes a model that establishes connections between related areas [8]. Consequently, the estimation outcomes are derived at the area level, compensating for inadequate sample sizes for direct estimation [9]. SAE can suppress significant variations from direct estimators [10].

The SAE can be categorized into two primary models: area-based and unit-based [6]. The area-level model relies on supporting data exclusively available at a specific area level and takes the form of a linear mixed model, commonly referred to as the Fay-Herriot model. On the other hand, unit-level models are characterized by accompanying data that may be individually associated with response data but might be limited in availability [4]. These models are also known as nested error regression (NER) models.

Fay-Herriot, a pioneering researcher in the field, was instrumental in developing small area statistics for area-level models, which are represented as mixed linear models. Within this framework, Empirical Best Linear Unbiased Prediction (EBLUP) is one of the parameter estimation methods employed [4]. The EBLUP method is an approach to estimating parameters in the General Linear Model (GLM) [11] that minimizes the *Mean Square Error* (MSE) generated by assuming unknown variance components [12]. EBLUP has been utilized to offer stable [13] and dependable estimates for small areas within normal linear mixed models [14].

Several studies on the use of the EBLUP-SAE method have been widely conducted. [15] have researched EBLUP-SAE for estimating per capita expenditure in Bali. Furthermore, [16] has conducted research on the application of the EBLUP-SAE method to estimate per capita expenditure in the Brebes Regency. Both studies show that the MSE value of EBLUP-SAE is smaller than the MSE of direct estimators, so it is concluded that the EBLUP method is better at producing accurate estimation results.

Thus, in order to attain a higher level of accuracy based on the data acquired from Susenas March 2019, a research study will be conducted using the EBLUP-SAE method to estimate per capita expenditure in the province of South Sulawesi. In addition, this study was also conducted to see the comparison of per capita expenditure generated through the direct estimator method and EBLUP-SAE in producing more precise estimations. This study not only contributes to the methodological advancement in SAE but also provides a novel perspective on optimizing per capita expenditure estimations. The findings of this research are expected to serve as a foundation for the government in presenting more accurate data, which will be used as a basis for crafting effective policies.

2. METHODS

2.1 Data dan Research Variables

This study used secondary data from the Central Bureau of Statistics of South Sulawesi Province. We categorized the variables into response and predictor. The response variable focused on the per capita expenditure of districts in South Sulawesi in 2019, which was acquired from the March 2019 National Socioeconomic Survey (Susenas). The predictor variables were derived from the 2020 Figures publication of South Sulawesi. The response and predictor variables are listed in Table 1.

Table 1. Response and predictor variables

Variable	Description
$\boldsymbol{Y}$	Average of per capita expenditure per month
$\boldsymbol{X_1}$	Number of population
$\boldsymbol{X_2}$	Number of primary schools per sub-district
$\boldsymbol{X_3}$	Number of hospitals per sub-district
$\boldsymbol{X_4}$	Percentage of the population who has BPJS
$\boldsymbol{X_5}$	Percentage distribution of households by non-electric lighting source
$\boldsymbol{X_6}$	Number of accommodations in non-classified hotel
$\boldsymbol{X_7}$	Percentage of population aged 5 year and over who have cellular phone (HP)

2.2 Analysis Methods

In this study, an area-based Small Area Estimation (SAE) is employed, defined as [6]:

$$\theta_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + v_i; i = 1, \dots, m, \tag{1}$$

where m is the number of areas, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ is a $p \times 1$ vector of regression coefficients for the auxiliary variable x_i, and v_i is the random effect of the small area assumed to follow a $N(0, \sigma_v^2)$ distribution. The estimator θ_i can be obtained by assuming that the direct estimator model $\hat{\theta}_i$ is available [6]:

$$\hat{\theta}_i = \theta_i + e_i;\ i = 1, \dots, m, \tag{2}$$

where $e_i \sim N(0, \psi_i)$ and ψ_i is known. Combining Equations (1) and (2) yields Equation (3), known as the Fay-Herriot model at the area level [6]:

$$\hat{\theta}_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + v_i + e_i;\ i = 1, \dots, m. \tag{3}$$

The data analysis conducted in this study involved several stages, which are outlined as follows:

1. Estimate the EBLUP-SAE model parameters using the Maximum Likelihood (ML) method.
2. Calculate direct estimator for Susenas 2019 data using following equation [16].
$$\hat{\theta}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i};\ i = 1, \dots, m;\ j = 1, \dots, n_i. \tag{4}$$
3. Prepare the predictor variables from South Sulawesi in 2020 figures data.
4. Test the correlation between response and predictor variables using Pearson Product Moment. The hypotheses for this test were formulated as [17]:
   H_0 : There is no significant relationship between the response and the predictor variables,
   H_1 : There is a significant relationship between the response and the predictor variables.
   The corellation is calculated using Equation (5):
$$r_{xy} = \frac{n\sum_{i=1}^{n} X_i Y_i - (\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} Y_i)}{\sqrt{\left\{(n\sum_{i=1}^{n} X_i^2) - (\sum_{i=1}^{n} X_i)^2\right\}\left\{(n\sum_{i=1}^{n} Y_i^2) - (\sum_{i=1}^{n} Y_i)^2\right\}}}. \tag{5}$$
   The H_0 is rejected if the significance value is less than α or if $|r_{xy}|$ value is greater than or equals to the critical value $\left(r_{\frac{\alpha}{2}, n-2}\right)$.
5. Test the multicollinearity using the Variance Inflation Factor (VIF) between predictor variables. The hypotheses for this test were formulated as [18]:
   H_0 : There is no multicollinearity,
   H_1 : There is multicollinearity.
   The H_0 is rejected if $VIF > 10.00$.
6. Calculate the s_i^2 values from the direct estimator using the following equation [19]:
$$s_i^2 = \frac{n_i \sum_{j=1}^{n_i} y_{ij}^2 - \left(\sum y_{ij}\right)^2}{n_i(n_i - 1)}. \tag{6}$$
   This value is used to obtain variations of the EBLUP estimation.
7. EBLUP estimation for response variables from the predictor variables using the following equation [6]:
$$\hat{\theta}_i^{EBLUP} = \gamma_i \hat{\theta}_i + (1 - \gamma_i)\boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}. \tag{7}$$
8. Test the normality assumptions using the Anderson-Darling (AD). The hypotheses are formulated as [20]:
   H_0 : The data are normally distributed
   H_1 : The data are not normally distributed
   The test statistics is
$$AD = -\sum_{i=1}^{n} \left[\frac{(2i-1)\{\log P_i + \log(1 - P_{n+1-i})\}}{n}\right] - n. \tag{8}$$

The H_0 is rejected if the $AD > CV$ or the significance value is less than α. The CV is calculated using [21]:

$$CV = \frac{0.752}{1+\frac{0.75}{m}+\frac{2.25}{m^2}}. \tag{9}$$

9. Calculate the MSE of direct estimator and the EBLUP model using the following Equation [6]:

$$MSE(\hat{\theta}_i) = \frac{s_i^2}{n_i}. \tag{10}$$

$$MSE(\hat{\theta}_i^{EBLUP}) = \gamma_i\psi_i + (1-\gamma_i)^2\boldsymbol{x_i^T}\left[\sum\frac{x_i x_i^T}{\sigma_v^2+\psi_i}\right]^{-1}\boldsymbol{x_i} + \frac{\psi_i}{(\sigma_v^2+\psi_i)^{-3}}2m^{-2}\sum(\sigma_v^2+\psi_i)^2. \tag{11}$$

10. Determine the best model between the direct estimator and the EBLUP model.

3. RESULTS AND DISCUSSION

3.1 SAE Model Parameter Estimation with EBLUP

The SAE is utilized as an indirect estimator model, whereby additional information from censuses or other administrative records is incorporated into the estimation of survey data. In this study, an area-based SAE is employed, as depicted in Equation (1). To enable inferences regarding the characteristics of small areas beneath Equation (2), it is necessary to first determine the direct estimates of $\hat{\theta}_i$, as specified in Equation (3). The combination of Equations (1) and (2) yields Equation (3), which is known as the Fay-Herriot model for the area level.

The Fay-Herriot model at the area level is an extension of the mixed linear model based on the availability of accompanying variables for the respective areas. In the estimation of Empirical Best Linear Unbiased Predictors (EBLUPs) within the Fay-Herriot framework, the β component is unknown. It must be estimated from the available data in Fay-Herriot's EBLUP estimation. The Maximum Likelihood (ML) method is one of the approaches used for this purpose. The ML method requires the likelihood function, which is the pdf of $\hat{\theta}_i$ with $E(\hat{\theta}_i) = \boldsymbol{x_i^T\beta}$ and $var(\hat{\theta}_i) = \sigma_v^2 + \psi_i$. Due to the normal distribution of v_i and e_i, $\hat{\theta}_i$ is also normally distributed and can be expressed as $\hat{\theta}_i \sim N\left(\boldsymbol{x_i^T\beta}, (\sigma_v^2+\psi_i)\right)$. The pdf of $\hat{\theta}_i$ is as follows:

$$\boldsymbol{f(\hat{\theta}_i) = \frac{1}{(2\pi)^{\frac{1}{2}}(\sigma_v^2+\psi_i)^{\frac{1}{2}}}\exp\left[-\frac{1}{2}\left(\left(\hat{\theta}_i - x_i^T\beta\right)^T(\sigma_v^2+\psi_i)^{-1}\left(\hat{\theta}_i - x_i^T\beta\right)\right)\right]}. \tag{12}$$

Based on Equation (12), the likelihood function as follows:

$$\begin{aligned} \boldsymbol{L(\hat{\theta}_i)} &= \boldsymbol{\prod_{i=1}^{m} f(\hat{\theta}_i)} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}\prod_{i=1}^{m}(\sigma_v^2+\psi_i)^{\frac{1}{2}}}\exp\left[-\frac{1}{2}\sum_{i=1}^{m}\left(\hat{\theta}_i - \boldsymbol{x_i^T\beta}\right)^T(\sigma_v^2+\psi_i)^{-1}\left(\hat{\theta}_i - \boldsymbol{x_i^T\beta}\right)\right]. \end{aligned} \tag{13}$$

The ln of the likelihood function is

$$\boldsymbol{l}(\widehat{\boldsymbol{\theta}}_i) = -\frac{1}{2}\left(\sum_{i=1}^{m}(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^T\left(\boldsymbol{\sigma}_v^2 + \boldsymbol{\psi}_i\right)^{-1}(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{x}_i^T\boldsymbol{\beta})\right) - \frac{n}{2}\ln(2\boldsymbol{\pi}) - \left(\sum_{i=1}^{m}\ln\left(\boldsymbol{\sigma}_v^2 + \boldsymbol{\psi}_i\right)^{\frac{1}{2}}\right). \quad (14)$$

Derive the Equation (14) to $\boldsymbol{\beta}$ and equal it to zero to get the maximum $\boldsymbol{\beta}$:

$$\frac{\partial \boldsymbol{l}(\widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\beta}} = \frac{\partial\left(-\frac{1}{2}\left(\sum_{i=1}^{m}(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^T\left(\boldsymbol{\sigma}_v^2 + \boldsymbol{\psi}_i\right)^{-1}(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{x}_i^T\boldsymbol{\beta})\right) - \frac{n}{2}\ln(2\pi) - \left(\sum_{i=1}^{m}\ln\left(\boldsymbol{\sigma}_v^2 + \boldsymbol{\psi}_i\right)^{\frac{1}{2}}\right)\right)}{\partial \boldsymbol{\beta}} = \mathbf{0}$$

$$-\frac{1}{2}\left(\sum_{i=1}^{m}(\sigma_v^2 + \psi_i)^{-1} \cdot 2\left(\hat{\theta}_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\right)(-\boldsymbol{x}_i)\right) = 0.$$

$$\left(\sum_{i=1}^{m}\boldsymbol{x}_i(\sigma_v^2 + \psi_i)^{-1}\hat{\theta}_i\right) = \left(\sum_{i=1}^{m}\boldsymbol{x}_i(\sigma_v^2 + \psi_i)^{-1}(\boldsymbol{x}_i^T\boldsymbol{\beta})\right).$$

Therefore, the β parameter is

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{m}\left(\boldsymbol{x}_i(\sigma_v^2 + \psi_i)^{-1}\boldsymbol{x}_i^T\right)\right)^{-1}\left(\sum_{i=1}^{m}\left(\boldsymbol{x}_i(\sigma_v^2 + \psi_i)^{-\mathbf{1}}\hat{\theta}_i\right)\right). \quad (15)$$

Similarly, we can estimate the $\boldsymbol{\sigma}_v^2$. Derive the Equation (14) to $\boldsymbol{\sigma}_v^2$ and equal it to zero to get the maximum $\boldsymbol{\sigma}_v^2$.

$$\frac{\partial \boldsymbol{l}(\widehat{\boldsymbol{\theta}}_i)}{\partial \boldsymbol{\sigma}_v^2} = \boldsymbol{s}_j(\widehat{\boldsymbol{\beta}}, \boldsymbol{\sigma}_v^2) = \mathbf{0}$$

$$\frac{\mathbf{1}}{\mathbf{2}}\left(\sum_{i=1}^{m}(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^T(\boldsymbol{\sigma}_v^2 + \boldsymbol{\psi}_i)^{-2}(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{x}_i^T\boldsymbol{\beta})\right) - \frac{\mathbf{1}}{\mathbf{2}}\left(\sum_{i=1}^{m}\frac{\mathbf{1}}{(\boldsymbol{\sigma}_v^2 + \boldsymbol{\psi}_i)}\right) = \mathbf{0}$$

$$\frac{1}{2}\left(\sum_{i=1}^{m}(\hat{\theta}_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^T(\sigma_v^2 + \psi_i)^{-2}(\hat{\theta}_i - \boldsymbol{x}_i^T\boldsymbol{\beta})\right) = \frac{1}{2}\left(\sum_{i=1}^{m}\frac{1}{(\sigma_v^2 + \psi_i)}\right) \quad (16)$$

The σ_v^2 cannot be completed analytically. Therefore, the estimation of σ_v^2 will be achieved numerically using the scoring algorithm based on the approach described by [6], with the $(a+1)$-th iteration as follows:

$$\boldsymbol{\sigma}_v^{2(a+1)} = \boldsymbol{\sigma}_v^{2(a)} + \left[\boldsymbol{\mathcal{L}}\left(\boldsymbol{\sigma}_v^{2(a)}\right)\right]^{-1}\boldsymbol{s}\left(\widehat{\boldsymbol{\beta}}(\boldsymbol{\sigma}_v^2)^{(a)}, \boldsymbol{\sigma}_v^{2(a)}\right), \quad (17)$$

with $\boldsymbol{\mathcal{L}}(\sigma_v^2) = \frac{1}{2}\left(\frac{1}{(\sigma_v^2+\psi_i)^2}\right)$.

$$\boldsymbol{s}(\widehat{\boldsymbol{\beta}}, \sigma_v^2) = -\frac{1}{2}\left(\frac{1}{(\sigma_v^2 + \psi_i)}\right) + \frac{1}{2}\left(\frac{\left(\hat{\theta}_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\right)^2}{(\sigma_v^2 + \psi_i)^2}\right).$$

The iteration process was determined by the condition $\left|(\hat{\sigma}_v^2)^{(a+1)} - (\hat{\sigma}_v^2)^{(a)}\right| < 10^{-4}$. This condition means that the process continued until the difference between the estimated variance $\hat{\sigma}_v^2$ in iteration $(a+1)$ and iteration a was smaller than 10^{-4}. Moreover, the value of $(\hat{\sigma}_v^2)^{(a+1)}$ can be considered as the estimator of the variance σ_v^2.

3.2. Per capita Expenditure Estimation

3.2.1. Direct Estimation

The average per capita expenditure of households in districts/cities in South Sulawesi was directly estimated using data from the March 2019 National Socioeconomic Survey (Susenas). A descriptive analysis based on the direct estimation was performed to provide a statistical description of the per capita expenditure data in South Sulawesi in 2019. The descriptive analysis results, outlining the characteristics of per capita expenditure, can be observed in Table 2.

Table 2. Descriptive statistic of the response variable

Variable	Minimum (Rp)	Maximum (Rp)	Average (Rp)
Average per capita expenditure per month	766,867.17	1,735,397.65	1,063,623.91

The calculations of the direct estimator presented in Table 2 solely rely on the estimates derived from a sample of households included in the March 2019 National Socioeconomic Survey (Susenas). According to Table 2, the direct estimation method reveals that Jeneponto Regency has the lowest per capita expenditure at Rp. 766,867.17, while Makassar City has the highest per capita expenditure at Rp. 1,735,397.65. Meanwhile, it can be inferred that the average monthly per capita expenditure in households in South Sulawesi in 2019 is Rp. 1,063,623.91. The details of per capita expenditure data in South Sulawesi in 2019 can be seen in Table 3.

Table 3. Per Capita Expenditure Data in South Sulawesi in 2019

Regency	EBLUP Estimator (Rp)	Regency	EBLUP Estimator (Rp)
Kep. Selayar	1,030,334.40	Wajo	1,123,318.03
Bulukumba	918,866.16	Sidrap	1,099,380.74
Bantaeng	1,010,192.10	Pinrang	1,007,523.55
Jeneponto	766,867.17	Enrekang	819,660.04
Takalar	988,536.90	Luwu	970,662.53
Gowa	1,002,294.08	Tana Toraja	1,033,005.43
Sinjai	961,457.24	Luwu Utara	1,047,306.29
Bone	1,151,686.12	Luwu Timur	1,269,099.26
Maros	1,046,122.29	Toraja Utara	996,089.49
Pangkep	947,037.00	Kota Makassar	1,735,397.65
Barru	876,180.60	Kota Parepare	1,332,340.46
Soppeng	973,118.02	Kota Palopo	1,420,498.20

3.2.2. Predictor Variables Selection

The variables were standardized before the estimation of response and predictor variables was conducted. The standardization was conducted to equalize the response and predictor variables, which consist of different units. After standardization, the subsequent step involves determining predictor variables using the Pearson Product Moment correlation test. The Pearson Product Moment correlation test results are presented in Table 4.

Table 4. Pearson Product Moment correlation test results for predictor variables

Variable	r_{xy}	$p-value$
X_1	0.5638	0.0041
X_2	-0.0260	0.9039
X_3	0.6406	0.0007
X_4	0.1491	0.4868
X_5	-0.3112	0.1388
X_6	0.6369	0.0008
X_7	0.8483	0.0000

From Table 4, it is evident that out of the seven predictor variables, four of them exhibit a significant correlation with per capita expenditure. This is supported by the p-values of these four variables, which are less than the predetermined significance level of $\alpha = 0.05$. Those four variables are population (X_1), hospitals (X_3), non-classified accommodations (X_6), and cellular phones (X_7). Based on the direction of the relationship, the four variables have a positive correlation value. This value means that the higher the value of the four variables, the higher the value of per capita expenditure.

The following correlation test is a multicollinearity test. This test was conducted to ensure no correlation between the predictor variables. The test conducted to determine the correlation between predictor variables and per capita expenditure utilized the Variance Inflation Factor (VIF) values for each predictor variable. The VIF values of each predictor variable can be seen in Table 5.

Table 5. VIF values

Variable	X_1	X_2	X_3	X_4	X_5	X_6	X_7
VIF	7.176	2.053	7.686	1.461	1.279	3.589	2.338

According to Table 5, all VIF values for the variables are below 10. This value indicates that the null hypothesis is accepted i.e. there is no multicollinearity among the predictor variables. Consequently, there is no significant issue of multicollinearity within the dataset. Upon identifying the predictor variables that exhibit correlation with the response variable, the subsequent step involves incorporating these predictor variables into the model. The outcomes of estimating the regression coefficients using these predictor variables are presented in Table 6. According to this table, variables X_1 and X_3 exhibit the greatest level of significance. As a result, further steps will be taken to include these variables in the model through stages of variable selection. These stages will involve assessing each model's Akaike Information Criterion (AIC) value.

Table 7 shows the model selection based on AIC. From Table 7, it can be inferred that during the third stage, significant variables with the lowest AIC (Akaike Information Criterion) value of -30.241 were identified. Therefore, X_6 and X_7 variables will be included in the model.

Table 6. Estimation of the regression coefficient of four variables

Variable	$\widehat{\boldsymbol{\beta}}$	$\boldsymbol{p-value}$	Decision
$\boldsymbol{X_1}$	0.000	0.455	Reject H_0
$\boldsymbol{X_3}$	0.000	0.931	Reject H_0
$\boldsymbol{X_6}$	0.039	0.019	Accept H_0
$\boldsymbol{X_7}$	-0.023	0.016	Accept H_0

Table 7. Model selection

Stage	Predictor Variables	AIC
1	X_1, X_3, X_6, X_7	-26.859
2	X_1, X_6, X_7	-28.852
3	X_6, X_7	-30.241

3.2.3. SAE-EBLUP Modeling

The regression coefficient estimation will be carried out after obtaining the variables to be included in the model. The estimated regression coefficients obtained from these variables will be employed to estimate the parameters related to household per capita expenditure in South Sulawesi in 2019 for a small area. The results of estimating the regression coefficient based on the variables obtained in the third stage are as follows:

Table 8. Estimation of regression coefficients of selected variables

Variable	$\widehat{\boldsymbol{\beta}}$	$\boldsymbol{t-value}$	$\boldsymbol{p-value}$
Intercept	0.096	1.059	0.290
$\boldsymbol{X_6}$	0.028	5.654	0.000
$\boldsymbol{X_7}$	-0.022	-2.659	0.008

Table 8 illustrates that, according to the estimated regression coefficients, variables X_6 and X_7 significantly influence per capita expenditure in South Sulawesi. Additionally, the estimated random effect variance $(\hat{\sigma}_v^2)$ was determined to be 0.010. The estimated variance of the random effect $(\hat{\sigma}_v^2)$ suggests that there is a limited amount of variability in the data that cannot be accounted for by the fixed effect model. This value of the estimated variance for the random effect $(\hat{\sigma}_v^2)$ will subsequently be utilized in the EBLUP estimation. The results of calculating per capita expenditure using the SAE-EBLUP model in South Sulawesi are presented in Table 9.

According to Table 9, the EBLUP estimation results reveal that Makassar City, Palopo City, and Parepare City were the three cities with the highest expenditure. Among them, Makassar City stood out as the city with the highest expenditure, amounting to Rp.1,206,352.79. On the other hand, based on the EBLUP estimates, the district with the lowest per capita expenditure was Jeneponto Regency, which recorded Rp.1,000,887.29. This disparity indicates that the level of welfare, consumption, and purchasing power among households in Makassar City is superior and greater compared to that of Jeneponto Regency.

The normality assumption test will be conducted using the Anderson-Darling (AD) following the per capita expenditure estimations using EBLUP. The results of the normality test are presented in Table 10. This table shows that the AD value is 0.475, and the p-value is 0.218. These results indicate that we accept the null hypothesis so that the residuals of the EBLUP model follow a normal distribution.

Table 9. Estimated per capita expenditure based on EBLUP

Regency	EBLUP Estimator (Rp)	Regency	EBLUP Estimator (Rp)
Kep. Selayar	1054083.08	Wajo	1074845.47
Bulukumba	1033575.36	Sidrap	1070048.77
Bantaeng	1050857.05	Pinrang	1051636.72
Jeneponto	1000887.29	Enrekang	1012677.13
Takalar	1052517.50	Luwu	1042125.51
Gowa	1050461.68	Tana Toraja	1054085.79
Sinjai	1041316.50	Luwu Utara	1056003.84
Bone	1076988.04	Luwu Timur	1096906.67
Maros	1056176.50	Toraja Utara	1047538.97
Pangkep	1039874.48	Kota Makassar	1206352.79
Barru	1024967.34	Kota Parepare	1111727.70
Soppeng	1044948.82	Kota Palopo	1130311.25

Table 10. Result of the Anderson-Darling test

Anderson-Darling	$p-value$	Decision
0.475	0.218	Accept H_0

3.3. Comparison of Direct Estimator and EBLUP

Figure 1 shows the comparison graph of MSE values from direct estimators and EBLUP in each district or city. Based on Figure 1, it is evident that the EBLUP method demonstrates a smaller average MSE compared to the direct estimator method. Specifically, the EBLUP method yields an average MSE value of 0.001, whereas the direct estimator method produces an average MSE value of 0.002. These findings highlight the superior accuracy of the EBLUP method in estimating per capita expenditure.

Furthermore, the estimation results obtained using the EBLUP method reveal that three regencies in South Sulawesi exhibit the lowest average per capita expenditure: Jeneponto Regency, Enrekang Regency, and Barru Regency. This data serves as a crucial indicator for local government decision-making, signaling persistently low poverty levels, consumption, and household purchasing power in these areas, necessitating targeted improvement initiatives.

This observation is supported by recent research by [22], which affirms the superior accuracy of the EBLUP method over direct estimation methods in estimating per capita household expenditure. Similarly, studies by [23], [24], and [25] demonstrate that the SAE method provides more precise and reliable estimation results. These findings collectively underscore the utility of the SAE-EBLUP method as a valuable tool for local governments in presenting effective and focused data for policy formulation and resource allocation.

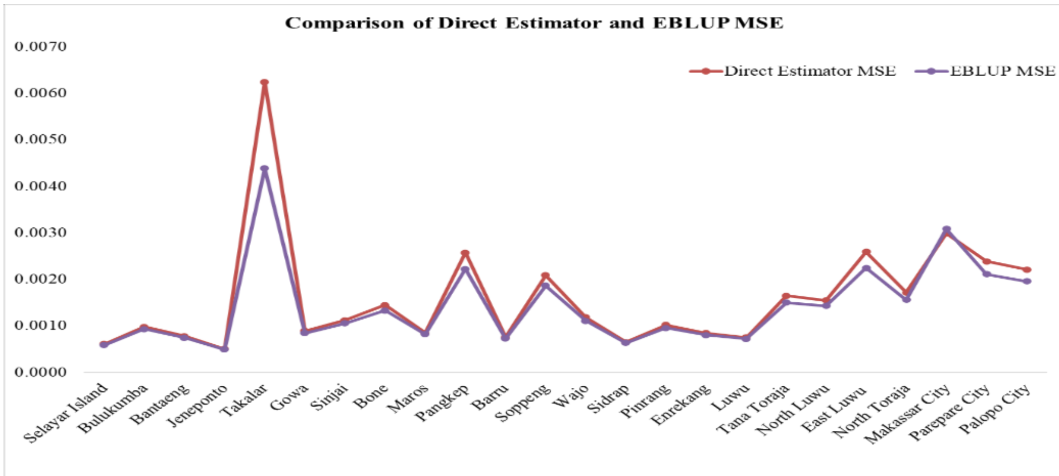

Figure 1. The MSE for direct estimators and EBLUP for every district

Moreover, considering the smaller MSE values reported in previous studies for the EBLUP method compared to direct estimator methods, it can be concluded that the EBLUP method consistently outperforms the direct estimator method in estimating per capita expenditure in South Sulawesi in 2019. These insights emphasize the importance of employing advanced statistical methodologies, such as the SAE-EBLUP method, for robust and accurate estimation of socio-economic indicators at local levels. This conclusion aligns with the findings of three prior studies conducted by [26], [12], and [16], which also reported superior performance of the EBLUP method in similar contexts.

4. CONCLUSIONS

Applying the EBLUP method, we found that the highest estimated per capita expenditure in South Sulawesi in 2019 was observed in Makassar City, with a value of Rp. 1,206,352.79. Meanwhile, using the EBLUP method, the lowest estimated per capita expenditure in South Sulawesi in 2019 occurred in Jeneponto Regency, which was Rp.1,000,887.29. The study demonstrates that the EBLUP method yields a smaller average Mean Squared Error (MSE) value of 0.001 compared to the average MSE value of direct estimators, which is 0.002. This value indicates that the EBLUP method provides more accurate and precise estimates of per capita expenditure in South Sulawesi in 2019 than the direct estimator method. In further studies, researchers suggested that estimates made using the EBLUP method could consider spatial effects, namely using the Spatial Empirical Best Linear Unbiased Prediction (SEBLUP) method.

REFERENCES

[1] D. S. Hutabarat, "Pengaruh Angka Harapan Hidup, Rata-Rata Lama Sekolah, Pengeluaran Riil Perkapita, Pertumbuhan Ekonomi dan Pengangguran terhadap Jumlah Penduduk Miskin di Provinsi Sumatera Utara," Universitas Sumatera Utara, Medan, 2018.

[2] BPS, *Statistik Kesejahteraan Rakyat Provinsi Sulawesi Selatan*. Makassar: BPS Provinsi Sulawesi Selatan, 2021.

[3] J. N. K. Rao and I. Molina, *Small Area Estimation*, Second edition. USA: John Wiley and Sons, 2015.

[4] N. H. Pusponegoro and R. N. Rachmawati, "Spatial Empirical Best Linear Unbiased Prediction in Small Area Estimation of Poverty," in *Procedia Computer Science*, Elsevier B.V., pp. 712–718, 2018.

[5] N. H. Pusponegoro, A. Kurnia, K. A. Notodiputro, A. M. Soleh, and E. T. Astuti, "Small Area Estimation of Sub-District's per Capita Expenditure through Area Effects Selection using LASSO Method," in *Procedia Computer Science*, Elsevier B.V., pp. 754–761, 2021.

[6] J. N. K. Rao, *Small Area Estimation*. USA: John Wiley and Sons, 2003.

[7] E. Ikhsan, C. A. Hidayat, and W. A. Nurizza, "Efisiensi Metode EBLUP pada Small Area Estimation Studi Kasus: Estimasi Persentase Penduduk Miskin di Provinsi Nusa Tenggara Timur Tahun 2017 (The Efficiency of EBLUP in Small Area Estimation Case Study: Estimation of Poor Pogpulation Percentage in East Nusa Tenggara Province in 2017)," *Jurnal Aplikasi Statistika & Komputasi Statistik*, vol. 10, no. 2, pp. 1–12, 2018.

[8] S. Iriyanto and M. Y. Darsyah, "Analysis of Poverty in Indonesia with Small Area Estimation : Case in Demak District," in *Kuala Lumpur International Business, Economics and Law Conference Vol.3*, Hotel Putra, pp. 52–57, 2014.

[9] S. Muchlisoh, "Pengembangan Model Pendugaan Area Kecil dengan Pengaruh Acak Waktu Mengikuti Proses Autoregresif Ordo Pertama," Disertasi, IPB, Bogor, 2017.

[10] M. A. Sidauruk and D. K. Sari, "Karakteristik Pendugaan Empirical Best Linear Unbiased Prediction (EBLUP) Pada Pendugaan Area Kecil," in *Prosiding SEMIRATA 2013*, 2013.

[11] M. I. Ilham and D. N. Hidayat, "Pemodelan Empirical Best Linear Unbiased Prediction (EBLUP) dalam Pendugaan Area Kecil: Studi Kasus Estimasi Rata-Rata Pengeluaran Per Kapita Menurut Kelurahan/Desa di Kabupaten Sukamara, Kalimantan Tengah," in *Seminar Nasional Official Statistics 2019*, 2019, pp. 117–123.

[12] D. Gartina and L. Khikmah, "Pendugaan Kemiskinan Menggunakan Small Area Estimation dengan Pendekatan Empirical Best Linier Unbiased Prediction (EBLUP)," *Statistika*, vol. 8, no. 2, pp. 158–164, 2020.

[13] T. Ito and T. Kubokawa, "Corrected Empirical Bayes Confidence Region in a Multivariate Fay–Herriot Model," *J Stat Plan Inference*, vol. 211, pp. 12–32, Mar. 2020, doi: 10.1016/j.jspi.2020.05.008.

[14] S. Sugasawa and T. Kubokawa, "On conditional prediction errors in mixed models with application to small area estimation," *J Multivar Anal*, vol. 148, pp. 18–33, Jun. 2016.

[15] L. Amaliana and M. L. S. Dewi, "EBLUP-Small Area Estimation Method for Per Capita Expenditures in Bali," *E-Jurnal Matematika*, vol. 11, no. 1, pp. 58–63, Jan. 2022.

[16] R. Ningtyas, R. Rahmawati, and Y. Wilandari, "Penerapan Metode Empirical Best Linear Unbiased Prediction (EBLUP) pada Model Penduga Area Kecil dalam Pendugaan Pengeluaran per Kapita di Kabupaten Brebes," *Jurnal Gaussian*, vol. 4, no. 4, pp. 977–986, 2015.

[17] H. Usman and R. P. Akbar, *Pengantar Statistika*. Jakarta: Bumi Aksara, 2008.

[18] G. Mardiatmoko, "Pentingnya Uji Asumsi Klasik Pada Analisis Regresi Linier Multiple (Studi Kasus Penyusunan Persamaan Allometrik Kenari Muda [Canarium Indicum L.])," *Jurnal Barekeng: Jurnal Ilmu Matematika dan Terapan*, vol. 14, no. 3, pp. 333–342, 2019.
[19] Sudjana, *Metoda Statistika*. Bandung: PT. Tarsito, 2005.
[20] B. W. Yap and C. H. Sim, "Comparisons of various types of normality tests," *J Stat Comput Simul*, vol. 81, no. 12, pp. 2141–2155, Dec. 2011.
[21] A. Setyawan, "Small Area Estimation Metode Spatial Empirical Best Linear Unbiased Predictor untuk Estimasi Persentase Wanita Usia Subur dengan Fertilitas Tinggi di Kabupaten Mamuju dan Mamuju Tengah," Tesis, Institut Teknologi Sepuluh Nopember, Surabaya, 2016.
[22] A. Desiyanti, I. Ginanjar, and T. Toharudin, "Application of an Empirical Best Linear Unbiased Prediction Fay–Herriot (EBLUP-FH) Multivariate Method with Cluster Information to Estimate Average Household Expenditure," *Mathematics*, vol. 11, no. 1, pp. 1–25, Jan. 2023.
[23] A. K. Kannaujiya *et al.*, "Spatial heterogeneity in intimate partner violence across the 640 districts of India: a secondary analysis of a cross-sectional, population-based survey by use of model-based small-area estimation," *Articles Lancet Glob Health*, vol. 11, pp. 1587–97, 2023.
[24] H. Wang, L. Smits, and P. Putrik, "Spatial variation in tobacco smoking among pregnant women in South Limburg, the Netherlands, 2016–2018: Small area estimations using a Bayesian approach," *Spat Spatiotemporal Epidemiol*, vol. 42, pp. 1–9, Aug. 2022.
[25] D. B. Rein *et al.*, "The prevalence of bilateral hearing loss in the United States in 2019: a small area estimation modelling approach for obtaining national, state, and county level estimates by demographic subgroup," *The Lancet Regional Health - Americas*, vol. 30, pp. 1–11, 2024.
[26] W. A. Nurizza, "Penerapan Model Fay-Herriot pada Small Area Estimation Studi Simulasi Pengeluaran Per Kapita Level Kabupaten/Kota Provinsi Kalimantan Timur Tahun 2020 (Application of the Fay-Herriot Model in Small Area Estimation: Simulation Study of Per Capita Expenditure at Regency/City Level in East Kalimantan Province in 2020)," *BESTARI: Buletin Statistika dan Aplikasi Terkini*, pp. 1–7, 2021.