BAREKENG: Journal of Mathematics and Its Applications
December 2023 Volume 17 Issue 4 Page 2059–2070
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol17iss4pp2059-2070

SIMULATION STUDY OF HIERARCHICAL BAYESIAN APPROACH FOR SMALL AREA ESTIMATION WITH MEASUREMENT ERROR

Leli Latifah [1], Kusman Sadik[2*], Indahwati[3]

[1,2,3]*Department of Statistics and Data Science, Faculty of Mathematics and Natural Sciences, IPB University*
St. Raya Dramaga, Bogor, 16680, Jawa Barat, Indonesia

*Corresponding author's e-mail: * **kusmans@apps.ipb.ac.id***

ABSTRACT

Article History:

Received: 19th June 2023
Revised: 21st September 2023
Accepted: 20th October 2023

Keywords:

Measurement Error;
EBLUP-FH;
Ybarra-Lohr SaeME;
Hierarchical Bayesian SaeME.

In small area estimation (SAE), the auxiliary variables used are commonly derived from registration data such as census and administrative data. It is assumed that the auxiliary variables are available for all areas. The limited availability of auxiliary variables can be an obstacle in SAE. The additional information from the survey can be alternative data, but it is assumed that the auxiliary variables will contain measurement errors. This study conducted a simulation of data that aims to handle when auxiliary variables are measured with errors. Two simulations were studied with some scenarios to the percentage area where the auxiliary variable is measured with error and scenarios to the generated auxiliary variables. Compare four methods: direct estimation, Fay-Herriot Empirical Best Linear Unbiased Prediction (EBLUP-FH), Ybarra-Lohr SAE with measurement error (SaeME), and Hierarchical Bayesian SaeME. The results show that, in both the simulation studies, the Hierarchical Bayesian SaeME method gives a smaller EMSE value than the other two methods when auxiliary information is measured with error.

How to cite this article:

L. Latifah, K. Sadik and Indahwati., “SIMULATION STUDY OF HIERARCHICAL BAYESIAN APPROACH FOR SMALL AREA ESTIMATION WITH MEASUREMENT ERROR,” *BAREKENG: J. Math. & App.*, vol. 17, iss. 4, pp. 2059-2070, December, 2023.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article • Open Access

1. INTRODUCTION

Small area estimation is an estimation method that can overcome the sample size problem. Rao and Molina define a small area to denote any domain that cannot accurately produce a direct estimate [1]. Additional information is needed or known as auxiliary variables, to obtain an adequate level of precision in indirect estimation with small area estimation. SAE can increase the effectiveness of sample size by borrowing the strength of neighboring areas and information from the auxiliary variables that have a strong relationship with observational variables [2]. Generally, additional information in estimating small areas is obtained from census calculations and administrative records [3]. Census may completely enumerate all units of its target population, so it is assumed that the auxiliary variables used do not contain measurement errors. However, the census has several drawbacks, such as a long time and, high costs, lack sufficient detail about the characteristic of interest [4], so surveys are activities that are often carried out compared to censuses. Survey activities use more cost-effectiveness and faster implementation time. Still, when additional information is obtained from the results of survey calculations, it is assumed that the auxiliary variables will contain measurement errors. Measurement error arises when a recorded measurement value is not exactly the same as the actual value [5]. The existence of measurement errors in the model causes parameter estimates to be biased and inconsistent, and conclusions may be drawn erroneously [6]. Therefore, the estimation of parameters in small area estimation with auxiliary variables containing errors will also be a biased estimator, and the mean squared error of the predictor may be enhanced [7]. So, we need a small area estimation method to accommodate auxiliary variables with measurement errors.

The development of a small area estimation model with auxiliary variables that contain measurement errors, in general, has been discussed by Hariyanto et al. [8] and Tanur [9]. In the unit-level small area estimation, Ghosh et al. [10] first developed a small area estimation with an auxiliary variable containing the error. Development is carried out by estimating a small area at the unit level with structural measurement errors in the auxiliary variables used. Torabi et al. [11] and Torkasvand et al. [12] estimated using the Bayesian method at the unit level with an auxiliary variable containing the error. At the area level, Ybarra and Lohr [13] modified the Fay-Herriot EBLUP method using auxiliary variables containing errors. Zhu and Zou [14]**,** Datta et al. [15], Datta et al [16], and Wulandari et al. [17] used the SAE method which accommodates structural measurement errors. Komalasari M. [18] applies SUSENAS data to estimate the average length of schooling by sub-district in Kampar Regency. The results show that the Ybarra-Lohr SaeME estimation model can predict a smaller MSE value than direct estimation. Aziz and Ubaidillah [19] used two SAE models SAE EBLUP Fay-Herriot model with auxiliary variables Podes data and SAE with Error Measurement with auxiliary variable Twitter data. Estimation results using the SAE method are better than direct estimates. Auxiliary variables that contain errors are sources from Big data which as Twitter. Hariyanto et al. [20] estimate parameters and develop empirical bates in small area estimation with measurement error in t distributed covariate variable. Tanur and Kurnia [21] conducted a study that developed an alternative small-area estimation for the autoregressive model with auxiliary variables containing measurement errors. Novkaniza et al. [22] estimate non-symmetrical count data in SAE for the Poisson-lognormal model with measurement error in covariate.

The Bayesian approach to estimating small areas at the area level by considering the auxiliary variables containing measurement errors was carried out by Arima et al. [23]. Method development from the Ybarra-Lohr SaeME method proposes an alternative estimator resulting from the Hierarchical Bayes measurement error model. Estimation in the simulation study is carried out with several scenarios that condition that there are unequal measurement errors in the generated areas. The simulations compared several SAE methods, including direct estimation, EBLUP-FH, EBLUP Bayesian, Ybarra-Lohr SaeME, and Hierarchical Bayesian SaeME. The simulation study results show that estimating with Hierarchical Bayes is more stable and has a smaller MSE value than the Ybarra–lohr SaeME estimation and others' estimation [23].

Arima et al. [23] only used one random effect variance and one auxiliary variable that contained errors in their simulation study. Based on this, in this study, an adaptation of the simulation conducted by Arima et al. [23] will be carried out by adding scenarios to the variance of random effect areas and scenarios to the auxiliary variables used. There are two simulations carried out in this study. These simulations were carried out by applying three small area estimation methods when the auxiliary variables have measurement errors, namely the SAE EBLUP-FH method, the Ybarra-Lohr SaeME method, and the Hierarchical Bayesian SaeME method. This study compares the four methods in some scenarios when the variance of random effect has different value. The first simulation was conducted to see each area's different measurement error conditions. The second simulation is carried out by creating scenarios on the auxiliary variables. The aim is to see the conditions when the auxiliary variables used are two auxiliary variables containing measurement errors and the conditions when some auxiliary variables used include measurement errors, and some do not. From these simulations can be used for the future research in applying empirical data with the method that can give the accurate result in estimating small area with auxiliary variables that containing measurement error.

2. RESEARCH METHODS

2.1 Fay-Herriot Model

Small area estimation is done to estimate parameters indirectly in a relatively small area in a pilot survey [24]. The availability of additional information and determining a good and suitable model is important in obtaining indirect estimates, especially in small areas [1]. In the area level model, the auxiliary variables $\boldsymbol{x_i} = (x_{i1}, \dots x_{iK})^T$available to a small level, and the observed variable is assumed to be a function of the average response variable $\theta_i = g(\bar{Y}_i)$ for $g(.)$. Where $z_i$is a known positive value constant and $\boldsymbol{\beta} = (\beta_i, \dots, \beta_K)^T$ is the regression coefficient, $u_i$is the small random effect (often assumed to be normal), $e_i$is the sampling error, $e_i \sim N(0, \sigma_e^2)$, so that the area level model (a model of linear mixed form or known as the Fay-Herriot model) can be written in **Equation (1)**:

$$\hat{\theta}_i = \boldsymbol{x_i}^{\boldsymbol{T}}\boldsymbol{\beta} + z_i u_i + e_i,\ \ i = 1, \dots, m \text{ (small area)} \qquad (1)$$

BLUP estimator (best linear unbiased prediction) to θ_i be formulated in **Equation (2)**:

$$\begin{aligned} \tilde{\theta}_i^{FH\ BLUP} &= \boldsymbol{x}_i^{\boldsymbol{T}}\tilde{\boldsymbol{\beta}} + \gamma_i(\hat{\theta}_i - \boldsymbol{x}_i^{\boldsymbol{T}}\tilde{\boldsymbol{\beta}}) \\ &= \gamma_i\hat{\theta}_i + (1-\gamma_i)\,\boldsymbol{x}_i^{\boldsymbol{T}}\tilde{\boldsymbol{\beta}} \end{aligned} \qquad (2)$$

In the BLUP estimator, the value σ_u^2 is assumed to be known. However, the random effect variance (σ_u^2) is unknown in practice, so it must be estimated first. The estimate σ_u^2will then be replaced by $\hat{\sigma}_u^2$, the BLUP estimator. Then a new EBLUP estimator (empirical best linear unbiased prediction) will be obtained in **Equation (3)**. Which is the weighted average (with weight $\hat{\gamma}_i$) of the direct estimator ($\hat{\theta}_i$) and the synthesis model $(\boldsymbol{x}_i^{\boldsymbol{T}}\hat{\boldsymbol{\beta}})$.

$$\hat{\theta}_i^{FH\ EBLUP} = \gamma_i\hat{\theta}_i + (1-\gamma_i)\,\boldsymbol{x}_i^{\boldsymbol{T}}\hat{\boldsymbol{\beta}} \qquad (3)$$

with $\hat{\gamma}_i = \hat{\sigma}_u^2 z_i^2 / (\hat{\sigma}_u^2 z_i^2 + \sigma_{e_i}^2)$ and value $\hat{\boldsymbol{\beta}}$ is EBLUP estimator for value $\boldsymbol{\beta}$.

2.2 Ybarra-Lohr SaeME Model

Ybarra and Lohr [13] developed this small area estimation because the survey has errors due to sampling and errors not due to sampling, so the additional information or auxiliary variables used contain errors. Therefore, this development is known as Small Area Estimation with Measurement Error (SaeME) or small Area estimation with auxiliary variables that contain errors. In small area estimation, $\boldsymbol{x_i}$ is the value of the auxiliary variable area i. If all components $\boldsymbol{x_i}$ is known, we use **Equation (4)**:

$$\hat{\theta}_i = \boldsymbol{x_i}^{\boldsymbol{T}}\boldsymbol{\beta} + z_i u_i + e_i \qquad (4)$$

where u_i and e_i are independent, e_i is the sampling error, $e_i \sim N(0, \sigma_e^2)$. However, when $\boldsymbol{x_i}$, there are auxiliary variables that are measured with errors, then the model with measurements that contain errors is as follows [13] in **Equation (5)**:

$$\hat{\theta}_i = \hat{\boldsymbol{x}}_{\boldsymbol{i}}^{\boldsymbol{T}}\boldsymbol{\beta} + r_i(\hat{\boldsymbol{x}}_{\boldsymbol{i}}, \boldsymbol{x_i}) + e_i \qquad (5)$$

Estimator $\hat{\boldsymbol{x}}_{\boldsymbol{i}}$ replaces the value for $\boldsymbol{x_i}$ with $r_i(\hat{\boldsymbol{x}}_{\boldsymbol{i}}, \boldsymbol{x_i}) = u_i + (\boldsymbol{x_i} - \hat{\boldsymbol{x}}_{\boldsymbol{i}})^T\beta$. It is assumed that the estimator for] $\boldsymbol{x_i}$ is available for all areas i. We can input or estimate multiple components $\boldsymbol{x_i}$ is not measurable. In this model, it is assumed that u_i and e_i are independent. Ybarra and Lohr SaeME formula can be written in **Equation (6):**

$$\theta_i^{ME} = \gamma_i y_i + (1-\gamma_i)\,\hat{\boldsymbol{x}}_{\boldsymbol{i}}^{\boldsymbol{T}}\boldsymbol{\beta} \qquad (6)$$

model is denoted as with $\gamma_i = \sigma_u^2 + \beta^T C_i \beta / (\sigma_u^2 + \beta^T C_i \beta + \sigma_e^2)$. Suppose $w_1, \dots, w_m$ is a set of finite weights bounded from 0. The estimated regression parameters are defined as follows in **Equation (7)**:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{w}} = \left(\sum_{i=1}^{m} w_i (\widehat{\boldsymbol{x}}_i \widehat{\boldsymbol{x}}_i^{\boldsymbol{T}} - C_i)\right)^{-1} \sum_{i=1}^{m} w_i \widehat{\boldsymbol{x}}_i y_i \qquad (7)$$

with $w_i = 1/(\sigma_u^2 + \beta^T C_i \beta + \sigma_e^2)$ for $i = 1, \dots, m$.

2.3 Hierarchical Bayesian SaeME Model

Arima et al**.** [23] examined a Hierarchical Bayesian model with measurement errors. The results show that the uncertainty of measuring the posterior variance of the Bayesian estimator is more stable than the EBLUP MSE proposed by Ybarra-Lohr [13]. In the Hierarchical Bayesian approach, the unknown model parameters (including the variance components) are treated as random components, each with a certain prior distribution. The posterior distribution for the parameter of interest is obtained based on the entire prior distribution. The prior that used in this study is non informative prior. So it requires the generation of sample data from each $\theta, \boldsymbol{x}, \boldsymbol{\beta}, \delta,$ dan σ_u^2 given parameter and the remaining data. Bayesian procedure implementation is used with the Markov chain Monte Carlo technique, especially the Gibbs sampler. In obtaining posterior distribution results, the trace plot is used to analyze MCMC convergence [25]. Trace plot is a plot of interaction against the resulting values. If there is a certain pattern then the Markov chain has not reached convergence. it is important to check algorithm convergence. Hierarchical Bayesian SaeME model can be witten in multi-stage model as follows in **Equation (8)**:

$$\text{i. } \theta_i | \boldsymbol{\beta}, u_i, \sigma_u^2, \theta_{(-i),} \boldsymbol{x_i}, y, \widehat{\boldsymbol{x}}_{\boldsymbol{i}} \sim N \left(\frac{\sigma_e^{2^{-1}} y_i + \sigma_u^{-2} (\boldsymbol{x_i}' \boldsymbol{\beta} + z'_i u_i)}{\sigma_e^{2^{-1}} + \sigma_u^{-2}}, \left(\sigma_e^{2^{-1}} + \sigma_u^{-2} \right)^{-1} \right); \qquad (8)$$

$$\text{ii. } \boldsymbol{x_i} | \boldsymbol{\beta}, u_i, \sigma_u^2 \theta, \boldsymbol{x_{-i}}, y, \widehat{\boldsymbol{x}}_{\boldsymbol{i}} \sim N \left(\widehat{\boldsymbol{x}}_{\boldsymbol{i}} + \frac{y_{i-} \boldsymbol{x_i}' \boldsymbol{\beta} - z'_i u_i}{\sigma_e^2 + \sigma_u^2 + \boldsymbol{\beta}' C_i \boldsymbol{\beta}} C_i \boldsymbol{\beta}, C_i - \frac{C_i \boldsymbol{\beta} \boldsymbol{\beta}' C_i}{\sigma_e^2 + \sigma_u^2 + \boldsymbol{\beta}' C_i \boldsymbol{\beta}} \right);$$

$$\text{iii. } \boldsymbol{\beta} \Big| u_i, \sigma_u^2 \theta, \boldsymbol{x_i}, y, \widehat{\boldsymbol{x}}_{\boldsymbol{i}} \sim N((\widehat{\boldsymbol{x}}_{\boldsymbol{i}}' \widehat{\boldsymbol{x}}_{\boldsymbol{i}})^{-\boldsymbol{1}} \boldsymbol{x}' (\theta - Z u_i), \sigma_u^2 (\boldsymbol{x}' \boldsymbol{x})^{-\boldsymbol{1}});$$

$$\text{iv. } u_i | \boldsymbol{\beta}, \sigma_u^2, \theta, \boldsymbol{x_i}, y, \widehat{\boldsymbol{x}}_{\boldsymbol{i}} \sim N((Z'Z)^{-\boldsymbol{1}} \boldsymbol{x}' (\theta - \boldsymbol{x_i} \boldsymbol{\beta}), \sigma_u^2 (Z'Z)^{-\boldsymbol{1}});$$

$$\text{v. } \sigma_u^2 | \boldsymbol{\beta}, u_i, \theta, \boldsymbol{x_i}, y, \widehat{\boldsymbol{x}}_{\boldsymbol{i}} \sim IG \left(\frac{\boldsymbol{1}}{\boldsymbol{2}} (\boldsymbol{m} - 2), \frac{\boldsymbol{1}}{\boldsymbol{2}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} (\theta_i - \boldsymbol{x}_i' \boldsymbol{\beta} - z'_i u_i) \right)$$

2.4 Data Analysis Procedure

The simulation data were analyzed using four estimation methods, including the direct estimation method. This SAE EBLUP-FH method assumes that the auxiliary variables containing errors are the true values, the Ybarra-Lohr SaeME method, and Hierarchical Bayesian SaeME. There were two simulations carried out in this study. Both simulations were carried out with scenarios on a variance of random effects ($\sigma_{u_1}^2 = 2$ dan $\sigma_{u_2}^2 = 4$) and scenarios on the measurement error ($c_i \epsilon \{0, d\}$ where d= 2 and 4). The first simulation is carried out with several k scenarios, $k \in \{0,20,50,80,100\}$ where k is the percentage of small areas where the auxiliary variables measured contain errors $c_i \epsilon \{0, d\}$, for example, k = 80% and c_i= 2 means that 80% of the small areas with auxiliary variables measured contained errors c_i= 2. The percentage of k was used to see the small area with different measurement error conditions. The greater the percentage of k, the greater the percentage of small area with auxiliary variables measured with error. The rest has a value of 0. In comparison, the second simulation has two scenarios on the auxiliary variables used. There are two auxiliary variables used in the second simulation. The first scenario of the second simulation is that one auxiliary variable contains an error, and one auxiliary variable does not contain an error. The second scenario of this second simulation is that the two auxiliary variables used contain errors Both simulations were carried out with 100 iterations to get an optimal result. Data processing was carried out using the R program. The following are the stages of the simulation study carried out:

1) Determine the number of small areas, namely m= 20.
2) Generating random effect area ($\mathrm{u_i}$) with two scenarios, namely $\mathrm{u_{1_i}} \sim \mathrm{N}(0, \sigma^2 = 2)$ and $\mathrm{u_{2_i}} \sim \mathrm{N}(0, \sigma^2 = 4)$ with $\mathrm{i} = 1,2,3 \dots . \mathrm{m}$.
3) Generating a sampling error $\mathrm{e_i} \sim \mathrm{N}(0,1)$.
4) Auxiliary variables that contain errors will later be generated with area error data of the auxiliary

variables ($\mathrm{v_i}$) from a normal distribution with zero expectation values and three scenarios of measurement error ($\mathrm{c_i}$), which are determined as constants $\mathrm{c_i} \in \{0, \mathrm{d}\}$ where d= 2 and 4.

5) Generating the first simulation data as follows:
   (i) Define values $\alpha = 1$ and $\beta = 3$,
   (ii) Generating one auxiliary variable from the Normal distribution, namely $\mathrm{x_i} \sim \mathrm{N}(5,9)$.
   (iii) From each iteration will be drawn $\mathrm{Y_i} = \alpha + \beta \mathrm{x_i} + \mathrm{u_i}$ and $\mathrm{y_i} = \mathrm{Y_i} + \mathrm{e_i}$
   (iv) Generating auxiliary variables containing errors from the equation $\hat{\mathrm{x}}_\mathrm{i} = \mathrm{x_i} + \mathrm{v_i}$.
   (v) Several scenarios will be carried out $k \epsilon \{0,20,50,80,100\}$ with k as a small percentage area with the measured auxiliary variables containing errors $\mathrm{c_i}$

6) Generating the second simulation data as follows:
   (i) Define value $\alpha = 1$ and value β with two scenarios $\beta_1 = 3$ and $\beta_2 = 5$
   (ii) Generating two auxiliary variables namely $\mathrm{x_{1_i}} \sim \mathrm{N}(5, \sigma = 3)$ and $\mathrm{x_{2_i}} \sim \mathrm{N}(3, \sigma = 5)$
   (iii) From each iteration will be drawn $\mathrm{Y_i} = \alpha + \beta_1 \mathrm{x_{1_i}} + \beta_2 \mathrm{x_{2_i}} + \mathrm{u_i}$ and $\mathrm{y_i} = \mathrm{Y_i} + \mathrm{e_i}$
   (iv) Generating auxiliary variables that contain errors from the equation, namely $\hat{\mathrm{x}}_{1_\mathrm{i}} = \mathrm{x_{1_i}} + \mathrm{v_i}$,

   and $\hat{\mathrm{x}}_{2_\mathrm{i}} = \mathrm{x_{2_i}} + \mathrm{v_i}$
   (v) Modeling is built with two scenarios:
      a. One auxiliary variable contains an error, and one variable does not have an error
      b. Both auxiliary variables have errors

7) Calculation of parameter estimators from both simulations is done by:
   (i) Direct estimation method $\mathrm{y_i}$
   (ii) SAE EBLUP-FH method assumes $\hat{\mathrm{x}}_\mathrm{i}$ as the true value.
   (iii) The Ybarra-Lohr SaeME method
   (iv) Hierarchical Bayesian SaeME Method. The Bayesian procedure is implemented with a monte carlo markov chain simulation.

8) The simulation results were evaluated by calculating the average bias and the average empirical mean squared error (EMSE) of 100 replicates for all scenarios. Then the results of all estimation methods are compared. The better method is the one with a smaller average bias (AB) and average EMSE (AEMSE) in **Equation (9)** and **Equation (10)** respectively.

$$\mathrm{AB} = \mathrm{K}^{-1} \sum_{\mathrm{k}=1}^{\mathrm{K}} (\mathrm{m}^{-1} \sum_{\mathrm{i}=1}^{\mathrm{m}} |\hat{\mathrm{y}}_\mathrm{i} - \mathrm{y_i}|) \tag{9}$$

$$\mathrm{AEMSE} = \mathrm{K}^{-1} \sum_{\mathrm{k}=1}^{\mathrm{K}} (\mathrm{m}^{-1} \sum_{\mathrm{i}=1}^{\mathrm{m}} (\hat{\mathrm{y}}_\mathrm{i} - \mathrm{y_i})^2) \tag{10}$$

3. RESULTS AND DISCUSSION

3.1 Simulation Study 1

From the results of processing with the R program, the average bias and average empirical mean squared error (EMSE) of several scenarios were obtained. The results of the four methods are then compared, the direct estimation method (y_i), the EBLUP-FH method, which assumes no errors in the auxiliary variables *(* $\hat{Y}_{iS}$*)*, the Ybarra-Lohr SaeME method *(*$\hat{Y}_{iME}$*)*, and the Hierarchical Bayesian SaeME method with the auxiliary variables containing errors *(*$\hat{Y}_{iB}$*)*. In obtaining good posterior distribution results for the Hierarchical Bayesian SaeME method, it is necessary to fulfill the convergence of the MCMC algorithm. The parameter estimation process for first simulation is carried out by generating sample data in 10,000 iterations with a burn-in of 5,000 and a thin of 20. Based on the MCMC algorithm, convergence occurs when the resulting Markov chain distribution approaches the posterior interest distribution. To evaluate the convergence is done by looking at the resulting Trace plot. From several scenarios, the resulting trace plot has a random pattern, and the plot is relatively stable at a specific value. So convergence has been achieved.

Table 1. Average Bias with Random Effect Area σ_u^2= 2 and c_i = 2

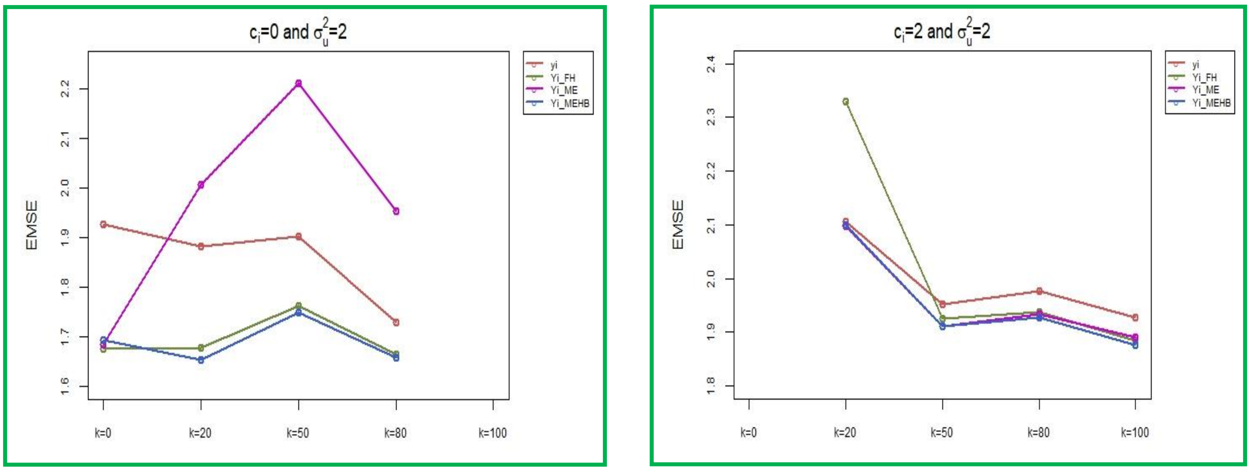
k	c_i	y_i	$\hat{Y}_{iS}$	$\hat{Y}_{iME}$	$\hat{Y}_{iB}$
0	0	-0.04009	-0.04009	-0.04009	-0.03910
20	0	-0.05907	-0.06272	-0.06159	-0.06125
	2	0.03584	0.05041	0.04467	0.04527
50	0	-0.01910	-0.02845	-0.02799	-0.02769
	2	-0.06109	-0.05173	-0.05301	-0.05255
80	0	-0.08365	-0.07789	-0.07420	-0.07885
	2	-0.02920	-0.03064	-0.03208	-0.03000
100	2	-0.04009	-0.04009	-0.04009	-0.03976

Table 1 above shows the average bias when m = 20, the random effect area σ_u^2= 2, and c_i= 2. The average bias is shown by separating the areas with c_i= 2 and c_i= 0. From these results it can be seen that the fourth method produces results that are not much different. However, in general it appears that direct estimation produces an average bias that is close to 0, especially when the percentage of area containing measurement error is k = 20 and k = 30 with c_i = (2, 3, or 4). However, when k the values are 0 and 100, estimating with Hierarchical Bayesian SaeME ($\hat{Y}_{iB}$) produces average bias, which is generally smaller than other methods. Table 2 shows the average bias when m = 20 and the random effect area σ_u^2= 4 and c_i= 2. Generally, when the variance of random effect is greater, σ_u^2= 4, the average bias is greater than when the variance of random effect is σ_u^2= 2. The result of average bias from four methods also close to 0. However, when k is large, Hierarchical Bayesian estimation with auxiliary variables containing errors ($\hat{Y}_{iB}$) produces an average bias generally smaller than other methods.

Table 2. Average Bias With Random Effect Area σ_u^2= 4 and c_i = 2

k	c_i	y_i	$\hat{Y}_{iS}$	$\hat{Y}_{iME}$	$\hat{Y}_{iB}$
0	0	-0.04009	-0.04009	-0.04009	-0.03957
20	0	-0.05907	-0.06210	-0.06125	-0.06135
	2	0.03584	0.04794	0.04383	0.04463
50	0	-0.01910	-0.02587	-0.02686	-0.02498
	2	-0.06109	-0.05431	-0.05395	-0.05532
80	0	-0.08365	-0.08012	-0.09375	-0.08423
	2	-0.02920	-0.03008	-0.02764	-0.02943
100	2	-0.04009	-0.04009	-0.04009	-0.03992

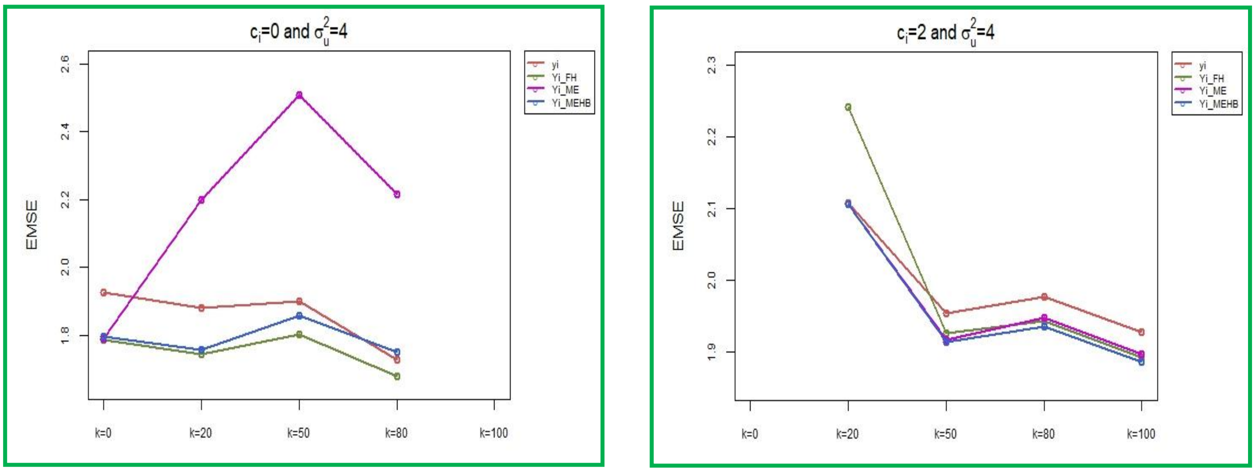
Based on the calculation of the Empirical Mean Squared Error (EMSE), the result is that in a scenario with a random effect area σ_u^2= 2 and c_i = 2 shown in Figure 1. The average EMSE is shown by separating the areas with c_i= 2 and c_i= 0. At k =0, the estimation results $\hat{Y}_{iS}$ provide the lowest average EMSE among the other methods. When c_i= 0, k =20, 50, and 80, $\hat{Y}_{iS}$ provide the average EMSE lower than y_i and $\hat{Y}_{iME}$. From two graphics, y_i provides the average EMSE, bigger than three other methods, except when c_i= 0 and k =20, 50, and 80. For c_i = 2, When k =20 and k =50, the estimation $\hat{Y}_{iME}$ gives the smallest average EMSE among other estimation methods. It means that $\hat{Y}_{iME}$ provides the smallest average EMSE when there is a measurement error in the auxiliary variable with a small k. In contrast, the smallest average EMSE is obtained by $\hat{Y}_{iB}$ when k values are 80 and 100. It can be said that the greater the percentage k and the value of c_i, the estimator $\hat{Y}_{iB}$is, the best estimator.



(a) **(b)**

Figure 1. Average EMSE with Random Effect Area σ_u^2= 2, when (a) $c_i = 0$, (b) $c_i = 2$

The comparison average EMSE at m = 20, c_i= 2, and the random effect area σ_u^2= 4 is shown in Figure 2. At times k =20, 50, and 80, c_i= 0, the smallest average EMSE was produced by estimating $\hat{Y}_{iS}$. For k =20 and c_i= 2, the estimation $\hat{Y}_{iME}$ gives the smallest average EMSE among other estimation methods. And when k =50, 80, and 100, c_i= 2, the estimation $\hat{Y}_{iB}$ gives the smallest average EMSE among other estimation methods. The average EMSE estimation y_i is the largest among other methods for almost all k and c_i. Figures 1 and 2 show a similar pattern, but it can be seen that the four estimation method give a bigger average EMSE when σ_u^2= 4 than σ_u^2= 2. It means the bigger σ_u^2, the bigger the average EMSE. From the bias average and EMSE average of the four methods, it can be said that $\hat{Y}_{iB}$ it gives the best results among the other methods



(a) **(b)**

Figure 2. Average EMSE with Random Effect Area σ_u^2= 4, when (a) $c_i = 0$, (b) $c_i = 2$

3.2 Simulation Study 2

In simulation study 2, two scenarios are carried out in the parameter estimation model. The first scenario is carried out with one auxiliary variable($\hat{x}_{1i}$) containing the measurement error and one auxiliary variable (x_{2i}) which does not include the measurement error. The second scenario has two auxiliary variables, each containing measurement errors ($\hat{x}_{1i}$ and $\hat{x}_{2i}$). The parameter estimation process with the Hierarchical Bayesian SaeME method is carried out by generating MCMC samples by generating sample data in 100,000 iterations with a burn-in of 5,000 and a thin of 20. The estimation is evaluated by comparing the average bias and the average EMSE.

The average bias results with random effect area σ_u^2= 2 and measurement error c_i= 2, 3, and 4 are shown in Table 3. It shows that the average bias produces in both scenarios 1 and 2 for each c_i is the same for the three methods which y_i, $\hat{Y}_{iS}$, and $\hat{Y}_{iME}$. Therefore, the average bias resulting from three methods, whether the auxiliary variable contains only one variable that containing measurement error or both containing measurement error, is not different. The results are different from the average bias of $\hat{Y}_{iB}$. The greater the number of auxiliary variables that contain measurement error, the greater the average bias. When in scenario 1, only one auxiliary variable contains measurement error, the average bias of $\hat{Y}_{iB}$ is smaller than other estimation methods. When the model contains auxiliary variables that both include errors, the resulting average bias is greater than the model where only one of the auxiliary variables has errors. In this simulation, the average bias of $\hat{Y}_{iME}$ is not greater than $\hat{Y}_{iS}$. It relates to Ybarra and Lohr [13] in certain conditions, $\hat{x}_i$ is ignored for mean squared error calculation, and the reported mean squared error of $\hat{Y}_{iS}$ will be too small, giving a misleading notion of precision.

Table 3. Average Bias With Random Effect Area σ_u^2= 2

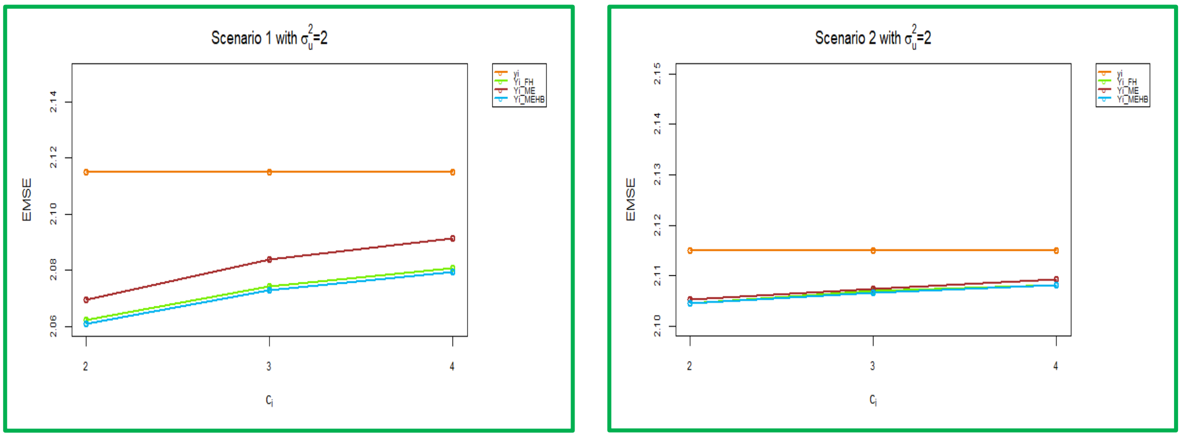
Scenario	c_i	y_i	$\hat{Y}_{iS}$	$\hat{Y}_{iME}$	$\hat{Y}_{iB}$
1	2	0.0424958	0.0424958	0.0424958	0.0419295
	3	0.0424958	0.0424958	0.0424958	0.0419235
	4	0.0424958	0.0424958	0.0424958	0.0419214
2	2	0.0424958	0.0424958	0.0424958	0.0425923
	3	0.0424958	0.0424958	0.0424958	0.0425620
	4	0.0424958	0.0424958	0.0424958	0.0426548

Table 4. Average Bias With Random Effect Area σ_u^2= 4

Scenario	c_i	y_i	$\hat{Y}_{iS}$	$\hat{Y}_{iME}$	$\hat{Y}_{iB}$
1	2	0.0424958	0.0424958	0.0424958	0.0419279
	3	0.0424958	0.0424958	0.0424958	0.0419239
	4	0.0424958	0.0424958	0.0424958	0.0419217
2	2	0.0424958	0.0424958	0.0424958	0.0425438
	3	0.0424958	0.0424958	0.0424958	0.0425974
	4	0.0424958	0.0424958	0.0424958	0.0426585

Table 4 shows the average bias when the random effect area σ_u^2= 4. In general, the lowest bias average is obtained by the estimation method $\hat{Y}_{iB}$. Three estimation method which are y_i, $\hat{Y}_{iS}$, and $\hat{Y}_{iME}$ give the result that is not different from Table 3. In general, the lowest average bias obtained by $\hat{Y}_{iB}$ estimation method is in scenario 1, only one of the auxiliary variables contains measurement error. From the two average bias tables, the four estimation methods produce an average bias that is close to 0.

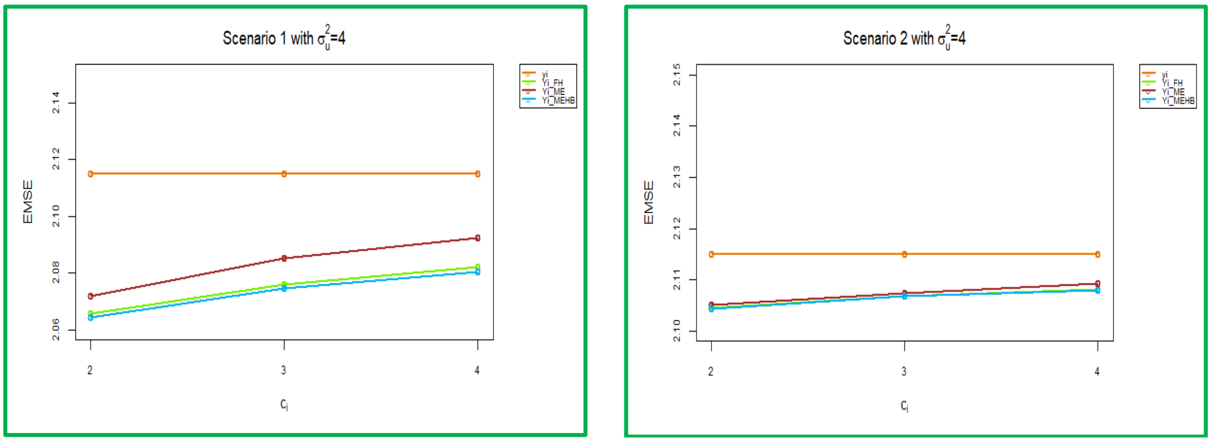
The estimation method is then evaluated by looking at the average EMSE output of the four estimation methods (Figure 3). The estimation y_i provides the average EMSE, which is lowest than other methods. When the random effect area σ_u^2= 2 and c_i= 2, 3, and 4, it can be seen that the larger c_i, the greater the average EMSE value. The average EMSE resulting from the four methods is not too different. However, in general, the average EMSE of the two scenarios with the smallest value is produced by the method $\hat{Y}_{iB}$ compared to method $\hat{Y}_{iS}$and $\hat{Y}_{iME}$.



(a) **(b)**

Figure 3. The Average EMSE with Random Effect Area σ_u^2= 2, (a) Scenario 1, (b) Scenario 2

When the random effect area σ_u^2= 4 and c_i= 2, 3, and 4, the average EMSE is shown in **Figure 4**. The average EMSE of the two scenarios with the smallest value is generally produced by the method $\hat{Y}_{iB}$ compared to method $\hat{Y}_{iS}$ and $\hat{Y}_{iME}$. Not much different as at σ_u^2= 2, the bigger it is, c_i, the bigger the average EMSE value. Although the resulting numbers are close enough, it is quite visible that the value increases as it increases c_i. However, it is different when compared, so the average EMSE value at σ_u^2= 4 is smaller than at σ_u^2= 2 in scenario 2. Based on the evaluation of the model estimation, both the average bias and the average EMSE, it can be said that the estimation of $\hat{Y}_{iB}$ gives the best results compared to the estimate of y_i, $\hat{Y}_{iS,}$and $\hat{Y}_{iME}$.



(a) **(b)**

Figure 4. The Average EMSE with Random Effect Area σ_u^2=4, (a) Scenario 1, (b) Scenario 2

4. CONCLUSIONS

The first simulation study was conducted to see each area's different measurement error conditions. When the percentage of areas containing measurement errors is zero, which means there are no measurement errors, the estimation with SAE EBLUP-FH is better. This can be seen from the smaller average EMSE. When the percentage of small areas containing measurement error is 20 and 50, the estimation with Ybarra-Lohr SaeME produces smaller EMSE averages. When the percentage of small areas containing measurement errors is 80 to 100, the mean bias and EMSE results with Hierarchical Bayesian SaeME are smaller. The second simulation study was carried out by creating scenarios on the auxiliary variables. When the auxiliary variables that contain errors increase, the average bias value and the resulting average EMSE from Hierarchical Bayesian SaeME method are also greater as the measurement error on the auxiliary variables increases. In general, for estimating small area estimation with measurement error, the Hierarchical Bayesian SaeME estimator was the best for the one and two simulation studies because the average bias and average EMSE were lower than the direct estimates, SAE EBLUP-FH and Ybarra-Lohr

SaeME when auxiliary variable is measured with error. For future research, the Hierarchical Bayesian SaeME method can be used to obtain small area estimation for empirical data to obtain better accuracy results which the auxiliary variables containing measurement errors such as from survey data.

ACKNOWLEDGMENT

If so, write an Acknowledgment or appreciation in this section. Acknowledgments can be addressed to funders (sponsors) who contributed to the article. It could be also to people who contributed to the article or data in the article.

REFERENCES

[1] J. N. K. Rao and I. Molina, Small Area Estimation Second Edition, New Jersey: John Willey & Sons, 2015.

[2] A. Ubaidillah, Simultaneous equation models for small area estimation, Bogor: IPB University, 2020.

[3] A. L. E., C. F. and P. Lahiri, "Use of administrative records in small area estimation," in *Administrative Records for Survey Methodology*, New Jersey, John Wiley & Sons, Inc., 2021, pp. 231-267.

[4] A. Kurniawan , E. Elmira , M. D. Anbarani, M. Rizky, N. S. Saputri and R. Al Izzati, Testing Small Area Estimation (SAE) Method for Generating Nutrion Maps in Indonesia: Rokan Hulu District, Jakarta: The SMERU Research Institute, 2019.

[5] P. P. Biemer, R. M. Groves, L. E. Lyberg and N. A. Ma, Measurement Errors in Surveys, John Willey & Sons, Inc, 2004.

[6] R. J. Carroll, D. Ruppert, L. A. Stefanski and C. M. Crainiceanu, Measurement Error in Nonlinear Models, CRC, 2006.

[7] T. Singh, S. Wang and R. J. Carroll, "Efficient Small Area Estimation When Covariates Are Measured With Error Using Simulation Extrapolation," in *The 60th ISI World Statistics Congress*, Rio De Janeiro, 2015.

[8] S. Hariyanto, K. A. Notodiputro, A. Kurnia and K. Sadik, "Measurement error in small area estimation: A literature review," in *IOP Conf. Ser. Earth Environ. Sci.*, 2018.

[9] E. Tanur, Pendugaan Area Kecil untuk Model Autoregresif dengan Kesalahan Pengukuran pada Peubah Penyerta, Bogor: IPB University, 2020.

[10] M. Ghosh, K. Sinha and D. Kim, "Empirical and hierarchical Bayesian estimation in finite population sampling under structural measurement error models," *Scand. J. Stat.*, vol. vol. 33, no. no. 3, p. pp. 591–608, 2006, doi: 10.1111/j.1467-9469.2006.00492.x.

[11] M. Torabi, G. S. Datta and J. N. K. Rao, "Empirical bayes estimation of small area means under a nested error linear regression model with measurement errors in the covariates," *Scand. J. Stat.*, vol. vol. 36, no. no. 2, p. pp. 355–369, 2009, doi: 10.1111/j.1467-9469.

[12] E. Torkashvand, On Small Area Estimation Problems with Measurement Errors and Clustering, The University of Manitoba, 2016.

[13] L. M. R. Ybarra and S. L. Lohr, "Small area estimation when auxiliary information is measured with error," *Biometrika*, vol. vol. 95, no. no. 4, p. pp. 919–931, 2008, doi: 10.1093/biomet/asn048.

[14] R. Zhu and G. H. Zou, "BLUP estimation of linear mixed-effects models with measurement errors and its applications to the estimation of small areas," *Acta Math. Sin. Engl. Ser.*, vol. vol. 30, no. no. 12, p. pp. 2027–2044, 2014, doi: 10.1007/s10114-014-2707-5.

[15] G. Datta, A. Delaigle, P. Hall and L. Wang, " Semiparametric prediction intervals in small area when auxiliary data are measured with error," *Stat. Sin*, pp. 28:232-2335, 2018a.

[16] G. Datta, M. Torabi, J. Rao and B. Liu, "Small area estimation with multiple covariates measured with error: a nested error linear regression approach of combining multiple surveys," *J. Mult. Anal.*, pp. 167:49-59, 2018b.

[17] I. Wulandari, A. Kurnia, K. A. Notodiputro and A. Fitrianto, "Small area estimation with multiple covariates under structural measurement error models," in *Procedia Computer Science 216 (2023) 168–176*, 2023.

[18] M. Komalasari, Kajian Pendugaan Area Kecil Menggunakan Peubah Penyerta yang Mengandung Galat (Studi Kasus: Ratarata Lama Sekolah di Kabupaten Kampar), Bogor: IPB University, 2019.

[19] S. D. Aziz and A. Ubaidillah, "Big Data for Small Area Estimation : Happiness Index with Twitter Data," in *The 1st International Conference On Data Science And Official Statistics ICDSOS*, 2021.

[20] S. Hariyanto, K. A. Notodiputro, A. Kurnia and K. Sadik, "Small Area Estimation with Measurement Error in t Distributed Covariate Small Area Estimation with Measurement Error in t Distributed Covariate Variable," *International Journal On Advanced Science Engineering Information Technology*, vol. vol. 10, no. no. 4, 2020, doi: 10.18517/ijaseit.10.4.9765.

[21] E. Tanur and A. Kurnia, "Small Area Estimation For Autoregressive Model With Measurement Error In The Auxiliary Variable," *CMBN*, p. pp. 1–29, 2022, https://doi.org/10.28919/cmbn/7577.

[22] F. Novkaniza, K. A. Notodiputro, K. Sadik and I. W. Mangku, "Poisson-lognormal model with measurement error in covariate for small area estimation of count data," *CMBN*, p. pp. 1–20, 2023, https://doi.org/10.28919/cmbn/7779.

[23] S. Arima, G. S. Datta and B. Liseo, "Bayesian Estimators for Small Area Models when Auxiliary Information is Measured with Error," *Scand. J. Stat.*, vol. vol. 42, no. no. 2, p. pp. 518–529, 2015, doi: 10.1111/sjos.12120.

[24] K. Sadik and K. A. Notodiputro, "Metode E-Blup Dalam Small Area Estimation Untuk Model Yang Mengandung Random Walk," *Forum Statitika dan Komputasi,* vol. vol. 11, no. no.2, pp. p: 37-41, 2006.

[25] F. Novkaniza, Poisson-Lognormal Model with Measurement Error in Its Covariates for Small Area Estimation of Count Data, Bogor: IPB University, 2021.