Application of Small Area Estimation for Estimating Households Living in Adequate Housing at the Subdistrict Level in DKI Jakarta

M Akbar[1,*] and N Istiana[2]

[1] Politeknik Statistika STIS, Jl, Otto Iskandardinata No.64C, Jakarta, Indonesia

*Corresponding author's email: m.akbar2209@gmail.com

Abstract. Access to adequate housing is a right of all Indonesian citizens guaranteed by the 1945 Constitution and is part of the Sustainable Development Goals (SDGs), specifically Goal 11. DKI Jakarta is the province with the second-lowest percentage of households living in adequate housing in Indonesia. Estimation at the subdistrict level is needed to support the policy on affordable vertical housing development initiated by the DKI Jakarta Department of Public Housing and Settlement Areas. Direct estimation at the subdistrict level based on the Susenas sampling design would result in inaccurate estimators. To address this issue, this study applies the Small Area Estimation (SAE) method using the Empirical Best Linear Unbiased Prediction (EBLUP) model and the Hierarchical Bayes (HB) Beta model, which leverage auxiliary variables to improve precision. The findings reveal that the HB Beta model provides the best estimates in measuring the percentage of households living in adequate housing in DKI Jakarta in 2024, producing accurate estimates across all subdistricts
Keyword: Adequate Housing, EBLUP, HB Beta, Small area Estimation

1. Introduction

Access to adequate housing is one of the fundamental human rights recognized by the United Nations (UN) and guaranteed by the 1945 Constitution of Indonesia. However, the realization of adequate housing remains a challenge that requires attention. Ensuring access to adequate housing is also one of the targets outlined in the Sustainable Development Goals (SDGs), specifically in Goal 11.1, which aims by 2030 to ensure access for all to adequate, safe, and affordable housing and basic services, and to upgrade slums. According to UN-Habitat, adequate housing is defined as housing that meets five key criteria: sufficient living space, structural durability, access to clean drinking water, availability of improved sanitation facilities, and secure tenure [1]. Meanwhile, Statistics Indonesia (BPS) adopts the same indicators as UN-Habitat but excludes the criterion of secure tenure. To be classified as adequate housing, all these indicators must be fulfilled [2]. One of the factors contributing to the existence of inadequate housing in major cities is the influx of urbanization and high population density [3]. According to Rijal and Tahir, urban areas are often the destination for migrants seeking a better life [4]. However, the high cost of housing in urban areas forces low-income migrants to settle in slum areas that are relatively more affordable, even at the expense of building safety and adequacy. Furthermore, the rapid pace of urbanization is not always matched by the government's capacity to provide adequate and affordable housing for all residents [5]

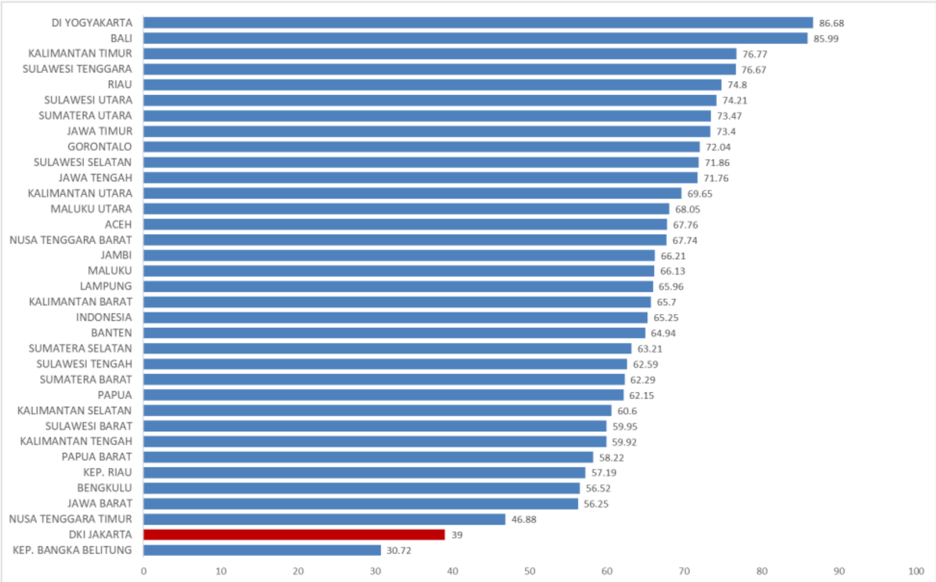


Figure 1. Percentage of Households Living in Adequate Housing by Province

Based on Figure 1, it can be seen that DKI Jakarta Province has the second lowest percentage of households living in adequate housing (RTRLH) after Bangka Belitung Islands. In 2024, this figure reached only 39 percent. This situation aligns with the high lifetime in-migration rate of 30.8% and a population density of 16,165 people per km². These challenges contribute to soaring land prices and the emergence of inadequate housing areas located along riversides, riverbanks, under bridges, and unmanaged open spaces [6]. These areas violate the Green Open Space (RTH) regulation set by the DKI Jakarta Province, which mandates a minimum of 20% public and 10% private green open space. This indicates that the issue of inadequate housing needs special attention from the government to develop more targeted and effective policies. Meanwhile, although Bangka Belitung Islands has a lower percentage of adequate housing than DKI Jakarta, there is currently no pressing urgency to estimate adequate housing at a small area level in that province.

To achieve the target set by the Department of Public Housing and Settlement Areas, which is 40.53% adequate housing coverage by 2025, DKI Jakarta has launched several programs aimed at increasing access to adequate housing and improving urban spatial planning. One of the policy directions is the development of affordable vertical housing (flats) in Transit Oriented Development (TOD) areas. TOD is an urban development concept that integrates urban design to connect people, activities, buildings, and public spaces through easy access to efficient public transportation throughout the city. Based on the study of the Housing and Settlement Area Development Plan (RP3KP), there are 21 subdistricts (Kecamatan) identified as potential TOD locations. Therefore, it becomes increasingly important to have detailed and accurate data at the subdistrict level to identify which areas most urgently require housing interventions, evaluate the effectiveness of existing programs, and ensure equitable development across regions.

To formulate effective and well-targeted policies, reliable and precise data is needed down to the subdistrict level. However, data on the percentage of households living in adequate housing is currently only available at the provincial level. This limitation is due to the insufficient sample size of the National Socioeconomic Survey (Susenas), making direct estimates at smaller geographic levels less precise. Consequently, the absence of subdistrict-level data hampers evidence-based policymaking, as spatial disparities in housing adequacy cannot be captured accurately. Generating small-area estimates would allow policymakers to allocate resources more efficiently, monitor progress locally, and design interventions tailored to specific community needs.

To overcome the limitations of direct estimation caused by small sample sizes, an indirect estimation method known as Small Area Estimation (SAE) has been developed. SAE methods improve the accuracy of estimates for small areas by utilizing auxiliary variables, thus addressing the shortcomings of survey sample sizes [7]. These auxiliary variables are sourced from the Village Potential Data Collection (Podes) and satellite imagery.

Previous research by Molina et al. compared direct and indirect estimations using SAE methods, and found that SAE significantly reduces the Mean Square Error (MSE) [8]. In addition, Utami and Ubaidillah conducted estimations of the percentage of households with access to clean drinking water, improved sanitation, and adequate housing using the Multivariate Fay-Herriot model. Their study concluded that the multivariate model produced lower MSE and Relative Standard Error (RSE) values compared to univariate models and direct estimates [9]. Another study by Buque et al. revealed that the level of adequate housing is influenced by socio-economic and health factors [10].

Therefore, this study aims to: (1) Perform direct estimation of the percentage of households living in adequate housing in DKI Jakarta Province in 2024; (2) Conduct indirect estimation of percentage of household living in adequate housing using the Small Area Estimation method; (3) Compare the results of the direct and indirect estimators; (4) Map the percentage of households living in adequate housing at the subdistrict level in DKI Jakarta Province for 2024 using the best-performing method.

2. Research Method

2.1. *Theoretical basis*

2.1.1. Adequate Housing

The concept of Adequate Housing according to Statistics Indonesia (BPS) is an adaptation of indicators developed by UN-Habitat. The indicator of secure tenure status, which is included in the UN-Habitat household concept, is not incorporated in the BPS definition of adequate housing. Therefore, the indicators used by BPS to classify adequate housing include the durability of the building, sufficient living space, access to safe drinking water, and access to improved sanitation. All of these indicators must be fulfilled for a house to be classified as adequate.

The percentage of households living in adequate housing refers to the proportion of households residing in homes that meet the requirements for building safety, sufficient living space, access to safe drinking water, and access to improved sanitation, relative to the total number of households in a given population.

$$PRLH = \frac{JRLH}{JRT} \times 100\% \tag{1}$$

explanation:

PRLH	:	Percentage of households living in adequate housing
JRLH	:	Number of households living in adequate housing
JRT	:	Total number of households

2.1.2. Small area estimation

Small Area Estimation (SAE) is a statistical method used to estimate parameters for subpopulations (small areas) that are part of a larger population, especially when the sample size in those areas is relatively small. The primary objective of SAE is to produce reliable parameter estimates—typically evaluated through standard error—for areas with insufficient or small sample sizes.

In traditional surveys, direct estimation relies solely on sample data from the area of interest without incorporating additional information from auxiliary variables. However, this approach often results in unstable or inaccurate estimates for small areas with limited or no sample data. Therefore, SAE is used as a solution by borrowing strength from auxiliary variables, which may be sourced from census data, administrative records, or other surveys, to enhance the accuracy and stability of the estimates.

2.1.3. Empirical Best Linear Unbiased Prediction (EBLUP)

In general, there are two basic models in SAE, namely the area-level model and the unit-level model. Since the availability of auxiliary variables is only at the area level, this study employs the area-level model. The area-level model equation is expressed as follows:

$$\hat{\theta}_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + b_i v_i + e_i, \quad i = 1,2,3, \dots, m \tag{2}$$

with e_i dan v_i independent, $v_i \sim N(0, \sigma_v^2)$, $e_i \sim N(0, \sigma_e^2)$

explanation:

$\hat{\theta}_i$	:	observed variable
$\boldsymbol{x}_i^T$	:	vector of auxiliary variables
$\boldsymbol{\beta}$	:	vector of regression coefficients
b_i	:	known positive constant
v_i	:	random effect area
m	:	number of observed small areas

EBLUP is one of the most commonly used methods in small area estimation. This method was developed from the Best Linear Unbiased Prediction (BLUP) approach. BLUP assumes that the variance component of the area random effect (σ_v^2) is known. However, in practice, the variance of the area random effect is difficult to determine. Therefore, the EBLUP model was developed by estimating this variance component using sample data. The estimation of small area parameters using the EBLUP method can be expressed as follows:

$$\hat{\theta}_i^{EBLUP} = \hat{\gamma}_i \hat{\theta}_i + (1 - \hat{\gamma}_i) \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}} \tag{3}$$

dengan:

$$\hat{\gamma}_i = \frac{\hat{\sigma}_v^2 b_i^2}{\hat{\sigma}_e^2 + \hat{\sigma}_v^2 b_i^2}$$

2.1.4. Hierarchical Bayes Beta

In this study, the variable of households living in adequate housing is expressed as a proportion within the range (0,1). Therefore, the SAE HB Beta model adopts a beta distribution in the sampling model, as it is suitable for modeling proportion data [11]. The hierarchical Bayesian (HB) Beta model combines a beta-distributed sampling model with a logit link function, ensuring that predicted values remain within the (0,1) interval.

In the sampling stage, the observed proportion for each small area $\hat{\theta}_i$is assumed to follow a Beta distribution as follows:

$$\hat{\theta}_i | \theta_i \sim Beta(a_i, b_i), \quad i = 1, \dots, m \tag{4}$$

where $\theta_i$denotes the true proportion parameter and $\hat{\theta}_i$is the direct estimator for area i. The parameters $a_i$and $b_i$represent the shape parameters of the Beta distribution, such that $E(\hat{\theta}_i) = \theta_i = \frac{a_i}{a_i + b_i}$. To account for area-level variation, the parameter k in the Beta distribution is modeled as $k \sim$ Gamma(g_1, g_2), where $g_1$and $g_2$are predetermined hyperparameters. The Gamma prior provides flexibility in modeling heterogeneity between areas, allowing the model to accommodate differences in precision or dispersion across small areas. While the logit link function is to assure outcome in range [0,1].

The linking model connects the proportion parameter $\theta_i$with a set of area-level auxiliary variables $\mathbf{x}_i$through a logit link, expressed as:

$$Logit(\theta_i) | \boldsymbol{\beta}, \sigma_v^2 \sim N\left(\boldsymbol{x}_i^T \boldsymbol{\beta}, \sigma_v^2\right) \quad i = 1, \dots, m \tag{5}$$

Here, the logit transformation ensures that predicted proportions stay bounded between 0 and 1 while enabling a linear relationship with the covariates. HB inference on $\theta_i$assumes flat priors for $\boldsymbol{\beta}$and σ_v^2.

Each regression coefficient β_j follows a prior distribution $\beta_j \sim N(\mu_{\beta_j}, \sigma^2_{\beta_j})$, while the area-level variance σ_v^2 follows an inverse gamma prior $\sigma_v^2 \sim IG(c_1, c_2)$, with c_1, c_2, μ_{β_j}, and $\sigma^2_{\beta_j}$ as fixed hyperparameters. The inverse gamma prior serves as a flexible choice to represent uncertainty in the variance component, allowing data to update the prior information through posterior inference.

The precision of the HB Beta estimator $\hat{\theta}_i^{HB}$ can be evaluated using the relative root mean square error (RRMSE), defined as:

2.2. *Data*

This study uses data sources from the 2024 National Socioeconomic Survey (Susenas), the 2024 Village Potential Data Collection (Podes), and satellite imagery data from the same year. The scope of this research includes all households in DKI Jakarta Province in 2024, which serve as the unit of observation for calculating the percentage of households living in adequate housing. The candidate auxiliary variables to be used are as follows:

Tabel 1. Kandidat variabel penyerta

Kode Variabel	Nama Variabel	Sumber
X_1	Population Density	Population Registry
X_2	Percentage of Households Using Electricity	Podes 2024
X_3	Ratio of Primary Schools (or equivalent)	Podes 2024
X_4	Ratio of Junior High Schools (or equivalent)	Podes 2024
X_5	Ratio of High Schools (or equivalent)	Podes 2024
X_6	Ratio of Hospitals	Podes 2024
X_7	Ratio of Community Health Centers (Puskesmas)	Podes 2024
X_8	Ratio of Midwife Practices	Podes 2024
X_9	Naturalized Difference Build-up Index	Sentinel 2
X_{10}	Air Pollution (CO)	Sentinel-5P
X_{11}	Nighttime Light Intensity	VIIRS

2.3. *Analysis Method*

The analysis follows a systematic process, beginning with Data Preprocessing. This step involves collecting satellite imagery through Google Earth Engine (GEE), with a focus on assessing cloud coverage to minimize obstructions. The imagery is then clipped and aggregated to the subdistrict level within DKI Jakarta Province, using the median statistic to represent each subdistrict from January 1, 2024, to December 31, 2024. The processed data is exported in both image and Excel formats to facilitate the creation of maps, graphs, and tables for understanding subdistrict characteristics. Furthermore, the Village Potential (Podes) data, initially at the village level, is also aggregated to the subdistrict level, and variable names are renamed from questionnaire codes to more interpretable labels.

Following preprocessing, the Direct Estimation Stage involves using the 2024 Susenas sampling design to directly estimate the percentage of households residing in adequate housing.

The subsequent Indirect Estimation Stage is where the SAE method is applied using the SaeHB package in R Studio. This stage starts with the selection of auxiliary variables using stepwise regression to ensure they have a significant influence on the outcome. Next, the Hierarchical Bayesian Beta SAE model (HB Beta) is constructed to estimate the proportion of households in adequate housing. Key parameters such as the number of updates, iterations, thinning, and burn-in are determined through trial and error until the algorithm achieves convergence.

Finally, the Model Evaluation Stage is conducted by comparing the Relative Standard Error (RSE) of the direct and indirect estimates. RSE serves as the metric for measuring the precision of an estimate. The process concludes with the Visualization Stage, where the results from the best-performing small area estimation model are used to generate maps illustrating the percentage of households living in adequate housing at the subdistrict level.

3. Result and Discussion

3.1. Descriptive Analysis of Direct Estimation

The direct estimation process of the percentage of households living in adequate housing at the sub-district level is based on the sampling design of the March 2024 SUSENAS survey. However, the SUSENAS survey is only designed to produce estimates at the provincial or district/municipality level. Therefore, generating estimates for smaller areas, such as sub-districts, will reduce the precision of the resulting estimates.

Tabel 2. Summary Statistics of Direct Estimation

Statistik Deskriptif	Direct estimation	RSE (%)
(1)	(2)	(3)
Minimum	10,01	9,31
Q1	27,95	18,35
Median	40,78	20,54
Mean	37,95	24,85
Q3	51,60	27,03
Maximum	68,80	61,47

Based on Table 2, the direct estimation results of the percentage of households living in adequate housing in DKI Jakarta Province for the year 2024 are presented. There are a total of 44 sub-districts in the province, with an average percentage of adequate housing occupancy (RTRLH) of 37.95%. The sub-district with the highest percentage is Cilandak, at 68.80%, while the lowest is Makasar, at only 10.01%.

In terms of Relative Standard Error (RSE), the maximum RSE is above 50% so the direct estimation is not reliable. To obtain statistically reliable estimates, indirect estimation methods such as Small Area Estimation (SAE) are required to produce more accurate results.

To statistically verify whether the percentage data of households living in adequate housing follows a normal distribution, a formal normality test using the Shapiro-Wilk method was conducted. This test serves as one of the assumption checks for applying the Empirical Best Linear Unbiased Predictor (EBLUP) method. The results of the normality test are presented in the following table:

Tabel 3. Shapiro-Wilk Test Results of Direct Estimator

Test Statistics (W)	*P-value*	Remarks
0,972	0,355	Estimasi langsung berdistribusi normal

Based on Table 3, the results of the normality test using the Shapiro-Wilk test indicate that the direct estimator of the percentage of households living in adequate housing is normally distributed. This suggests that the use of the EBLUP model is appropriate for estimating the percentage of households living in adequate housing.

3.2. Indirect Estimation Using the EBLUP Method

The indirect estimation process using the EBLUP method begins with the selection of candidate auxiliary variables. This step aims to reduce the total number of candidates (19 variables) and identify the most influential variables. According to Rao and Molina [7], the success of indirect estimation greatly depends on the availability of high-quality auxiliary variables.

The selection of auxiliary variables is conducted using stepwise regression to determine the best subset of variables to include in the model. In this process, all candidate variables are regressed against the direct estimates. The goal of stepwise regression is to obtain the most parsimonious model that still explains the maximum variation. This procedure identified three key variables: (1) the ratio of senior high school equivalent per 1,000 population (X_5), (2) the ratio of midwife practices per 1,000 population (X_8), and (3) the Naturalized Difference Built-up Index from Sentinel-2 (X_9).

The EBLUP model is one of the most widely used Small Area Estimation (SAE) techniques for indirect estimation. The estimation results from the EBLUP model can serve as a benchmark for comparing the accuracy and precision of other SAE models. Based on the variable selection process, the three chosen auxiliary variables were included in the EBLUP model. The results of the indirect estimation using the EBLUP model are presented in Table 5.

Tabel 4. Parameter estimation of model EBLUP

Variabel	Coefficient Estimation	*Standard error*	*P-value*
Intercept	0,364	0,021	0,000*
X_5	0,081	0,023	0,000*
X_8	0,058	0,023	0,010*
X_9	0,068	0,021	0,001*

Remarks: *) Significant at $\alpha = 5\%$

As shown in Table 4, the estimated beta coefficients ($\boldsymbol{\beta}$) of the EBLUP model are all statistically significant. These auxiliary variables contribute to improving the precision of the estimates. Furthermore, to ensure that the EBLUP estimator is valid for comparison with other SAE estimators, it is necessary to test the normality of both the area-level random effects (v_i) and the model residuals (e_i). However, as emphasized by Rao and Molina (2015), the EBLUP model primarily relies on the normality assumption of the area-level random effects. The results of the normality test for the random effects of the EBLUP model are presented as follows.

Tabel 5. Normality Test Results of Random Effect Area

Variabel	Test Statistics	*P-value*	Remarks
v_i	0,981	0,676	*Random effect area follows a normal distribution*

Based on Table 6, the normality tests for both the area-level random effects and model residuals yielded p-values greater than 0.05, indicating that both components are normally distributed. Therefore, the EBLUP estimator is considered suitable for comparison with other SAE models. The summary statistics of the estimation results using the EBLUP model are presented in the following table:

Tabel 6. Summary Statistics of EBLUP Estimation

Deskriptive Statistics	EBLUP Estimate	RSE EBLUP
Minimum	10,01	8,808
Q1	27,95	16,147
Median	40,78	17,841
Mean	37,95	20,525
Q3	51,60	22,695
Maximum	68,80	57,782

As shown in Table 7, the estimation results using the EBLUP estimator still include areas with Relative Standard Error (RSE) values exceeding 25 percent. This implies that the EBLUP estimates are not yet fully accurate, despite showing improvements compared to the direct estimates. To obtain more optimal results, this study proceeds with the application of the Hierarchical Bayes model.

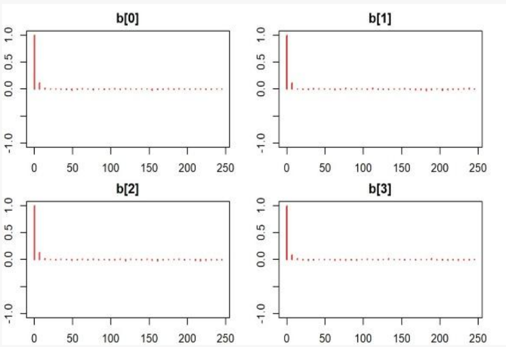
3.3. Indirect estimation with HB Beta model

Based on the direct estimation results of the percentage of households living in adequate housing, it is known that the data is in the form of proportions ranging between (0,1). Therefore, the appropriate Small Area Estimation (SAE) model to be used is the Hierarchical Bayes (HB) model with a beta distribution. According to Daamgard and Irvine [12], the beta distribution is suitable for modeling data constrained between 0 and 1. Thus, this study employs the HB Beta model, as it is considered more appropriate for estimating the percentage of households living in adequate housing compared to other distributions.

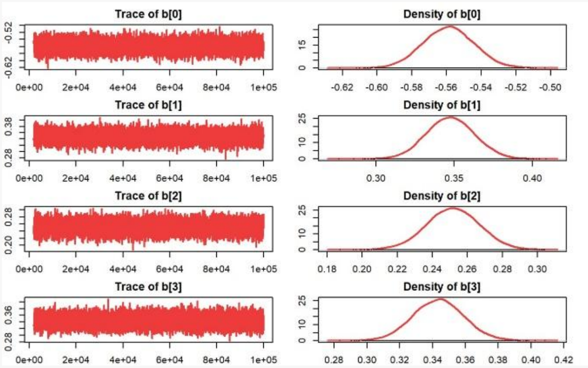
Similar to the EBLUP model, the HB Beta model also involves the selection of auxiliary variables using stepwise regression. However, a key difference in the HB Beta approach is that the candidate variables are regressed against the logit of the direct estimates. This is because the HB Beta model uses a logit link function, which ensures that the resulting estimates remain within the valid range of (0,1). The stepwise regression between the logit-transformed direct estimates and the auxiliary variables resulted in the selection of three significant variables: (1) the ratio of senior high school equivalent per 100,000 population (X_5), (2) the ratio of midwife practices per 100,000 population (X_8), and (3) the Naturalized Difference Built-up Index from Sentinel-2 (X_{17}).

The parameter estimation process under the HB Beta model is performed using a Markov Chain Monte Carlo (MCMC) algorithm that has reached convergence. This technique is applied to solve complex integration problems and to generate appropriate posterior distributions in Bayesian inference, such as for estimating the parameters of the HB Beta model. The analysis is conducted based on posterior sample data generated by the MCMC algorithm; hence, achieving convergence is critical.

In the HB Beta modeling, convergence was achieved using 100,000 MCMC iterations, 35 update iterations, a thinning interval of 7, and a burn-in of 2,000. The convergence of the model was confirmed through diagnostic plots: the autocorrelation plot showed a cut-off at the first lag, the trace plot demonstrated stationarity with no periodic patterns, and the density plots appeared smooth for all parameters.



(a)



(b) (c)

Figure 2. *Autocorrelation plot* (a), *Trace plot* (b), *density plot* (c)

After examining the diagnostic plots, the next step is to observe the estimation results of the HB Beta model parameters. Based on Table 7, it can be seen that all parameters are significant. This is indicated by the 2.5% to 97.5% credible interval of each parameter, which does not include the value zero. The estimation results of the HB Beta model parameters are presented as follows.

Tabel 7. Parameter Estimation HB Beta model

Parameter Estimate	Mean	Standard Deviation	2,5%	97,5%
$\hat{\beta}_0$	-0,5588	0,0146	-0,5872	-0,5301
$\hat{\beta}_1$	0,3476	0,0154	0,3179	0,3777
$\hat{\beta}_2$	0,2519	0,0151	0,2218	0,2814

$\hat{\beta}_3$	0,3436	0,0154	0,3138	0,3742

Based on the indirect estimation results using the SAE HB Beta method, the average percentage of households living in adequate housing in DKI Jakarta Province in 2024 is 37.95 percent. The subdistrict with the highest percentage is Cilandak, with a value of 68.33 percent, while the lowest percentage is found in Makasar Subdistrict, with 10.76 percent. A general overview of the indirect estimation results using HB Beta can be seen in the following table

Table 8. Summary Statistics of model HB Beta

Descriptive Statistics	RTRLH (%)	RSE (%)
Minimum	10,76	3,208
Q1	28,34	4,661
Median	40,74	5,708
Mean	37,95	6,738
Q3	51,30	7,509
Maximum	68,33	14,918

3.4. Model Evaluation

After estimating the percentage of households living in adequate housing in DKI Jakarta Province in 2024, the next step is to conduct an evaluation based on the precision values to determine the best model. The measure used in this study to assess the level of precision and estimation results is the RSE. The smaller the RSE value, the more precise the estimation results, and thus the best model is selected as the one with the lowest RSE.

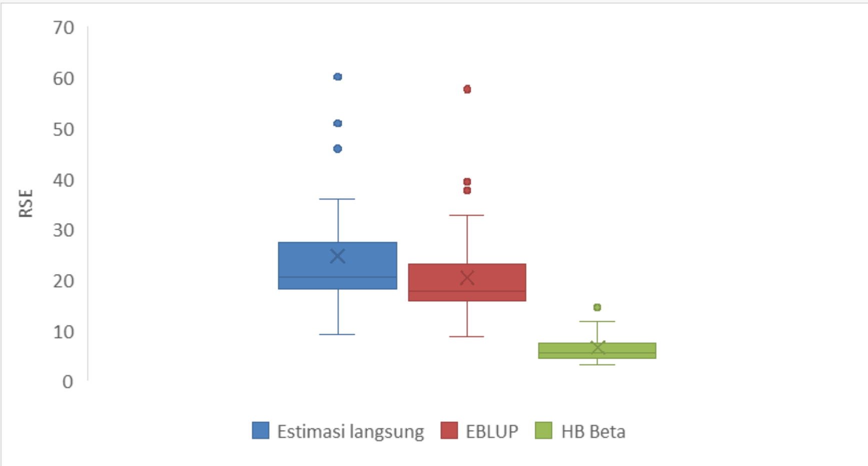


Figure 3. Boxplot of RSE for Direct Estimator, EBLUP, and HB Beta

Based on the boxplot above, it can be observed that the HB Beta model has the lowest RSE values. There is also a noticeable difference between the direct estimator, the EBLUP estimator, and the HB Beta estimator. The RSE values of the direct estimator have a relatively wide range, whereas the HB Beta estimator shows a narrower range and shorter whiskers. When viewed by subdistrict, it is evident that all RSE values of the HB Beta estimator are lower compared to those of the direct and EBLUP estimators.

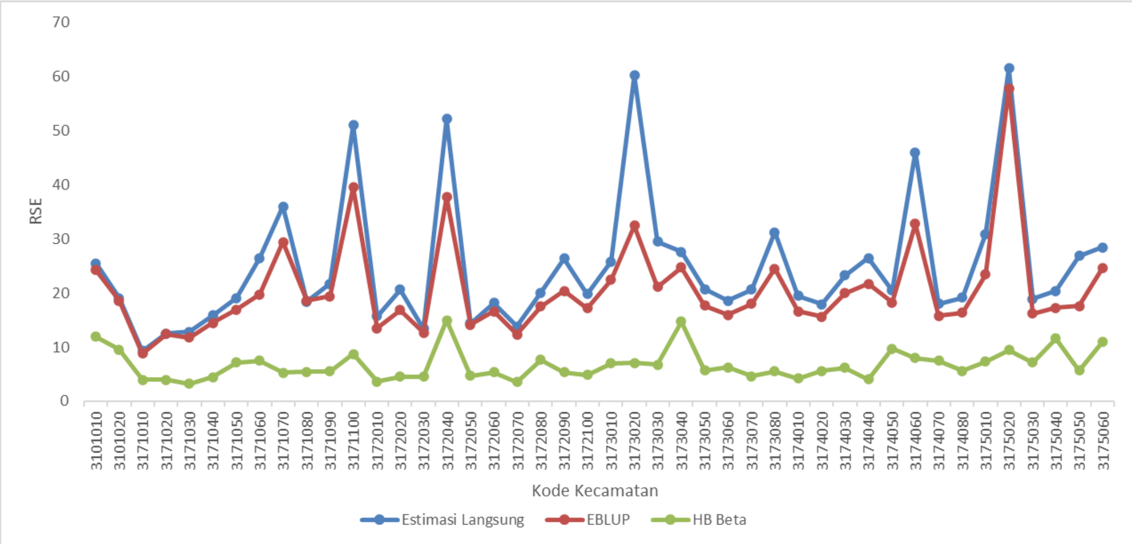


Figure 4. Dot Plot of RSE for Direct Estimator, EBLUP, and HB Beta

Furthermore, a tabulation of the number of subdistricts categorized by their RSE values is presented. Table 9 shows that the HB Beta estimator achieves superior precision, as indicated by all RSE values being below 25 percent. Therefore, it can be concluded that the indirect estimation using the SAE HB Beta method produces an accurate estimator. Consequently, the HB Beta estimator can be considered the best estimator in this study, as it demonstrates the highest precision, and will serve as the basis for mapping the percentage of households living in adequate housing across subdistricts.

In terms of Relative Standard Error (RSE), 17 sub-districts have high RSE values—exceeding 25% (Table 9). Furthermore, the direct estimation results for the percentage of households living in adequate housing cannot be considered reliable, as 4 sub-districts have RSE values above 50%. This lack of precision is due to insufficient sample sizes in the SUSENAS survey design.

Table 9. Number of Sub-districts by RSE Category

Condition	Direct Estimation	EBLUP Estimate	HB Beta Estimate
(1)	(2)	(3)	(4)
$RSE \leq 25\%$	27	38	44
$25\% < RSE \leq 50\%$	13	5	0
$RSE > 50\%$	4	1	0

3.5. Mapping of Estimation Results

To facilitate the interpretation of the percentage of households living in adequate housing in DKI Jakarta Province, a thematic map in the form of a choropleth map is presented. The Provincial Government of DKI, as stated in the Strategic Plan of the Department of Public Housing and Settlement Areas (DPRKP) 2023–2026, has set a performance target of at least 40.53% of households living in adequate housing by 2025. A thematic map related to the DPRKP target is presented below. In this map, areas that have not yet met the target (<40.53%) are marked in red, while areas that have achieved the target (≥40.53%) are marked in green.

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

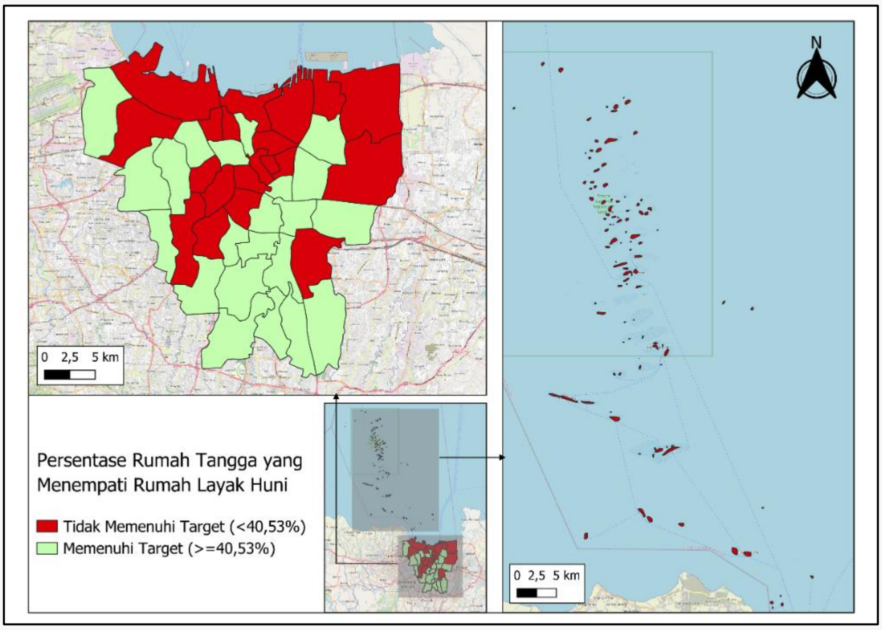


Figure 5. Thematic Map Based on DPRKP 2025 Target

Based on Figure 5, it can be seen that there are still 22 subdistricts that have not met the target for households living in adequate housing (RTRLH), with values below 40.53 percent (marked in red). The subdistrict with the lowest percentage is Makasar, at 10.76 percent, while the highest is Cilandak, at 68.88 percent. Considering the wide range of values, this indicates a significant disparity in meeting adequate housing needs across subdistricts in DKI Jakarta.

Furthermore, Figure 5 shows a clear spatial pattern in the distribution of subdistricts based on the percentage of households living in adequate housing (RTRLH). The red-marked areas, which have not yet reached the 40.53% target, tend to cluster in the northern and central parts of DKI Jakarta, particularly in North Jakarta and Central Jakarta. In contrast, South Jakarta and most of East Jakarta predominantly appear in green, indicating that the majority of their subdistricts have already met the target. This spatial pattern suggests the presence of geographic or spatial correlation in the percentage of households living in adequate housing.

4. Conclusion

Based on the research findings, the direct estimation of the percentage of households living in adequate housing in DKI Jakarta Province in 2024 proved to have low precision, with 17 out of 44 subdistricts showing a Relative Standard Error (RSE) greater than 25%, thus indicating inaccurate results. While the EBLUP method offered an improvement in precision, it still failed to produce the best results, as six subdistricts retained inaccurate estimates. The HB Beta method ultimately emerged as the best estimator, successfully improving upon both direct and EBLUP estimates to yield accurate results in all subdistricts. The subsequent mapping of the results revealed that 22 subdistricts have yet to meet the 2025 target of 40.53% set by the Department of Public Housing and Settlement Areas, and importantly, the analysis indicated a spatial relationship characterized by the clustering of neighboring subdistricts within the same housing adequacy category.

References

[1] UN-Habitat, "Slums: Some Definitions," 2007.

[2] BPS, "Indikator Perumahan dan Kesehatan Lingkungan 2023," 2023.

[3] T. W. Prayojana, Mardhatil, A. N. Fazri, and B. Saputra, "DAMPAK URBANISASI TERHADAP PEMUKIMAN KUMUH (SLUM AREA)," *J. Kependud. dan Pembang. Lingkung.*, 2020.

[4] S. Rijal and T. Tahir, "Analisis Faktor Pendorong Terjadinya Urbanisasi di Wilayah Perkotaan (Studi Kasus Wilayah Kota Makassar)," vol. 3, no. 1, 2022.

[5] T. I. Praganingrum, “Identifikasi karakteristik permukiman kumuh di kawasan kota denpasar (studi kasus permukiman jalan karya makmur, ubung kaja, denpasar utara),” *Ganec Swara*, vol. 18, no. 3, 2024.

[6] M. Ihwan, C. Fadillah, S. N. Hidayah, and B. Sumardiana, “Pemenuhan Hak Atas Rumah Layak Huni Bagi Masyarakat Miskin,” *J. Pengabdi. Huk. Indones. (Indonesian J. Leg. Community Engag. JPHI*, vol. 5, no. 1, pp. 89–101, 2022, doi: 10.15294/jphi.v5i1.50011.

[7] J. Rao and I. Molina, *Small area estimation*, vol. 88. 2015. doi: 10.1007/978-3-662-46008-5_11.

[8] I. Molina and J. Rao, “Small area estimation of poverty indicators,” *Can. J. Stat.*, p. n/a-n/a, Sep. 2014, doi: 10.1002/cjs.

[9] M. S. Utami and A. Ubaidillah, “Pendugaan Persentase Rumah Tangga yang Memiliki Akses Terhadap Air Minum Layak, Sanitasi Layak, serta Rumah Layak Huni dan Terjangkau pada Level Kecamatan Di Provinsi Papua Tahun 2019 Menggunakan Model Fay Herriot Multivariat,” *Semin. Nas. Off. Stat.*, vol. 2022, no. 1, pp. 437–448, 2022, doi: 10.34123/semnasoffstat.v2022i1.1498.

[10] T. Crocker-Buque, G. Mindra, R. Duncan, and S. Mounier-Jack, “Immunization, urbanization and slums - A systematic review of factors and interventions,” *BMC Public Health*, vol. 17, no. 1, pp. 1–16, 2017, doi: 10.1186/s12889-017-4473-7.

[11] B. Liu, *Hierarchical Bayes estimation and empirical best prediction of small-area proportions*, vol. 11, no. 1. 2009.

[12] C. F. Damgaard and K. M. Irvine, “Using the beta distribution to analyse plant cover data,” *J. Ecol.*, vol. 107, no. 6, pp. 2747–2759, 2019, doi: 10.1111/1365-2745.13200.