Ingénierie des Systèmes d'Information
Vol. 31, No. 6, June, 2026, pp. 1811-1823
Journal homepage: http://iieta.org/journals/isi

A Random Forest–Augmented Fay–Herriot Model for Nonlinear Area-Level Small Area Estimation

Ari Shobri Bukhari[1,2*], Khairil Anwar Notodiputro[1], Indahwati[1], Anwar Fitrianto[1]

[1] Statistics and Data Science Study Program, School of Data Science, Mathematics, and Informatics, IPB University, Bogor 16680, Indonesia
[2] Directorate of Statistical Analysis and Satellite Accounts, Statistics Indonesia (BPS), Jakarta 10440, Indonesia

Corresponding Author Email: ari_shobri@apps.ipb.ac.id

https://doi.org/10.18280/isi.310604

Received: 15 December 2025
Revised: 10 March 2026
Accepted: 24 March 2026
Available online: 30 June 2026

Keywords:
small area estimation, Fay–Herriot model, Random Forest, nonlinear estimation, model-assisted survey estimation, area-level model, agricultural household expenditure

ABSTRACT

Small area estimation (SAE) provides reliable estimates for domains with limited sample sizes. However, conventional area-level models, particularly the Fay–Herriot model, may lose efficiency when the relationship between auxiliary information and the target variable is nonlinear. This study develops a Random Forest–augmented Fay–Herriot (FHRF) model to extend the traditional Fay–Herriot framework for nonlinear area-level SAE. Two model formulations are proposed. The first incorporates Random Forest modelling into the Fay–Herriot residual structure, whereas the second replaces the linear predictor with a nonlinear machine-learning predictor while retaining Fay–Herriot-based empirical variance estimation. The proposed models were evaluated using simulation experiments under different nonlinear settings and an empirical application involving agricultural household per capita expenditure in Indonesia. The simulation results showed that the proposed models achieve lower estimation errors than the conventional Fay–Herriot estimator, with reductions in relative root mean squared error (RRMSE) of 11%–21% under nonlinear conditions. The empirical results further showed that the Random Forest–based specification decreased average relative standard error (RSE) from 7.46% to 6.35% and reduced the number of domains exceeding the 25% reliability threshold. Overall, the proposed framework provides a flexible extension of the Fay–Herriot model and offers a practical solution for SAE problems where nonlinear associations limit the performance of traditional linear approaches.

1. INTRODUCTION

Small area estimation (SAE) is a statistical method for producing reliable estimates for subpopulations with limited or no direct samples [1]. It is used when area-specific sample sizes are too small to achieve adequate precision with direct estimates [2]. SAE addresses this limitation by "borrowing strength'' across domains using auxiliary information—typically from censuses or administrative registers—and by modeling inter-area variability to stabilize estimates [2, 3]. In official statistics, SAE is crucial because policymakers require disaggregated indicators for evidence-based decision-making [4, 5]. However, standard SAE models based on linear mixed models may perform poorly when the assumptions of linearity, normality, and homoscedasticity are violated, leading to biased or inefficient estimators [2]. To address these issues, various extensions have been developed, including flexible distributions, such as GB2 [6], response transformations [7], and nonparametric or semiparametric methods [8, 9].

More recently, machine learning has been introduced to enhance robustness against model misspecification and to capture complex nonlinear relationships. Machine learning methods have also been applied successfully in area-level contexts, such as regency-level production modelling in Indonesia using Random Forest [10, 11]. A notable contribution to integrating machine learning into SAE is the mixed effects random forest (MERF) framework by Krennmair and Schmid [12], which improves prediction under nonlinear structures. Extensions such as the generalized mixed effects random forest (GMERF) for count data further illustrate the potential of machine learning-assisted SAE models [13]. However, most existing applications of machine learning in SAE are based on unit-level models, which require access to microdata.

In practice, reliance on unit-level data poses significant challenges for official statistics due to confidentiality constraints, infrequent collection cycles, and higher operational costs. Consequently, there is a strong demand for methodologies that operate directly on area-level covariates, which are more readily available and routinely maintained by statistical agencies. Despite the growing interest in machine learning-assisted SAE, the literature on nonlinear area-level SAE remains limited. Existing area-level SAE with machine learning approaches typically or implicitly treat prediction as a direct regression problem and do not preserve the hierarchical structure, sampling-error variance, or empirical

Bayes shrinkage inherent in the Fay–Herriot framework. As a result, such approaches do not provide area-level mean squared error (MSE) estimation or compatibility with likelihood-based inference.

Beyond unit-level frameworks such as MERF, several recent studies have explored machine learning for area-level SAE. Viljanen et al. [14] employed extreme gradient boosting (XGBoost) to predict social and health indicators at the domain level, demonstrating the capacity of machine learning to capture nonlinear relationships across areas. Michal et al. [15] investigated Random Forest and LASSO for areal prediction combined with split-conformal uncertainty intervals. Tzavidis [16] discusses the broader context of machine learning use for small area estimates across diverse data sources, underscoring both the promise and the methodological challenges of moving beyond classical models. Although these methods offer enhanced nonlinear predictive power, they generally do not retain the sampling-error structure, shrinkage behavior, and expectation–maximization (EM)-based variance estimation that characterize the classical Fay–Herriot model. Consequently, their direct applicability within official statistics workflows remains limited.

To address these gaps, this study extends the classical Fay–Herriot model by integrating Random Forest at the area level while preserving the inferential structure of the Fay–Herriot framework. The Fay–Herriot model effectively combines direct survey estimates with area-level auxiliary variables. However, its linear specification may perform poorly when relationships are nonlinear or heterogeneous across areas [1, 3]. In contrast, Random Forests are capable of capturing complex nonlinear interactions, are robust to multicollinearity, and perform well in high-dimensional settings [15-18]. Recent research has demonstrated the feasibility of integrating flexible machine learning methods into the Fay–Herriot framework to model nonlinear relationships while retaining the hierarchical structure of area-level small area estimation [19]. Therefore, combining Random Forest with the Fay–Herriot model offers a promising strategy to enhance robustness while maintaining the advantages of area-level SAE.

Random Forest is chosen as the nonlinear learner for several methodological and practical reasons. First, Random Forest provides flexible nonlinear function approximation without imposing strong parametric assumptions and exhibits stability under multicollinearity, heteroskedasticity, and noise conditions commonly encountered in area-level SAE [17]. Second, Random Forest requires relatively limited hyperparameter tuning and performs reliably in tabular data settings with a modest number of areas, whereas boosting algorithms and neural networks may require more extensive tuning and optimization to achieve reliable performance and avoid overfitting [18-21]. Third, Random Forest offers built-in out-of-bag (OOB) error estimation and variable importance measures, facilitating diagnostic assessment and interpretability [15, 16]. These properties make Random Forest particularly well aligned with the operational constraints of official statistics.

The proposed Random Forest–Augmented Fay–Herriot (FHRF) model embeds Random Forest within the Fay–Herriot framework, with variance components estimated using the expectation–maximization (EM) algorithm. EM-based estimation mitigates the risk of zero or negative variance component estimates, a common issue in Fay–Herriot estimation [22], and supports likelihood-consistent inference. Two complementary extensions are introduced: (i) a correction-based approach (FHRF-C), which models nonlinear residual structures while retaining the linear fixed-effects predictor, and (ii) a replacement-based approach (FHRF-R), which substitutes the linear predictor with a Random Forest function. This study is the first to integrate Random Forest into the Fay–Herriot model estimated via the expectation–maximization algorithm (hereafter referred to as the FH–EM framework) for nonlinear area-level SAE, while preserving area-level MSE quantification and compatibility with official statistics workflows.

The methodological contributions of this study are threefold. First, we propose two Random Forest–augmented Fay–Herriot models that operate at the area level while retaining the core inferential properties of the classical Fay–Herriot framework. Second, we provide likelihood-consistent variance-components estimation via the EM algorithm, enabling valid uncertainty quantification. Third, we evaluate the proposed methods through controlled nonlinear simulation studies and an empirical application based on official statistics. The practical relevance of the proposed approach is illustrated by estimating per capita expenditure among agricultural households in Indonesia, where area-level auxiliary variables derived from Village Potential Statistics (Potensi Desa) may exhibit inherently nonlinear relationships with the target variable.

2. METHOD

2.1 Conceptual basis of the proposed Random Forest–augmented Fay–Herriot model

Machine learning has recently been integrated into SAE, most notably through the MERF of Krennmair and Schmid [12], which extends the mixed-model framework of Hajjem et al. [23]. MERF modifies the EM algorithm of Wu and Zhang [24] by replacing the fixed-effect term $\boldsymbol{X_i^T\beta}$ with OOB Random Forest predictions. While MERF works effectively for unit-level SAE, it cannot be directly transferred to the area-level Fay–Herriot model due to two reasons: (1) Fay–Herriot responses are area means, not unit-level outcomes; and (2) Random Forest minimizes prediction error [25], whereas the EM algorithm maximizes likelihood; thus, their objectives are not naturally aligned.

To retain the likelihood principles of Fay–Herriot while enabling nonlinear prediction, the proposed approach applied Random Forest after the FH–EM algorithm converges. This ensures that variance estimation follows the standard Fay–Herriot likelihood and the fixed-effect predictor is either replaced or corrected using Random Forest, depending on the variant considered.

The original Fay–Herriot model [26] is given in Eq. (1):

$$\hat{\theta}_i = \boldsymbol{X_i^T\beta} + z_i v_i + e_i, i = 1, \cdots, m \quad (1)$$

where, i represents the area; $\boldsymbol{X}_i$ is a $p \times 1$ vector of area-level covariates; $\boldsymbol{\beta}$ is the corresponding vector of regression parameters; $v_i \sim N(0, \sigma_v^2)$ is a random component which captures the unobserved heterogeneity among areas; and $e_i \sim N(0, \sigma_{e_i}^2)$ is sampling error, where $\sigma_{e_i}^2$ is known from the sampling design; and $z_i = 1$ in the classical Fay–Herriot specification. The covariance structure is $\Omega = \boldsymbol{ZGZ}^T + \boldsymbol{S}$,

with $\boldsymbol{G} = \boldsymbol{I_m}\sigma_v^2$ and $\boldsymbol{S} = \boldsymbol{I_m}\boldsymbol{\sigma_e^2}$, where $\boldsymbol{I_m}$ denotes the $m \times m$ identity matrix.

The key idea of the FH–RF approach is to retain the random-effects structure as expressed in Eq. (2):

$$\hat{\theta}_i = (\boldsymbol{X_i^T\beta}) + (z_i v_i + e_i) \tag{2}$$

While substituting or correcting the linear fixed-effect component with a Random Forest predictor to capture nonlinear relationships that may be present in socioeconomic and area-level data [27].

2.2 Expectation–maximization algorithm for the Fay–Herriot model

Valdez et al. [22] proposed a Fay–Herriot-specific EM algorithm to ensure stable estimation of σ_v^2 and avoid zero/negative values. The algorithm iterations are shown in Eqs. (3)–(5):

- E-step (Eq. (3)):

$$\widehat{\boldsymbol{v}}^{(r+1)} = \left(\frac{1}{\hat{\sigma}_v^{2,(r)}}\boldsymbol{I_m} + \boldsymbol{S}^{-1}\right)^{-1}\boldsymbol{S}^{-1}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{X}^T\boldsymbol{\beta}^{(r)}\right) \tag{3}$$

- M-step (Eq. (4) and Eq. (5)):

$$\widehat{\boldsymbol{v}}^{(r+1)} = \left(\frac{1}{\hat{\sigma}_v^{2,(r)}}\boldsymbol{I_m} + \boldsymbol{S}^{-1}\right)^{-1}\boldsymbol{S}^{-1}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{X}^T\boldsymbol{\beta}^{(r)}\right) \tag{4}$$

$$\hat{\sigma}_v^{2,(r+1)} = \frac{1}{m}\left(\widehat{\boldsymbol{v}}^{(r+1)T}\widehat{\boldsymbol{v}}^{(r+1)} + tr\left(\frac{1}{\hat{\sigma}_v^{2,(r)}}\boldsymbol{I_m} + \boldsymbol{S}^{-1}\right)^{-1}\right) \tag{5}$$

The updating process is repeated until the change in the parameter estimates becomes very small, for instance, below 10^{-5}. The EM output is then used as the foundation for integrating Random Forest to produce the two model variants below.

2.3 Proposed model 1: Fay–Herriot with Random Forest correction

FHRF-C retains the linear Fay–Herriot predictor but augments it with a nonlinear Random Forest correction, as shown in Eq. (6):

$$\hat{\theta}_i = f^*(\boldsymbol{X_i}) + v_i + e_i \tag{6}$$

where, the corrected fixed-effect component is defined in Eq. (7):

$$f^*(\boldsymbol{X_i}) = \boldsymbol{X_i^T}\widehat{\boldsymbol{\beta}} + \hat{f}(\boldsymbol{X_i}) \tag{7}$$

Steps:

1) Perform FH–EM to obtain $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{v}}_{\boldsymbol{i}}$, and $\hat{\sigma}_v^2$.
2) Construct residuals using Eq. (8):

$$\hat{\boldsymbol{\varepsilon}}_i = \hat{\theta}_i - \boldsymbol{X_i^T}\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{v}}_{\boldsymbol{i}},\ i = 1, \cdots, m \tag{8}$$

3) Fit Random Forest to $(\hat{\boldsymbol{\varepsilon}}_i, \boldsymbol{X}_{(i)})$ using bootstrap sampling (SRSWR) to construct the Random Forest model. Predictions $\hat{f}\left(\boldsymbol{X}_{(i)}\right)$, as shown in Eq. (9), rely solely on OOB estimates to avoid overfitting:

$$\hat{f}(\boldsymbol{X_i}) = \left[\hat{f}(\boldsymbol{X_1}), \cdots, \hat{f}(\boldsymbol{X_m})\right]^T \tag{9}$$

4) The corrected fixed-effect $f^*(\boldsymbol{X_i})$ is then computed using Eq. (7).
5) SAE is carried out using the Empirical Best Linear Unbiased Predictor (EBLUP) [1] in Eq. (10):

$$\begin{aligned}\hat{\theta}_i^{FHRF-C} &= \left(T\left(\hat{\theta}_i\right)\right)(\sigma_v^2) \\ &= f^*(\boldsymbol{X_i}) + \frac{\sigma_v^2}{\left(\sigma_v^2 + \sigma_{e_i}^2\right)}\left(\hat{\theta}_i - f^*(\boldsymbol{X_i})\right)\end{aligned} \tag{10}$$

This model is suitable when the Fay–Herriot linear predictor is moderately misspecified but still informative.

2.4 Proposed model 2: Fay–Herriot with Random Forest replacement

FHRF-R replaces the entire Fay–Herriot linear predictor with a nonparametric Random Forest function, yielding Eq. (11):

$$\hat{\theta}_i = \hat{f}(\boldsymbol{X_i}) + v_i + e_i \tag{11}$$

Steps:

1) Perform FH–EM to obtain $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{v}}_{\boldsymbol{i}}$, and $\hat{\sigma}_v^2$.
2) Construct an adjusted response using Eq. (12):

$$\hat{\theta}_i^* = \hat{\theta}_i - \hat{v}_i,\ i = 1, \cdots, m \tag{12}$$

3) Fit Random Forest to $(\hat{\theta}_i^*, \boldsymbol{X}_{(i)})$ using OOB predictions.
4) Obtain Random Forest fixed-effect component $\hat{f}(\boldsymbol{X_i})$ as in Eq. (13):

$$\hat{f}(\boldsymbol{X_i}) = \left[\hat{f}(\boldsymbol{X_1}), \cdots, \hat{f}(\boldsymbol{X_m})\right]^T \tag{13}$$

5) Computation of small area estimates using the EBLUP as shown in Eq. (14):

$$\begin{aligned}\hat{\theta}_i^{FHRF-R} &= \left(T\left(\hat{\theta}_i\right)\right)(\sigma_v^2) \\ &= \hat{f}(\boldsymbol{X_i}) + \frac{\sigma_v^2}{\left(\sigma_v^2 + \sigma_{e_i}^2\right)}\left(\hat{\theta}_i - \hat{f}(\boldsymbol{X_i})\right)\end{aligned} \tag{14}$$

This variant is appropriate when the Fay–Herriot linearity assumption is strongly violated.

2.5 Practical selection between Fay–Herriot with Random Forest correction and Fay–Herriot with Random Forest replacement

Although FHRF-C and FHRF-R differ conceptually in how Random Forest is incorporated into the Fay–Herriot framework, practical applications require guidance on model selection. In general, FHRF-C is preferable when the relationship between area-level covariates and the target parameter is predominantly linear with localized nonlinear residual patterns, allowing the classical Fay–Herriot predictor to remain effective. The Random Forest component captures departures from linearity. In contrast, FHRF-R is more suitable when the mean structure itself is strongly nonlinear,

so that replacing the linear predictor with a full Random Forest function provides a better fit. Diagnostic evidence such as significant RESET or Generalized Additive Model (GAM) test results, pronounced systematic residual patterns, or substantial reductions in Random Forest OOB error relative to the linear Fay–Herriot model may indicate preference for FHRF-R, whereas milder nonlinear residual structures favor FHRF-C. From a computational perspective, FHRF-C is typically more efficient because it retains the closed-form Fay–Herriot predictor, while FHRF-R requires repeated Random Forest estimation. A summary of practical diagnostic indicators and trade-offs is provided in Table 1.

Table 1. Summarizes practical diagnostic indicators and trade-offs for selecting between Fay–Herriot with Random Forest correction (FHRF-C) and Fay–Herriot with Random Forest replacement (FHRF-R)

Criterion	FHRF-C	FHRF-R
Mean structure	Mostly linear, mild nonlinear residuals	Strongly nonlinear mean response
Residual diagnostics	Mild RESET/Generalized Additive Model (GAM) deviations	Strong RESET/GAM deviations
Out-of-bag (OOB) improvement over Fay–Herriot	Small/moderate	Large
Computational cost	Lower	Higher
Risk of overfitting	Lower	Higher (needs tuning)
Recommended when	Interpretability + shrinkage dominant	Predictive accuracy dominant

2.6 Bootstrap mean squared error estimation for Random Forest–Augmented Fay–Herriot models

Analytical MSE formulas are not available once Random Forest components replace or correct the linear predictor. Building on the methodology proposed by Rao and Molina [1], the bootstrap method is employed to evaluate the MSE of the FHRF models. The main steps of this procedure are summarized below:

1) Fit FHRF to obtain the fixed component $\hat{f}(\boldsymbol{X_i})$ and $\hat{\sigma}_v^2$.
2) For b = 1, …, B:
   a) Generate pseudo-true area values using (15):

$$\theta_{i*}^{(b)} \sim N\left(\hat{f}(\boldsymbol{X_i}), \hat{\sigma}_v^2\right) \tag{15}$$

   b) Generate synthetic direct estimates using (16):

$$\hat{\theta}_{i*}^{(b)} \sim N\left(\theta_{i*}^{(b)}, \sigma_{e_i}^2\right) \tag{16}$$

   c) Refit FHRF and obtain $\hat{\theta}_{i*}^{H^{(b)}}$.
3) Compute bootstrap MSE using Eq. (17):

$$\widehat{mse}(\theta_i) = mse_B\left(\hat{\theta}_i\right) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\theta}_{i*}^{H^{(b)}} - \theta_{i*}^{(b)}\right)^2 \tag{17}$$

4) Compute the relative standard error (RSE) using Eq. (18):

$$\widehat{RSE}(\theta_i) = \frac{\sqrt{\widehat{mse}(\theta_i)}}{\hat{\theta}_{i*}^{H^{(b)}}} \times 100\% \tag{18}$$

Bootstrap MSE captures uncertainty from both the sampling design and the Random Forest-enhanced model structure.

3. SIMULATION AND EMPIRICAL EVALUATION

3.1 Simulation design

3.1.1 Simulation design

Model performance was assessed through a simulation study to evaluate the proposed FHRF-C and FHRF-R models, relative to the direct estimator and the FH–EM model. The design follows general recommendations for SAE simulation frameworks [1, 10].

Five key data–generation factors were varied to assess robustness under both ideal and misspecified conditions:

1) Functional form (linear vs. nonlinear): The response–covariate relationship was set to either linear (L) or nonlinear (NL) to test whether Random Forest-based models improve accuracy when the true process departs from linearity [14, 22].
2) Multicollinearity (no multicollinearity vs. high multicollinearity): Covariates were generated with either no correlation (NM, $\rho = 0$) or high correlation (HM, $\rho = 0{,}9$) to mimic typical socioeconomic auxiliary data.
3) Between-area variance small/large (BVS vs. BVL): The random-effect variance σ_v^2 was set to represent small vs large domain heterogeneity, using values 3/6 for linear scenarios and 6/12 for nonlinear scenarios. These magnitudes follow typical SAE evaluations [6].
4) Within-area variation small/large (WVS vs. WVL): Unit dispersion, expressed through the coefficient of variation (CV), was set to 2/4 for linear cases and 4/8 for nonlinear cases, with a larger CV in NL scenarios to ensure sufficient nonlinear complexity.
5) Sampling-error distribution (symmetric error vs. nonsymmetric error): Errors followed either a symmetric normal distribution (S) or a skewed Pareto distribution (NS), enabling evaluation under asymmetric noise structures.

A combination of these five factors yields ten scenarios is summarized in Table 2.

Table 2. Simulation scenarios

No.	Scenario Code	Scenario ID
1	L-NM-BVS-WVS-S	L1
2	L-HM-BVS-WVS-S	L2
3	L-NM-BVL-WVS-S	L3
4	L-NM-BVS-WVL-S	L4
5	L-NM-BVS-WVS-NS	L5
6	NL-NM-BVS-WVS-S	NL1
7	NL-HM-BVS-WVS-S	NL2
8	NL-NM-BVL-WVS-S	NL3
9	NL-NM-BVS-WVL-S	NL4
10	NL-NM-BVS-WVL-S	NL4

Notes: L = Linear; NL = Nonlinear; NM = No multicollinearity; HM = High multicollinearity; BVS = Between-area variance (small); BVL = Between-area variance (large); WVS = Within-area variance (small); WVL = Within-area variance (large); S = Symmetric error; NS = Nonsymmetric error.

3.1.2 Data generation process

The simulation mimics the data collection process of Indonesia's National Socioeconomic Survey (Susenas), but with a simplified one-stage sampling design (the actual survey uses two-stage sampling). The process consisted of generating a pseudo-population and then obtaining direct estimates through the following steps:

1) Area-level population parameters.
   a) The simulation considers $m = 250$ small areas, indexed by $i = 1, 2, \dots, m$.
   b) For each area i, four covariates $(x_{1i}, x_{2i}, x_{3i}, x_{4i})$ are generated jointly, with their correlation structure varied to induce different levels of multicollinearity.
   c) The coefficient vector β determines the underlying association linking covariates to the response.
   d) Random effects (v_i) are generated from $N(0, \sigma_v^2)$.
   e) Errors (e_i) are generated from $N(0, \sigma_e^2)$.
   f) The response variable (y_i) for each area is then formed according to the model type:
      - Linear specification (Eq. (19))

$$y_i = x_{1i} + x_{2i} + x_{3i} + x_{4i} + v_i + e_i \quad (19)$$

      - Nonlinear specification (Eq. (20))

$$y_i = x_{1i}^2 + 2x_{2i}^3 + 3ln(|x_{3i}| + 1) + 4ln(|x_{4i}|) + v_i + e_i \quad (20)$$

2) Pseudo-population construction at the unit level.
   a) The population sizes N_i were allowed to vary and sampled uniformly over [7500,15000].
   b) For each area, unit-level values y_{ij} were generated around the mean y_i following a Gaussian distribution with a variance of σ_i^2, where j indexes units.
3) Sampling.
   a) From each area, units were selected through a simple random sampling (SRS) design.
   b) The sampling fraction was set to 0.0042, consistent with Susenas.
   c) The resulting sample size per area was n_i, and sample data y_{ij} were obtained for all areas.
4) Direct estimates.
   a) The direct estimator for area *i* was computed as $\hat{\theta}_i$.
   b) The sampling variance was estimated as $\psi_i = (SE(\hat{\theta}_i))^2$, where the standard error $SE(\hat{\theta}_i)$ was derived under the SRS design.

3.1.3 Performance metrics

Following standard SAE simulation practice [12], estimator performance is evaluated using the relative bias (RB) as shown in Eq. (21) and the relative root mean squared error (RRMSE) as shown in Eq. (22). Results are reported for areas $i = 1, \dots, M$ across the simulation replications $b = 1, \dots, B$.

$$RB_i = \frac{1}{B}\sum_{b=1}^{B}\left(\frac{\hat{\theta}_i^{(b)} - \theta_I}{\theta_I}\right) \quad (21)$$

$$RRMSE_i = \frac{\sqrt{\frac{1}{B}\sum_{b=1}^{B}\left(\hat{\theta}_i^{(b)} - \theta_I\right)^2}}{\theta_I} \quad (22)$$

Estimator stability across replications is evaluated using the mean relative absolute error (MRAE) and median relative absolute error (MdRAE). For each replication $b = 1, \dots, B$, MRAE summarizes relative absolute errors by their mean across areas (Eq. (23)), while MdRAE uses the median (Eq. (24)). Tracking these measures over replications provides insight into the stability of the resulting error magnitudes.

$$MRAE_b = \frac{1}{M}\sum_{m=1}^{M}\frac{\sqrt{\left(\hat{\theta}_b^{(m)} - \theta^{(m)}\right)^2}}{\theta^{(m)}} \quad (23)$$

$$MdRAE_b = median_{m=1,\cdots,M}\left(\frac{\left(\hat{\theta}_b^{(m)} - \theta^{(m)}\right)^2}{\theta^{(m)}}\right) \quad (24)$$

Combined with RB and RRMSE, MRAE and MdRAE summarize bias, precision, and replication-level stability of the estimators.

3.2 Empirical data description

The data used in this study come from Susenas 2019 (March) and Podes 2018 microdata, obtained by IPB University from Statistics Indonesia (BPS) through an official request via the SILASTIK data-access system. Due to BPS access and confidentiality restrictions, the microdata cannot be publicly disseminated.

The analysis is limited to households with heads who work in agriculture. Average per capita expenditure at the kabupaten/kota level, calculated from Susenas microdata, is used as the response variable.

From the Podes microdata, 45 auxiliary variables were aggregated to the same geographic level. To ensure interpretability and avoid redundant descriptions, the variables were grouped into five thematic categories commonly used in socioeconomic modelling:

1) Infrastructure & basic services (5 variables): indicators of electricity access (PLN), communication devices, educational and health facility availability.
2) Health & education (6 variables): presence of malnutrition cases, PAUD, elementary schools, health posts (puskesmas, polindes), midwives, and reported disease outbreaks.
3) Access & economic amenities (20 variables): coverage of roads, transport, mobile network, markets, commercial services, financial access (ATMs, bank agents, BMT, KUR/KUK), and other local service facilities.
4) Disability indicators (9 variables): village-level counts of residents with various types of disabilities, including multiple impairments.
5) Social protection and local production (5 variables): BPJS-PBI and SKTM coverage, and the presence of small-scale industries (wood processing, ceramics, pottery).

These indicators were included because they are commonly linked to household expenditure in rural and agriculture-based areas. Their combination provides area-level inputs for evaluating the proposed FHRF models.

4. RESULTS AND DISCUSSION

4.1 Simulation results

Algorithmic convergence is an essential consideration in

EM-based SAE models [28]. To ensure stable initialization, the proposed method adopts starting values from an EBLUP-based Fay–Herriot formulation [1], which is computationally efficient and structurally consistent with the FH–EM framework. The simulation consisted of 100 replications, each covering population generation, sampling, direct estimation, and model fitting. Performance is evaluated using RRMSE (precision), RB (bias), and replication-based stability measures MRAE and MdRAE.

Table 3 presents the RRMSE values across all simulation settings (L1–L5, NL1–NL5). Both FHRF variants consistently dominate direct estimation, achieving uniformly lower mean and median RRMSE. Because the RRMSE differences violated normality (Shapiro–Wilk), the Wilcoxon Signed-Rank test [29] was employed. All comparisons yield extremely small p-values (7.19×10^{-33} – 5.25×10^{-43}), confirming the statistical significance of the improvements. Beyond statistical significance, the magnitude of the gains is also substantial: the median RRMSE reduction from direct estimation to FHRF-R/FHRF-C ranged from 47.4% to 63.9% across scenarios. The largest reduction (63.9%) occurred in scenario L4, where the RRMSE decreased from 0.5357 under direct estimation to 0.2069 for the FHRF-C model, indicating a practically meaningful improvement in estimation accuracy.

Under nonlinear scenarios, FHRF-C and FHRF-R outperform FH–EM, as indicated by consistently lower mean and median RRMSE across all evaluated settings (Table 3). Specifically, both models achieved average RRMSE reductions of 11%–21% and median reductions of 14%–27% relative to the FH–EM estimator. Pairwise Wilcoxon signed-rank tests comparing FHRF-C and FHRF-R with FH–EM confirm that these improvements are statistically significant, with p-values ranging from 1.34×10^{-3} to 8.78×10^{-13}. Beyond statistical significance, the magnitude of the improvements is also meaningful: the median RRMSE reduction from FH–EM to the FHRF models ranged from 14.3% to 27.2% across nonlinear scenarios. The largest reduction (27.2%) occurred in scenario NL5, where the RRMSE decreased from 0.2192 under FH–EM to 0.1595 for the FHRF-R model, highlighting the practical advantage of Random Forest in capturing nonlinear relationships without requiring explicit functional form specification [14, 15, 22].

Table 4 and Table 5 further examine robustness when the number of areas (m) and the number of covariates (p) are varied while preserving the same nonlinear data-generating mechanism (scenario NL1). Increasing the number of areas substantially strengthens the relative efficiency of the FHRF estimators: the average RRMSE reduction relative to FH–EM rises from about 16% at $m = 100$ to approximately 30–40% at $m = 500$, indicating that the machine-learning component benefits from richer cross-area information and more stable learning of nonlinear relationships.

Table 3. Relative root mean squared error (RRMSE) results across 100 simulated replications under linear and nonlinear scenarios

		Linear					Nonlinear		
Scenario	**Model**	**Mean**	**Median**	**SD**	**Scenario**	**Model**	**Mean**	**Median**	**SD**
L1	Direct	0.3369	0.3325	0.053	NL1	Direct	0.3307	0.332	0.087
	FH–EM	**0.1181**	**0.1006**	0.066		FH–EM	0.2369	0.224	0.127
	FHRF-R	0.162	0.1397	0.088		FHRF-R	0.1912	**0.1747**	0.091
	FHRF-C	0.1499	0.1296	0.064		FHRF-C	**0.1887**	0.1756	0.094
L2	Direct	0.3287	0.3315	0.091	NL2	Direct	0.3006	0.3201	0.154
	FH–EM	**0.1139**	**0.1016**	0.089		FH–EM	0.2049	0.2084	0.168
	FHRF-R	0.1543	0.1506	0.092		FHRF-R	0.1647	**0.1526**	0.139
	FHRF-C	0.1553	0.1489	0.086		FHRF-C	**0.1627**	0.1528	0.137
L3	Direct	0.3418	0.3369	0.054	NL3	Direct	0.3286	0.3311	0.089
	FH–EM	**0.1556**	**0.1353**	0.074		FH–EM	0.2222	0.2027	0.128
	FHRF-R	0.1745	0.1479	0.084		FHRF-R	0.1917	0.1737	0.099
	FHRF-C	0.1707	0.1491	0.079		FHRF-C	**0.1889**	**0.1675**	0.102
L4	Direct	0.5357	0.5169	0.085	NL4	Direct	0.5185	0.5273	0.155
	FH–EM	**0.1389**	**0.1243**	0.068		FH–EM	0.2811	0.279	0.179
	FHRF-R	0.2191	0.1915	0.103		FHRF-R	**0.2483**	0.2132	0.194
	FHRF-C	0.2069	0.1868	0.075		FHRF-C	0.2488	**0.2077**	0.189
L5	Direct	0.3408	0.3291	0.054	NL5	Direct	0.3315	0.3275	0.083
	FH–EM	**0.1143**	**0.0936**	0.077		FH–EM	0.2269	0.2192	0.122
	FHRF-R	0.2083	0.1394	0.313		FHRF-R	**0.1785**	**0.1595**	0.088
	FHRF-C	0.1926	0.1336	0.279		FHRF-C	0.1786	0.161	0.09

Note: SD = standard deviation. The smallest are marked in bold; FH–EM: Fay–Herriot model estimated via the expectation–maximization algorithm; FHRF-C: Fay–Herriot with Random Forest correction; FHRF-R: Fay–Herriot with Random Forest replacement.

Table 4. Relative root mean squared error (RRMSE) across 100 simulation replications under varying numbers of areas and covariates

Scenario	**Setting**	**Mean RRMSE**				**Median RRMSE**			
		Direct	**FH–EM**	**FHRF-R**	**FHRF-C**	**Direct**	**FH–EM**	**FHRF-R**	**FHRF-C**
Number of areas (m)	100	0.341	0.260	0.225	0.224	0.336	0.223	0.203	0.196
	250	0.331	0.237	0.191	0.189	0.332	0.224	0.175	0.176
	500	0.335	0.255	0.195	0.196	0.329	0.238	0.171	0.170
Number of covariates (p)	3	0.337	0.275	0.225	0.208	0.329	0.265	0.219	0.190
	4	0.331	0.237	0.191	0.189	0.332	0.224	0.175	0.176
	7	0.337	0.198	0.191	0.169	0.332	0.170	0.159	0.151

Note: FH–EM: Fay–Herriot model estimated via the expectation–maximization algorithm; FHRF-C: Fay–Herriot with Random Forest correction; FHRF-R: Fay–Herriot with Random Forest replacement.

Table 5. Percentage reduction in relative root mean squared error (RRMSE) relative to FH–EM under varying dimensional settings

Scenario	Setting	Mean Reduction (%)		Median Reduction (%)	
		FHRF-R	FHRF-C	FHRF-R	FHRF-C
Number of areas (m)	100	15.53	15.98	10.15	13.76
	250	23.90	25.54	28.22	27.56
	500	30.48	30.24	39.59	39.99
Number of covariates (p)	3	22.45	31.96	20.78	39.29
	4	23.90	25.54	28.22	27.56
	7	3.54	16.77	6.83	12.85

Note: FH–EM: Fay–Herriot model estimated via the expectation–maximization algorithm; FHRF-C: Fay–Herriot with Random Forest correction; FHRF-R: Fay–Herriot with Random Forest replacement.

Table 6. Summarizes of replication-level stability metrics for each nonlinear scenario, where the reported values represent the means across simulation replications

Scenario	MRAE			MdRAE		
	FH–EM	FHRF-R	FHRF-C	FH–EM	FHRF-R	FHRF-C
NL1	0.21	0.16	0.15	0.17	0.12	0.12
NL2	0.18	0.13	0.13	0.16	0.11	0.11
NL3	0.19	0.16	0.15	0.16	0.12	0.12
NL4	0.25	0.2	0.2	0.22	0.16	0.16
NL5	0.2	0.15	0.15	0.17	0.12	0.11

Note: MRAE: mean relative absolute error; MdRAE: median relative absolute error; FH–EM: Fay–Herriot model estimated via the expectation–maximization algorithm; FHRF-C: Fay–Herriot with Random Forest correction; FHRF-R: Fay–Herriot with Random Forest replacement.

In contrast, increasing the number of covariates slightly attenuates the relative gains, particularly at $p = 7$ (Table 5), where the improvement becomes more moderate. This pattern reflects the typical dimensionality effect in area-level SAE, in which the effective information per parameter decreases as the covariate dimension grows. The attenuation was also partly related to the simulation design, in which additional covariates were entered through quadratic transformations that alter the degree of nonlinearity. Nevertheless, both FHRF specifications consistently outperformed FH–EM across all dimensional settings, with FHRF-C showing the most stable improvements, which confirmed the robustness of the proposed framework.

Replication-level stability metrics further support the superiority of the FHRF models under nonlinear settings (NL1–NL5). Figure 1 displays MRAE values computed across 250 areas for each of the 100 replications, showing that both FHRF-C and FHRF-R consistently yield lower absolute errors than FH–EM and direct estimation. To complement the graphical interpretation, Table 6 summarizes the replication-level stability metrics using mean MRAE and MdRAE values across nonlinear scenarios. The quantitative comparison confirms the visual patterns observed in Figure 1. Across NL1–NL5, the FHRF models reduce MRAE by approximately 19%–27% relative to FH–EM, with the largest reduction observed in scenario NL5, where the MRAE decreases from 0.20 (FH–EM) to 0.15 (FHRF-R and FHRF-C). Similar improvements are observed for MdRAE, with reductions ranging from 20% to 31% across nonlinear scenarios; for example, in NL5, MdRAE decreases from 0.17 to approximately 0.12–0.11.

The MdRAE results mirror the MRAE patterns, confirming that both FHRF variants consistently deliver more minor replication-level errors than FH–EM across nonlinear scenarios. This consistent performance across replications demonstrated the robustness of the FHRF approach in capturing nonlinear dependencies, reinforcing its suitability as a reliable alternative to conventional SAE methods.

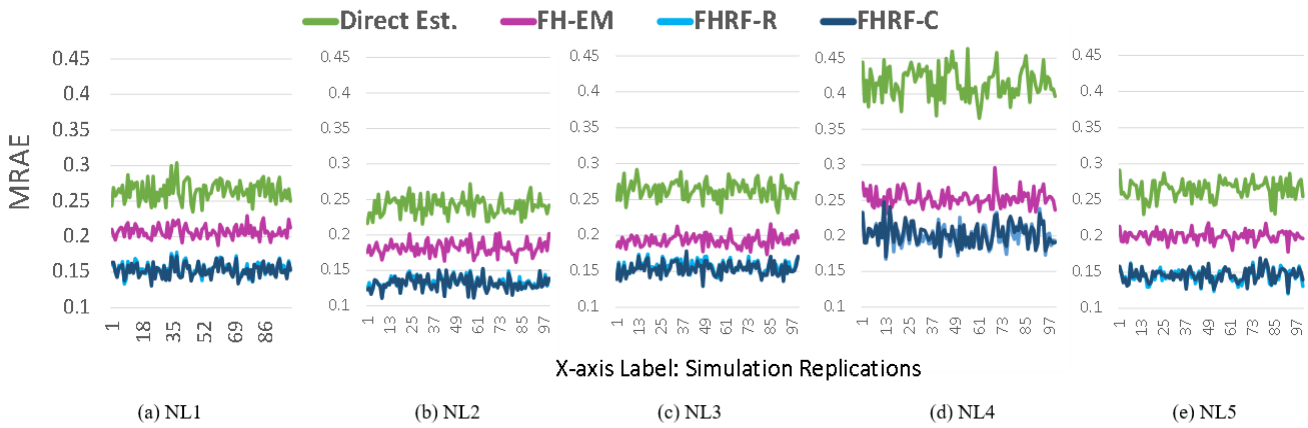


Figure 1. Mean relative absolute error (MRAE) across 100 simulation replications (250 areas each) under the nonlinear scenario group

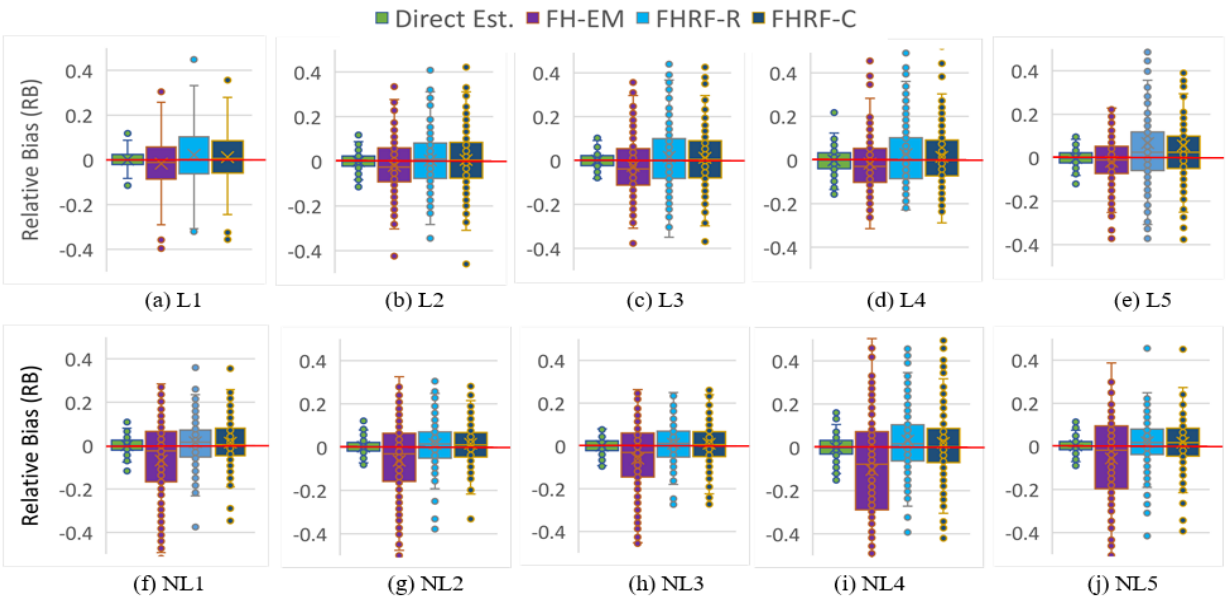


Figure 2. Boxplots of the relative bias (RB) distribution under the simulation scenarios
Note: The red line indicates zero relative bias (RB) (reference line).

In linear scenarios, FH–EM remains the best-performing model, producing the lowest RRMSE values across L1–L5 (Table 3), indicating that FH–EM remained the best-performing estimator when the linear mixed-model assumptions were correctly specified. This result was theoretically expected because FH–EM is a maximum likelihood estimator under the Gaussian–linear mixed model, thereby enjoying optimal properties such as consistency and efficiency. This pattern was consistent with the results of Krennmair and Schmid [12], who found that unit-level SAE models incorporating Random Forest (SAE-MERF) outperform classical SAE only when the underlying relationships are nonlinear. When the response–covariate relationship is truly linear, the inclusion of nonparametric learners such as Random Forest does not provide additional benefits. It may have introduced approximation noise due to data-driven partitioning and bootstrap aggregation. Hence, an important limitation of the proposed FHRF approach was that it was not designed to improve performance under correctly specified linear settings, but rather to address cases where linear specifications are inadequate.

A direct comparison between FHRF-C and FHRF-R shows that both models produced similar RRMSE distributions and consistently outperformed the benchmarks in all nonlinear scenarios (NL1–NL5). Although Table 3 reports marginally lower mean or median RRMSE for FHRF-R in a few scenarios (NL1, NL2, NL4, NL5), the Wilcoxon Signed-Rank tests indicated that FHRF-C achieved significantly lower RRMSE in almost all cases, except NL2, where the two models are statistically indistinguishable. In terms of effect size, the differences between the two variants are generally small: the median RRMSE of FHRF-R is lower than that of FHRF-C by approximately 0.13%–0.94% in scenarios L3, NL1, NL2, and NL4, whereas FHRF-C shows larger advantages of about 1.13%–7.23% in scenarios L1, L2, L4, L5, NL3, and NL5. Replication-level error patterns (Figure 1) further show that both variants displayed comparable stability in terms of MRAE. Overall, these results suggested that while FHRF-C exhibits a modest statistical advantage, both FHRF variants remain equally dependable under nonlinear conditions.

While RRMSE captures both estimator accuracy and precision, examining RB remains essential for understanding systematic deviation. Interpreting RB alongside RRMSE reflects the classical bias–variance trade-off, in which an effective estimator must control both systematic bias and sampling variability. The RB patterns in Figure 2 indicate that direct estimates have a median close to zero, with a narrow interquartile range (IQR), because positive and negative sampling errors tend to offset each other under SRS [30]. However, the relatively large RRMSE of direct estimates reflects their high sampling variability, making direct estimation inefficient for small or heterogeneous domains despite their near-zero RB.

Under linear scenarios (L1–L5), FH–EM exhibits the most favorable RB, with medians near zero and smaller IQRs than both FHRF variants (Figures 2(a)–(e)), confirming its suitability when the actual relationship is linear. In contrast, under nonlinear scenarios (NL1–NL5), FH–EM exhibited systematic underestimation (medians below zero) and a wider IQR (Figures 2(f)–(j)), indicating difficulty in capturing nonlinear structure. Both Random Forest-based models addressed this limitation: FHRF-C and FHRF-R yielded RB distributions that are more tightly concentrated around zero, with IQRs reduced by approximately 41%–59% relative to FH–EM across all nonlinear scenarios, and exhibit nearly identical bias patterns.

Taken together, these results showed that the relative performance of FH–EM and FHRF models is fundamentally driven by the underlying functional form. FH–EM remained the preferred estimator under linear data-generating processes, whereas FHRF-C and FHRF-R provided clear advantages only when nonlinear relationships are present. Accordingly, the proposed FHRF framework should be viewed not as a universal replacement for the classical Fay–Herriot model, but as a complementary extension designed for nonlinear and heterogeneous settings frequently encountered in socioeconomic applications.

4.2 Analysis of empirical data

Building on the simulation evidence, the proposed methods were applied to an empirical case study of agricultural

household expenditure in Indonesia. At the regency/city level, Statistics Indonesia commonly publishes average per capita expenditure estimates based on direct estimation [31]. While this approach is adequate for the general population, where RSE values typically fall within acceptable limits [1, 26], it becomes unreliable for subpopulations, such as agricultural households, where sample sizes are substantially smaller. Consistent with standard publication criteria (RSE $\leq$ 25%) [32], 15 regencies/cities exceeded this threshold, indicating the necessity of model-based SAE methods [28, 29, 33, 34].

Several diagnostic issues further confirmed the limitations of classical linear SAE approaches. First, clear nonlinear patterns were observed in the association between the outcome variable and the auxiliary covariates: the residuals failed the Shapiro–Wilk normality test, and both the GAM and Ramsey's RESET tests rejected linear functional form assumptions ($p < 0.05$). Second, severe multicollinearity was observed among the auxiliary variables [35]. Third, the response variable was highly skewed (Z-skew = 35.548) [36]. Fourth, the unavailability of recent unit-level census covariates [37], due to Indonesia's 10-year census cycle, restricted the feasibility of unit-level SAE. These issues collectively indicated that (i) direct estimation is imprecise for small domains, (ii) FH–EM is mis-specified for nonlinear structures, and (iii) area-level, machine-learning–enhanced models are operationally preferred.

Given these conditions, Random Forest offered several advantages: it captures nonlinear patterns [25], mitigates multicollinearity through randomized feature selection [38], and remains robust to skewness and non-normality [21, 33, 34]. Combined with the Fay–Herriot framework [26], these properties motivate the FHRF approach for this empirical context.

Both FHRF-C and FHRF-R yielded similar average RSE levels, at 6.56% (median 5.37%) and 6.35% (median 5.35%), respectively. The normality of paired differences was rejected by the Shapiro–Wilk test, necessitating the use of the Wilcoxon signed-rank test, which did not indicate a significant difference in median RSE ($p = 0.144$). In terms of effect size, the difference between the two models is small: FHRF-R achieved slightly lower RSE values, with a reduction of about 0.02 percentage points in the median (approximately 0.5% relative reduction) and 0.21 percentage points in the mean (approximately 3.2% relative reduction), indicating practically similar performance between the two approaches. Despite this, visual inspection showed that FHRF-R tended to produce more uniform gains, particularly among areas with initially high RSE. Most direct estimates fell below the 45° line when compared with FHRF-C (Figure 3(a)), indicating lower RSEs for FHRF-C, although a small number of points lay above the line. In contrast, almost all high-RSE direct estimates fell below the 45° line when compared with FHRF-R (Figure 3(b)), highlighting the superior performance of FHRF-R in these cases. Figure 3(c) shows that FH–EM produced higher RSEs than the direct estimator for many areas. The comparison in Figure 3(d) indicates that FHRF-R generally achieved lower RSEs than FHRF-C, particularly in areas with high sampling variability. Comparisons with FH–EM (Figures 3(e) and 3(f)) further show that both FHRF-C and FHRF-R generally outperformed FH–EM, with the advantage being more pronounced for FHRF-R.

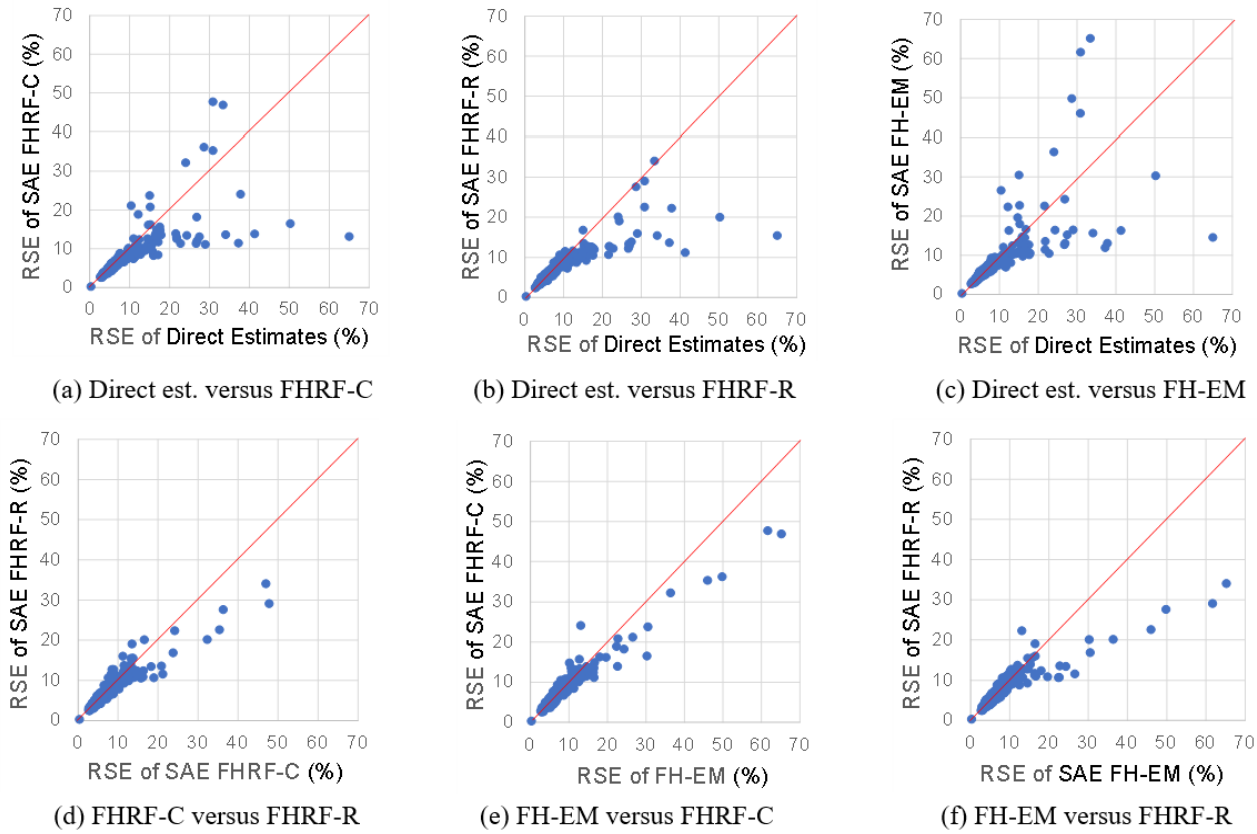


Figure 3. Pairwise comparisons of relative standard error (RSE) across estimation methods
Note: The 45° line indicates equal performance between the two methods. Points falling below this line indicate that the method on the vertical axis produces lower RSE values than the method on the horizontal axis.

The divergence between empirical and simulated findings may result from the stronger nonlinear structure in the real data and the concentration of problematic areas among high-RSE domains—features that were not explicitly emphasized in the simulations. Because the main empirical objective is to reduce the proportion of areas exceeding the 25% RSE threshold, subsequent analysis focuses on FHRF-R.

The SAE-FHRF-R model delivered clear gains in precision. The average RSE decreased from 7.46% (median 5.57%) under direct estimation to 6.35% (median 5.35%) with FHRF-R, corresponding to an approximate 15% reduction. The statistical significance of this reduction was confirmed by the Wilcoxon signed-rank test ($p = 2.93 \times 10^{-38}$). In terms of effect size, this corresponds to an absolute reduction of 1.11 percentage points in the mean RSE (approximately 14.9% relative reduction) and 0.22 percentage points in the median (approximately 4.0% relative reduction), indicating a practically meaningful improvement in precision. The association between direct and model-based RSE is also strong (Pearson $r = 0.852$). This pattern is visible in Figure 3b, where most points fall below the 45° benchmark, reflecting systematic improvement. The most significant gains occurred in areas with high direct-estimate RSE (15%–60%), where the RSE values were consistently reduced to below the 25% publication threshold. In this context, the number of areas exceeding the 25% threshold decreases from 15 under direct estimation to 3 under FHRF-R. In contrast, areas with already low RSE exhibited minimal change. This selective improvement reflects the model's adaptive behavior: As implied by Eq. (14), the fixed-effect component exerts greater influence when sampling-error variance (and thus RSE) is large, yielding marked reductions where they are most needed.

Comparisons with FH–EM further underscore this advantage. FH–EM produced a higher average RSE (6.80%) than FHRF-R (6.35%), and several areas exhibited RSE values even larger than their direct estimates (Figure 3c), indicating a lack of model fit. Across the full distribution, FHRF-R consistently outperformed FH–EM (Figure 3f), aligning with evidence that machine-learning methods are more effective at capturing complex and nonlinear relationships [25].

When areas are sorted by sample size, high RSE values from direct estimation—often above the 25% publication threshold—were concentrated in areas with very small agricultural household samples (Figure 4). In these domains, direct estimates became unstable, whereas FHRF-R consistently reduced RSEs to more acceptable levels. In such high-variability domains with small effective sample sizes, the reduction in RSE achieved by the FHRF-R model can reach up to 76%.

Figure 5 highlights that the most significant discrepancies between direct and FHRF-R estimates occur in areas with unusually high per capita expenditure relative to their neighbors. A clear example is Kota Manado (7171): the direct estimate was IDR 2,206,176 with an RSE of 34.07%, far higher than other districts in the province (IDR 671,533–1,153,002; RSE 4.18%–8.61%). After applying FHRF-R, the estimate decreased to IDR 1,055,405, and the RSE dropped to 15.50%, resulting in values more consistent with the regional pattern. This finding illustrates the model's ability to correct extreme direct estimates arising from small sample sizes by borrowing strength across auxiliary variables and inter-area variability [1, 28].

The pattern illustrated by the case of Kota Manado is not isolated. Table 7 summarizes the ten districts/cities exhibiting the largest absolute differences between direct and SAE-FHRF-R point estimates. These areas are predominantly urban districts with very small agricultural household samples, ranging from 15 to 201 observations, which is far below the national average of 962 households per district/city. They are also characterized by high direct-estimate RSE values, indicating substantial sampling uncertainty.

In these domains, the SAE-FHRF-R model produces systematic adjustments to point estimates, accompanied by marked reductions in RSE—for example, the 76% reduction observed in Kota Surakarta. By contrast, districts with adequate sample sizes and low direct-estimate RSEs exhibit negligible changes. This selective behavior explains the high overall correlation between direct and SAE-FHRF-R estimates ($r = 0.87$), which arises from the stability of well-sampled areas combined with targeted corrections in a limited number of small-sample urban domains.

Taken together, Figures 4 and 5 along with the summary in Table 7 indicate that substantial differences between direct and FHRF-R estimates arise almost exclusively in small-sample urban districts (area codes with third digit "7"), where agricultural households are scarce. Several cities with RSEs above the 25% threshold—such as Surakarta, Manado, Bandung, and Cirebon—exhibit substantial reductions in RSE under FHRF-R. In contrast, cities with moderate direct RSEs display smaller or mixed changes. In contrast, the majority of districts with adequate sample sizes and low direct-estimate RSEs remain largely unchanged, resulting in only localized differences in the spatial maps (Figure 6).

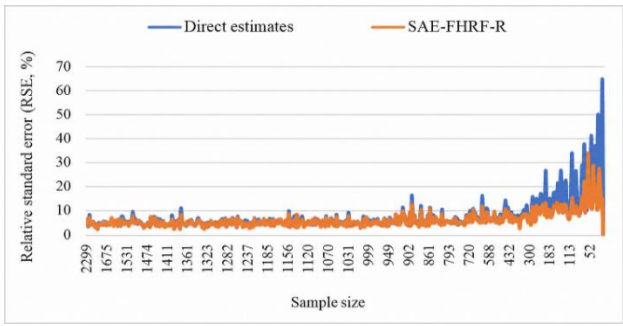


Figure 4. Comparison of relative standard error (RSE) values from direct estimation and SAE-FHRF-R across decreasing sample sizes

Note: SAE: Small area estimation; FHRF-R: Fay–Herriot with Random Forest replacement.

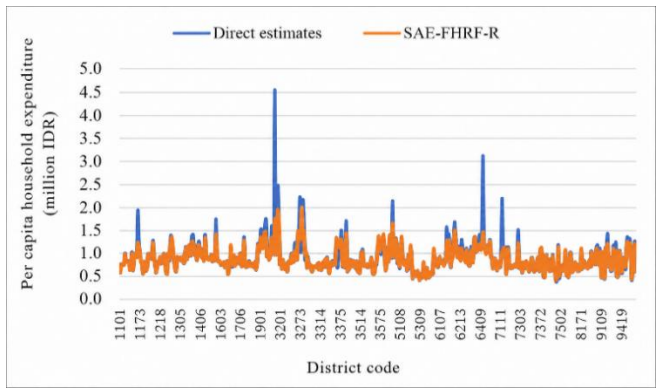


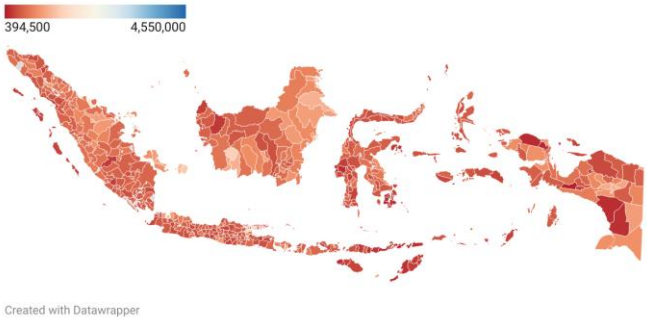
Figure 5. Estimated per capita household expenditure of agricultural households by district: direct estimates vs. SAE-FHRF-R

Note: SAE: Small area estimation; FHRF-R: Fay–Herriot with Random Forest replacement.

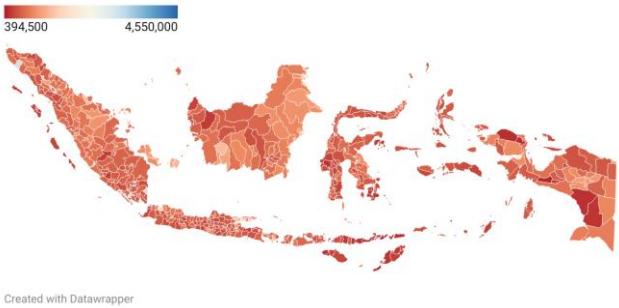
Table 7. Districts with the largest changes in point estimates between direct estimation and SAE-FHRF-R

Area Code	City	Point Estimate (Indonesian Rupiah)		Relative Standard Error (RSE) (%)		Sample Size (n)
		Direct	FHRF-R	Direct	FHRF-R	
3372	Surakarta	691,911	1,294,624	64.98	15.48	15
3171	Jakarta Selatan	4,551,182	1,751,303	33.37	34.06	57
6471	Balikpapan	3,116,754	1,459,451	26.77	13.41	201
7171	Manado	2,206,176	1,055,405	34.07	15.50	102
3578	Surabaya	1,005,785	1,425,182	15.00	16.85	52
1171	Banda Aceh	1,949,388	1,232,938	15.90	10.94	46
3273	Bandung	2,230,516	1,429,727	37.79	22.33	65
3276	Depok	2,171,051	1,529,311	24.01	20.17	63
3274	Cirebon	1,027,863	1,325,024	27.45	13.97	37
1971	Pangkal Pinang	1,759,144	1,262,291	17.63	12.33	166

Note: SAE: Small area estimation; FHRF-C: Fay–Herriot with Random Forest correction; FHRF-R: Fay–Herriot with Random Forest replacement.



(a) Direct estimates



(b) SAE-FHRF-R estimates

Figure 6. Estimated average per capita household expenditure of agricultural households by district, Indonesia (2019, Indonesian rupiah)
Note: SAE: Small area estimation; FHRF-R: Fay–Herriot with Random Forest replacement.

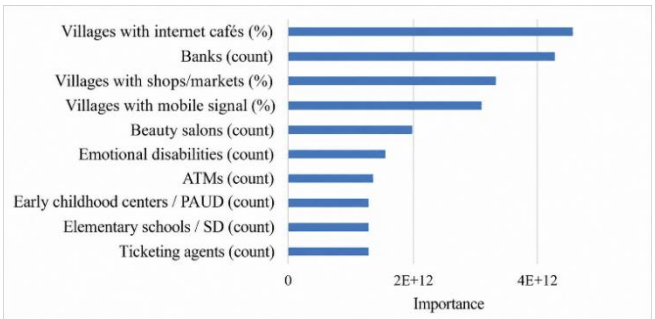


Figure 7. Random Forest variable importance measured by mean decrease in accuracy (top ten)

In addition to improving precision, FHRF-R yields interpretable information from its variable-importance results. As shown in Figure 7, the share of villages with internet cafés was the strongest predictor of agricultural household expenditure, suggesting the role of digital infrastructure in local economic conditions [39]. This information complements the model's predictive advantages by providing empirically grounded insights that can support targeted policy interventions [14, 21].

Overall, the empirical results showed that SAE-FHRF-R outperforms both direct estimation and FH–EM under nonlinear conditions, yielding higher precision and stable performance in small-sample urban areas. These findings are consistent with prior literature [4, 14, 21] and are reinforced by recent developments in integrating machine learning methods into the SAE framework [4, 10]. The variable-importance results further identify the most influential determinants of agricultural household expenditure, enhancing the model's policy relevance.

Consistent with this pattern, the spatial maps in Figure 6 exhibit only localized differences between direct estimates and FHRF-R estimates. This pattern occurs because most adjustments are concentrated in small urban domains with high RSEs, whereas the majority of districts—with already low RSEs—receive minimal correction.

5. CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Conclusions

This study proposes two Random Forest–enhanced extensions of the Fay–Herriot model—FHRF-R and FHRF-C—to address nonlinearity and model misspecification in area-level SAE. Both variants incorporate machine-learning components into the classical FH–EM framework, enabling the capture of nonlinear fixed-effect structures while preserving the stability of likelihood-based variance estimation.

Simulation experiments demonstrated that FHRF-R and FHRF-C consistently outperformed the standard FH–EM estimator under nonlinear data-generating processes, yielding lower RB and RRMSE. Across repeated replications, both models also showed stable error behavior, indicating reliable performance under data-generating processes characterized by strong nonlinear relationships. Sensitivity analyses further indicated that these efficiency gains are robust to dimensional changes, increasing with larger numbers of areas and remaining consistently superior to FH–EM even as the number of covariates grows. At the same time, it is important to acknowledge that, in simulation scenarios where the true data-generating process is strictly linear, the classical Fay–Herriot model estimated via EM consistently outperforms the proposed machine-learning-augmented variants in terms of

both point-estimation accuracy and RSE. This behavior is expected, as the Fay–Herriot model is optimal under correct linear specification and benefits from its parsimonious structure and likelihood-based estimation.

Empirical results on agricultural household per capita expenditure in Indonesia further reinforce the advantages of the proposed approach in applied settings. In particular, FHRF-R achieved substantial reductions in estimation uncertainty (RSE) relative to both direct estimates and FH–EM. The largest improvements occur in urban districts with small samples, where direct estimates were unstable and frequently exceeded the 25% publication threshold. By contrast, areas with already low RSEs showed little change, indicating that the method enhances precision primarily where it is most needed without introducing unnecessary distortion.

Overall, the proposed FHRF framework should be viewed not as a replacement for the classical Fay–Herriot model, but as a complementary, diagnostic-driven extension. While FH–EM remains preferable in well-specified linear settings, integrating Random Forest into the FH–EM framework extends the applicability of area-level SAE to situations where linear assumptions are violated. In this sense, FHRF—particularly FHRF-R—offers a flexible and methodologically sound enhancement to traditional area-level SAE, providing more precise and policy-relevant estimates for official statistics in the presence of nonlinear and heterogeneous relationships.

5.2 Future directions

The following research directions arise directly from the limitations identified in the present study. First, the simulation results indicate that FHRF does not outperform FH–EM when the linear mixed-model assumptions hold; therefore, future research should develop diagnostic or model-selection procedures capable of detecting nonlinear auxiliary relationships before estimation. Second, although sensitivity analyses with respect to the number of areas and covariate dimensions were conducted, they cover only a moderate range of configurations; extending the evaluation to more extreme settings would provide a more comprehensive robustness assessment. Third, integrating nonlinear learners into EM produced non-monotonic likelihood trajectories in some settings, motivating further theoretical investigation into convergence behavior and alternative stopping criteria. Finally, extending the framework to non-Gaussian outcomes and additional official-statistics applications would broaden the operational applicability of the proposed approach.

ACKNOWLEDGMENT

This research was supported by the Directorate General of Higher Education, Research, and Technology of the Ministry of Education, Culture, Research, and Technology, Republic of Indonesia, for funding this research through the 2024 Doctoral Research Scheme (Contract No. 027/E5/PG.02.00.PL/2024, June 11, 2024).

REFERENCES

[1] Rao, J.N.K., Molina, I. (2015). Small Area Estimation, 2nd Edition. Wiley, Hoboken, New Jersey, USA. https://doi.org/10.1002/9781118735855

[2] Jiang, J., Lahiri, P. (2006). Mixed model prediction and small area estimation. TEST, 15(1): 1-96. https://doi.org/10.1007/BF02595419

[3] Pfeffermann, D. (2013). New important developments in small area estimation. Statistical Science, 28(1): 40-68. https://doi.org/10.1214/12-STS395

[4] Tzavidis, N., Zhang, L.C., Luna, A., Schmid, T., Rojas-Perilla, N. (2018). From Start to Finish: A Framework for the Production of Small Area Official Statistics. Journal of the Royal Statistical Society Series A: Statistics in Society, 181(4): 927-979. https://doi.org/10.1111/rssa.12364

[5] Sunandi, E., Kurnia, A., Sadik, K., Notodiputro, K.A. (2021). A Bayesian logit-normal model in small area estimation. Journal of Physics: Conference Series, 1863(1): 012039. https://doi.org/10.1088/1742-6596/1863/1/012039

[6] Graf, M., Marín, J.M., Molina, I. (2018). A generalized mixed model for skewed distributions applied to small area estimation. TEST, 28(2): 565-597. https://doi.org/10.1007/s11749-018-0594-2

[7] Sugasawa, S., Kubokawa, T., Rao, J.N.K. (2018). Small area estimation via unmatched sampling and linking models. TEST, 27(2): 407-427. https://doi.org/10.1007/s11749-017-0551-5

[8] Opsomer, J.D., Claeskens, G., Ranalli, M.G., Kauermann, G., Breidt, F.J. (2008). Non-parametric small area estimation using penalized spline regression. Journal of the Royal Statistical Society, Series B - Statistical Methodology, 70(1): 265-286. https://doi.org/10.1111/j.1467-9868.2007.00635.x

[9] Tzavidis, N., Salvati, N., Pratesi, M., Chambers, R. (2008). M-quantile models with application to poverty mapping. Statistical Methods & Applications, 17(3): 343-366. https://doi.org/10.1007/s10260-007-0070-8

[10] Santoni, M.M., Widiyanto, D., Prasvita, D.S., Suryani, W., Awang, W. (2024). Prediction of horticultural production using machine learning regression models: A case study from Indramayu regency, Indonesia. Mathematical Modelling of Engineering Problems, 11(11): 3015-3024. https://doi.org/10.18280/mmep.111114

[11] Unik, M., Sukaesih, I., Syaufina, L., Jaya Surati, I.N. (2025). Application of random forest algorithm to analyze the confidence level of forest fire hotspots in Riau Peatland. Journal of Natural Resources and Environmental Management, 15(2): 255-266. https://doi.org/10.29244/jpsl.15.2.255

[12] Krennmair, P., Schmid, T. (2022). Flexible domain prediction using mixed effects random forests. Journal of the Royal Statistical Society, Series C: Applied Statistics, 71(5): 1865-1894. https://doi.org/10.1111/rssc.12600

[13] Frink, N., Schmid, T. (2025). Small area prediction of counts under machine learning-type mixed models. Computational Statistics & Data Analysis, 211: 108218. https://doi.org/10.1016/j.csda.2025.108218

[14] Viljanen, M., Meijerink, L., Zwakhals, L., van de Kassteele, J. (2022). A machine learning approach to small area estimation: Predicting the health, housing and well-being of the population of Netherlands. International Journal of Health Geographics, 21(1): 4. https://doi.org/10.1186/s12942-022-00304-5

[15] Michal, V., Wakefield, J., Schmidt, A.M., Cavanaugh,

A., Robinson, B., Baumgartner, J. (2023). Small area estimation with random forests and the lasso. arXiv preprint arXiv:2308.15180. https://doi.org/10.48550/arXiv.2308.15180
[16] Tzavidis, N. (2025). Small area estimation in the era of machine learning and alternative data sources: Opportunities, challenges, and outlook. Journal of Official Statistics, 41(3): 921-929. https://doi.org/10.1177/0282423X251342004
[17] Biau, G., Scornet, E. (2016). A random forest guided tour. TEST, 25(2): 197-227. https://doi.org/10.1007/s11749-016-0481-7
[18] Wright, M.N., Ziegler, A. (2017). Ranger: A fast implementation of random forests for high dimensional data in C++ and R. Journal of Statistical Software, 77(1): 1-17. https://doi.org/10.18637/jss.v077.i01
[19] Parker, P.A. (2024). Nonlinear Fay–Herriot models for small area estimation using random weight neural networks. Journal of Official Statistics, 40(2): 317-332. https://doi.org/10.1177/0282423X241244671
[20] Probst, P., Wright, M.N., Boulesteix, A.L. (2019). Hyperparameters and tuning strategies for random forest. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(3): e1301. https://doi.org/10.1002/widm.1301
[21] Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep Learning. MIT Press, Cambridge, MA, USA.
[22] Ávila-Valdez, J.L., Huerta, M., Leiva, V., Riquelme, M., Trujillo, L. (2020). The Fay-Herriot model in small area estimation: EM algorithm and application to official data. REVSTAT-Statistical Journal, 18(5): 613-635. https://doi.org/10.57805/revstat.v18i5.323
[23] Hajjem, A., Bellavance, F., Larocque, D. (2012). Mixed-effects random forest for clustered data. Journal of Statistical Computation and Simulation, 84(6): 1313-1328. https://doi.org/10.1080/00949655.2012.741599
[24] Wu, H.L., Zhang, J.T. (2006). Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modeling Approaches. Wiley, Hoboken, New Jersey, USA. https://doi.org/10.1002/0470009675
[25] Breiman, L. (2001). Random forests. Machine Learning, 45(1): 5-32. https://doi.org/10.1023/A:1010933404324
[26] Fay III, R.E., Herriot, R.A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. Journal of the American Statistical Association, 74(366a): 269-277. https://doi.org/10.1080/01621459.1979.10482505
[27] Wertis, L., Sugg, M.M., Runkle, J. D., Rao, D. (2023). Socio-environmental determinants of mental and behavioral disorders in youth: A machine learning approach. GeoHealth, 7(9): e2023GH000839. https://doi.org/10.1029/2023GH000839
[28] Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B (Methodological), 39(1): 1-22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x
[29] Hollander, E., Wolfe, D.A., Chicken, E. (2014). Nonparametric Statistical Methods. John Wiley & Sons, Hoboken, NJ, USA.
[30] Cochran, W.G. (1977). Sampling Techniques. John Wiley & Sons, New York, USA.
[31] Badan Pusat Statistik (BPS - Statistics Indonesia). (2025). Pengeluaran untuk Konsumsi Penduduk Indonesia September 2024. https://www.bps.go.id/assets/publication/2025/05/28/b67b4702334f3123221372ba/pengeluaran-untuk-konsumsi-penduduk-indonesia--september-2024.html, accessed on Oct. 17, 2025.
[32] Badan Pusat Statistik (BPS - Statistics Indonesia). (2024). Laporan Kinerja Direktorat Statistik Ketahanan Sosial 2024. https://ppid.bps.go.id/upload/doc/LAKIN_DIR_HANSOS_2024_1745291237.pdf, accessed on Oct. 17, 2025.
[33] Ghosh, M., Rao, J.N.K. (1994). Small area estimation: An appraisal. Statistical Science, 9(1): 55-76. https://doi.org/10.1214/ss/1177010647
[34] Chambers, R., Clark, R. (2012). An Introduction to Model-Based Survey Sampling with Applications. Oxford University Press, Oxford, UK. https://doi.org/10.1093/acprof:oso/9780198566625.001.0001
[35] Gujarati, D.N., Porter, D.C. (2009). Basic Econometrics. McGraw-Hill/Irwin, New York, USA.
[36] Doane, D.P., Seward, L.E. (2011). Measuring skewness: A forgotten statistic? Journal of Statistics Education, 19(2): 1-18. https://doi.org/10.1080/10691898.2011.11889611
[37] Deville, J.C., Särndal, C.E. (1992). Calibration estimators in survey sampling. Survey Methodology, 87(418): 376-382. https://doi.org/10.2307/2290268
[38] Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J. (2007). Random forests for classification in ecology. Ecology, 88(11): 2783-2792. https://doi.org/10.1890/07-0539.1
[39] Liaw, A., Wiener, M. (2002). Classification and regression by random forest. R News, 2(3): 18-22. https://journal.r-project.org/articles/RN-2002-022/.