HIERARCHICAL BAYES APPLICATION FOR SMALL AREA ESTIMATION WITH ERROR MEASUREMENT ON CHILD POVERTY IN SUMATERA ISLAND

Aisha Arthamevia[1)]
Zahra Rizky Fadilah[2)]
Rasya Az Zahra[3)]
Sausan Salsabila[4)]
Renandika Salsabiila Agistasari[5)]
Nofita Istiana[6)]
[1,2,3,4,5,6)] Politeknik Statistika STIS, Jakarta, Indonesia
e-mail: nofita@stis.ac.id

ABSTRACT

Child poverty on Sumatera Island remains a significant issue, as four provinces recorded child poverty rates above the national average in 2021, increasing to five provinces in 2022. To support more effective and targeted policies, reliable estimates at the district/city level are required; however, direct estimates from the March 2023 *Susenas* data showed low precision, with 37 of 154 districts/cities having a Relative Standard Error (RSE) greater than 25%. To improve accuracy, this study applied Small Area Estimation using a Hierarchical Bayes model with Measurement Error on the Beta distribution (SAE HB ME Beta). Empirical findings revealed serious precision problems, particularly in *Kepulauan* Bangka Belitung, where all districts had RSE values above 25%, and in West Sumatera, which ranked second with more districts exceeding the threshold than those below it, including the highest overall RSE. When the model was initially estimated jointly for all provinces, one district in West Sumatera still had an RSE above 25% and estimates for *Kepulauan* Bangka Belitung failed to satisfy the internal consistency criterion. To address this heterogeneity, the model was re-estimated separately for West Sumatera and *Kepulauan* Bangka Belitung and for the remaining provinces. The final results show that all districts/cities achieved RSE ≤ 25% and met internal consistency requirements, indicating that the proposed approach improves the precision and reliability of district-level child poverty estimates across Sumatera Island.

Keywords: child poverty, small area estimation, SAE HB ME Beta.

INTRODUCTION

Poverty is one of the most important problems in developing countries like Indonesia. Poverty is defined as a person's economic inability to meet their needs, both food and non-food, which is measured in terms of expenditure (BPS, 2024). Poverty is a measure of social and economic conditions in society as well as the success of development carried out by the government (Priseptian & Primandhana, 2022). Poverty eradication is a top priority in the Sustainable Development Goals (SDGs) program, namely achieving zero poverty.

The problem of poverty violates human rights. If poverty occurs in the family environment, children will also be vulnerable to problems of hunger, health, and so on (Darjono et al., 2019). Child poverty needs special attention because more than half of children in developing countries grow up in poverty. Poverty in children can make it difficult for them to survive and grow (Kertayana, 2017). In the future, children living in poverty will reduce the quality of human resources because they will have difficulty competing for jobs, so they get low wages and remain in poverty. This causes a cycle of poverty in society ((Salis & Ubaidillah, 2023). Therefore, the government is trying to overcome the problem of child poverty by creating one indicator, namely "the percentage of children aged 0-17 years living below the poverty line" which can later measure the success of policies made by the government (Riany et al., 2022).

In 2021, there were four provinces in Sumatera Island that had a percentage of children aged 0-17 years living below the poverty line above the national value. However, in 2022, there were five additional provinces that were above the national value, namely Aceh, Bengkulu, Sumatera Selatan, Lampung, and Sumatera Utara. This indicates that there is a problem of child poverty in Sumatera Island that has not been optimally addressed (Riany et al., 2022). Therefore, an estimate of child poverty at a lower level is needed so that the government can make more optimal and targeted policies.

However, the results of child poverty rate estimation based on the National Socio-Economic Survey (*Susenas*) for district/city level produce an unreliable Relative Standard Error (RSE). Estimation for smaller areas have limited number of samples so that direct estimation will produce a large RSE of more than 25 percent. To overcome these conditions, an indirect estimation method can be used, namely Small Area Estimation (SAE) by utilizing additional information (Permatasari & Larasati, 2022).

In small area estimation of child poverty, the selection of auxiliary variables should reflect key dimensions of household welfare that are theoretically and empirically linked to poverty. Housing quality and access to clean water and sanitation represent basic living standards and are consistently associated with improved economic well-being (Azizah et al., 2022; Juwita & Pertiwi, 2022; Lake, 2020). Access to education and information technology reflects human capital development and economic participation, both of which are important for long-term poverty reduction (Christiani & Nainupu, 2021; Hofmarcher, 2021; Nisa & Budiarti, 2020; Sholeh, 2022; Wanka & Rena, 2019; Wulansari et al., 2022). Health-related factors, including the availability of health facilities, health personnel, morbidity rates, and nutritional conditions, capture vulnerability aspects that may limit productivity and increase poverty risk (Adetola & Olufemi, 2012; Akbar & Amaliah, 2024; Olagunju et al., 2018; Salis & Ubaidillah, 2023). Based on these multidimensional considerations, variables representing housing conditions, sanitation, education access, information access, and health were selected as auxiliary variables to strengthen the precision and reliability of small area child poverty estimates.

In this study, the auxiliary variables are derived from the *Susenas* KOR March 2023 and PODES 2021 data. Since several auxiliary variables—particularly those obtained from *Susenas* KOR—are survey-based estimates, they are subject to sampling variability. Ignoring this uncertainty may bias parameter estimates and underestimate the variability of the small area estimates. Therefore, a measurement error component is incorporated into the Hierarchical Bayes Beta SAE model to explicitly account for the sampling error in the auxiliary variables. Accordingly, this study aims to estimate the percentage of children aged 0–17 years living below the poverty line at the district/city level on Sumatera Island in 2023 by utilising auxiliary information while accounting for measurement error.

METHOD

Child Poverty

According to BPS, poor children are those who are 0-17 years old and live in households that have an average expenditure per capita per month below the poverty line. The calculation of child poverty generally uses the same method as the calculation of poverty in general, which starts by calculating the poverty line and comparing the average monthly per capita expenditure of a household to the poverty line. If the average monthly per capita expenditure of the household is below the poverty line, then all members of the household are considered poor. Poor children are children aged 0-17 years who live in the household (BPS, 2017).

BPS calculates the percentage of poor people (P_0), which is the percentage of the population living below the poverty line, with the following formula.

$$P_\alpha = \frac{1}{n}\sum_{i=1}^{q}\left[\frac{z-y_i}{z}\right]^\alpha, i = 1, 2, \dots, q, y_i < z \quad (1)$$

with,
α = 0
z = poverty line
y_i = average monthly per capita expenditure of the population below the poverty line
q = the number of people below the poverty line
n = total population

To calculate the percentage of child poverty, the population data used is only the population aged 0-17 years.

Direct Estimation

Direct estimation is a design-based approach used to obtain population parameter estimates from survey data by applying sampling weights, where statistical inference is derived from the probability distribution induced by the sampling design while treating population values as fixed. In this study, the direct estimator of child poverty is defined as the weighted proportion of children aged 0–17 years whose per capita household expenditure falls below the official poverty line, calculated using the *Susenas* March 2023 data. The estimation follows the *Susenas* March 2023 sampling design, which applies a two-stage, one-phase sampling design. As with any design-based estimator, adequate precision at the domain (area) level depends on having a sufficiently large sample size within the respective domain.

Hierarchical Bayes (HB) Model for Proportional Data

To accommodate non-normal sampling distribution on proportion parameter, Beta distribution assumption is used as an alternative because the Beta distribution corresponds to the nature of the proportion data which is in the range of 0 to 1 (Liu, 2009). Beta-logistic model with unknown sampling variance or also known as SAE HB Beta can be written as follows.

a. Sampling model

$$\hat{\theta}_i|\theta_i, k_i \sim^{ind} Beta(a_i, b_i), \qquad i = 1, \dots, m \quad (2)$$

where θ_i denotes the true proportion parameter in small area i, $\hat{\theta}_i$ is the direct estimator of the proportion in area i, $a_i = \theta_i k_i$ and $b_i = (1-\theta_i)k_i$ are the parameters of the Beta distribution, $E(\hat{\theta}_i|\theta_i, k_i) = \theta_i = \frac{a_i}{a_i+b_i}$, and k_i is a precision (or concentration) parameter that is related to the sampling variability of the design-based direct estimator. In particular, k_i can be interpreted as being associated with the effective sample size or the design-based variance in area i, ensuring coherence between the model-based Beta assumption and the complex survey design framework. To allow heterogeneity across areas, k_i is assumed to follow a Gamma distribution, namely $k_i \sim Gamma(g_1, g_2)$.

b. Linking model

$$logit(\theta_i)|\beta, \sigma_v^2 \sim^{ind} N(x_i^T\beta, \sigma_v^2), \qquad i = 1, \dots, m \tag{3}$$

or can be written,

$$logit\left(\frac{a_i}{a_i+b_i}\right)|\beta, \sigma_v^2 \sim^{ind} N(x_i^T\beta, \sigma_v^2), \qquad i = 1, \dots, m \tag{4}$$

Bayesian inference on parameters θ_i applied to β dan σ_v^2, namely $\beta_j \sim N(\mu_{\beta_j}, \sigma_{\beta_j}^2)$ and $\sigma_v^2 \sim IG(t_1, t_2)$ where IG is the *invers Gamma*, μ_{β_j} is the mean of prior β_j, $\sigma_{\beta_j}^2$ is the variance of prior β_j, t_1 and t_2 are the parameters of prior σ_v^2. In this case, $g_1, g_2, \mu_{\beta_j}, \sigma_{\beta_j}^2, t_1$, and t_2 are fixed.

Hierarchical Bayes (HB) Model with Measurement Error (ME) on Beta Distribution

The SAE HB model with measurement error based on Beta distribution is the development of SAE HB Beta model and SAE model with measurement error (Liu, 2009; Ybarra & Lohr, 2008). In this study, this model is hereafter referred to as the SAE HB ME Beta model.

The SAE model with measurement error can be written as follows.

$$y_i = \hat{x}_i^T + r_i(\hat{x}_i, x_i) + e_i, \qquad i = 1, \dots, m \tag{5}$$

with $r_i(\hat{x}_i, x_i) = (x_i + \hat{x}_i)^T\beta + v_i$, m is the number of areas, β s the vector of regression coefficients, $\hat{x}_i$ is the unbiased estimator of x_i in the form of a matrix of size $m \times p$ that is independent of the area random effects $v_i \sim N(0, \sigma_v^2)$ and sampling error $e_i \sim N(0, \psi_i)$.

The posterior distribution in Bayesian inference is obtained by combining prior information with information from the data used (likelihood). The likelihood function and prior distribution used in SAE HB ME Beta are as follows.

a. Sampling model

$$\hat{\theta}_i|\theta_i, k_i \sim^{ind} Beta(a_i, b_i), \qquad i = 1, \dots, m \tag{6}$$

where θ_i denotes the true proportion parameter in small area i, $\hat{\theta}_i$ is the direct estimator of the proportion in area i, $a_i = \theta_i k_i$ and $b_i = (1 - \theta_i)k_i$ are the parameters of the Beta distribution, $E(\hat{\theta}_i|\theta_i, k_i) = \theta_i = \frac{a_i}{a_i+b_i}$, and k_i is a precision (or concentration) parameter that is related to the sampling variability of the design-based direct estimator. In particular, k_i can be interpreted as being associated with the effective sample size or the design-based variance in area i, ensuring coherence between the model-based Beta assumption and the complex survey design framework. To allow heterogeneity across areas, k_i is assumed to follow a Gamma distribution, namely $k_i \sim Gamma(g_1, g_2)$.

b. Linking model

$$logit(\theta_i)|\beta, \sigma_v^2 \sim^{ind} N(x_i^T \beta, \sigma_v^2), \qquad i = 1, \dots, m \tag{7}$$

or can be written,

$$logit\left(\frac{a_i}{a_i + b_i}\right) |\beta, \sigma_v^2 \sim^{ind} N(x_i^T \beta, \sigma_v^2), \; i = 1, \dots, m \tag{8}$$

c. Measurement error effect

$$\hat{x}_i | x_i \sim N(x_i, C_i), \qquad i = 1, \dots, m \tag{9}$$

with $\hat{x}_i$ representing the effect of measurement error in the accompanying variables, where $C_i = \sigma_{x_i}^{-1}$ is known as the variance of the auxiliary variable x_i.

d. β, σ_v^2 are mutually independent with $\beta_L \sim N(\mu_{\beta_L}, \tau_{\beta_L}^2)$ and $\sigma_v^2 \sim IG(t_1, t_2)$.
where β s the vector of regression coefficients and σ_v^2 is the variance of area random effects. μ_{β_L} is the mean of the priors β_L, $\tau_{\beta_L}^2$ s the variance of the priors β_L, t_1 and t_2 is the prior parameter σ_v^2. In this case, $g_1, g_2, \mu_{\beta_L}, \tau_{\beta_L}^2, t_1$, and t_2 are fixed.

In the process of calculating the posterior distribution, it is often found that the solution is not in closed form or very complicated so that numerical assistance is needed to facilitate its completion. In Bayesian analysis, this numerical calculation is done using Markov Chain Monte Carlo (MCMC) with the Gibbs sampling algorithm so that the posterior distribution is produced as follows.

$$f(\theta_i | \hat{\theta}_i) \propto f(\theta_i, \beta, k, \sigma_v^2) f(\beta_L) f(x_i) f(\sigma_v^2) f(v) \tag{10}$$

Data

The study uses secondary data derived from the raw *Susenas* KOR March 2023 and PODES 2021 datasets. In addition to providing direct estimates of the variables of interest, the *Susenas* KOR March 2023 survey is also used to construct the auxiliary variables. As a survey based source, it is subject to measurement error. The list of variables and their data sources is presented in Table 1.

Table 1. List of variables and data source

Variable	Description	Data Source
Y	Percentage of children aged 0-17 years living below the poverty line	*Susenas* KOR 2023
X1	Percentage of households with clean drinking water sources	*Susenas* KOR 2023
X2	Percentage of households with access to safe drinking water	*Susenas* KOR 2023
X3	Percentage of households with disposal facilities self-defecation	*Susenas* KOR 2023

Variable	Description	Data Source
X4	Percentage of households with a disposal site end in the form of a septic tank	*Susenas* KOR 2023
X5	Percentage of villages where the facilities for most families are their own toilets	PODES 2021
X6	Number of villages with facilities for defecation in most families not a latrine	PODES 2021
X7	Proportion of villages with family subscriptions wired phone	PODES 2021
X8	Percentage of population aged 5 years and above who access the internet	*Susenas* KOR 2023
X9	Percentage of population aged 5 years and above who using a computer	*Susenas* KOR 2023
X10	Number of people suffering from malnutrition	PODES 2021
X11	Proportion of villages that have primary school/elementary school/equivalent	PODES 2021
X12	Proportion of villages that have junior high school/ secondary school/equivalent	PODES 2021
X13	Proportion of villages that have senior high school/equivalent	PODES 2021
X14	Proportion of villages with academies/colleges	PODES 2021
X15	Number of families using electricity	PODES 2021
X16	Percentage of villages with families living in riverside	PODES 2021
X17	Percentage of villages with families living in slums	PODES 2021
X18	Number of health facilities	PODES 2021
X19	Number of health workers living in the village	PODES 2021
X20	Morbidity rate	*Susenas* KOR 2023

Data Preparation

In this stage, the data filtering process is carried out in accordance with the information needed for direct estimation and aggregation of 20 candidate auxiliary variables at the district/city level from the *Susenas* KOR March 2023 and PODES 2021 data.

Direct Estimation

Direct estimation calculations were conducted using sampling design-based *Susenas* KOR March 2023 to produce direct estimates of the percentage of children aged 0-17 years living below the poverty line.

Indirect Estimation

The indirect estimation stage consists of the following steps:

1. Detecting multicollinearity. Variables with correlations exceeding 0.8 indicates multicollinearity, so variable selection is performed.
2. Selecting the auxiliary variables using the stepwise method.
3. The SAE HB ME Beta model was constructed with the auxiliary variables from the stepwise results. SAE HB ME Beta is used because the direct estimator is assumed not to be normally distributed and there is additional information containing errors. In forming the SAE HB ME Beta model, the child poverty rate data which is the direct estimator needs to be converted into a proportion with an interval of 0 to 1. Then, it is also necessary to determine the number of updates, iterations, thinning, and burn-in. Update is useful in updating the prior information of

the initial value based on the likelihood function or data, thin to reduce the autocorrelation between the generated samples, and iteration and burn in to remove the influence of the initial value. The parameters are selected by trial and error until the algorithm converges.

Model Selection

The model selection was based on the results of the model evaluation process and the internal consistency criterion of the indirect estimator values generated from the model. Model evaluation was done by comparing the RSE of the direct estimator with the indirect estimator. Meanwhile, validation of the indirect estimator value is done by comparing the direct estimator value in each province with the indirect estimator value in each district/city to assess consistency between provincial and district-level estimates. Provincial direct estimators were considered reliable benchmarks due to sufficient sample sizes and low RSE. The best model chosen is the one that produces indirect estimator values with the smallest RSE and is valid. The model is considered valid if the direct estimator value in each province falls within the range of the minimum and maximum indirect estimator values in the districts/cities within that province.

RESULTS AND DISCUSSION

Direct Estimation

The following is a description of the direct estimation results of the percentage of children aged 0-17 years on Sumatera Island living below the poverty line in 2023.

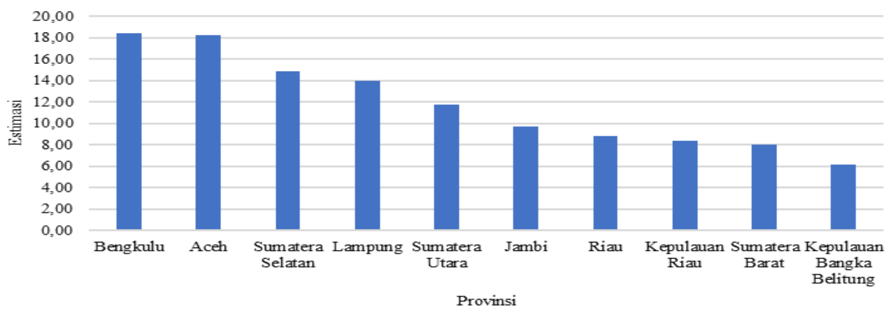

Figure 1. Direct estimation of the percentage of children aged 0-17 in Sumatera Island living below the poverty line, 2023

Source: BPS, processed

Based on Figure 1, Bengkulu and Aceh are the provinces with the highest percentage of children aged 0-17 years living below the poverty line in 2023. Meanwhile, West Sumatera and Kep Bangka Belitung are the provinces with the lowest percentage of children aged 0-17 living below the poverty line in 2023.

The direct estimation results of the percentage of children aged 0-17 years living below the poverty line in Sumatera Island in 2023 are presented summarized along with their RSE values.

Table 2. Summary statistics of direct estimation and RSE

Descriptive Statistics	Direct Estimation	RSE
Minimum	2,58	10,47
Q1	8,85	16,49
Median	12,46	19,96
Mean	13,12	21,54

Descriptive Statistics	Direct Estimation	RSE
Q3	16,50	25,21
Maximum	31,35	46,59
Variance	32,15	50,95

Based on Table 2, the variance of child poverty percentage in Sumatera Island in 2023 is 32.15, indicating substantial variability across districts/cities. The highest and lowest values of child poverty percentage in Sumatera Island in 2023 are 31.35% and 2.58%, respectively, suggesting considerable inequality in the distribution of child poverty within the island. This wide range may reflect spatial disparities, where certain areas experience much higher levels of child poverty than others. These differences may be associated with variations in regional economic structures, levels of educational attainment, infrastructure development, and access to health services, which can influence household welfare and child well-being. In addition, several districts/cities in Sumatera Island that have a child poverty percentage with RSE ≥ 25%, indicating low precision in the direct estimates. The large RSE values reduce the reliability of these estimates. Therefore, indirect estimation is needed to produce smaller RSE values and more accurate and reliable estimates. High RSE values imply substantial sampling variability, which reduces confidence in the direct estimates for small areas. In this context, Small Area Estimation (SAE) is methodologically advantageous because it borrows strength from related areas and incorporates auxiliary information to improve estimation precision. Therefore, indirect estimation through SAE is needed to reduce the RSE and produce more reliable and stable estimates of child poverty at the district/city level.

Indirect Estimation of the Percentage of Children 0-17 Years Old Living Below the Poverty Line in Sumatera Island in 2023 Using SAE HB ME Beta

a. Initial Model

The first step was to check the multicollinearity of all the auxiliary variables used. Then, variables with correlation value greater than 0.8 were not selected as candidate variables to be modelled (0,8 was chosen to avoid redundancy and inflated variance) The variables that were not selected were the proportion of villages with family subscriptions wired phone, Proportion of villages that have senior high school/equivalent, and the number of families that use electricity. Next, the selected auxiliary variables were selected using the stepwise method. Based on the stepwise results, five accompanying variables were selected to be used in the modelling, namely the percentage of households with a clean drinking water source, the percentage of the population aged 5 years and above who accessed the internet, proportion of villages that have primary school/elementary school/equivalent, the percentage of villages where families lived in riverside, and percentage of villages where families lived in slums. These accompanying variables are sourced from *Susenas* KOR 2023 and PODES 2021. The auxiliary variables derived from PODES 2021 mainly represent structural village-level characteristics, such as sanitation infrastructure, educational and health facilities, electricity access, settlement conditions, and availability of health personnel. These characteristics tend to change gradually rather than abruptly within a short two-year period. Therefore, although there is a temporal gap between 2021 and 2023, the selected variables are conceptually stable and continue to reflect the underlying socio-economic structure relevant for the 2023 estimation. Nonetheless, we acknowledge this temporal mismatch as a limitation of the study.

The modelling process was carried out using R software by implementing the 'meHBbeta' function in the 'saeHB.ME.beta' package. In modelling with this function, convergent conditions were achieved for the number of MCMC samples of 10,000, the number of updates of 20, burn.in of 2,000, and thinning number of 10. Convergence in the Markov chain was checked based on the trace plot, density plot, and autocorrelation plot presented in Figure 2.

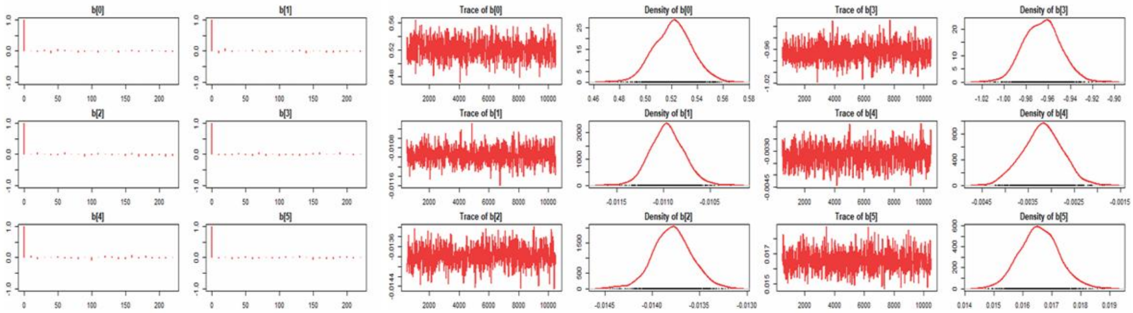

Figure 2. Autocorrelation plot, trace plot, and density plot of the initial SAE HB ME Beta model
Source: BPS, processed

Based on Figure 2, it can be seen that the trace plot does not form a certain periodic pattern and the density plot for all parameters tends to be smooth. In addition, the resulting autocorrelation plot is cut off at the initial lag. This indicates that the convergence of the MCMC algorithm has been achieved, suggesting adequate mixing of the Markov chain. Formal quantitative diagnostics such as the Gelman–Rubin statistic require multiple chains and were not implemented in this study, as the estimation procedure in saeHB.ME.beta was conducted using a single-chain MCMC algorithm. This is recognized as a limitation of the study. Nevertheless, the visual diagnostics consistently indicated stable posterior behavior.

The parameter coefficient estimation results of the SAE HB ME Beta model and 95% credible interval are presented in the following table.

Table 3. Parameter estimates for the initial SAE HB ME Beta model

Descriptive Statistics	Mean	2,5%	97,5%
Constant	0,5212	0,4937	0,5490
X1	-0,0109	-0,0113	-0,0106
X8	-0,0137	-0,0141	-0,0133
X11	-0,9655	-0,9957	-0,9324
X16	-0,0031	-0,0040	-0,0023
X17	0,0165	0,0153	0,0179

Based on Table 3, it can be seen that all variables affect child poverty with a credible interval of 95%. This is indicated by the absence of estimation numbers that cross zero. This means that the variables of the percentage of households with a clean drinking water source (X1), the percentage of the population aged 5 years and above who access the internet (X8), the proportion of villages with primary schools (X11), the percentage of villages where families live in riverside (X16), and the percentage of villages where families live in slums (X17) provide additional information to the estimation process of the proportion of child poverty in Sumatera Island. In addition, the variance value of the random effect is 0.109.

b. Initial Model Evaluation

To see the comparison of estimation precision, the RSE calculation of the estimation results with the SAE HB ME Beta method is carried out. Based on Table 4, it was found that the SAE HB ME Beta method was able to minimise the RSE. However, there is still one district/city that produces RSE greater than 25%, namely Sawah Lunto City, West Sumatera. Thus, it can be concluded that this estimation has not produced the best level of precision.

Table 4. Comparison of RSE between direct estimation and the initial SAE HB-ME Beta model

Descriptive Statistics	Direct Estimation RSE (%)	SAE HB ME Beta RSE (%)
Minimum	10,47	9,00
Q1	16,49	12,98
Median	19,96	14,54
Mean	21,54	14,86
Q3	25,21	16,07
Maximum	46,59	26,40

The next step is to conduct an internal consistency check of the SAE HB ME Beta model. This procedure begins by converting the estimated proportions into percentages. The results are considered consistent if the provincial direct estimate lies within the range (minimum–maximum) of the corresponding district/city SAE estimates within the province. The results show that *Kepulauan* Bangka Belitung does not satisfy this consistency criterion. A summary of the internal consistency check results is presented in Table 5.

Table 5. Internal consistency check of the initial SAE HB ME Beta model

Province	Direct Estimation	Minimum	Maximum	Consistency
Aceh	18,2524	9,785	23,245	Consistent
Sumatera Utara	11,7654	6,652	30,489	Consistent
West Sumatera	8,0553	4,866	19,834	Consistent
Riau	8,8329	5,848	27,404	Consistent
Jambi	9,7211	5,963	14,690	Consistent
Sumatera Selatan	14,8697	11,532	19,053	Consistent
Bengkulu	18,3765	13,468	22,476	Consistent
Lampung	13,9499	8,132	20,022	Consistent
Kep. Bangka Belitung	6,1652	6,371	9,912	Inconsistent
Kepulauan Riau	8,3923	7,055	13,917	Consistent

While the initial SAE HB ME Beta model considerably enhanced overall estimation precision, a small number of areas still exhibited precision and consistency issues, as one district/city (Kota Sawah Lunto, West Sumatera) still had an RSE greater than 25%, and the internal consistency criterion was not satisfied for *Kepulauan* Bangka Belitung.

The substantial reduction in mean RSE (from 21.54 percent to 14.86 percent) reflects the shrinkage effect inherent in the hierarchical Bayesian framework. In areas with small sample sizes, the direct estimator tends to have large sampling variance, leading to inflated RSE values. By incorporating auxiliary village-level structural variables from PODES and introducing random effects, the model borrows strength across areas, effectively stabilizing extreme estimates and reducing posterior variance. This mechanism explains the observed improvement in overall precision.

Nevertheless, some areas remained problematic. For instance, one district in West Sumatera exhibited an RSE greater than 25 percent, likely due to very small effective sample size or weak association between auxiliary variables and the target variable. Additionally, estimates for *Kepulauan* Bangka Belitung did not satisfy the internal consistency check, reflecting structural heterogeneity not fully captured by the model because its data exhibited different structural characteristics compared to other provinces. This heterogeneity likely results in poor alignment between the auxiliary variables and the response variable in that province. In small area estimation, reliable model-based estimates rely on both sufficient effective sample size and meaningful statistical relationships between auxiliary information and the target variable. When these conditions are not met — for example due to a weak correlation between predictors and response — the SAE model may not adequately improve precision, leading to invalid or unstable estimates for that area.

To address this heterogeneity, the estimation was conducted using a partitioned approach, allowing the model to better accommodate regional variation in precision levels. This adjustment improved stability while maintaining compliance with internal consistency criteria.

Therefore, further modelling was conducted by separating West Sumatera and *Kepulauan* Bangka Belitung—where these estimation issues were observed—from the remaining provinces, and estimating the model for these two groups separately.

Final Model

Based on the results of variable selection using the correlation matrix and the stepwise method, the estimation of the percentage of child poverty for the provinces of West Sumatera and *Kep.* Bangka Belitung was carried out with auxiliary variables sourced from the *Susenas* KOR, namely the percentage of households with a clean drinking water source and the percentage of the population aged 5 years and over who use a computer.

Then, with the SAE HB ME Beta method, the estimation of equation parameters was carried out with 20 update iterations, MCMC iterations per update of 10,000, thinning number of 10, and burn.in of 2,000. Diagnostic plots in the form of autocorrelation function plots, trace plots, and density plots that can be seen in Figure 3 show that the MCMC algorithm has converged.

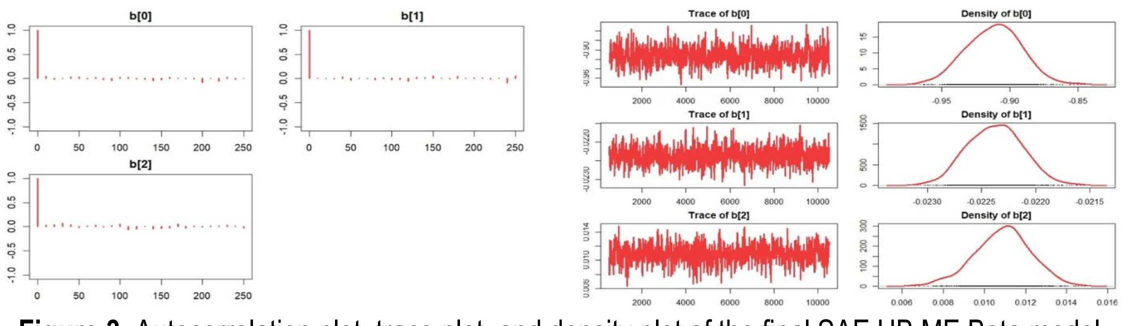

Figure 3. Autocorrelation plot, trace plot, and density plot of the final SAE HB ME Beta model (West Sumatera and *Kep.* Bangka Belitung)
Source: BPS, processed

The results of the parameter estimation equation can be seen in the following table:

Table 6. Parameter estimates for the final SAE HB ME Beta model (West Sumatera and *Kep.* Bangka Belitung)

	Mean	**2,5%**	**97,5%**
Constant	-0,9117	-0,9503	-0,8726
X1	-0,0223	-0,0228	-0,0218
X9	0,0108	0,0079	0,0134

Based on Table 6, it can be seen that all variables affect the proportion of child poverty with a credible interval of 95%. This is indicated by the absence of estimation numbers that exceed 0. In addition, the variance value of random effect is 0.068.

Based on the results of variable selection using the correlation matrix and the stepwise method, the estimation of the percentage of child poverty for provinces other than West Sumatera and *Kep.* Bangka Belitung was carried out with auxiliary variables sourced from both *Susenas* KOR 2023 and PODES 2021, namely the percentage of households with clean drinking water sources, the percentage of the population aged 5 years and over who access the internet, the percentage of villages where families live in riverside, the percentage of villages where families live in slums, the number of health facilities, and the morbidity rate.

By using the SAE HB ME Beta method, the estimation of equation parameters is carried out with 20 update iterations, 10,000 MCMC iterations per update, 10 thinning numbers, and 2,000 burn.in. Diagnostic plots in the form of autocorrelation function plots, trace plots, and density plots that can be seen in Figure 4 show that the MCMC algorithm has converged.

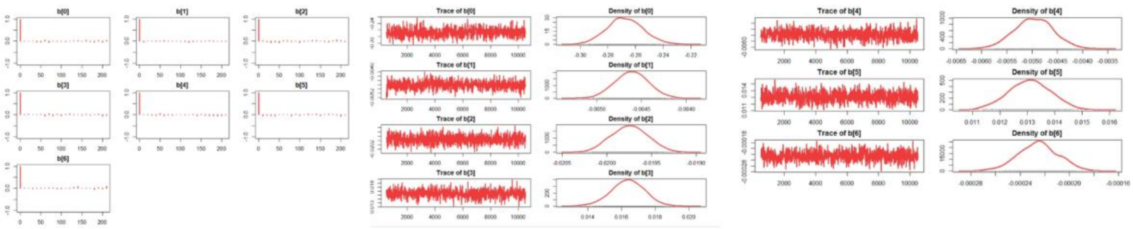

Figure 4. Autocorrelation plot, trace plot, and density plot of the final SAE HB ME Beta model (other than West Sumatera and *Kep.* Bangka Belitung)

Source: BPS, processed

The results of the parameter estimation equation can be seen in the following table.

Table 7. Parameter estimates for the final SAE HB ME Beta model (other than West Sumatera and *Kep.* Bangka Belitung)

	Mean	**2,5%**	**97,5%**
Constant	-0,2674	-0,2921	-0,2419
X1	-0,0046	-0,0050	-0,0042
X8	-0,0197	-0,0201	-0,0194
X16	0,0164	0,0144	0,0183
X17	-0,0050	-0,0057	-0,0042
X18	0,0131	0,0116	0,0145
X20	-0,0002	-0,0003	-0,0002

Based on Table 7, it can be seen that all variables affect the proportion of child poverty with a credible interval of 95%. This is indicated by the absence of estimation numbers that exceed 0. In addition, the variance value of random effect is 0.073.

Final Model Evaluation

The estimation results with the SAE HB ME Beta model applied to separate observations, namely West Sumatera – *Kep.* Bangka Belitung and the other provinces located on the island of Sumatera combined, have generally reduced the RSE of the direct estimation of the percentage of child poverty. This can be shown through the comparison table below.

Table 8. Comparison of RSE categories between direct estimation and the final SAE HB ME Beta model

Province	Method	RSE ≤ 25%	25% < RSE ≤ 50%	RSE > 50%
Aceh	Direct Estimation	23	0	0
	SAE HB ME Beta	23	0	0
Sumatera Utara	Direct Estimation	30	3	0
	SAE HB ME Beta	33	0	0
West Sumatera	Direct Estimation	6	13	0
	SAE HB ME Beta	19	0	0
Riau	Direct Estimation	8	4	0
	SAE HB ME Beta	12	0	0
Jambi	Direct Estimation	6	5	0
	SAE HB ME Beta	11	0	0
Sumatera Selatan	Direct Estimation	17	0	0
	SAE HB ME Beta	17	0	0
Bengkulu	Direct Estimation	10	0	0
	SAE HB ME Beta	10	0	0
Lampung	Direct Estimation	14	1	0
	SAE HB ME Beta	15	0	0
Kep. Bangka Belitung	Direct Estimation	0	7	0

Province	Method	RSE ≤ 25%	25% < RSE ≤ 50%	RSE > 50%
	SAE HB ME Beta	7	0	0
Kep. Riau	Direct Estimation	3	4	0
	SAE HB ME Beta	7	0	0
Pulau Sumatera	Direct Estimation	117	37	0
	SAE HB ME Beta	154	0	0

Based on Table 8, it can be seen that after the SAE HB ME Beta is applied, there are no more district/city estimates in Sumatera Island that have RSE above 25%. The following is a visualization of the RSE of the direct and indirect estimates of the percentage of child poverty.

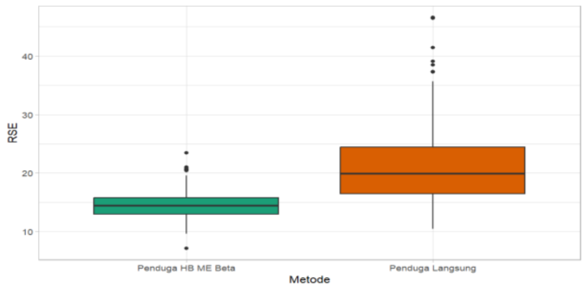

Figure 5. Distribution of direct and indirect estimated RSE
Source: BPS, processed

Based on Figure 5, it can be seen that, in general, the RSE of the SAE HB ME Beta estimator is smaller than that of the direct estimator. This is reflected in the comparison of the central tendency and spread shown in the boxplot. The estimated values are then converted from proportions into percentages and assessed using the internal consistency criterion. As summarised in Table 9, the SAE HB ME Beta estimation results satisfy the internal consistency criterion for all provinces on Sumatera Island. The following table presents the results of the internal consistency assessment along with the visualization of the direct and indirect estimates of the percentage of child poverty on Sumatera Island.

Table 9. Internal consistency check of the final SAE HB ME Beta model

Province	Direct Estimation	Minimum	Maksimum	Consistency
Aceh	18,2524	10,250	22,880	Consistent
Sumatera Utara	11,7654	6,748	29,798	Consistent
West Sumatera	8,0553	3,583	19,753	Consistent
Riau	8,8329	6,907	26,314	Consistent
Jambi	9,7211	8,254	14,190	Consistent
Sumatera Selatan	14,8697	11,910	19,210	Consistent
Bengkulu	18,3765	13,450	22,080	Consistent
Lampung	13,9499	9,268	19,374	Consistent

Province	Direct Estimation	Minimum	Maksimum	Consistency
Kep. Bangka Belitung	6,1652	4,776	9,428	Consistent
Kep. Riau	8,3923	7,918	14,790	Consistent

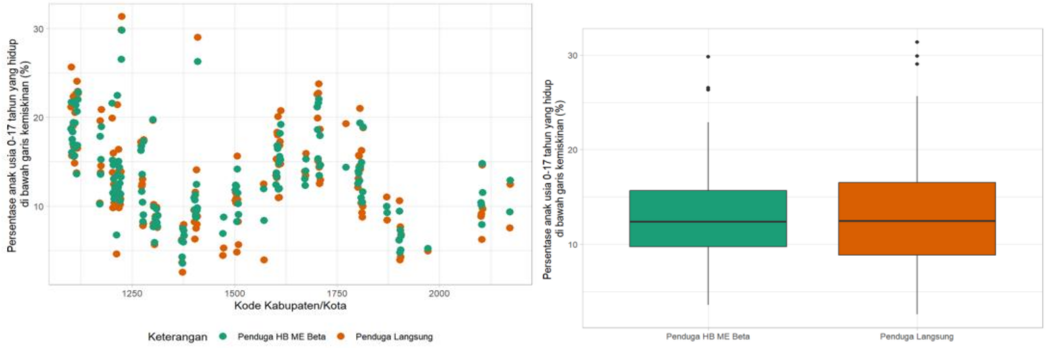

Figure 6. Comparison of direct and indirect estimation results
Source: BPS, processed

Based on Figure 6, the SAE HB ME Beta estimates exhibit a narrower range than the direct estimates, while maintaining a similar overall distribution pattern. Through the evaluation process, it can be concluded that the SAE HB ME Beta model—applied separately to West Sumatera and *Kepulauan* Bangka Belitung and jointly to the remaining provinces on Sumatera Island—improves estimation quality. This improvement is evidenced by lower RSE values compared to the direct estimator, with overall RSE values below 25%.

In addition, the model satisfies the internal consistency criterion, as the provincial direct estimates lie within the minimum–maximum range of the corresponding district/city indirect estimates in each province on Sumatera Island. This improvement reflects the suitability of the auxiliary variables and the partitioned model specification, which allows the estimation framework to better account for inter-provincial heterogeneity and thereby produce more stable and precise estimates.

The satisfaction of the internal consistency criterion indicates that the indirect estimates at the district/city level are now coherent with the corresponding provincial direct estimates. This occurs because the refined model specification and partitioned estimation approach strengthen the statistical relationship between auxiliary variables and the response variable within each provincial group. By reducing potential model misspecification and better capturing regional heterogeneity, the resulting estimates become both more precise and internally consistent.

Mapping of Indirect Estimation Results

Based on the estimation results with the SAE HB ME Beta method in the final model, the percentage of child poverty in Sumatera Island was grouped with natural breaks. Then, a thematic map was prepared to see its geographical distribution. The map can be seen in the figure below.

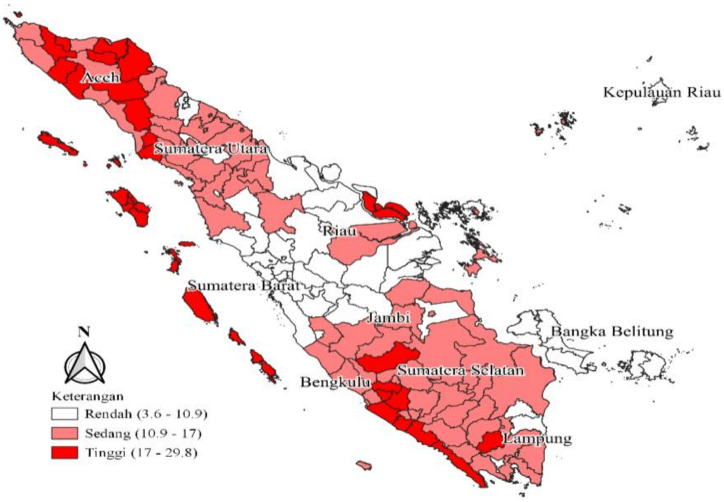

Figure 7. Map of indirect estimation of the percentage of poor children in Sumatera Island in 2023

Based on Figure 7, the percentage of poor children on the island of Sumatera is grouped into three categories: low (3.6% ≤ x < 10.9%), medium (10.9% ≤ x < 17%), and high (17% ≤ x ≤ 29.8%). It can be seen that districts/municipalities with a low percentage of poor children are clustered in the central part of Sumatera Island and the islands. Meanwhile, the northern and southern parts of Sumatera Island mostly have medium and high percentages of poor children. The provinces where most of the districts/cities have a low percentage of poor children are West Sumatera, Riau, *Kep.* Bangka Belitung, and *Kep.* Riau. This may be associated with a relatively greater average of percentage of households with a clean drinking water source than other provinces. Provinces where most of the districts/cities have a medium percentage of poor children are Sumatera Utara, Jambi, Sumatera Selatan, and Lampung. The province with most of its districts/cities having a high percentage of poor children is Aceh. This may be associated with a relatively smaller average of the percentage of the population aged 5 years and over who access the internet and a relatively greater average of percentage of villages where families live in slums than other provinces. In addition, the districts/municipalities in Bengkulu Province are equally divided into two categories, namely medium and high.

Beyond methodological performance, these findings have broader implications for poverty measurement and policy formulation. The application of SAE enables reliable estimation of child poverty at the district/city level, where direct survey estimates are often unstable due to small sample sizes. This enhances the granularity of poverty statistics. From a policy perspective, more precise small-area estimates can support evidence-based allocation of social assistance and development programs. For statistical practice, this study demonstrates the practical value of hierarchical Bayesian SAE models with auxiliary information in improving precision while maintaining internal consistency. It also underscores the importance of accounting for regional heterogeneity through appropriate model specification.

CONCLUSION

This study shows that the direct estimation of the percentage of child poverty on Sumatera Island in 2023 is not sufficiently reliable, as several districts/cities still have RSE values greater than 25%. To address this issue, Small Area Estimation using the SAE HB ME Beta method was applied and substantially reduced the RSE values, with 153 out of 154 districts/cities achieving RSE below 25%. However, the model estimated jointly for all provinces still resulted in one district/city with RSE greater than 25% and failed to satisfy the internal consistency criterion in *Kepulauan* Bangka Belitung, despite the substantial overall improvement in precision. In contrast, when the SAE HB ME Beta model was estimated separately for West Sumatera and *Kepulauan* Bangka Belitung and jointly for the remaining provinces, all districts/cities achieved RSE below 25% and satisfied the internal consistency requirement across all provinces on Sumatera Island.

These findings demonstrate that the hierarchical Bayesian SAE framework provides a reliable approach for producing precise and internally consistent child poverty estimates at the district/city level, particularly in the presence of small sample sizes. By incorporating relevant auxiliary information and accounting for regional heterogeneity, the model enhances the quality and granularity of poverty statistics. More accurate small-area estimates can support better-targeted social protection programs and resource allocation, especially in regions exhibiting persistently high child poverty rates. Overall, this study highlights the practical value of model-based small area estimation in improving official poverty statistics and underscores the importance of appropriate model specification when addressing regional structural differences.

REFERENCE

Adetola, A., & Olufemi, P. (2012). Determinants of child poverty in Rural Nigeria: A multi-dimensional approach. *Global Journal of Human Social Science Arts & Humanities*, *12*, 38–52.

Akbar, R. A., & Amaliah, I. (2024). Pengaruh pengeluaran per kapita, angka kesakitan, angka perceraian, dan persentase penerima kredit usaha rakyat terhadap tingkat kemiskinan di Jawa Barat tahun 2010-2022. *Bandung Conference Series: Economics Studies*, 74–81.

Azizah, S. P. N., Pratiwi, L. S., Amaliah, I., & Fitriyana, F. (2022). Sanitasi dan kepadatan penduduk sebagai dinamika kemiskinan kota studi kasus Provinsi Jawa Barat. *Nuansa Akademik: Jurnal Pembangunan Masyarakat*, *7*(1), 55–70. https://doi.org/10.47200/jnajpm.v7i1.1148

BPS. (2017). *Analisis kemiskinan anak dan deprivasi hak-hak dasar anak di Indonesia*. Badan Pusat Statistik.

BPS. (2024). *Penduduk di bawah garis kemiskinan*. https://sirusa.web.bps.go.id/metadata/indikator/3704

Christiani, N. V., & Nainupu, A. E. (2021). Pengaruh akses terhadap internet, listrik dan PDRB per kapita terhadap tingkat kemiskinan di Nusa Tenggara Timur Tahun 2015-2019. *Jurnal Statistika Terapan*, *1*(1), 37–52.

Darjono, A. H., Aprilasani, Z., & Munandar, A. I. (2019). *Pembangunan berkelanjutan: studi kasus di Indonesia*. Bypass.

Hofmarcher, T. (2021). The effect of education on poverty: A European perspective. *Economics of Education Review*, *83*, 102124. https://doi.org/10.1016/j.econedurev.2021.102124

Juwita, A., & Pertiwi, A. (2022). Kemiskinan dan kesehatan lingkungan bersinergi dalam pembangunan berkelanjutan. *Lakeisha*, 45–59.

Kertayana, I. K. (2017). Kondisi kemiskinan anak dan harapan sustainable development goals (SDGs) di Indonesia. *Seminar Nasional Kependudukan & Kebijakan Publik*, 40–51.

Lake, I. (2020). Analisis status sosial ekonomi orang tua terhadap kemiskinan Di Kecamatan Insana Kabupaten Timor Tengah Utara. *EKOPEM : Jurnal Ekonomi Pembangunan*, *5*(4), 1–11.

Liu, B. (2009). *Hierarchical bayes estimation and empirical best prediction of small-area proportions* [Dissertations]. University of Maryland.

Nisa, K., & Budiarti, W. (2020). Pengaruh teknologi informasi dan komunikasi terhadap tingkat kemiskinan di Indonesia tahun 2012-2017. *Seminar Nasional Official Statistics*, *2019*(1), 759–768. https://doi.org/10.34123/semnasoffstat.v2019i1.186

Olagunju, K. O., Ogunniyi, A., & Olufadewa, M. S. (2018). Spatial analysis of structural determinants of child poverty incidence in Nigeria. *30th International Conference of Agricultural Economists*.

Permatasari, N., & Larasati, W. (2022). Perbandingan metode SAE EBLUP dan SAE HB pada pendugaan area kecil (studi kasus pendugaan kemiskinan di Provinsi Jawa Timur). *Jurnal Statistika Dan Aplikasinya*, *6*(1), 96–108. https://doi.org/10.21009/JSA.06109

Priseptian, L., & Primandhana, W. P. (2022). Analisis faktor-faktor yang mempengaruhi kemiskinan. *FORUM EKONOMI*, *24*(1), 45–53. https://doi.org/10.30872/jfor.v24i1.10362

Riany, Y. E., Fauziah, H., & Putri, D. K. (2022). *Indeks Perlindungan Anak Tahun 2022*.

Salis, D. R., & Ubaidillah, A. (2023). Estimasi Tingkat Kemiskinan Anak Level Kabupaten/Kota di Provinsi Banten Tahun 2018-2021 dengan Small Area Estimation (SAE) Rao-Yu Pendekatan Hierarchical Bayes. *Seminar Nasional Official Statistics*, *2023*(1), 515–524. https://doi.org/10.34123/semnasoffstat.v2023i1.1709

Sholeh, M. (2022). Pengaruh pendidikan, jumlah anggota keluarga, dan akses informasi terhadap kemiskinan di Indonesia. *Jurnal Ekonomi Dan Pendidikan*, *19*(1), 61–72. https://doi.org/10.21831/jep.v19i1.55021

Wanka, F. A., & Rena, R. (2019). The impact of educational attainment on household poverty in South Africa: A case study of Limpopo province. *African Journal of Science, Technology, Innovation and Development*, *11*(5), 597–609. https://doi.org/10.1080/20421338.2018.1557368

Wulansari, J., Permatasari, N., & Ubaidillah, A. (2022). Pendugaan Area Kecil Persentase Anak-anak Usia Kurang dari 18 Tahun yang Hidup di Bawah Garis Kemiskinan Tingkat Kabupaten/Kota di Indonesia Tahun 2020. *Seminar Nasional Official Statistics*, *2022*(1), 383–394. https://doi.org/10.34123/semnasoffstat.v2022i1.1467

Ybarra, L. M. R., & Lohr, S. L. (2008). Small area estimation when auxiliary information is measured with error. *Biometrika*, *95*(4), 919–931. https://doi.org/10.1093/biomet/asn048