Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 10, Number 2, 2026
DOI: 10.30829/zero.v10i2.28716
Page: 793-804

Nighttime Lights in Unit-Level Small Area Estimation for Estimating per Capita Expenditure

[1]Diaztri Hazam
School of Data Science, Mathematics, and Informatics, IPB University, Bogor, Indonesia

[2]Erfiani
School of Data Science, Mathematics, and Informatics, IPB University, Bogor, Indonesia

[3]Anang Kurnia
School of Data Science, Mathematics, and Informatics, IPB University, Bogor, Indonesia

Article Info

Article history:

Accepted: 03 August 2026

Keywords:

Nighttime lights;
Per capita expenditure;
Small area estimation.

ABSTRACT

The most widely used auxiliary variables for estimating per capita expenditure using small area estimation (SAE) are from National Socio-Economic Survey (SUSENAS) or Village Potential (PODES) data. Another alternative is remote sensing, which can quickly and cheaply identify area characteristics, such as nighttime lights (NTL). This study will compare the SAE model with auxiliary variables using PODES data, NTL data, and a combination of both. The method used is a unit-level SAE model with log-transformation to estimate per capita expenditure at the subdistrict level in Bandung Regency. Model performance was assessed using Relative Root Mean Squared Error (RRMSE), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). Estimation from all models have a similar range. The model with auxiliary variables using PODES data and combined data has similar RRMSE, AIC, and BIC. The model with auxiliary variables using only NTL data has the smallest RRMSE, AIC, and BIC.

Corresponding Author:

Diaztri Hazam,
School of Data Science, Mathematics, and Informatics,
IPB University, Bogor, Indonesia
Email: diaztrihazam@apps.ipb.ac.id

1. INTRODUCTION

Surveys are one way to obtain information from a large population at a lower cost. Surveys are widely used not only for estimating the total population under study, but also for subpopulations. Subpopulations can be geographical areas such as provinces, sub-districts, neighborhoods, or villages. An area is considered large if the sample size of the subpopulation is large enough to obtain results based on direct estimation with good precision. An area is considered small if the available sample size is not large enough to support direct estimation. If direct estimation is carried out in small areas that have too few examples, it will produce estimates with large standard errors and become unreliable [1].

Small area estimation (SAE) is a technique used to obtain estimates in these small areas. The problem is solved by conducting an indirect estimation that borrows the strength of information from the surrounding area and is linked to the model approach, such as administrative data and census data [2]. SAE relies on a mixed model relating the response variable in small areas to each other and to covariates [3]. SAE model is a linear model that combines fixed effects and area-specific random effects proposed by [4]. The fixed effects model is based on the assumption that the variability of the response variable within a small area can be fully explained by the corresponding variability relationship with additional information. The random effects model is based on the assumption that the variability specific to the small area cannot be explained by additional information [5].

SAE models can be divided into two types: area-level models and unit-level models [6]. Area-level models use the direct estimate as the response, but unit-level models use individual survey units as the response data. At the area-level, the areas with smaller sample sizes may "borrow strength" from areas with larger samples. Area-level models link predictors directly within a small area with covariates appropriate to that area. While unit-level models can overcome some of the limitations of area-level models. Unit-level models link unit values of the response variable with covariates specific to each unit. Estimation can be made at the same unit level or aggregated up to any desired level [7].

The concept of the linear model in model-based SAE is the main concept, so linear assumptions must be met and become an important requirement. [8] estimated per capita expenditure with one of the methods is the empirical best linear unbiased prediction (EBLUP) approach, and found that transformations were needed on response variables, auxiliary variables, or both. Transformation is performed on the response variable before using the SAE technique in estimating the parameter. Log-transformation is used by [9] with an area-level model to accommodate the linear assumptions. However, the EBLUP method is an inaccurate estimator because it violates the assumption of normality; [10] estimated the parameter using the empirical best predictor (EBP) method with log-transformation in the response variable. [11] compared the EBLUP method and EBP method in a unit-level model, and the result is EBP method provides more stable results than EBLUP.

SAE is often used to estimate socio-economic indicators, such as poverty in regions with limited survey data. According to Badan Pusat Statistik (BPS), poverty is the economic inability to meet basic food and non-food needs, measured by average monthly per capita expenditure below the Poverty Line. BPS defines per capita expenditure as the cost of food and non-food consumption for all household members during a month divided by the number of household members. Household expenditure can be used as a measure to assess the level of economic welfare [12].

Estimating per capita expenditure with SAE often uses data from SUSENAS or PODES as auxiliary variables. SUSENAS and PODES are used in [12], [13], [14], [15] as auxiliary variables for estimating per capita expenditure in district level. Due to the limited number of samples in SUSENAS, there is not much information available for smaller area levels, such as subdistricts or villages. PODES provides data at the subdistrict or village level, but it is mostly updated only every several years. That data may not be up to date with the current socio-economic context.

Remote sensing refers to a technique used to capture or obtain information from Earth's surface; one kind of example is Nighttime light (NTL). The use of NTL for the extraction of socio-economic data has increased [16]. NTL data indicate the light intensity related to economic activity in an area. High light intensity in some areas indicates more economic activity than low light intensity. NTL is used for estimating socio-economic cases, such as poverty [17], urban growth [18], Gross Regional Domestic Product (GRDP) [19], per capita expenditure [20], [21], [22]. A model with NTL data produces smaller errors and is more efficient than other models. NTL data is preferable because it has detailed resolution, which provides subdistrict-level or village-level data and is updated regularly.

[20], [21], [22] used the EBLUP method for estimating per capita expenditure. Usually in such cases, the normality assumption is not met, so EBLUP becomes an inaccurate estimator. EBP method was chosen because it overcomes the failure to fulfill the assumption of normal distribution, which often occurs in per capita expenditure, by adding a logarithmic transformation to the response variable. This study will use SAE EBP unit-level models with a log-transformation model to estimate per capita expenditure at the subdistrict level in Bandung Regency. NTL and PODES data will be included as auxiliary variables. This study will compare a model with PODES as auxiliary variables, a model with NTL as auxiliary variables, and a model with a combination of both data as auxiliary variables. We aim to study the performance of NTL data for estimating per capita expenditure. Model performance is assessed using Relative Root Mean Squared Error (RRMSE), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). The model with the best estimation is indicated by the smallest RRMSE. AIC and BIC are used to assess model complexity, which means the simpler model is the best model.

2. RESEARCH METHOD

2.1 Unit-Level SAE EBP Model

$y_{ij} \sim LN(\mu_i, \sigma^2) y_{ij}$[23] developed an empirical best predictor (EBP) method, which assumes that the response variable follows a lognormal distribution . Response variable with logarithmic transformation as follows

$$\log y_{ij} = y_{ij}^{*} = \boldsymbol{x}_{ij}^{T}\boldsymbol{\beta} + z_{ij}v_i + e_{ij};\ i = 1,2,\dots,m,\ j = 1,2,\dots,n_i \tag{1}$$

where $y_{ij}^{*} \sim N\left(x_{ij}^{T}\beta, z_{ij}^{2}\sigma_v^2 + \sigma_e^2\right)$; $\boldsymbol{\beta}$ is the vector of regression coefficients for each auxiliary variable; z_{ij} is a known positive constant; v_i is the random effect of area i; e_{ij} is the sampling error for the area i and unit j; assumes that $v_i \sim {}_{iid}N(0, \sigma_v^2)$ $e_{ij} \sim_{iid} N(0, \sigma_e^2)$ are independent. The response variable follows a lognormal distribution, so the mean of y_{ij} is $E(y_{ij}) = \mu_i = \exp\left(x_{ij}^{T}\beta + \frac{1}{2}\left(z_{ij}^{2}\sigma_v^2 + \sigma_e^2\right)\right)$ and the variance of y_{ij} is $Var(y_{ij}) = \sigma^2 = \exp\left(2\left[x_{ij}^{T}\beta + \frac{1}{2}\left(z_{ij}^{2}\sigma_v^2 + \sigma_e^2\right)\right]\right)\left[\exp\left(\exp(z_{ij}^{2}\sigma_v^2 + \sigma_e^2)\right) - 1\right]$.

Suppose the response variable in the small area is $\mu_i = \frac{1}{N_i}\sum_{i=1}^{N_i} y_{ij} = \frac{1}{N_i}\left[\sum_{j\in s_i} y_{ij} + \sum_{j\in r_i} y_{ij}\right]$ where s_i is the number of units sampled in the small area and r_i is the number of unsampled units. To obtain $\sum_{j\in r_i} y_{ij}$ we estimate the unsampled value of y_{ij} as $\hat{y}_{ij} = E(y_{ij})$. Estimator for μ_i is

$$\hat{\mu}_i^{BP} = \frac{1}{N_i}\left[\sum_{j\in s_i} y_{ij} + \sum_{j\in r_i} \hat{y}_{ij}^{BP}\right] \tag{2}$$

where $\hat{y}_{ij}^{BP} = \exp\left(x_{ij}^{T}\hat{\beta} + v_i + \frac{1}{2}\left(z_{ij}^{2}\sigma_v^2 + \sigma_e^2\right)\right)$. In reality σ_v^2 and σ_e^2 are usually unknown and estimated by $\hat{\sigma}_v^2$ and $\hat{\sigma}_e^2$ using Maximum Likelihood (ML) or Restricted Maximum Likelihood (REML). In this study, variance was estimated by REML. EBP of μ_i is obtained by replacing unknown variance with the estimate as follows

$$\hat{\mu}_i^{EBP} = \frac{1}{N_i}\left[\sum_{j\in s_i} y_{ij} + \sum_{j\in r_i} \hat{y}_{ij}^{EBP}\right] \tag{3}$$

where $\hat{y}_{ij}^{EBP} = \exp\left(x_{ij}^{T}\hat{\beta} + v_i + \frac{1}{2}\left(z_{ij}^{2}\hat{\sigma}_v^2 + \hat{\sigma}_e^2\right)\right)$. If y_{ij} is not lognormally distributed, then $\hat{\mu}_i^{EBP}$ is biased. The bias correction is $c_{ij} = 1 + \frac{1}{2}x_{ij}V(\hat{\beta})x_{ij}^{T} + \frac{1}{8}\nabla\left(z_{ij}^{2}\hat{\sigma}_v^2 + \hat{\sigma}_e^2\right)$ where $V(\hat{\beta}) = [(X^{T}V^{-1}X)^{T}]^{-1}$ or variance of $\hat{\beta}$ and $\nabla\cdot$ is asymptotic variance-covariance matrix of total variability of response. The bias-corrected EBP of μ_i is

$$\hat{\mu}_i^{EBP} = \frac{1}{N_i}\left[\sum_{j\in s_i} y_{ij} + \sum_{j\in r_i} \hat{y}_{ij}^{EBP_c}\right] \tag{4}$$

$\hat{y}_{ij}^{EBP_c} = \left(c_{ij}\right)^{-1}\exp\left(x_{ij}^{T}\hat{\beta} + v_i + \frac{1}{2}\left(z_{ij}^{2}\hat{\sigma}_v^2 + \hat{\sigma}_e^2\right)\right)$ where [10].

2.2 Data Source and Variables

Direct estimation of per capita expenditure was derived from a household expenditure survey in Bandung Regency by the Human Development Index (HDI) Assessment Team, Department of Statistics, Faculty of Mathematics and Natural Sciences, IPB University in 2023. The observation area in this study is 31 subdistricts, with a sample of 3-4 villages selected from each district. Subsequently, 4-6 households were sampled within each village, resulting in a total dataset of 500 households. PODES data in 2021 are from Badan Pusat Statistik (BPS). Bandung Regency map at the subdistrict level is from BPS. The Visible and Infrared Imaging Suite (VIIRS) Nighttime Lights (VNL) is one such imagery product provided by the Earth Observation Group (EOG) from the National Oceanic and Atmospheric Administration (NOAA). The resulting imagery has been filtered for extraneous features such as biomass burning, auroras, and background light [24]. Average nighttime light intensity was obtained from feature extraction in Annual VNL 2.2 in 2022. For a unit-level model, auxiliary variables should be at the unit level. However, in this study, auxiliary variables at the unit level were not available, so it borrowed strength from a larger area at the village level.

There are twelve auxiliary variables, with eleven variables from 2021 PODES data and one variable from NTL data. The twelve auxiliary variables are urban area or rural area status (X_1), lighting status on main roads (X_2), proportion of households using electricity (X_3), proportion of households in slums (X_4), proportion of households that have Certificate of Inability to Pay (SKTM) a certificate indicating economic incapacity issued by local authorities (X_5), number of educational facilities (X_6), number of health facilities (X_7), distance from subdistrict's office to city (X_8), number of banks (X_9), number of traditional markets (X_{10}), number of minimarkets or supermarkets (X_{11}), and average nighttime light intensity (A).

2.3 Analytical Methods

The analysis process in this study consists of the following steps:

a. Preparing average per capita expenditure as the response variable and all auxiliary variables from PODES and average nighttime light intensity from NTL data. To obtain average nighttime light intensity data, a process known as feature extraction is required. Feature extraction is performed by converting nighttime lights satellite imagery raster data into a vector table using zonal statistics. Shapefile data is used as a zone layer to determine regional boundaries and to obtain average light-intensity values from each pixel in nighttime-lights satellite imagery raster data [21].

b. Performing exploratory data analysis on the response variable and auxiliary variables.
c. Testing the response variable per capita expenditure using the Kolmogorov-Smirnov test and a histogram to assess the pattern to meet the assumptions of a lognormal distribution.
d. Estimating using the unit-level SAE EBP method with three models, which will be compared using different auxiliary variables. The first model uses only PODES data, the second one uses a combination of PODES data and NTL data, and the third one uses only NTL data. The models are shown as follows

$$\text{Model 1: } \hat{\mu}_i^{(1)} = \frac{1}{N_i}\left[\sum_{j\in s_i} y_{ij} + \sum_{j\in r_i} \hat{y}_{ij}^{*}\right];\ \hat{y}_{ij}^{*} = \exp\left(\hat{\beta}_0 + \sum_{k=1}^{p=11} x_{ijk}\hat{\beta}_k + \frac{1}{2}\left(z_{ij}^2\hat{\sigma}_v^2 + \hat{\sigma}_e^2\right)\right) \quad (5)$$

$$\text{Model 2: } \hat{\mu}_i^{(2)} = \frac{1}{N_i}\left[\sum_{j\in s_i} y_{ij} + \sum_{j\in r_i} \hat{y}_{ij}^{*}\right];\ \hat{y}_{ij}^{*} = \exp\left(\hat{\beta}_0 + \sum_{k=1}^{p=11} x_{ijk}\hat{\beta}_k + A_{ij}\hat{\beta} + \frac{1}{2}\left(z_{ij}^2\hat{\sigma}_v^2 + \hat{\sigma}_e^2\right)\right) \quad (6)$$

$$\text{Model 3: } \hat{\mu}_i^{(3)} = \frac{1}{N_i}\left[\sum_{j\in s_i} y_{ij} + \sum_{j\in r_i} \hat{y}_{ij}^{*}\right];\ \hat{y}_{ij}^{*} = \exp\left(\hat{\beta}_0 + A_{ij}\hat{\beta} + \frac{1}{2}\left(z_{ij}^2\hat{\sigma}_v^2 + \hat{\sigma}_e^2\right)\right) \quad (7)$$

e. Evaluating the estimation results of average per capita expenditure based on Mean Squared Error (MSE) for the EBP method using from [8] is

$$MSE(\hat{\mu}_i^{EBP}) = (1-f_i)^2\left\{e^{2x_i^T\hat{\beta}+\hat{\sigma}_e^2+\hat{\sigma}_v^2}\left(x_i^T\hat{V}(\hat{\beta})x_i + \frac{1}{4}\hat{\nabla}\left(\hat{\sigma}_v^2\right)\right) + \hat{\sigma}_v^2\right\} + N_i^{-2}(N_i - n_i)\hat{\sigma}_e^2 \quad (8)$$

where n_i is the sample size in small area i; N_i is the population size in small area i; $f_i = \frac{n_i}{N_i}$ and $\nabla\left(\hat{\sigma}_v^2\right)$ is the asymptotic variance of $\hat{\sigma}_v^2$. To simplify the interpretation, RRMSE will be used with the formula

$$RRMSE(\hat{\mu}_i^{EBP}) = \frac{\sqrt{MSE(\hat{\mu}_i^{EBP})}}{\hat{\mu}_i^{EBP}} \times 100 \quad (9)$$

f. Examining model performance through the evaluation of RRMSE, AIC, and BIC. RRMSE is used to evaluate the accuracy of estimation values, while AIC and BIC are used to assess model efficiency. Lower values of RRMSE, AIC, and BIC indicate the model has smaller errors between actual and estimated values and the most efficient model for estimating per capita expenditure.

2.4 Software and Tools

In this study, we used two computational tools: QGIS and R. QGIS was used to perform a clipping process on the NTL raster data to match the shapefile boundary of Bandung Regency. R was used for statistical analysis with readxl, foreign, and dplyr packages for exploratory data analysis; sf, terra, raster, and dplyr packages for calculating zonal statistics and visualizing the map of Bandung Regency; writexl, dplyr, and sae packages for estimating using the unit-level SAE EBP method.

3. RESULT AND ANALYSIS

The histogram of survey data per capita expenditure in Bandung Regency is shown in Figure 1. Based on the plot, the distribution of per capita expenditure extends to the right and does not appear to follow a normal distribution. The largest units are in the per capita expenditure range of IDR 100.000,00 to IDR 200.000,00. The average per capita expenditure is IDR 174.281,41 and the median is IDR 149.780,08. The highest per capita expenditure is in Arjasari Subdistricts with IDR 907.888,89 and the lowest is in Cangkuang Subdistrict with IDR 11.416,67.

The plot shows that per capita expenditure in Bandung Regency looks right-skewed, which indicates a lognormal distribution. It means the data is not normally distributed and is also supported by the Kolmogorov-Smirnov test, which shows a *p*-value less than $2.2\ x\ 10^{-6}$. This case meets the assumptions of using the unit-level SAE EBP method with log-transformation in the response variable.

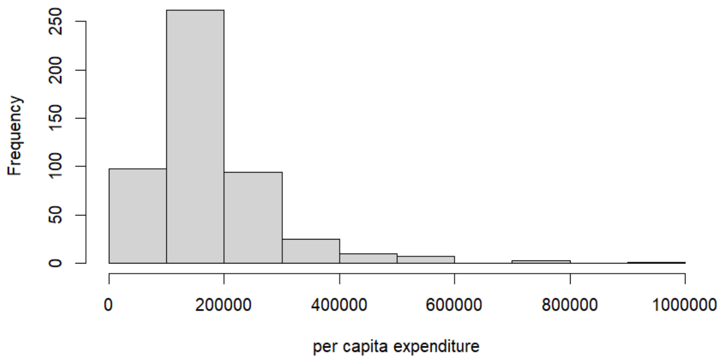


Figure 1. Histogram of per capita expenditure in Bandung Regency

The nighttime light intensity map of Bandung Regency by subdistricts is shown in Figure 2. Based on this map, Bandung Regency is dominated by areas that have low light intensity. Light intensity indicates nighttime community activity, which requires the use of light in public places [19]. The upper part of Bandung Regency has higher light intensity than the lower part. Soreang, the center of government in Bandung Regency, has high light intensity. Some subdistricts with high light intensity are Cileunyi, Margaasih, Margahayu, Dayeuhkolot, Bojongsoang, Katapang, Soreang, and Baleendah. This indicates that subdistricts have more nighttime community activity than in other subdistricts.

The correlation between all variables is shown in Figure 3. All auxiliary variables have low correlation values and there are several variables that are highly correlated with each other. Number of educational facilities and health facilities are correlated by 0.60; number of health facilities and number of banks are correlated by 0.61; number of health facilities and number of minimarkets are correlated by 0.70; number of health facilities and nighttime lights intensity are correlated by 0.61; number of banks and number of minimarkets are correlated by 0.72; number of minimarkets and nighttime lights intensity are correlated by 0.60.

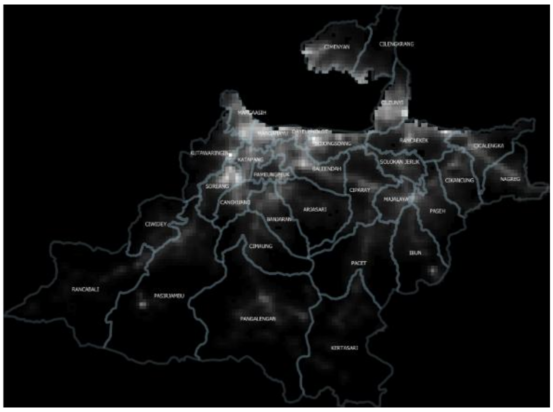


Figure 2. Map nighttime lights intensity in Bandung Regency

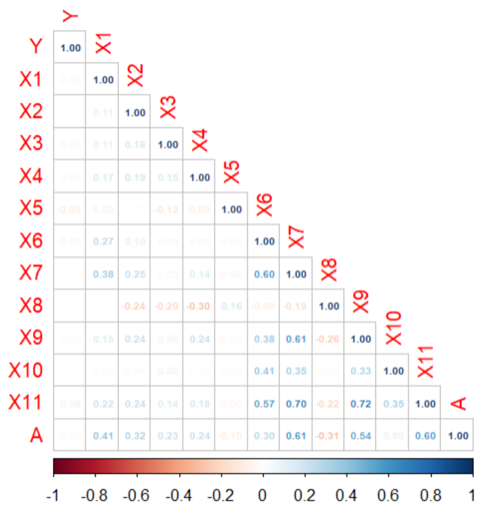


Figure 3. Correlation between all variables

Table 1. Summary of auxiliary variables from X_3 to A

Aux. Var.	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	A
Mean	1.058	0.006	0.089	10.85	9.402	23.81	1.056	0.33	2.564	5.612
Minimum	0.664	0	0.002	2	2	1	0	0	0	0.094
Maximum	1.877	0.078	0.749	47	29	60	12	4	17	16.838
Standard Deviation	0.181	0.015	0.097	6.146	6.357	14.712	1.966	0.736	3.040	4.072

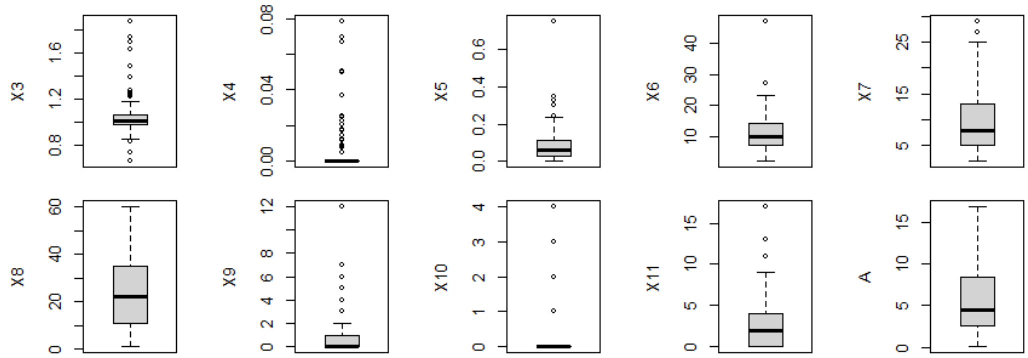


Figure 4. Boxplot of auxiliary variables from X_3 to A

There are three types of data in auxiliary variables, namely categorical, discrete, and continuous. In this study, there is no transformation in auxiliary variables even though the data types are different. For X_1 and X_2 have categorical data. X_1 shows the status of each village. There are 415 areas with urban status and 85 areas with rural status. X_2 shows the status of lighting on main roads. There are 375 areas where most of their main roads have lighting, while in 125 areas, only a small portion of their main roads have lighting. This indicates that even at night, community activities can still take place, supported by the lighting on main roads.

Plots for auxiliary variables X_3 to A are shown in Figure 3 and the summary in Table 1. X_3 shows the proportion of households using electricity for each village. The data shows that most villages have all households using electricity. X_4 shows proportion of households in slums and the data shows that most villages have zero proportion of households in slums. X_5 shows proportion of households have SKTM in each village. The data shows a proportion mostly zero, but there is an outlier which shows a value of 0.749. It means more than half of households in that village have SKTM.

X_6 shows number of educational facilities which include primary school, secondary school, high school and university. The mean of number of educational facilities is 10,85 with one outlier which shows a value of 47. It shows that there is one village has more educational facilities than the others. X_7 shows number of health facilities, with an average of 9,402. Based on the table, educational facilities and health facilities shows a very wide range. Educational facilities have 2 as minimum value and 47 as maximum value. Health facilities have 2 as minimum value and 29 as maximum value. There are signs of significant disparities between villages.

X_8 shows the distance from village's office to the city. The range of variable X_9 shows wide from one kilometre to sixty kilometres. It means some villages are located quite far from the city, while others are located close to the city. X_9 shows number of banks and the data dominates 0. It means in Bandung Regency, only a few villages have banks. There is one village has twelve banks thus creating an outlier in the data. The same pattern looks in X_{10} that shows number of traditional markets. The range in variable X_{10} is not too far, from zero to four traditional markets.

X_{11} shows number of minimarkets or supermarkets. The data has domination under 5 minimarkets or supermarkets for each village and there is only one village that has more than 15 minimarkets or supermarkets. Last auxiliary variable is A that shows the average nighttime lights intensity from NTL data. The range of data is between 0.094 and 16.838 with the average of data is 5.612, meaning there are some villages that have high light intensity but some villages have low light intensity approaching zero. Low light intensity indicates that there are a few or less community activities in the village at night.

X_5X_{11}Parameter estimation using unit-level SAE EBP method for Model 1, Model 2, Model 3 is shown in Table 2. Based on the results, many variables in parameter estimation do not show statistically significant results in all models. In Model 1 and Model 2, variables have statistically significant are (proportion of households have SKTM) and (number of minimarkets or supermarkets). Model 3 has no variables that are statistically significant. However, in SAE, interpretation of a statistically significant parameter is unnecessary [25]. This is because the focus of SAE is estimating the value of μ, in this study it is per capita expenditure. Therefore, all auxiliary variables can be used for estimating per capita expenditure in Bandung Regency.

Table 2. Parameter estimation using unit-level SAE EBP Model 1

	Estimation	**Standard Error**	**T- value**
Model 1			
Intercept	11.995	0.195	61.614*
X_1	0.017	0.076	0.224
X_2	-0.040	0.065	-0.618
X_3	-0.027	0.154	-0.173
X_4	1.006	1.847	0.545
X_5	-0.639	0.265	-2.412*
X_6	0.002	0.005	0.277
X_7	-0.008	0.006	-1.219
X_8	0.001	0.002	0.262
X_9	-0.021	0.019	-1.115
X_{10}	-0.028	0.038	-0.719
X_{11}	0.039	0.014	2.707*
Model 2			
Intercept	11.999	0.196	61.301*
X_1	0.007	0.081	0.089
X_2	-0.043	0.066	-0.660
X_3	-0.039	0.157	-0.247
X_4	0.995	1.855	0.537
X_5	-0.639	0.266	-2.404*
X_6	0.002	0.006	0.337
X_7	-0.008	0.007	-1.277
X_8	0.001	0.002	0.306
X_9	-0.023	0.019	-1.158
X_{10}	-0.026	0.039	-0.654
X_{11}	0.037	0.015	2.513*
A	0.004	0.010	0.406
Model 3			
Intercept	11.8875	0.0491	242.017*
A	0.0007	0.0067	1.019

*significant at the 5% level

As a result, parameter estimation for X_5 and X_{11} in Model 1 and Model 2 show similar values. Variable X_5 has a negative sign for estimating per capita expenditure. It means that if the proportion of households with SKTM increases, per capita expenditure will decrease. Meanwhile, variable X_{11} has a positive sign for estimating per capita expenditure. It means an increase in the number of minimarkets and supermarkets will also increase per capita expenditure.

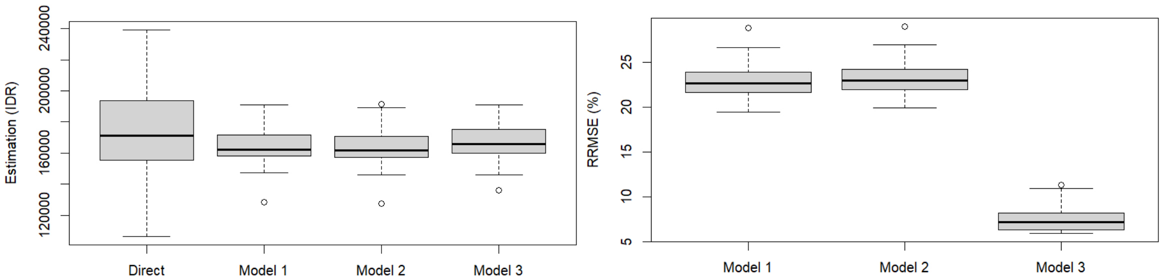
Table 3. Comparison between direct estimation and unit-level SAE EBP model

	Direct Estimation (IDR)	**Model 1** Estimation (IDR)	**Model 1** RRMSE (%)	**Model 2** Estimation (IDR)	**Model 2** RRMSE (%)	**Model 3** Estimation (IDR)	**Model 3** RRMSE (%)
Average	174,281.41	164,906.03	23.00	164,406.92	23.34	166,996.67	7.55
Minimum	106,163.70	128,564.62	19.44	127,370.10	19.92	136,086.70	5.92
Maximum	239,364.71	191,040.57	28.86	191,653.21	29.03	190,900.43	11.32

The results of the comparison between direct estimation and the unit-level SAE EBP method are shown in Table 3 and the boxplot in Figure 5. The table contains a summary of per capita expenditure estimation and RRMSE from all models at the subdistrict level. RRMSE is used to evaluate how well the estimation values are based on the model. Lower values of RRMSE indicate better model performance. There is not much difference in estimation between the model that has NTL data as an auxiliary variable and one that does not have NTL, besides being a single variable.

Model 1 and Model 2 have similar estimation. The second model has a smaller average and minimum estimation than the first model, although the second model has the addition of NTL data as an auxiliary variable. Model 1 has a range of IDR 128,564.62 to IDR 191,040.57 with an average of IDR 164,906.03. Model 2 has a arrange of IDR 127,370.10 to IDR 191,653.21 with an average of IDR 164,406.92. The third model with only NTL data as an auxiliary variable has the highest average and minimum estimation, but the smallest maximum estimation. Model 3 has a range of IDR 136,086.70 to IDR 190,900.43 with an average of IDR 166,996.67.

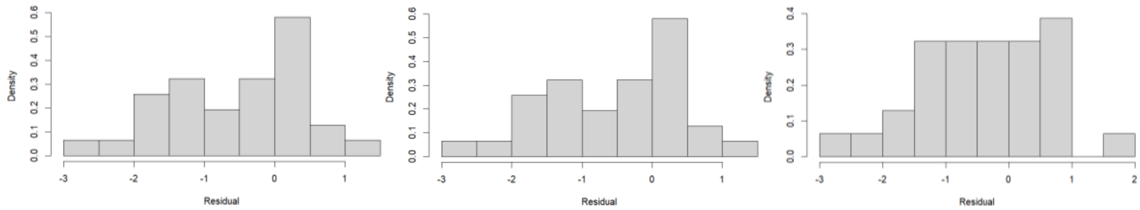
However, RRMSE from three models has different interpretations. Model 1 and Model 2 have similar values, but the first model has a smaller value than the second model. Both Model 1 and Model 2 have a range of 19% to 29% with an average 23% of RRMSE. Model 3 has quite a big difference from the other two models. The third model has the smallest average, minimum, and maximum values with a range of 5.92% to 11.32% and an average of 7.55%. This indicates the third model has the smallest error for estimating per capita expenditure in Bandung Regency. If we look at the MSE formula in equation (8), we can find that the formula has a matrix of auxiliary variables. This means that more auxiliary variables in the model will increase MSE.



(a) (b)

Figure 5. Boxplot comparison between (a) direct estimation and unit-level SAE EBP model (b) RRMSE estimation with unit-level SAE EBP model

In Figure 5, we can see that direct estimation has a large range in IDR 106,163.70 to IDR 239,364.71 with an average of IDR 174,281.41. All estimation with unit-level SAE EBP models narrows the range and make it look more stable. There is one outlier in Model 1 and Model 3 under minimum data and two outliers in Model 2 at the upper maximum and under minimum data. Model 3 looks more stable, with the median line almost dividing the middle value of the data with outliers not too far. In the boxplot of RRMSE, we can see that the boxplots of Model 1 and Model 2 are located above than the boxplot of Model 3. The outlier in the boxplot for Model 3 is not too far from the maximum data point.



(a) (b) (c)

Figure 6. Histogram of residuals (a) Model 1, (b) Model 2, and (c) Model 3

The histogram of residuals from the estimation results of all models is shown in Figure 6. The assumption that must be met is that the residuals must follow a normal distribution. The histogram of all models shows that the residuals have a normal distribution and is also supported by the Kolmogorov-Smirnov test, which shows a *p*-value of 0.155 for Model 1, 0.216 for Model 2, and 0.577 for Model 3. All models meet the assumption.

Other values used to see effectiveness of the model are AIC and BIC. The comparison of AIC and BIC between each model is shown in Table 4. The results for AIC and BIC yield the same conclusion, namely that the third model has the smallest AIC at 794.67 and BIC at 803.10. This shows that the third model is the most efficient for estimating per capita expenditure in Bandung Regency.

Table 4. Comparison of AIC and BIC between each model

	Model 1	**Model 2**	**Model 3**
AIC	858.87	867.87	794.67
BIC	909.44	922.66	803.10

The plot comparing direct estimation and the unit-level SAE EBP model is shown in Figure 7. All models show a stable line in estimating per capita expenditure in Bandung Regency. The black line shows direct estimation, the red line shows estimation from Model 1, the green line shows estimation from Model 2, and the blue line shows estimation from Model 3. Based on the plot, we can say Model 1 and Model 2 have similar estimation. Model 3 looks more stable and in a few spots has quite different estimation from the other two models. Figure 8 shows RRMSE of the estimation from each unit-level SAE EBP model. In this plot, Model 1 and Model 2 have similar RRMSE but are quite different from Model 3, which has the smallest RRMSE.

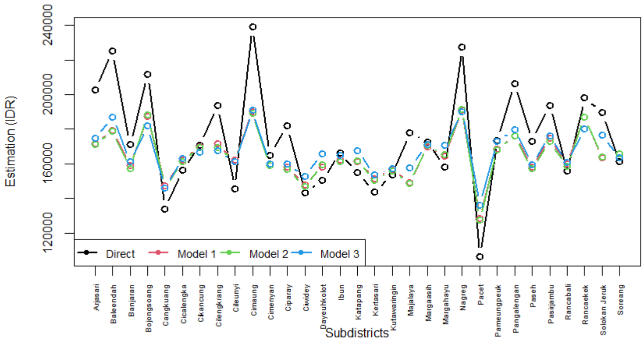


Figure 7. Plot comparison between direct estimation and unit-level SAE EBP model

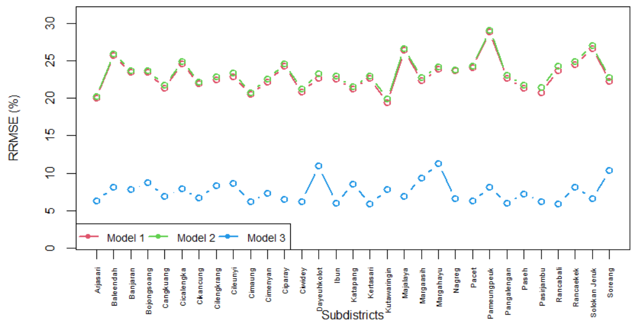


Figure 8. Plot RRMSE estimation with unit-level SAE EBP model

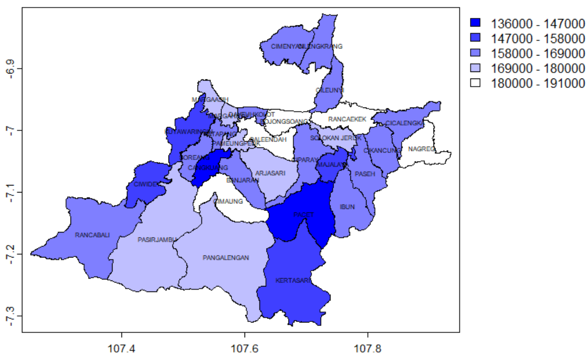


Figure 9. Map per capita expenditure estimation using Model 3

Estimation for per capita expenditure in Bandung Regency using unit-level SAE EBP Model 3 is shown in Figure 9. Dark blue indicates low per capita expenditure and white indicates high per capita expenditure. The subdistricts with low per capita expenditure in a range of IDR 136,000 to IDR 147,000 are Pacet and Cangkuang. The subdistricts with high per capita expenditure in a range of IDR 180,000 to IDR 191,000 are Bojongsoang, Rancaekek, Nagreg, Baleendah, and Cimaung. The center of government, Soreang has a light blue which is in the middle of the list with a range of IDR 158,000 to IDR 169,000.

Based on the results of this study, Model 3 which has only NTL data as an auxiliary variable, showed better estimation using the unit-level SAE EBP method. Model 3 shows more stable estimation, with the smallest error and a more efficient model. These results are consistent with [20], [21], [22], which model with night light intensity as auxiliary variables produce better results than models with PODES using the area-level SAE EBLUP method. Even though the methods used are different, the results obtained remain the same.

Although NTL in Model 3 was not shown to be statistically significant, NTL can still be used as an auxiliary variable for estimating per capita expenditure. NTL data reflects socio-economic activity using the concept of brightness of the lights. The results of those studies show NTL becomes a preferable covariate due to its lower cost and more comprehensive coverage. An easier way to obtain NTL data is one of the considerations for policymakers to be able to estimate per capita expenditure by only using night satellite imagery. NTL data has more frequent updates, so it can save costs in estimating per capita expenditure at the subdistrict level. Meanwhile, NTL data alone did not cover area characteristics in terms of economic cases. Combination with PODES data as an auxiliary variable might still be valuable despite having a higher RRMSE in this case.

In this study, we did not consider spatial dependence that may occur in small areas. Based on Figure 7, we can see that a subdistrict with high per capita expenditure may affect other subdistricts around it. Spatial modelling improves estimation through borrowing strength, where information from neighboring regions helps improve estimates for small areas [26]. The performance of the spatial EBP (SEBP) method in a simulation study reported by [27] concludes that SEBP has a smaller RRMSE than direct estimation and the EBP method. Adding consideration of spatial dependence in small areas with the SEBP method could be a future recommendation in estimating per capita expenditure.

4. CONCLUSION

Estimating per capita expenditure in Bandung Regency using the tunit-level SAE EBP method with log-transformation gives stable estimation. Using nighttime lights data as an auxiliary variable to estimate per capita expenditure yields efficient results with the smallest RRMSE, AIC, and BIC. Adding NTL data as an auxiliary variable along with PODES data does not yield a small RRMSE. The model with the most efficient estimation result has only NTL data as an auxiliary variable.

A simpler method of obtaining NTL data than PODES or SUSENAS data can be considered as an option for using it as an auxiliary variable in estimating per capita expenditure. This can facilitate the assessment and estimation of per capita expenditure at subdistrict level. Recommendations for further research include considering spatial dependence among small areas using the SEBP method.

5. REFERENCES

[1] J. N. K. Rao and I. Molina, "*Small Area Estimation,*" New Jersey: John Wiley & Sons, 2015. https://doi.org/10.1002/9781118735855

[2] L. Latifah, K Sadik, and Indahwati, "Simulation study of hierarchical Bayesian approach for small area estimation with measurement error," *BAREKENG: Journal of Mathematics and Its Applications*, vol. 17, no. 4, pp. 2059–2070, 2023. https://doi.org/10.30598/barekengvol17iss4pp2059-2070

[3] S. L. Lohr and N. G. N. Prasad, "Small Area Estimation with Auxiliary Survey," [Online]. Available: https://www.stat.ualberta.ca/people/prasad/preprints/Tech%20Report_01_08.pdf

[4] R. E. Fay III and R. A. Herriot, "Estimates of income for small places: an application of James-Stein procedures to census data," *Journal of the American Statistical Association*, vol. 74, no. 366a, pp. 269–277, 1979. https://doi.org/10.1080/01621459.1979.10482505

[5] R. Ananda, K. A. Notodiputro, and M. N. Aidi, "Modified mixed effects random forest in small area estimation using PCA and rotation forest with correlated auxiliary variables," *Scientific Journal of Informatics*, vol. 11, no. 3, pp. 705–720, 2024. https://doi.org/10.15294/sji.v11i3.10633

[6] J. N. K. Rao, "Small Area Estimation," New Jersey: John Wiley & Sons, 2003. https://doi.org/10.1002/0471722189

[7] P. A. Parker, R. Janicki, and S. H. Holan, "A comprehensive overview of unit-level modelling of survey data for small area estimation under informative sampling," *Journal of Survey Statistics and Methodology*, vol. 11, no. 4, pp. 829–857, 2023. https://doi.org/10.1093/jssam/smad020

[8] A. Kurnia, "Prediksi terbaik empirik untuk model transformasi logaritma di dalam pendugaan area kecil dengan penerapan pada data SUSENAS," dissertation, Bogor: IPB University, 2009.

[9] N. Permatasari and A. Ubaidillah, "Small area estimation of poverty using remote sensing data," *Statistical Journal of the IAOS*, vol. 41, no. 1, pp. 180–190, 2025. https://doi.org/10.1177/18747655241308390

[10] D. Handayani, H. Folmer, A. Kurnia, and K. A. Notodiputro, "The spatial empirical bayes predictor of the small area mean for a lognormal variable of interest and spatially correlated random effects," *Empir Econ*, vol. 55, pp. 147–167, 2018. https://doi.org/10.1007/s00181-018-1452-5

[11] G. Fauziah, A. Kurnia, and A. Djuraidah, "The empirical best linear unbiased prediction and the empirical best predictor unit-level approaches in estimating per capita expenditure at the subdistrict level," *Scientific Journal of Informatics*, vol. 12, no. 2, pp 283–294, 2025. https://doi.org/10.15294/sji.v12i2.25037

[12] W. A. Nurriza, "Penerapan Model Fay-Herriot pada Small Area Estimation Studi Simulasi Pengeluaran Per Kapita Level Kabupaten/Kota Provinsi Kalimantan Timur tahun 2020," *BESTARI: Buletin Statistika dan Aplikasi Terkini*, vol. 1, no. 1, pp. 29–35, 2021.

[13] A. H. Hakim and N. Hajarisman, "Pendugaan Rata-rata Pengeluaran Per Kapita Menurut Kabupaten/Kota di Provinsi Jawa Barat Melalui Empirical Best Linear Unbiased Prediction dalam Pendugaan Area Kecil," *Bandung Conferences Series: Statistics*, vol. 2, no. 2, pp. 474–481, 2022. https://doi.org/10.29313/bcss.v2i2.4747

[14] G. H. Zakiya, "Penerapan Model Fay-Herriot pada Small Area Estimation Studi Pengeluaran Per Kapita Level Kabupaten/Kota Provinsi Jawa Barat 2020," *Bandung Conferences Series: Statistics*, vol. 2, no. 2, pp. 343–350, 2022. https://doi.org/10.29313/bcss.v2i2.4323

[15] N. S. Belinda, K.A. Notodiputro, and A. M. Soleh, "BHF and copula models in small area estimation for household per capita expenditure in Bogor District," *Jurnal Natural*, vol. 24, no. 2, pp. 61–72, 2024. https://doi.org/10.24815/jn.v24i2.37278

[16] K. Ivan, I-H Holobaca, J. Benedek, and I. Torok, "VIIRS nighttime light data for income estimation at local level," *Remote Sensing*, vol. 12, no. 18, 2020. https://doi.org/10.3390/rs12182950

[17] S. P. Subash, R. R. Kumar, and K. S. Aditya, "Satellite data and machine learning tools for predicting poverty in rural India," *Agricultural Economic Research Review*, vol. 31, no. 2, pp. 231–240, 2018. https://doi.org/10.5958/0974-0279.2018.00040.X

[18] M. I. Habibie and N. Purwono, "Identification of socio-economic activities as urban growth based on nighttime light data (study on Kendal District-Indonesia)," *2022 IEEE Asia-Pacific Conference on Geoscience, Electronics, and Remote Sensing Technology (AGERS)*, 2022. [Online]. Available: https://doi.org/10.1109/AGERS56232.2022.10093456

[19] F. Afrianto, "East Java Province GRDP projection model using night-time light imagery," *East Java Economic Journal*, vol. 6, no. 2, pp. 208–223, 2022. https://doi.org/10.53572/ejavec.v6i1.83

[20] A. Surya, Indahwati, and Erfiani, "Study of small area estimation when nighttime lights as an auxiliary information is measured with error," *Indonesian Journal of Statistics and Its Applications*, vol. 8, no. 1, pp. 47–57, 2024. https://doi.org/10.29244/ijsa.v8i1p47-57

[21] P. A. Kaban, B. I. Nasution, R. E. Caraka, and R. Kurniawan, "Implementing night light data as auxiliary variable of small area estimation," *Communications in Statistics—Theory and Methods*, vol. 53, no. 1, pp. 310–327, 2024. https://doi.org/10.1080/03610926.2022.2077963

[22] M. Feriyanto, A. W. Wijayanto, I. Y. Wulansari, and N. B. Parwanto, “Small area estimation approaches using satellite imageries auxiliary data for estimating per capita expenditure in West Java, Indonesia,” *Jurnal Aplikasi Statistika & Komputasi Statistik*, vol. 16, no. 2, pp. 205–221, 2024. https://doi.org/10.34123/jurnalasks.v16i2.799
[23] E. Berg, H. Chandra, “Small area prediction for a unit-level lognormal model,” *Computational Statistics and Data Analysis*, vol 78, pp. 159–175, 2014. https://doi.org/10.1016/j.csda.2014.03.007
[24] C. D. Elvidge, M. Zhizhin, T. Ghosh, F-C Hsu, and J. Taneja, “Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019,” *Remote Sensing*, vol. 13, no. 5, 2021. https://doi.org/10.3390/rs13050922
[25] R. D. Aditya and A. Ubaidillah, “Application of small area estimation for global hunger index at regency/municipality level in Papua Island,” *Jurnal Matematika, Statistika, dan Komputasi*, vol. 22, no. 2, pp. 276–290, 2026. https://doi.org/10.20956/j.v22i2.46784
[26] B. D. Arianingsih, K. Sadik, and Indahwati, “Comparative performance of spatial robust small area estimation methods: a simulation study,” *Zero: Jurnal Sains, Matematika, dan Terapan*, vol. 9, no. 3, pp. 921–931, 2025. https://doi.org/10.30829/zero.v9i3.26477
[27] D. Handayani, K. A. Notodipuro, A. Saefuddin, I. W. Mangku, A. Kurnia, “Spatial empirical best predictor of small area poverty indicator,” *Int. J. Advance Soft Compu. Appl*, vol. 16, no. 2, pp. 103122, 2024. https://doi.org/10.15849/ijasca.240730.07