Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)

APPLICATION OF SMALL AREA ESTIMATION WITH HB BETA MODEL TO ESTIMATE CHILD LABOR PROPORTION IN KALIMANTAN

Rika Lusiana Simbolon[1)]
Amara Putri Shabrina[2)]
Fatima Azzahro Binti Fatihah[3)]
Immanuel Nicholas Fransepta Samosir[4)]
Muhammad Aswan Aziz[5)]
Nofita Istiana[6]

[1,2,3,4,5,6)]Study Program D-IV Statistics, Politeknik Statistika STIS, DKI Jakarta, Indonesia
e-mail: 212112323@stis.ac.id

ABSTRACT

The global commitment to eliminate the number child labor is reflected in Target 8.7 of the Sustainable Development Goals (SDGs), which aims to end all forms of child labor by 2025. One of the key indicators used to measure the prevalence of child labor is the percentage of child engaged in labor. Kalimantan Island is one of the regions with a high level of vulnerability to child labor practices, particularly in plantation areas that are difficult to access and not covered by adequate monitoring systems. To ensure that the number of child labor elimination policies in Kalimantan are accurately targeted, data at the small-area level are needed to precisely identify locations with a high risk of child labor involvement. However, the direct estimates of child labor percentages at the regencies/municipalities level produced by Badan Pusat Statistik (BPS) lack sufficient precision to be used as a reliable basis for policymaking. Therefore, this study aims to produce indirect estimates of child labor proportion at the regencies/municipalities level in Kalimantan to improve the accuracy of the direct estimation. This research applies the Small Area Estimation (SAE) method using the Hierarchical Bayes Beta (HB-Beta) modeling approach. The SAE HB Beta method was used because the normality assumption for proportion data is often violated, while the Beta distribution is naturally suited to variable constrained to the interval (0,1). The exploratory analysis identified nine auxiliary variables that are appropriate for use in the estimation process. The results from the SAE HB-Beta model indicate that the relative standard Error (RSE) values for all regencies and municipalities in Kalimantan are less than 25 percent. Thus, the indirect estimates obtained are considered statistically reliable and suitable for use as a basis for evidence-based policymaking.

Keywords: child labor, indirect estimation, small area estimation, area-level model, HB Beta.

INTRODUCTION

Child labor is a form of economic activity that must be eliminated (BPS, 2024). The international agreement on this matter is embodied in Sustainable Development Goal (SDG) target 8.7, which seeks to eradicate all forms of child labor by 2025. Child labor adversely impacts children's physical, psychological, and cognitive development (Mardiyanti & Handayani, 2020). Children involved in labor often experience health problems, injuries, reduced self-esteem, and emotional instability due to excessive workloads and limited educational opportunities (Foenay et al., 2023). These conditions negatively impact their educational access and exacerbate poverty (Darmika & Usman, 2021). Child laborers typically exhibit diminished educational achievement relative to their non-working counterparts (Hamdani et al., 2023), hindering the cultivation of a proficient workforce and impeding overall economic advancement. The eradication of child labor is both a human rights imperative and a vital measure for attaining sustainable economic

development. The Indonesian government has enacted UU No. 35 Tahun 2014 on Child Protection, which explicitly forbids the employment of minors as a commitment to combating child labor. Per this legal framework, child labor has been identified as a primary concern in the Ministry of Women's Empowerment and Child Protection's 2019–2024 Medium-Term Development Plan (Kemen PPPA, 2024). To support this, the government also implemented specific measures like establishing the framework of the "Roadmap Towards a Child Labor Free Indonesia 2022" (OECD, 2018) . Despite all those efforts, child labor continues to be an important issue in the country.

Data from Badan Pusat Statistik (BPS) indicates that the prevalence of child labor has generally decreased over the past five years (refer to Figure 1). In 2024, the proportion of child labor increased by 0.45 percentage points to 2.17 percent, indicating that approximately 1.27 million children aged 5–17 are involved in labor. It tells that in 2024, two to three out of every 100 children in Indonesia were employed. Most of these child laborers are dominated by boys aged 5 to 12 who live in rural regions. This finding indicates the presence of challenges in fulfilling children's welfare, particularly in areas with specific economic characteristics, such as the informal sector and agriculture.

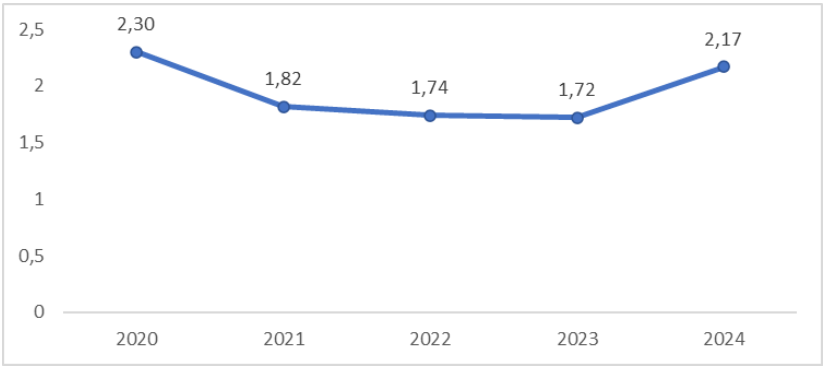

Figure 1. Percentage of Population Aged 5-17 Years Who Are Employed
Source: Badan Pusat Statistik (BPS)

ILO (2017) stated that the agricultural sector is the primary sector employing the highest number of child laborers compared to the industrial and service sectors. Moreover, the ILO also noted that most children working in agriculture are unpaid, as they are engaged in family-based labor. One of the subsectors in agriculture that is particularly vulnerable to child labor practices is the oil palm plantation industry, especially small-scale plantations managed by families. This vulnerability is largely driven by the rapid expansion of the palm oil industry in Indonesia (Atem, 2017). The Bureau of International Labor Affairs of United States has even included Indonesian palm oil in its annual list of goods produced with child labor. As one of the country's leading palm oil-producing regions (Figure 2), Kalimantan Island is highly vulnerable to child labor, particularly in plantation areas with limited access and inadequate oversight systems. These circumstances highlight Kalimantan Island as a key focus area in national strategies to eradicate child labor. To ensure that child labor elimination policies in Kalimantan are well-targeted, small area data are needed to accurately identify specific locations with a high risk of child labor involvement.

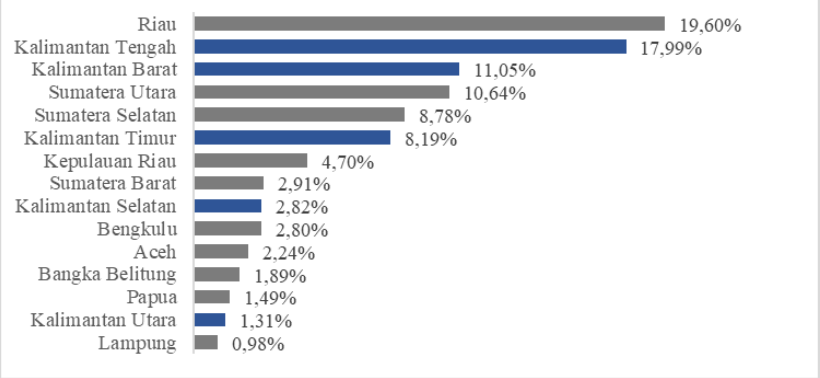

Figure 2. 15 Highest Palm Oil Producing Provinces in Indonesia by 2023
Source: World Resources Institute

The availability of child labor proportion data at the provincial level is provided by Badan Pusat Statistik (BPS) through direct estimation from Survei Angkatan Kerja Nasional (SAKERNAS). Several publications have also included direct estimates of child labor proportion at the regency/municipality level, such as The Profile of Child Labor in East Kalimantan. The precision of an estimate is measured by the Relative Standard Error (RSE) value where the estimate is considered precise if the RSE value is less than 25 percent (BPS, 2023b). However, based on direct estimation method used by BPS, 53 out of 56 regencies/municipalities still have RSE values above 25 percent. Only three regencies, namely Pulang Pisau, Penajam Paser Utara, and Nunukan, achieved the acceptable precision criterion. This indicates that the direct estimation approach is insufficiently precise to be presented. For the policy on the elimination of child labor in Kalimantan Island to be well-targeted, proper and accurate small area data is needed to identify locations with a high potential risk of child labor involvement. The Small Area Estimation (SAE) method can be used to obtain more precise estimates by utilizing additional information without having to increase survey costs (Rao & Molina, 2015).

A number of studies have identified factors that determined the emergence of child labor, both theoretically and empirically. Webbink et al. (2013) developed a new theoretical framework that explains the involvement of child labor in developing countries due to three main factors that include resources, culture, and structure. ILO & ECLAC (2022) also identified that child labor is influenced by 2 main factors, namely contextual factors related to the area where children and their parents live, and individual factors related to individual characteristics. Then, Busquet et al. (2021) that child labor is determined by a multidimensional perspective including natural capital, social capital, physical capital, human capital, economic or financial capital. Then, Thévenon & Edmonds (2019) also suggested that economic factors such as ownership of productive assets and technological advancement also affect the emergence of child labor. UNICEF (2023) suggest that social protection also plays a role in the elimination of child labor. In line with some of these theoretical studies, Kim et al. (2020) conducted an indirect estimation of child labor in India using a normal Poisson-log model by including several auxiliary variables including the child's gender, age, classification of region of residence, type of household head, social status, and sector of employment of the household head. Several studies using logistic regression also show that individual-level factors have a significant effect on child labor (Al Muflih & Wijaya, 2024).

However, there are also studies showing that factors at the individual level, household factors, and the environment simultaneously have a significant effect on the incidence of labor (Abebe & Fikre, 2021; Magdalena et al., 2021). The results of previous studies can be taken into consideration in determining the auxiliary variables in the model with the SAE method to obtain

indirect estimates. Based on this explanation, this study aims to explore the direct estimation of the variable percentage of child labor in Kalimantan Island in 2024, explore auxiliary variables for the indirect estimation of child labor proportion in Kalimantan Island in 2024, and perform indirect estimation of child labor proportion in Kalimantan Island in 2024 using SAE HB Beta method. The SAE HB Beta method is employed in this study considering that the normality assumption on the sampling distribution of proportion is difficult to satisfy when the sample size is small and or when the true proportion is around 0 or 1. Therefore, the Beta distribution is more appropriate as it naturally has a (0,1) range which inherently suitable for proportion data.

METHOD

Child Labor

According to Badan Pusat Statistik (BPS), work is an activity carried out to earn or help earn income or profit with a minimum duration of one hour during the past week. Child labor includes people aged 5-12 years who work, people aged 13-14 years who work more than 15 hours per week, and people aged 15-17 years who work more than 40 hours per week (BPS, 2023a).

Direct Estimation

Direct estimation is a method to estimate parameters only based on sample data from a specific domain or area (Rao & Molina, 2015). However, small sample sizes can lead to high standard error values. In some cases, there are also areas that have no samples. This has an impact on the accuracy of the estimation results. Relative standard error (RSE) value is generally used to assess the accuracy of the estimation result. The estimation value is considered accurate if the RSE value is less than 25 percent (BPS, 2023b). If the RSE value is in the range of 25 percent < RSE≤ 50 percent, then the estimation results should be used with caution. However, if the RSE > 50 percent then the estimation result is said to be inaccurate.

There are two approaches in direct estimation, namely design based and model (Rao & Molina, 2015). The model based approach employs a statistical model that treats population values as random variables. Estimation result under model based approach are generally independent of survey weights and the corresponded inferences are derived from the probability distribution of the assumed model. The design based approach employs survey weights in the estimation process, wherein the resulting inferences, including standard errors and confidence intervalas, are grounded in the probability distribution generated by the sampling design.

The Horvitz-Thompson (H-T) estimator is on of the most frequently used estimators within the design based approach. The population total (Y) is estimated through the following estimator:

$$\hat{Y}_{HT} = \sum_{j \in s} w_j y_j = \sum_{j \in s} \frac{y_j}{\pi_j} \tag{1}$$

where s denotes the selected sample, $w_j = \frac{1}{\pi_j}$ represents the design weight assigned to unit j, π_j is the inclusion probability of unit j, and y_j is the observed value of the variable of the interest for unit j. This estimator is design-unbiased for the population total Y.

In the case where the variable y takes binary values $y_j \in \{0,1\}$, the domain mean $(\bar{Y}_i)$ simplifies to the domain proportion (P_i). The H-T estimator for the domain proportion is subsequently obtained from the domain mean estimator and is expressed as follows:

$$\widehat{P}_i = \hat{\bar{Y}}_i = \frac{\hat{Y}_i}{\widehat{N}_i} = \frac{\sum_{j \epsilon s_i} w_j y_j}{\sum_{j \epsilon s_i} w_j} \quad (2)$$

where $s_i$denotes the sample drawn form domain U_i, $\hat{Y}_i$is the estimated total of domain i, $\widehat{N}_i$is the estimated population size of domain i, and $y_j = 1$ if unit j possesses the characteristic of interest, and $y_j = 0$ otherwise.

Small Area Estimation (SAE)

Small area refers to a domain where direct estimates with sufficient precision cannot be generated due to sample size limitations (Rao & Molina, 2015). Small Area Estimation (SAE) is an indirect estimation method that resolves the issue of limited or zero sample sizes in subpopulations, thereby enabling the acquisition of parameter estimates that are statistically reliable. The basic principle of SAE is to borrow strengths from observations of adjacent area samples and other data sources that have a wider coverage such as census and administrative data to improve the accuracy of direct estimates from surveys (Rao & Molina, 2015). Based on the availability of auxiliary variables, there are two basic models in SAE, namely area level and unit level. This study uses an area level model because the availability of auxiliary variables is only up to the area level.

In the area level model, it is assumed that θ_i is a function of the mean of the interest variable in a small area i with $i = 1, \dots, m$ and can be written as $g(\overline{Y}_i)$ for a given function $g(.)$. Furthermore, θ_i is assumed to be correlated with auxiliary variables in each small area i or can be denoted as $\boldsymbol{x_i} = \left(x_{i1}, \dots, x_{ip}\right)^T$, through a linear model as given in equation (3) called linking model.

$$\theta_i = \boldsymbol{x_i^T \beta} + b_i v_i, i = 1,2, \dots, m \quad (3)$$

where b_i is a positive constant that is generally equal to 1, $\boldsymbol{\beta} = \left(\beta_1, \dots, \beta_p\right)^T$ is a vector of regression coefficients with dimensions $p \times 1$, $\boldsymbol{x_i^T}$ is a vector of auxiliary variables, v_i is the random effect area which is assumed $v_i \sim N(0, \sigma_v^2)$, and σ_v^2 is the variance of the random effect area. Then, to obtain inferences about the small area mean $(\overline{Y}_i)$, it is assumed that the direct estimator $\left(\hat{\bar{Y}}_i\right)$ is known and can be expressed by equation (4), which is referred to the sampling model.

$$\hat{\theta}_i = g\left(\hat{\bar{Y}}_i\right) = \theta_i + e_i, \quad i = 1, \dots, m \quad (4)$$

where e_i is the sampling error and it is assumed that $e_i \sim N(0, \sigma_i^2)$. The combination of equation (3) and equation (4) is called a *linear mixed model* and can be written as equation (5).

$$\hat{\theta}_i = \boldsymbol{x_i^T \beta} + b_i v_i + e_i, \; i = 1, \dots, m \quad (5)$$

Small Area Estimation (SAE) with Hierarchical Bayes Beta (SAE HB-Beta)

Bayesian is a statistical method for estimating model parameters. The Bayesian approach assumes that the parameters are random variables that have certain distribution (Iriawan et al.,

2020). A Bayesian model has three components, namely prior, likelihood, and posterior function. Prior distribution represents initial information about the parameter before observing the reality of the data, while the likelihood function reflects the information derived from the observed data. Posterior distribution is a combination of the initial parameter information and information from the observed data. The relationship between these three components is expressed in equation (6).

$$f(\boldsymbol{\theta}|\boldsymbol{y}) \propto f(\boldsymbol{\theta}) f(\boldsymbol{y}|\boldsymbol{\theta}) \tag{6}$$

In Hierarchical Bayesian (HB) method, the prior distribution ($f(\boldsymbol{\theta})$) is determined by the researcher to obtain the posterior distribution ($f(\boldsymbol{\theta}|\boldsymbol{y})$) of the small area parameter ($\boldsymbol{\theta}$), based on the available data ($\boldsymbol{y}$). For a model with hierarchical parameteres in the HB approach, with parameter $\boldsymbol{\lambda}$ and the subjective prior distribution $f(\boldsymbol{\lambda})$, the posterior distribution ($f(\boldsymbol{\theta}|\boldsymbol{y})$) can be formulated as equation (7) (Rao & Molina, 2015).

$$f(\boldsymbol{\theta}|\boldsymbol{y}) = \int f(\boldsymbol{\theta}, \boldsymbol{\lambda}|\boldsymbol{y}) d\boldsymbol{\lambda} \tag{7}$$

$$= \int f(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{\lambda}) f(\boldsymbol{\lambda}|\boldsymbol{y}) d\boldsymbol{\lambda} \tag{8}$$

In the HB method, where $\boldsymbol{\theta}$ represents the parameter of the interest variable, the inference can be expressed in the form ($\boldsymbol{\phi} = h(\boldsymbol{\theta})$). The parameter$\boldsymbol{\phi}$ is estimated using mean of the posterior ($\hat{\boldsymbol{\phi}}^{\boldsymbol{HB}} = E[h(\boldsymbol{\theta})|\boldsymbol{y}]$) and variance of the posterior ($V[h(\boldsymbol{\theta})|\boldsymbol{y}]$) used as a measure of precision of the estimators. Refers to equation (6), obtaining the posterior distribution and deriving its quantities, such as the mean of the posterior, requires multidimensional integration. As an alternative to solving integral in equation (6), the Markov Chain Monte Carlo (MCMC) algorithm can be applied (Rao & Molina, 2015). The MCMC method generates samples from the posterior distribution and uses simulated samples to approximate quantities from the posterior distribution.

The HB Beta is a SAE model that assumes the area-level proportions follow a Beta distribution (Liu et al., 2014). SAE with HB Beta approach uses beta distribution as a sampling model and logit as the link function in the linking model. Beta distribution is considered suitable for modeling proportion data with intervals $0 - 1$ and when the distribution of the data tends to be asymmetrical. The model of SAE with HB Beta approach follows the equation as listed below.

i. Sampling Model

$$\hat{\theta}_i|\theta_i \sim Beta\ (a_i, b_i), \quad i = 1, .., m \tag{9}$$

where θ_i is the parameter of proportion in small area i, $\hat{\theta}_i$ is the estimator of proportion in small area i, $a_i = \theta_i k$ and $b_i = (1 - \theta_i)k$ are the parameters of beta distribution. k is defined as $\left(\frac{n_i}{deff_i} - 1\right)$ where n_i is the sample size within area i and $deff_i$ is approximates design effect associated with $\hat{\theta}_i$. k is assumed to follow a gamma distribution or $k \sim Gamma\ (g_1, g_2)$ where the values of g_1 and g_2 are fixed.

ii. Linking Model

$$\hat{\theta}_i|\theta_i \sim Beta\ (a_i, b_i), \quad i = 1,..,m \tag{10}$$

$$logit(\theta_i)\ |\boldsymbol{\beta}, \sigma_v^2 \sim N\ (\boldsymbol{x_i^T\beta}, \sigma_v^2)\ , \qquad i = 1,\ldots,m \tag{11}$$

The model in equation (9) can be written as equation (10):

$$logit(\theta_i) = \boldsymbol{x_i^T\beta} + v_i\ , \qquad i = 1,\ldots,m \tag{12}$$

where $\boldsymbol{\beta}$ is the vector of regression coefficients and σ_v^2 is the variance of random effects. $\boldsymbol{\beta}$ and σ_v^2 are independent with $\beta \sim N\ (\mu_\beta, \sigma_\beta^2)$ and $v_i \sim N\ (0, \sigma_v^2)$ with $\sigma_v^2 \sim IG(c_1, c_2)$, IG is the inverse gamma distribution. The value of $c_1, c_2, \mu_\beta, \sigma_\beta^2$ is fixed.

Evaluation of SAE with HB-Beta approach is based on the relative standard error (RSE) value which can be compared with other models and direct estimation. RSE in HB-Beta can be formulated as equation (13).

$$RSE = \left(\hat{\theta}_i^{HB}\right) = \frac{\sqrt{V[h(\boldsymbol{\theta}|\boldsymbol{y})]}}{\hat{\theta}_i^{HB}} \times 100\% \tag{13}$$

Data Source

The variable of interest in this study is the proportion of child labor at the regencies/municipalities level in Kalimantan Island in 2024. The direct estimation value of child labor proportion is obtained from the data of Survei Angkatan Kerja Nasional (Sakernas) Agustus 2024. The auxiliary variables are obtained from Potensi Desa (Podes) 2024 data and Registrasi Sosial Ekonomi (Regsosek) 2022 data. The candidate of auxiliary variable that have correlation with the proportion of child labor in Kalimantan Island are presented in Table 1.

Stages of Analysis

This research conducts several stages of SAE model development which include:

1. Conducting direct estimation of child labor proportion at the level of regencies/municipalities in Kalimantan Island using Sakernas data.
2. Carrying out preliminary data exploration, includes checking data completeness, identifying outliers, and checking normality assumptions.
3. Identifying potential auxiliary variables through a literature review, followed by the selection of candidate variables based on their correlation with the logit-transformed proportion of child labor. The optimal combination of variables is then determined using the backward elimination method.
4. Performing small area estimation with the HB Beta approach using R software version 4.3.0. The packages used in modeling are "sae" and "saeHB". Estimation with the HB Beta approach begins by determining the number of iterations updates, burn in, MCMC iterations, and thin interval to be used. Then, convergence is checked based on the results of the density plot, trace plot, and autocorrelation plot.
5. Evaluating the performance of the SAE HB Beta model by comparing its estimation results with those obtained through direct estimation. The assessment of model accuracy is based on the

Relative Standard Error (RSE); the model is considered to produce accurate estimates if the RSE value is below 25 percent.

6. Assessing the consistency of the estimation results by comparing the SAE HB Beta estimates at the regencies/municipalities level with the direct estimates at the provincial level.

Table 1. Description of variables and data sources used

Variable	Description	Data Source
X1	Percentage of families living under Extra High Voltage Transmission Line	Podes 2024
X2	Percentage of families subscribed to cable telephone	Podes 2024
X3	Ratio of polyclinics per 1000 population	Podes 2024
X4	Percentage of families receiving cash transfers	Regsosek 2022
X5	Percentage of families without moveable assets	Regsosek 2022
X6	Percentage of population with chronic kidney disease	Regsosek 2022
X7	Proportion of population have employer health coverage	Regsosek 2022
X8	Proportion of population have both employer and other health coverage	Regsosek 2022
X9	Proportion of families with other main source of drinking water	Regsosek 2022
X10	Proportion of families with protected well as the main source of drinking water	Regsosek 2022
X11	Proportion of families with unprotected well as the main source of drinking water	Regsosek 2022
X12	Proportion of families with parquet/vinyls as the largest floor type	Regsosek 2022
X13	Proportion of families with rainwater as the main source of drinking water	Regsosek 2022
X14	Proportion of families with marble as the largest floor type	Regsosek 2022
X15	Percentage of villages that have business credit facilities	Podes 2024
X16	Percentage of population with chronic disease	Regsosek 2022
X17	Percentage of population without a primary school degree	Regsosek 2022
X18	Percentage of population that is vulnerable to poverty	Regsosek 2022

RESULTS AND DISCUSSION

Exploration of Direct Estimators

Direct estimation was conducted to estimate the proportion of child labor at the level of regencies/municipalities in Kalimantan Island in 2024. The summary statistics of the estimation results are presented in Table 2. There is a high disparity in the estimation of child labor in Kalimantan. The average percentage of child labor on Kalimantan Island is 1.8 percent. Lamandau, located in Central Kalimantan Province, has the lowest percentage of child labor at 0.1 percent. On the other hand, Nunukan in North Kalimantan Province, possesses the highest percentage, which is 7.2 percent.

Table 2. Summary Statistics of Direct Estimates of Child Labor Proportion in Regencies/Municipalitie Level in Kalimantan Island in 2024

Statistics	Description
N	56
Minimum	0,1
Mean	1,8
Median	1,7
Maximum	7,2

As shown on the histogram and boxplot (Figure 3), the direct estimates of child labor proportion in Kalimantan Island does not follow a normal distribution and tend to be right-skewed. This finding is supported by the results of the normality test using the Jarque-Bera method, which produced a test statistic value of 145.78 with a p-value of 0.0000. It can be concluded that at 5% significance level, the direct estimates of child labor proportion at the level of regency/municipality in Kalimantan Island does not follow a normal distribution. The violance of normality assumption in direct estimates indicates that a normality assumption is untenable for proportion data that is naturally restricted to the interval (0,1). Consequently, this finding strengthens the rationale for applying the HB Beta model, which is specifically constructed to handle the bounded and skewed distribution of proportion data, while also mitigating the unreliability of direct estimates that arises from small domain sample sizes.

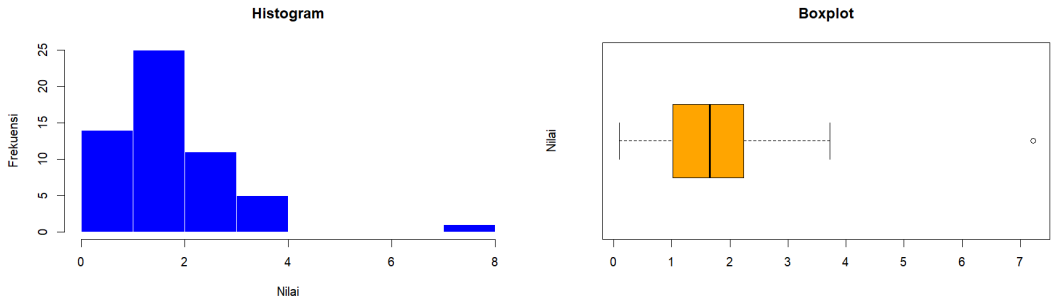

Figure 3. Histogram and Boxplot of Child Labor Proportion in Regency/Municipality Level in Kalimantan Island in 2024

To assess the quality of both direct and indirect estimation results, the Relative Standard Error (RSE) is used as an indicator of estimation precision. An estimate is considered precise when the RSE does not exceed 25 percent. If the RSE falls between 25 and 50 percent, the estimation result is considered less reliable and should be interpreted carefully. On the other hand, an RSE above 50 percent indicates a highly inaccurate estimate (Wulansari et al., 2022). With direct estimation, there are only three regencies/municipalities that have RSE smaller than 25 percent (Table 3). This indicates that only three regions have child labour indicators with sufficient estimation accuracy to be published. These regions are Pulang Pisau, Penajam Pasir Utara, and Nunukan.

Table 3. Number of Regency/Municipality by RSE Category of Direct Estimates for Child Labor Proportion in Kalimantan Island in 2024

RSE Category	Number of Regency/Municipality
RSE ≤ 25 percent	3
25 percent < RSE ≤ 50 percent	34
RSE > 50 percent	19
Total	56

Auxiliary Variable Selection

Before estimating the proportion of child labor as the target indicator, the first step is selecting relevant auxiliary variables. This process is done by analyzing the correlation between the auxiliary variables and the target variable that has been transformed into logit form. The candidate of auxiliary variables that have a correlation of more than |0.2| are then further selected by Backward Elimination. Table 4 presented the Backward Elimination results and the correlation values between the target indicators and the selected auxiliary variables.

Table 4. The Correlation Between Logit of Child Labor Proportion and Selected Auxiliary Variable by Backward Elimination

Variable	Description	Correlation
X1	Percentage of families living under Extra High Voltage Transmission Line	-0,3444
X2	Percentage of families subscribed to cable telephone	-0,2477
X3	Ratio of polyclinics per 1000 population	-0,2826
X4	Percentage of families receiving cash transfers	0,3045
X5	Percentage of families without moveable assets	0,2245
X6	Percentage of population with chronic kidney disease	0,3235
X7	Proportion of population have employer health coverage	-0,2039
X8	Proportion of population have both employer and other health coverage	-0,2279
X9	Proportion of families with other main source of drinking water	-0,2932

The variables of percentage of families receiving cash transfers, percentage of families without movable assets, and percentage of population with chronic kidney disease have a positive correlation to the logit of child labor proportion, with the highest correlation is percentage of population with chronic kidney disease of 0.3235 (Table 4). Meanwhile, the percentage of families living under extra high voltage transmission line, percentage of families subscribed the cable telephones, ratio of polyclinics per 1000 population, the proportion of the population have employer health coverage, the proportion of the population both employer health coverage and other health insurance, and the proportion of families with other main source of drinking water have a negative correlation with the logit of child labor proportion, with the highest correlation is percentage of families living under extra high voltage transmission line of -0.3444. The auxiliary variables selected in this study align with the results of previous research on child labour, including studies Busquet et al. (2021) and ILO & ECLAC (2022).

According to Busquet et al. (2021), understanding household dimensions is essential for identifying various factors, both internal and external to the value chain, that contribute to the prevalence of child labor. One relevant framework for categorizing these factors is the Sustainable Livelihood Approach (SLA), which classifies determinants into five types of capital: natural capital, human capital, economic capital, physical capital, and social capital. In this study, the auxiliary variables are structured based on this framework. Variables such as X6, X7, and X8 are categorised under human capital. The X3 is classified as physical capital, while the X5 represents economic capital. Meanwhile, the X9 is associated with natural capital. In addition, ILO & ECLAC (2022) assert that poverty is a significant factor contributing to child labor. This study identifies numerous relevant variables, including X1, X2, and X4, that can reflect poverty levels across different locations. The integration of these variables offers a thorough comprehension of the socioeconomic factors that may lead to child labour.

SAE HB Beta

To obtain a reliable estimate of child labor proportion as a target indicator, the modeling is continued with the Hierarchical Bayesian-based Small Area Estimation (SAE HB) approach with Beta distribution. This model accommodates all selected auxiliary variables as listed in Table 4. The estimation process was conducted until the model converged, which was obtained through 90 update iterations, with 2,000 burn-ins from a total of 10,000 Markov Chain Monte Carlo (MCMC) iterations and setting the thinning value at 5. The convergence of the SAE HB Beta model is shown in the following diagnostic graphs.

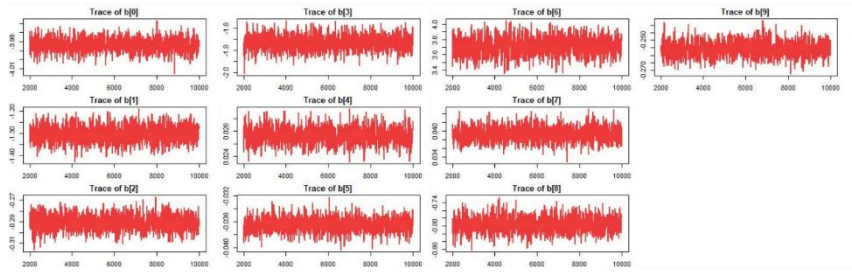

Figure 4. Trace Plot of SAE HB Beta Parameters

The trace plot is a time series graph describing the generation of a Markov chain (parameters) at each iteration, where the y-axis represents the parameter value and the x-axis the number of iterations. This graph is used to evaluate the convergence of parameters after the estimation process with the MCMC method (Iriawan et al., 2020). Trace plots that show good performance generally resemble random patterns, without a clear trend. The average value of parameter estimation has reached stability, and there is no periodic pattern, which ind busq icates that the entire Markov chain has been in a stationary condition (Figure 4).

Density plot is a graph that shows the distribution of the parameter values generated (Iriawan et al., 2020). The model has reached convergence as shown in the density plot with a smooth and symmetrical distribution pattern, resembling the normal distribution for all parameters in the model (Figure 5). In addition, the convergence of the model is also supported by the results of the autocorrelation plot. Autocorrelation plot is a graph of Autocorrelation Function (ACF) values that shows the correlation between samples for each iteration (Iriawan et al., 2020). In Figure 6, the autocorrelation plot shows a value close to one at lag 0 and a sharp decrease (cut off) immediately afterwards. This pattern indicates the absence of interparameter autocorrelation and the rapid

convergence process. Thus, the three diagnostic plots provide evidence that all parameters in the SAE HB Beta model have reached convergence.

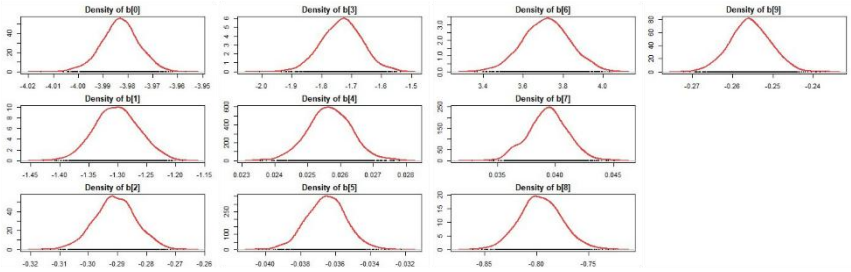

Figure 5. Density Plot of SAE HB Beta Parameters

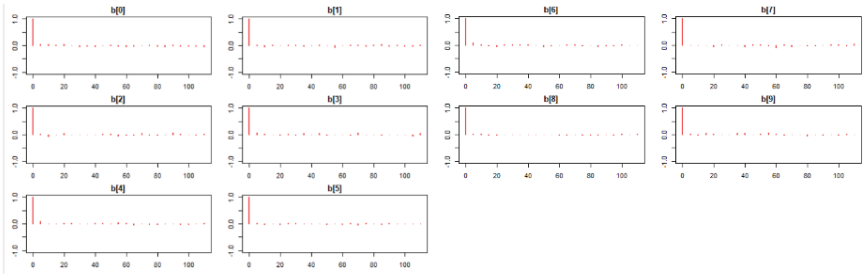

Figure 6. Autocorrelation Plot of SAE HB Beta Parameters

Coefficient estimates of the SAE HB Beta model parameters that have reached convergence are presented in Table 5. In the Bayesian inference framework, a parameter is considered significant if its credible interval does not cross zero (Iriawan et al., 2020). Thus, all the auxiliary variables used in the SAE HB Beta model are proven to have a significant influence on the direct estimation of child labor proportion in the Kalimantan Island (Table 5).

Table 5. Coefficients Estimates of SAE HB Beta Model

Variable	Mean	Std. Deviation	2,5%	97,5%
Intercept	-3.9831	0.0074	-3.9986	-3.9688
X1	-1.3017	0.0377	-1.3730	-1.2267
X2	-0.2908	0.0071	-0.3049	-0.2769
X3	-1.7329	0.0669	-1.8623	-1.5994
X4	0.0257	0.0007	0.0243	0.0270
X5	-0.0365	0.0011	-0.0388	-0.0343
X6	3.7157	0.1155	3.4896	3.9412
X7	0.0393	0.0017	0.0360	0.0426
X8	-0.7961	0.0201	-0.8344	-0.7553
X9	-0.2556	0.0050	-0.2653	-0.2457

The next step is to evaluate the reliability of the model through the Relative Standard Error (RSE) value of the indirect estimate of the indicator, based on the SAE HB Beta model estimation results. SAE HB Beta model can significantly reduce the RSE value compared to the direct

estimation results (Figure 7). The smallest RSE value obtained from the indirect estimation is 3.616 percent, namely Kabupaten Nunukan, while the highest value reaches 23.884 percent, namely Kabupaten Lamandau, with an average of 8.300 percent. This finding shows that the SAE HB Beta model is not only effective in reducing the RSE value substantially but also succeeds in producing a reliable estimation of the indicator of child labor proportion at level regency/municipality in Kalimantan Island. Considering that all RSE values are below the 25 percent threshold, so the estimation results can be presented.

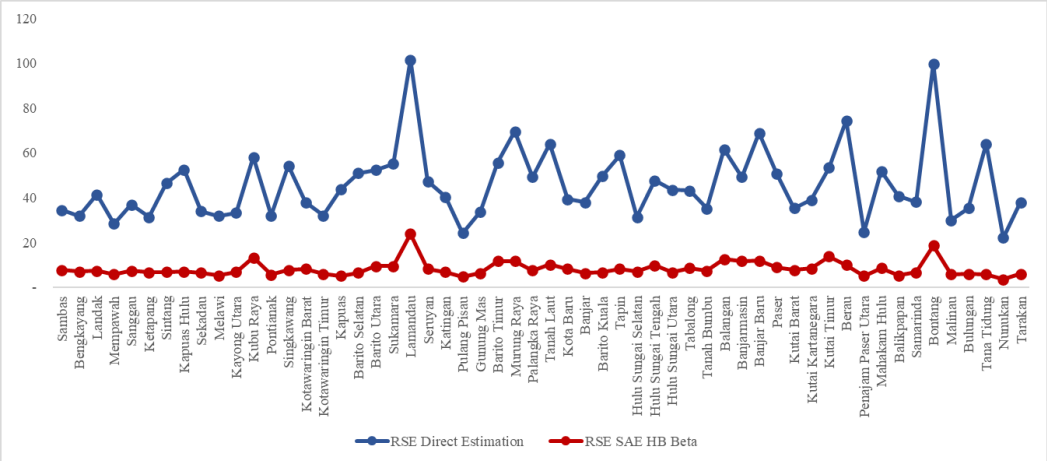

Figure 7. Comparison of RSE Value from Direct Estimation and SAE HB Beta

Once the estimator is confirmed to be sufficiently reliable, the next step is to assess the consistency of the model. A model is considered consistent if the direct estimate at the provincial level lies within the range of the indirect estimates across the regencies/municipalities within that province. It is important to note that the provincial-level direct estimates, obtained from BPS, have all met the reliability standard, with RSE values below 25 percent.

Table 6. Consistency Check of Indirect Estimates of Child Labor Proportions Using SAE HB Beta Model for at The Provinces Level in Kalimantan Island 2024

Provinces	Minimum	Maximum	Provinces Value	Description
West Kalimantan	0.49	3.28	1,8	Consistent
Central Kalimantan	0.11	3.67	2	Consistent
South Kalimantan	0.51	2.18	1,3	Consistent
East Kalimantan	0.25	3.44	1,7	Consistent
North Kalimantan	2.30	7.15	3,9	Consistent

All provincial direct estimator values are within the range of the distribution of the minimum and maximum values of the corresponding regency/municipality indirect estimators (Table 6). This finding indicates that the SAE HB Beta model can produce consistent estimates for 56 regencies/municipalities in Kalimantan Island.

CONCLUSION

Based on the results and discussion presented earlier, there are several conclusions can be drawn. First, the direct estimation results show that Nunukan has the highest percentage of child labor in Kalimantan Island in 2024, while Lamandau has the lowest. Second, the evaluation of the

direct estimation results shows that only 3 out of 56 regencies/municipalities have RSE values below 25 percent. This indicates that the direct estimation data in most areas of Kalimantan Island have not reached an adequate level of accuracy to be used as a basis for publication or policy formulation. Third, the exploration results show that nine auxiliary variables are theoretically and statistically related to the percentage of child labor. Fourth, the estimation results using the SAE HB Beta method show that all regencies/municipalities have RSE values below 25 percent, with the lowest value recorded at 3.616 percent and the highest at 23.884 percent. Model evaluation shows that the estimates are both convergent and consistent. Therefore, the estimation results can be used to support decision-making in child labor elimination strategies at the regencies/municipalities level across Kalimantan Island. It is expected that relevant stakeholders, particularly local governments in Kalimantan, will conduct ground-checking activities to ensure that the estimation results reflect actual field conditions.

REFERENCE

Abebe, G., & Fikre, S. (2021). Individual, household, and community level factors of child labor in rural Ethiopia. *Cogent Social Sciences*, *7*(1). https://doi.org/10.1080/23311886.2021.1961402

Al Muflih, M. R., & Wijaya, Y. T. (2024). Pekerja anak di kawasan timur indonesia tahun 2022: kondisi dan faktor yang memengaruhi. *Jurnal Ekonomi Kependudukan Dan Keluarga*, *1*(2). https://doi.org/10.7454/jekk.v1i2.06

Atem, A. (2017). Fenomena pekerja anak perkebunan kelapa sawit di wilayah perbatasan Kabupaten Sambas. *Raheema: Jurnal Studi Gender Dan Anak*, *4*(2), 137–147.

BPS. (2023a). *Pekerja anak di indonesia 2022: sebelum dan semasa pandemi COVID-19*.

BPS. (2023b). *Sampling Error Survei Angkatan Kerja Nasional (Sakernas)*.

BPS. (2024). *Analisis Isu Terkini 2024*.

Bureau of International Labor Affairs United States Department of Labor. (2024). *List of goods produced by child labour or forced labour*. 1–330.

Busquet, M., Bosma, N., & Hummels, H. (2021). A multidimensional perspective on child labor in the value chain: The case of the cocoa value chain in West Africa. *World Development*, *146*, 105601. https://doi.org/10.1016/j.worlddev.2021.105601

Darmika, W. D. B., & Usman, H. (2021). Pengaruh karakteristik kepala rumah tangga dan rumah tangga terhadap munculnya pekerja anak di Indonesia tahun 2018. *Seminar Nasional Official Statistics 2020*.

ILO & ECLAC. (2022). *Child labour risk identification model: methodology for designing preventives strategies at the local level.*

Foenay, E. Y., Radja, G. G., Tafui, I. S., Bait, R. N. D. P., & Killing, I. Y. (2023). Dampak psikologis pada anak yang bekerja dibawah umur. *Dampak Psikologis Pada Anak Yang Bekerja Dibawah Umur*, 91–96.

Hamdani, F., Nooraeni, R., & Lumaksono, A. (2023). Pekerja Anak dan Pendidikannya Di Masa Depan. *Jurnal Pendidikan Nonformal*, *18*(1), 24. https://doi.org/10.17977/um041v18i12023p24-35

ILO. (2017). *Global estimates of child labour: results and trends, 2012-2016*.

Iriawan, N., Ulama, B. S. S., Fithriasari, K., Chor, A. S., & Prasetyo, R. B. (2020). *Bayesian Neural Network dalam Pemodelan Small Area Estimation*. Penerbit Andi.

Kim, J., Olsen, W., & Wiśniowski, A. (2020). A bayesian estimation of child labour in India. *Child Indicators Research*, *13*(6), 1975–2001. https://doi.org/10.1007/s12187-020-09740-w

Liu, B., Lahiri, P., & Kalton, G. (2014). Hierarchical bayes modeling of survey-weighted small area proportions. *Survey Methodology*, *40*(1).

Magdalena, F., Sukamdi, S., & Rofi, A. (2021). The determinants of child labor participation in indonesia: a multilevel approach. *Southeast Asian Journal of Economics*, *9*(3), 75–108.

Mardiyanti, D., & Handayani, D. (2020). Bekerja, baik atau buruk bagi kesehatan anak?. *INOVASI: Jurnal Ekonomi, Keuangan, Dan Manajemen*, *16*(1), 167–177.

OECD. (2018). *Towards the implementation of the G20 roadmap for digitalisation: skills, business dynamics and competition*.

Peraturan Menteri Pemberdayaan Perempuan Dan Perlindungan Anak Republik Indonesia Nomor 6 Tahun 2024, 4 (2024).

Rao, J. N. K., & Molina, I. (2015). *Small area estimation* (2nd ed.). Wiley Series in Survey Methodology.

Thévenon, O., & Edmonds, E. (2019). *Child labour: causes, consequences and policies to tackle it.*

UNICEF. (2023). *Child work and child labour: the impact of educational policies and programmes in low-and middle-income countries*.

Webbink, E., Smits, J., & de Jong, E. (2013). Household and Context Determinants of Child Labor in 221 Districts of 18 Developing Countries. *Social Indicators Research*, *110*(2), 819–836. https://doi.org/10.1007/s11205-011-9960-0