Enhancing Poverty Rates Reliability Using Small Area Estimation

N Permatasari[1*]

[1] BPS Statistics, Indonesia

*Corresponding author's email: noviaprmatasari@gmail.com

Abstract. This study systematically compares the performance of three Small Area Estimation (SAE) methods—Empirical Best Linear Unbiased Predictor (EBLUP), Hierarchical Bayes (HB) Beta, and HB Flexible Beta—using two different auxiliary data sources-Village Potential (*Podes*) and Socio-Economic Registration data (*Regsosek*). The SAE methodologies were applied in a case study focusing on Java Island, Indonesia. Direct estimates remain has high Relative Standard Errors (RSE) above 25%, indicating low reliability. EBLUP methods improved estimate reliability but still produced some unreliable estimates. The HB Beta method further reduced RSE values, while the HB Flexible Beta model achieved the lowest RSE, eliminating all unreliable estimates. Moreover, Socio-Economic Registration data consistently resulted in lower RSE values compared to Village Potential data, particularly when used with the HB Flexible Beta model. These result highlight that integrating advanced SAE models such as HB Flexible Beta with high-quality administrative data such as Socio-Economic Registration data is crucial for producing reliable and precise poverty estimates for more targeted and effective poverty alleviation policies.

Keyword: HB Beta, HB Flexible Beta, Poverty, Small Area Estimation

1. Introduction

Poverty remains a persistent challenge not only in developing countries but also in many developed nations, including Indonesia [1] [2]. Having precise, reliable, and consistent poverty data is becoming increasingly important for the success of any poverty reduction program [3] [4] [5]. Accurate macro poverty data are crucial for planning the poverty alleviation policies, determining the allocation of poverty alleviation programs, monitoring and evaluating these programs, and evaluating the performance of both central and regional governments [3]. Over the past three decades, the growing awareness of the importance of data has led to a significant increase in both the quantity and frequency of poverty data collection [5].

In formulating poverty alleviation policies through evidence-based decision-making, ensuring the reliability of the data is essential. One of statistical reliability measurement is the Relative Standard Error (RSE) of estimates. Australia Bureau of Statistics recommends that estimates with RSE between 25% and 50% should be utilized with caution, while the estimates with RSE higher than 50% classified as unreliable estimates and should be aggregated with other domain to decrease the error [6]. These criteria are adopted by BPS Statistics Indonesia, estimates whose RSE above 25% need for careful interpretation and consideration in their application (Badan Pusat Statistik, 2019).

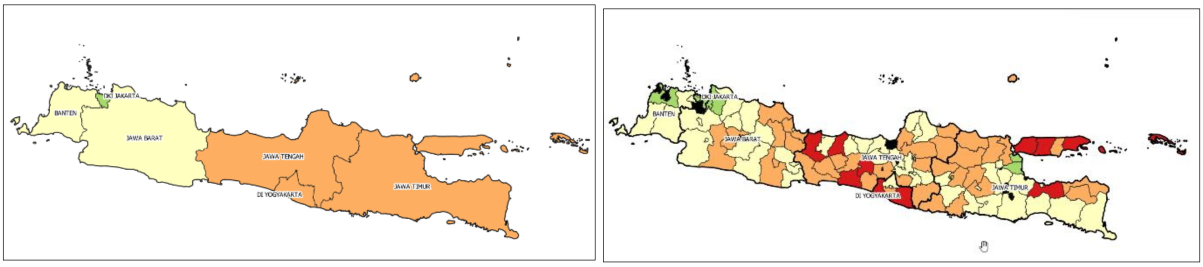

Figure 1. Comparison between poverty estimation in province level and regency/municipality level, in 2023

The poverty rate by province and regency/municipality levels in Java Island in 2023, as shown in figure 1, highlights the importance of poverty estimates for smaller areas. The estimates at the district level shows that several areas require increased attention due to their high poverty rates, including Sampang, Bangkalan, Sumenep, Probolinggo, Kebumen, Brebes, Kulon Progo, Gunung Kidul, Wonosobo, and Pemalang, with the percentage of people living in poverty exceeds 15%.

Table 1. Number of regency/municipality based on their RSE of 2023 Poverty Estimates.

No	Province	RSE < 25%	RSE >= 25%
1	DKI Jakarta	5	1
2	West Java	25	2
3	Central Java	32	3
4	DI Jogjakarta	5	0
5	East Java	34	4
6	Banten	5	3
	Java Island	106	13

Based on the 2023 poverty rates across 119 regencies/municipalities in Java Island, 13 of them have RSE greater than 25%. It indicates that these estimation results should be used with caution. Based on table 1, the regencies/municipalities with less reliable data are spread across DKI Jakarta, West Java, Central Java, East Java, and Banten. This is higher than in previous years, which only three regencies/municipalities had RSE above 25% in 2021 and six in 2022, distributed across East Java, West Java, and Banten.

The figure 2 shows the poverty rates and its confidence interval (CI). The upper and lower bounds of the interval are quite wide, which becoming wider at high poverty rate estimates. A wider CI indicates the uncertainty about the value of the parameter, which are the actual value of the poverty rate parameter could lie within the upper and lower bounds. For instance, in Sumenep District (3529), the true value of the poverty percentage could fall between 15% and 22%.

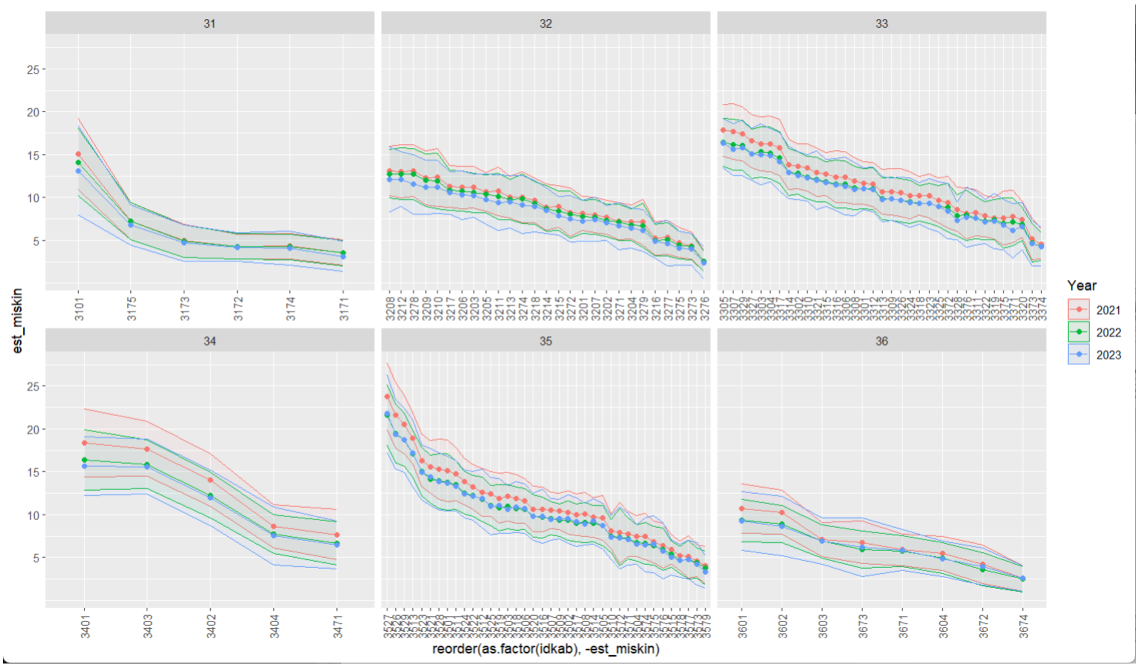

Figure 2. Poverty estimation and their 95% of confidence interval, in 2021-2023

The challenges of high RSE and wide CI complicate in comparing estimates across different time periods. Shown on the figure 2, poverty rate in 2021, 2022, and 2023 are mostly overlapping in confidence intervals (CIs). Although the point estimates show a decline in poverty, it was crucial to determine whether this is due to statistical noise (standard error) or a genuine reduction in poverty [7] .

The use of SAE to improve the reliability of poverty estimates has grown significantly in recent years. Several studies have demonstrated its application, including in Brazil [8], Palestine [9], Uganda [10], Tanzania and Sri Lanka [11], Mexico [12], Italy [13] [14], Germany [15], Spain [16], and Vietnam [17].

In Indonesia, research on the use of Small Area Estimates (SAE) for poverty estimates has increased. Several studies have utilized the Empirical Best Linear Unbiased Prediction (EBLUP) method to estimate poverty rates [18], [19], [20]. However, the effectiveness of the EBLUP method depends on the assumption of normally distributed sampling errors [21], [22], [23]. To address the normality issue, several alternative methods have been developed, including Geoadditive Models [24], Empirical Bayes [25], and Hierarchical Bayes [26], [27] [28]. Research indicates that Hierarchical Bayes models outperform direct estimates [26], EBLUP [27], and Empirical Bayes models [29].

Poverty rates, defined as the proportion of people whose expenditure below the poverty line to the total population, are commonly modelled using the beta distribution. [30] [27]. However, poverty data frequently has high skewness and presence of outlier, which the beta distribution model may not perform well. A promising advancement to address this limitation is the HB Flexible Beta model [31], which models data as a mixture of two Beta distributions and robust for skewness in poverty data. Yet, it remains lack of research in Indonesia.

Beyond methodological choices, the reliability of SAE is critically dependent on the quality and granularity of auxiliary data. The traditional small area estimation models require non-error auxiliary variables. Most SAE research in Indonesia uses Village Potential data as auxiliary variables [18] (Arisona & Pascasarjana, 2018) (Permatasari & Ubaidillah, n.d.) [33] [34] and decadal census data [35] [36]. In 2023, the Indonesian government conducted the Socio-Economic Registration (*Registrasi Sosial Ekonomi - Regsosek*), which collected socio-economic data for the entire population of Indonesia.

These data could be a powerful potential auxiliary variable for SAE because they do not have sampling errors and include household-level socio-economic data.

While previous studies have advanced the use of SAE in Indonesia, a systematic comparison that simultaneously evaluates the advance methodological and auxiliary data is lacking. Specifically, to prove the advantages of the HB Flexible Beta model over the widely used EBLUP and standard HB Beta approaches. Furthermore, the recent Socio-Economic Registration provides a powerful source of household-level auxiliary data without sampling error, yet its effectiveness compared to the commonly used Village Potential data remains unsystematically tested.

This research aims to fill these gaps by systematically comparing the reliability of poverty estimates produced by EBLUP, HB Beta, and HB Flexible Beta model in estimating poverty rates, across Java Island, Indonesia. We evaluate the performance of each model using both the Village Potential data and the Socio-Economic Registration data. Our analysis assesses the reliability of estimates based on their Relative Standard Error (RSE), with the goal of identifying the model-data combination that most effectively produces reliable estimates (RSE < 25%) for precise and effective policy targeting.

2. Data

2.1. Poverty Data

In Indonesia, poverty data is collected by Socio-Economic Survey (*Susenas*), which is conducted twice in a year. The Socio-Economic Survey conducted in March is designed to produce estimation at the regency/municipality level, while survey conducted in September is designed to result estimation at the province level.

A person is considered as poor if their expenditure falls below the poverty line. The poverty rate shown the number of people whose expenditure below the poverty line per total population. This paper studies the poverty rates in Java Island, in 2023.

Table 2 show the minimum, mean, and maximum poverty rates and their RSE for each province in Java Island. It shown that Banten province has lowest mean of district level poverty rates than other provinces. The highest poverty rate is 21.8% in one municipality in East Java Province. Although the mean of RSE is under 25%, however, there are several regencies/municipalities whose RSE more than 25%, which a maximum of 40%.

Table 2. Summary of Poverty rate and RSE of poverty rate in Java Island for each province

No	Province	Poverty Rate			RSE of Poverty Rate		
		Min	Mean	Max	Min	Mean	Max
1	[31] DKI Jakarta	3,10	6,00	13,13	17,72	22,43	28,92
2	[32] West Java	2,38	8,18	12,13	12,19	18,30	40,56
3	[33] Central Java	4,23	10,40	16,34	8,95	15,74	29,07
4	[34] DI Jogjakarta	6,49	11,44	15,64	10,48	16,07	22,70
5	[35] East Java	3,31	10,29	21,76	10,15	16,54	29,29
[redacted]	[36] Banten	2,57	6,05	9,27	18,65	23,38	29,41

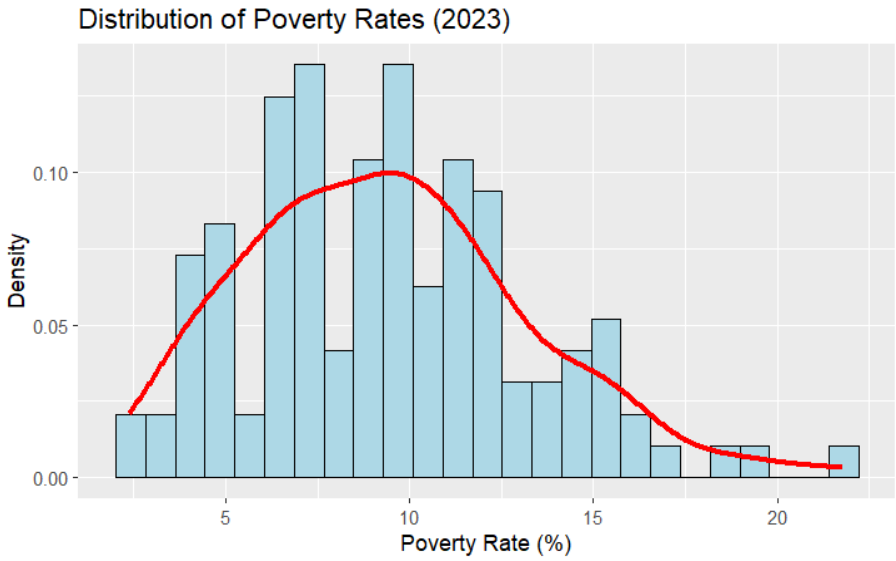

Figure 3. Distribution of Poverty Rates in Java Island, 2023.

Figure 3 illustrates the distribution of regency level poverty rates in Java Island, showing a right-skewed distribution with a few districts having poverty rates exceeding 15%. The Shapiro–Wilk normality test was conducted, and the result (W = 0.975, p = 0.028) confirms that the data do not follow a normal distribution.

Village Potential Data

This study utilizes Village Potential data (*Podes*), a comprehensive periodic census conducted by BPS-Statistics Indonesia that collects data on resources and infrastructure at the village level. The potential auxiliary variables consist of economic infrastructure, educational infrastructure, health infrastructure, and population conditions. To identify the most relevant and non-redundant predictors for our poverty model, we employed the Least Absolute Shrinkage and Selection Operator (LASSO) regression. From the initial set of indicators, LASSO selected the four most significant variables, detailed in table 3.

Table 3. Auxiliary variables from Village Potential Data

Variables	Village Potential Indicators
X_VP_1	Ratio of public schools to the number of villages
X_VP_2	Percentage of people with disabilities
X_VP_3	Percentage of people under restraint or institutionalization
X_VP_4	Ratio of modern markets to the number of villages

2.2. Socio-Economy Registration Data

In 2023, BPS-Statistics Indonesia under the coordination of Ministry of National Development Planning (*Bappenas*) conducted Socio-Economy Registration (*Registrasi Sosial Ekonomi / Regsosek*). This census is designed to collect detailed data of socio-economic conditions of all household across Indonesia, such as: socio-economic and demographic conditions, housing conditions, sanitation and clean water conditions, asset ownership, vulnerability conditions of specific population groups, population information, elderly (senior citizens), persons with disabilities, and employment. As a full coverage population census, Socio-Economy Registration Data has no sampling error and provides more informative data than village potential data for auxiliary variables. By utilizing LASSO variable selection, we have selected the six significant variables, detailed in table 4.

Table 4. Auxiliary Variables from Socio-Economy Registration Data

Variables	Village Potential Data
X_SER_1	The percentage of households with rent-free houses.
X_SER _2	The percentage of households with bamboo houses.
X_SER _3	The percentage of households with uncovered embankments.
X_SER _4	The percentage of the population receiving direct cash assistance (*BLT*).
X_SER _5	The percentage of households without internet access.
X_SER_6	The percentage of the population without elementary school certificates.

3. Methodology

In this study we compare three SAE models, which are EBLUP and two Bayesian approaches. The EBLUP is a most common SAE model and well-established best linear predictor. The HB Beta model is specifically designed for proportional data like poverty rates. The HB Flexible Beta further extends this by modelling the data as a mixture of two Beta distributions, making it more robust to skewness and extreme values, which are common in poverty data.

2.1. SAE EBLUP

SAE using Empirical Best Linear Unbiased Prediction (EBLUP), known as The Fay Herriot Model, is commonly used as a benchmark in any SAE method studies. The Fay-Herriot model[37] is a foundational and widely used SAE model due to its computational simplicity. The Fay-Heriot model is as follows:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e} \tag{1}$$

$$\boldsymbol{u} \sim iid\ (0, \boldsymbol{G}), \qquad \boldsymbol{e} \sim iid\ (0, \boldsymbol{R})$$

Where $\boldsymbol{G} = \boldsymbol{I}_m \sigma_u^2$and $\boldsymbol{R} = \boldsymbol{I}_m \sigma_e^2$, where $\boldsymbol{I}_m$ It is a matrix identity. Matrix covariance from $\boldsymbol{y}$ stated with $\boldsymbol{\Omega} = \mathbf{ZGZ}^{\mathrm{T}} + \boldsymbol{R}$

The most common approach of $\hat{\mu}$ is BLUP proposed by Henderson (1953), and EBLUP if the variance component is not known and must be carried out. The formula of EBLUP is as follows:

$$\hat{\mu}_i^H(\hat{\sigma}_u^2) = \boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}} + \frac{\hat{\sigma}_u^2 b_i^2}{\hat{\sigma}_u^2 b_i^2 + \sigma_e^2}\left(y_i - \boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}\right) = \hat{\gamma}_i y_i + (1-\hat{\gamma}_i)\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}, \qquad i = 1, \dots, m \tag{2}$$

$$\hat{\gamma}_i = \frac{\hat{\sigma}_u^2 b_i^2}{\hat{\sigma}_u^2 b_i^2 + \sigma_e^2} \tag{3}$$

The shrinkage factor, $\hat{\gamma}_i$ determines the weight of direct estimate y_i compared to the regression-synthetic estimate $\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}$. While robust and popular, the EBLUP relies on asymptotic normality assumptions for its inference, which may be suboptimal for proportional data like poverty rates. It serves as our baseline linear benchmark.

2.2. SAE HB Beta

Beta distribution is the natural choice for modelling proportion data bounded between 0 and 1. The HB Beta models potentially offering superior performance over the linear EBLUP model, since it ensures all predictions adhere to the [0, 1] interval, which the EBLUP does not guarantee.

The Beta distribution random variable $Y \sim Beta(\mu\phi, (1-\mu)\phi)$ has probability density function:

$$f_B(y;\mu,\phi) = \frac{\Gamma[\phi]}{\Gamma[\mu\phi]\,\Gamma[(1-\mu)\phi]} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} \quad (4)$$

With $0 < y < 1$, $0 < \mu < 1$, and $\phi > 0$.

In Beta small area model for direct estimates y_d of domain d, the direct estimator's conditional distribution is represented as below:

$$y_d |\, \theta_d, \phi_d \sim Beta(\theta_d \phi_d, (1-\theta_d)\theta_d, \phi_d) \quad (5)$$

The target parameter $E(y_d|\, \theta_d, \phi_d) = \theta_d$ is estimated using logit regression:

$$logit(\theta_d |\, \beta, v_d) = x_d^T \beta + v_d \quad (6)$$

Where $v_d | \sigma_e^2 \sim N(0, \sigma_e^2)$ is the area random effect

2.3. SAE HB Beta Flex

While the standard Beta model is an improvement, it may not optimally handle the complexities of poverty data, such as high skewness or outliers. The HB Flexible Beta model (Nicolo, Ferrante, & Pacei, 2024) handle the issues by modelling the data as a mixture of two Beta distributions. The Flexible Beta distribution combine two Beta random variables with different location but one dispersion parameter, with the probability density function:

$$f_{FB}(\lambda_1, \lambda_2. \phi, p) = p \,.\, f_B(y; \lambda_1, \phi) + (1-p).\, p \,.\, f_B(y; \lambda_2, \phi) \quad (7)$$

The direct estimator's of Flexible Beta small area model has conditional distribution represented as below:

$$y_d |\, \lambda_{1d}, \lambda_{2d}. \phi_d, p \sim FB\ (\lambda_{1d}, \lambda_{2d}. \phi_d, p) \quad (8)$$

Implementation and Model Evaluation

All models were implemented in the R statistical environment [38]. The EBLUP models were fitted using the emdi package [39], while the Bayesian models (HB Beta and HB Flexible Beta) were implemented using the tipsae package [40], which is specifically designed for small area estimation of proportions and provides a robust framework for fitting complex Beta models. The reliability of the estimates from all models—direct, EBLUP, HB Beta, and HB Flexible Beta—was evaluated using the Relative Standard Error (RSE). The RSE is calculated as the ratio of the standard error of the estimate to the estimate itself, expressed as a percentage:

$$RSE = \frac{se(\hat{y})}{\hat{y}} * 100\% \quad (9)$$

The RSE is considered reliable if RSE < 25%, in line with Australian Bureau of Statistics and BPS Statistics Indonesia treshold.

4. Result

4.1 Model Coefficients and Interpretation

Table 5-10 present the coefficients for the six models: EBLUP, Hierarchical Bayesian (HB) Beta, and HB Flexible Beta model, each applied for both the Village Potential and Socio-Economic Registration data. For the EBLUP models, coefficients are presented with their standard errors, t-values, and p-values to assess statistical significance. For the Bayesian models (HB Beta and HB Flexible Beta), the posterior mean of the coefficients is reported along with the 95% Credible Interval (CrI). A coefficient is considered significant if its 95% CrI does not contain zero.

Table 5. Model Coefficient of EBLUP using Village Potential Data

Variables	coefficients	std.error	t.value	p.value
(Intercept)	6.64630	1.09700	6.0586	1.373e-09
X_VP _1	0.50543	0.25052	2.0175	0.0436457
X_VP _2	6.90464	2.20627	3.1296	0.0017507
X_VP _3	1405.49623	423.52276	3.3186	0.0009047
X_VP _4	-1.10795	0.27669	-4.0043	6.220e-05

Table 6. Model Coefficient of EBLUP using Socio-Economic Registration Data

Variables	coefficients	std.error	t.value	p.value
(Intercept)	-2.503259	1.382452	-1.8107	0.070181
X_SER _1	0.094720	0.037754	2.5089	0.012112
X_SER _2	-0.548898	0.186468	-2.9437	0.003244
X_SER _3	0.691071	0.337442	2.0480	0.040563
X_SER _4	0.111003	0.058503	1.8974	0.057776
X_SER _5	0.224332	0.031648	7.0883	1.358e-12
X_SER _6	0.204785	0.084598	2.4207	0.015492

Table 7. Model Coefficient of SAE HB Beta using Village Potential Data

Variables	mean	std.error	2.5% CI	97.5% CI
(Intercept)	-2.363	0.029	-2.420	-2.307
X_VP _1	0.120	0.049	0.024	0.217
X_VP _2	0.138	0.044	0.052	0.224
X_VP _3	0.087	0.031	0.026	0.148
X_VP _4	-0.150	0.036	-0.220	-0.079

Table 8. Model Coefficient of SAE HB Beta using Socio-Economic Registration Data

Variables	mean	std.error	2.5% CI	97.5% CI
(Intercept)	-2.369	0.027	-2.422	-2.316
X_SER _1	0.089	0.030	0.031	0.148
X_SER _2	-0.070	0.032	-0.134	-0.010
X_SER _3	0.059	0.031	-0.004	0.120
X_SER _4	0.106	0.039	0.028	0.182
X_SER _5	0.261	0.036	0.191	0.331
X_SER _6	0.065	0.034	-0.003	0.132

Table 9. Model Coefficient of SAE HB Flexible Beta using Village Potential Data

Variables	mean	std.error	2.5% CI	97.5% CI
(Intercept)	-3.305	0.206	-3.715	-2.893
X_VP _1	0.231	0.101	0.045	0.443
X_VP _2	0.208	0.086	0.046	0.385
X_VP _3	0.145	0.058	0.036	0.265
X_VP _4	-0.325	0.087	-0.512	-0.171

Table 10. Model Coefficient of SAE HB Flexible Beta using Socio-Economic Registration Data

Variables	mean	std.error	2.5% CI	97.5% CI
(Intercept)	-3.448	0.201	-3.861	-3.067
X_SER _1	0.146	0.070	0.015	0.288
X_SER _2	-0.143	0.073	-0.307	-0.023
X_SER _3	0.118	0.059	0.007	0.238
X_SER _4	0.270	0.090	0.110	0.461
X_SER _5	0.506	0.094	0.339	0.706
X_SER _6	0.088	0.064	-0.037	0.212

Generally, the significance and direction of the auxiliary variables were consistent across all three models. In the EBLUP model, most variables were statistically significant at the 5% level (p-value < 0.05), confirming they are informative auxiliary variables for estimating poverty rates. This finding aligns with the results from the HB Beta and HB Flexible Beta models, where the 95% Credible Intervals for the majority of coefficients do not contain zero.

For the Village Potential data, all three models show significant positive coefficients for variables such as the ratio of public schools, the percentage of people with disabilities, and the percentage of people under restraint or institutionalization. However, the ratio of modern markets has a significant negative coefficient. For the Socio-Economic Registration data, the EBLUP, Beta, and Flexible Beta models also show significant positive coefficients for variables, including the percentage of households with rent-free houses, the percentage with uncovered embankments, the percentage of the population receiving direct cash assistance (*BLT*), the percentage of households without internet access, and the percentage of the population without elementary school certificates. Conversely, the percentage of households with bamboo houses shows a significant negative value.

4.2 Model Performance: Area-Level Variance

The performance of the auxiliary variables on a SAE model is assessed by examining the estimated area-level variance (σ_u^2) and the shrinkage factor (γ), which indicates the balance between the direct estimates and the regression-synthetic estimates.

Table 11. Area Level Variance of EBLUP Model

Model	Area Level Variance
EBLUP – Village Potential Data	3.311274
EBLUP - Socio-Economic Registration	2.602953

ICDSOS
The 3rd International Conference on Data Science and Official Statistics
November 27 - 28, 2025

Table 12. Shrinkage Factor of EBLUP Model

Model	Min	Mean	Median	Max
EBLUP – Village Potential Data	0.3239	0.6152	0.6121	0.8529
EBLUP - Socio-Economic Registration	0.2736	0.5591	0.5537	0.8200

Table 11 and table 12 show the comparison of estimated area-level variance (σ_u^2) and the shrinkage factor between models utilizing Village Potential Data and Socio-Economic Registration data. The area-level variance for the EBLUP model using Village Potential data is higher than that using Socio-Economic Registration data. This smaller variance in the Socio-Economic Registration model indicates a better fit, as the auxiliary data more effectively captures variation across areas compared to the Village Potential model. The slightly smaller shrinkage factor value for the EBLUP model with Socio-Economic Registration data suggests it relies more on synthetic model predictions than the Village Potential model, which consistent with lower the estimated variance.

Table 13. Area Level Variance of Beta and Flexible Beta Model

Model	Mean	SD	2.5%	97.5%
Beta – Village Potential Data	0.2398987	0.02483402	0.1941111	0.2919382
Flexible Beta - Village Potential Data	0.4298035	0.07438405	0.3000637	0.5918306
Beta – Socio-Economic Registration	0.2176898	0.02490098	0.1719214	0.2699082
Flexible Beta - Socio-Economic Registration	0.3970800	0.06901481	0.2774635	0.5467915

For the Bayesian models, shown at table 13, Flexible Beta models consistently produce higher mean of area-level variance compared to their Beta model. Comparing the data sources, models using Socio-Economic Registration data demonstrate lower area level variance, which is consistent with the EBLUP model. Overall, the three models show that Socio-Economic Registration data perform better due to its smaller area-level variance, which makes the model rely more on auxiliary variables, reflecting the richness and greater informativeness of Socio-Economic Registration data.

4.3 SAE Estimates

In this section, we assess the reliability of the estimates by calculating the Relative Standard Error (RSE) and comparing the credible interval range of the estimates. Estimates with an RSE below 25% were considered sufficiently reliable for analysis, indicating a lower likelihood of significant sampling error.

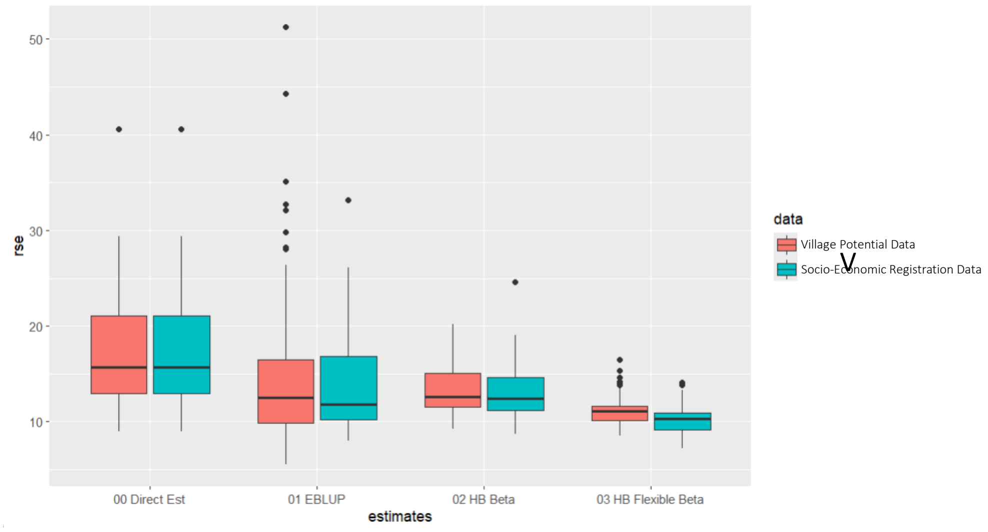

Figure 4. Comparison of RSE between direct estimates and SAE Model

The boxplot on figure 4 visualizes the distribution of Relative Standard Errors (RSE) for four different estimation methods on both Village Potential and Socio-Economic Registration data. The RSE of direct estimation remains high, with some observations with RSE more than 25%, indicating unreliable estimates. The EBLUP significantly reduce the RSE than the direct estimates, though the observation with RSE above 25% still exist. The HB Beta model further reduces the RSE values for both Village Potential and Socio-Economic Registration data, which there is no observation with RSE more than 25%. The HB Flexible Beta model shows the lowest RSE values for both data sources. A comparison of the data sources shows that the utilization of Socio-Economic Registration data consistently getting lower RSE than Village Potential data in EBLUP, HB Beta, and HB Flexible Beta model.

Table 14. Comparison of RSE between direct estimates and SAE Models

Province	Statistics	RSE			
		Direct estimates	EBLUP	HB Beta	HB Flexible Beta
[31] DKI Jakarta	Village Potential	22,43	27,93	15,21	10,83
	Socio-Economic Registration	22,43	19,57	16,41	8,49
[32] West Java	Village Potential	18,30	18,12	13,87	11,37
	Socio-Economic Registration	18,30	14,62	13,34	10,39
[33] Central Java	Village Potential	15,74	12,76	12,62	11,14
	Socio-Economic Registration	15,74	12,06	12,10	10,15
[34] DI Jogjakarta	Village Potential	16,07	12,83	15,22	10,68
	Socio-Economic Registration	16,07	13,31	14,13	10,06
[35] East Java	Village Potential	16,54	11,99	12,76	10,81

	Socio-Economic Registration	16,54	12,60	12,41	10,26
[36] Banten	Village Potential	23,38	15,00	14,92	11,70
	Socio-Economic Registration	23,38	18,97	15,58	9,31
Java Island	Village Potential	17,44	14.65	13.35	11.09
	Socio-Economic Registration	17.44	13.71	13.02	10.10

Table 14 presents province-specific RSE values across Java Island, show a significant improvement from the use of HB methods. While all SAE methods generally outperform direct estimation, the results in DKI Jakarta are an exception, where the EBLUP method produces a higher RSE than the direct estimate, though both HB methods is lower.

In EBLUP and HB Beta, there is no significant differences of RSE between Village Potential data and Socio-Economic Registration data. However, in HB Flexible Beta method, the estimates in all provinces shows that Socio-Economic Registration outperform the Village Potential data to produce more reliable estimates.

These province-level results are consistent with the island-wide result. The HB Flexible Beta model produce the most reliable estimates compared to the direct estimates, EBLUP, and HB Beta, shown by the significant reduction in the RSE values. In Flexible Beta model, the Socio-Economic Registration data also consistently shows more reliable estimates than Village Potential data. Overall, the HB Flexible Beta using Socio-Economic Registration data is the best model with the highest RSE lower than 15%.

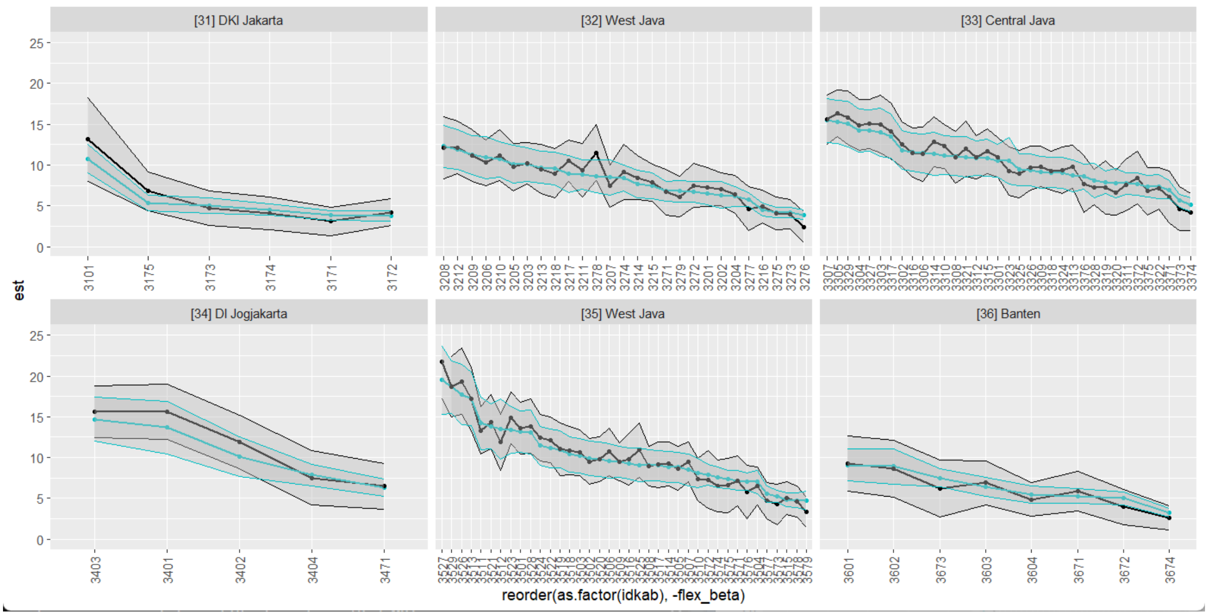

Figure 5. Comparison between direct estimates and SAE Model, as well as their confidence interval

Finally, figure 5 provides visual proof of the enhanced reliability. The 95% credible intervals for the HB Flexible Beta estimates using Socio-Economic Registration are narrower than the direct estimates across all districts. This reduction in uncertainty is critical for policymakers, as it allows for confident distinction between true changes in poverty over time and mere statistical noise.

5. Discussion

5.1 Main Findings and Comparison with Previous Studies

The performance of SAE methods—from direct estimation to EBLUP to HB Beta, and to HB Flexible Beta—highlights an important lesson that model choice must be driven by the nature of the data. Poverty rates are proportional and often skewed, violating the normality assumptions that crucial for the traditional EBLUP approach. While EBLUP offered a significant improvement over direct estimation, its failure to eliminate unreliable (RSE > 25%) estimates limits its utility for precise policy targeting. The better performance of the Bayesian methods, especially the HB Flexible Beta model, shows that considering the proportional and skewed nature of poverty data is essential to get reliable estimates.

Our results also align with the existing body of literature. The finding that Hierarchical Bayes models outperform direct estimates is consistent with global studies [26] and previous work in Indonesia [27], [29]. Our results demonstrating that the Flexible Beta offers an improvement over the standard HB Beta model, validating its theoretical advantages for handling skewed data (Nicolò et al., 2024) in a real-world policy context.

The consistent advantage of Socio-Economic Registration data over the Village Potential in all models highlights important implications for Small Area Estimation (SAE) practitioners. The lower area-level variance and RSE values indicate that Socio-Economic Registration data are more effective at explaining the social and economic variation between districts than the village-level infrastructure in Village Potential data. The model using Socio-Economic Registration data results in greater precision and reliability on the estimates.

The primary output of this research recommend the HB Flexible Beta model with Socio-Economic registration data should be adopted for producing sub-national poverty estimates. This combination produces reliable estimates, enabling confident and precise policy actions in targeting, monitoring, and evaluating poverty programs.

5.2 Limitations and Future Research

This study has limitations that point toward valuable future research. First, our analysis was limited to Java Island. While Java contains a diverse range of economic conditions, future work should validate these models across other Indonesian area to ensure their national applicability.

Second, the current models do not incorporate spatial dependency, yet poverty in one district may be influenced by conditions of their neighbours. Future research could integrate spatial effects into the HB Flexible Beta model, potentially unlocking further gains in precision. For example, the R package tip_sae provides a function to include spatially structured random effects by supplying spatial data frames to the model.

Future research also could explore utilizing non-traditional data sources, such as satellite imagery, social media data, and mobile phone data, to further enhance the timeliness and granularity of poverty measurement. Additionally, the utilization of the National Single Socio-Economic Data (*Data Tunggal Sosial Ekonomi Nasional* / DTSEN), which provides comprehensive and annually updated socio-economic data, is a high potential auxiliary data source. Combining these diverse and rich auxiliary data sources within a SAE model is a new potential for creating detailed, timely, and precise poverty estimates that can better inform targeted social and economic policies.

6. Conclusion

Our findings demonstrate the varying reliability of poverty rate estimates across different Small Area Estimation (SAE) methods and data sources. Direct estimates have RSE above 25% indicating the unreliability. The Empirical Best Linear Unbiased Predictor (EBLUP) model shows improved reliability but still includes observations with RSE over 25%. The Hierarchical Bayes (HB) Beta method further reduced RSE values, and the HB Flexible Beta method produced the lowest RSE values, showing the highest precision and reliability. Additionally, Socio-Economic Registration data consistently shows lower RSE values compared to Village Potential data for all models. These results highlight the importance of utilizing advanced estimation techniques and up-to-date, detailed data sources to achieve reliable poverty measurement.

References

[1] S. N. Fathia, R. Majidah, A. D. Octary, and N. Nurdiono, "Democracy, Political Stability, and Level of Corruption to Cure Poverty in Asia," in *International Conference of Economics, Business, and Entrepreneur 2024 T*, European Alliance for Innovation n.o., Dec. 2023. doi: 10.4108/eai.13-9-2023.2341203.

[2] L. Sugiharti and M. R. Primanthi, "The Determinants of Poverty: Case of Indonesia," *Global Journal of Business and Social Science Review*, vol. 5, no. 3, pp. 58–68, 2017.

[3] A. Adji, T. Hidayat, H. Tuhiman, S. Kurniawati, and A. Maulana, "Measurement of Poverty Line in Indonesia: Theoretical Review and Proposed Improvements," 2020. [Online]. Available: www.tnp2k.go.id

[4] B. Cahyadie, B. Juanda, A. Fauzi, and R. A. Kinseng, "Distribution Analysis and Ranking Analysis of Poverty Data From Three Data Sources in Bekasi Regency," *Jurnal Bina Praja*, vol. 15, no. 3, pp. 453–466, Dec. 2023, doi: 10.21787/jbp.15.2023.453-466.

[5] S. Alkire, "Towards Frequent and Accurate Poverty Data," 2016. [Online]. Available: http://www.ophi.org.uk/multidimensional-poverty-index/mpi-2014/

[6] Australian Bureau of Statistics, " Reliability of Estimates - Survey of Income and Housing, User Guide, Australia." Accessed: Jul. 07, 2024. [Online]. Available: https://www.abs.gov.au/statistics/detailed-methodology-information/concepts-sources-methods/survey-income-and-housing-user-guide-australia/2019-20/reliability-estimates#cite-window2

[7] I. P. David, "POVERTY STATISTICS AND INDICATORS: HOW OFTEN SHOULD THEY BE MEASURED?," 2000.

[8] S. Lange, U. J. Pape, and P. Pütz, "Small Area Estimation of Poverty Under Structural Change," *Review of Income and Wealth*, vol. 68, no. S2, 2022, doi: 10.1111/roiw.12558.

[9] A. Bikauskaite, I. Molina, and D. Morales, "Multivariate mixture model for small area estimation of poverty indicators," *J R Stat Soc Ser A Stat Soc*, vol. 185, no. S2, 2022, doi: 10.1111/rssa.12965.

[10] J. Farris, C. Larochelle, J. Alwang, G. W. Norton, and C. King, "Poverty analysis using small area estimation: an application to conservation agriculture in Uganda," *Agricultural Economics (United Kingdom)*, vol. 48, no. 6, 2017, doi: 10.1111/agec.12365.

[11] T. Masaki, D. Newhouse, A. R. Silwal, A. Bedada, and R. Engstrom, "Small area estimation of non-monetary poverty with geospatial data," *Stat J IAOS*, vol. 38, no. 3, 2022, doi: 10.3233/SJI-210902.

[12] D. L. Newhouse, J. D. Merfeld, A. Ramakrishnan, T. Swartz, and P. Lahiri, "Small Area Estimation of Monetary Poverty in Mexico Using Satellite Imagery and Machine Learning," *SSRN Electronic Journal*, 2022, doi: 10.2139/ssrn.4235976.

[13] A. Gardini, E. Fabrizi, and C. Trivisano, "Poverty and inequality mapping based on a unit-level log-normal mixture model," *J R Stat Soc Ser A Stat Soc*, vol. 185, no. 4, pp. 2073–2096, Oct. 2022, doi: 10.1111/rssa.12872.

[14] G. Tonutti, G. Bertarelli, C. Giusti, and M. Pratesi, "Disaggregation of poverty indicators by small area methods for assessing the targeting of the 'Reddito di Cittadinanza' national policy in Italy," *Socioecon Plann Sci*, vol. 82, Aug. 2022, doi: 10.1016/j.seps.2022.101327.

[15] I. Benedetti and F. Crescenzi, "The role of income poverty and inequality indicators at regional level: An evaluation for Italy and Germany," *Socioecon Plann Sci*, vol. 87, Jun. 2023, doi: 10.1016/j.seps.2023.101540.

[16] L. Marcis, D. Morales, M. C. Pagliarella, and R. Salvatore, "Three-fold Fay–Herriot model for small area estimation and its diagnostics," Dec. 01, 2023, *Institute for Ionics*. doi: 10.1007/s10260-023-00700-6.

[17] C. V. Nguyen, "Combining samples in small area estimation: An application to poverty mapping in Vietnam," *Economics of Transition and Institutional Change*, 2024, doi: 10.1111/ecot.12413.

[18] A. Syafira and N. Hajarisman, "Penerapan Penduga Area Kecil Melalui Metode Empirical Best Linear Unbiased Prediction (EBLUP) untuk Estimasi Persentase Penduduk Miskin Level Kabupaten/Kota di Provinsi Jawa Barat," *Bandung Conference Series: Statistics*, vol. 2, no. 2, pp. 401–408, Aug. 2022, doi: 10.29313/bcss.v2i2.4507.

[19] D. Gartina, L. Khikmah, and A. Statistika Muhammadiyah Semarang, "PENDUGAAN KEMISKINAN MENGGUNAKAN SMALL AREA ESTIMATION DENGAN PENDEKATAN EMPERICAL BEST LINIER UNBIASED PREDICTION (EBLUP)," *Statistika*, vol. 8, no. 2, 2020.

[20] Y. Ramadhan and A. Ubaidillah, "Pendugaan Persentase Penduduk Miskin Ekstrem Menggunakan Small Area Estimationdengan Partitioning Around Medoids Clustering," in *Seminar Nasional Official Statistics 2023*, 2023.

[21] A. Salma, K. Sadik, and K. A. Notodiputro, "Small area estimation of per capita expenditures using robust empirical best linear unbiased prediction (REBLUP)," in *AIP Conference Proceedings*, American Institute of Physics Inc., Mar. 2017. doi: 10.1063/1.4979443.

[22] S. K. Sinha and J. N. K. Rao, "Robust Small Area Estomation," *Canadian Journal of Statistics*, p. n/a-n/a, Sep. 2014, doi: 10.1002/cjs.

[23] Titin Yuniarty, Indahwati, and Aji Hamim Wigena, "A Comparison Small Area Estimation for Skewed Data with EBLUP and Hierarchical Bayes Approaching using Rao-Yu Model," *Int J Sci Res Sci Eng Technol*, pp. 132–143, Nov. 2023, doi: 10.32628/ijsrset231063.

[24] N. H. Pusponegoro, A. Djuraidah, A. Fitrianto, and I. M. Sumertajaya, "Geo-additive Models in Small Area Estimation of Poverty," *Journal of Data Science and Its Applications*, vol. 2, no. 1, 2019, doi: 10.21108/jdsa.2019.2.15.

[25] N. A. Yensy, "The Small Area Estimation by Using Empirical Bayes Method," 2021.

[26] I. Molina, B. Nandram, and J. N. K. Rao, "Small area estimation of general parameters with application to poverty indicators: A hierarchical bayes approach," *Annals of Applied Statistics*, vol. 8, no. 2, pp. 852–885, 2014, doi: 10.1214/13-AOAS702.

[27] W. Larasati and N. Permatasari, "Perbandingan Metode SAE EBLUP dan SAE HB Pada Pendugaan Area Kecil (Studi Kasus Pendugaan Kemiskinan di Provinsi Jawa Timur)," *Jurnal Statistika dan Aplikasinya*, vol. 6, no. 1, pp. 96–108, 2022.

[28] E. Budiatmodjo, A. T. Rumiati, and D. D. Prastyo, "Benchmarking hierarchical Bayesian small area estimators in the percentage of poverty at sub-districts level in Central Java," in *THE 3RD INTERNATIONAL CONFERENCE ON SCIENCE, MATHEMATICS, ENVIRONMENT, AND EDUCATION: Flexibility in Research and Innovation on Science, Mathematics, Environment, and education for sustainable development*, AIP Publishing, Jan. 2023, p. 080026. doi: 10.1063/5.0106437.

[29] W. Kusuma and N. Iriawan, "Small Area Estimation of Expenditure Per-capita in Banyuwangi with Hierarchical Bayesian and Empirical Bayes Methods," 2017.

[30] R. Janicki, "Properties of the beta regression model for small area estimation of proportions and application to estimation of poverty rates," *Commun Stat Theory Methods*, vol. 49, no. 9, pp. 2264–2284, May 2020, doi: 10.1080/03610926.2019.1570266.

[31] S. De Nicolò, M. R. Ferrante, and S. Pacei, "Small Area Estimation of Inequality Measures using Mixtures of Beta," Sep. 2022, doi: 10.1093/jrsssa/qnad083.

[32] D. C. Arisona and S. Pascasarjana, "KAJIAN PENDUGAAN AREA KECIL PADA DATA OVERDISPERSI MENGGUNAKAN REGRESI ZERO-INFLATED POISSON," 2018.

[33] A. Ubaidillah, K. A. Notodiputro, A. Kurnia, and I. W. Mangku, "Multivariate Fay-Herriot models for small area estimation with application to household consumption per capita expenditure in Indonesia," *J Appl Stat*, vol. 46, no. 15, pp. 2845–2861, Nov. 2019, doi: 10.1080/02664763.2019.1615420.

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

[35] Q. Makrufa Ayesha Tantriana and Q. Makrufa, "IARIW-World Bank Analysis of Poverty in Village Level of Indonesia with Small Area Estimation: Case in Bangkalan District Analysis of Poverty in Village Level of Indonesia with Small Area Estimation: Case in Bangkalan District."

[36] D. Widyawati and S. Muchlisoh, "Seminar Nasional Official Statistics 2020:Statistics in the New Normal : a Challenge of Big Data and Official Statistics PENERAPAN METODE ELL-COUNTERFACTUAL UNTUK PEMETAAN KEMISKINAN LEVEL KECAMATAN DAN DESA/KELURAHAN Studi Kasus di Kota Yogyakarta Tahun 2018 (The Application of ELL-Counterfactual Method for Poverty Mapping at Sub-District and Village Level: Case Study in Kota Yogyakarta 2018)."

[37] R. E. Fay and R. A. Herriot, "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data," *J Am Stat Assoc*, vol. 74, no. 366a, pp. 269–277, Jun. 1979, doi: 10.1080/01621459.1979.10482505.

[38] R Core Team, "R: A language and environment for statistical ## computing," 2021, *R Foundation for Statistical Computing, Vienna, Austria*.

[39] A.-K. Kreutzmann, S. Pannier, N. Rojas-Perilla, T. Schmid, M. Templ, and N. Tzavidis, "The R Package emd for Estimating and Mapping Regionally Disaggregated Indicators," *J Stat Softw*, vol. 91, no. 7, 2019, doi: 10.18637/jss.v091.i07.

[40] S. De Nicolò and A. Gardini, "The R Package tipsae: Tools for Mapping Proportions and Indicators on the Unit Interval," *J Stat Softw*, vol. 108, no. 1, 2024, doi: 10.18637/jss.v108.i01.