*ICIC Express Letters* ICIC International ©2021 ISSN 1881-803X
Volume **15**, Number **5**, May **2021** pp. **449–456**

# LOG-NORMAL SMALL AREA ESTIMATION WITH MEASUREMENT ERROR – AN APPLICATION FOR ESTIMATING HOUSEHOLD CONSUMPTION EXPENDITURE

Erwin Tanur[1,2], Anang Kurnia[2], Khairil Anwar Notodiputro[2] and Agus Mohamad Soleh[2]

[1]Education and Training Centre
Statistics Indonesia
Jalan Raya Jagakarsa 70, Lenteng Agung, Jakarta 12620, Indonesia
wintanoer@bps.go.id

[2]Department of Statistics
IPB University
Jalan Meranti, Kampus IPB Darmaga, Bogor 16680, Indonesia
{ anangk; khairil; agusms }@apps.ipb.ac.id

Received September 2020; accepted December 2020

Abstract. *We propose an alternative method for estimating average per capita consumption in a small area with information from repeated surveys. This method is a modified method of the empirical Bayes estimation of a small area mean in a nested linear regression model with measurement errors on the covariates. The problem is how to deal with the use of additional variables with measurement error to estimate the provincial average per capita consumption. Estimator values for the district/city level were obtained after estimating information at the unit level that was not the sample unit in the survey. This case is found in the National Socioeconomic Survey. At the application stage with the survey data, the estimation with measurement error on the auxiliary variables can give better results.*
**Keywords:** Subsample survey, Small area estimation, Restricted maximum likelihood, Structural measurement error, Mean square error bootstrap

1. **Introduction.** Any survey design generally faces a challenge either to make conclusions at the desired level of precision at minimal cost or to maximize precision at a fixed cost. This challenge continues to increase as the reduction in response rates in the most recent annual survey increases the risk of bias. If the answers in the survey differ from the actual values, a measurement error occurs. One survey that is routinely conducted in Indonesia is the National Socioeconomic Survey (*Susenas*). In its implementation, the survey collected data regarding the socioeconomic conditions of the community. There is some important information in the implementation of the 2015 *Susenas* namely: 1) data collection was performed in March and September; 2) data collection in March was carried out with a large number of samples to produce representative data for the district/city level; 3) the data collection in September was conducted with a small sample size and produced representative data only for the provincial and national levels [1]. A few available sample information in September made it impossible to estimate per capita expenditure/consumption at the district/city level or in smaller areas.

In a research study, researchers often deal with asymmetrical populations, which tend to be rightward dominant, usually occurring in business surveys and agricultural companies, as well as surveys on personal income and wealth. According to Karlberg, if using the direct method, the total population of survey variables with a small number of samples

DOI: 10.24507/icicel.15.05.449

would be a problem for two reasons: 1) when there are no extreme values in the sample, too small estimators are obtained, and 2) if there are extreme values in the sample, the estimation becomes very large [2]. Rao and Molina, use annual net income information as a welfare variable for individuals [3]. The variable has been changed by adding a fixed quantity to always be positive and then taking the logarithmic value. Tanur et al. [4] have implemented a model proposed by Torabi et al. [5] using the 2015 *Susenas* data to estimate the average per capita consumption in districts/cities. Research on more specific objects has also been carried out by Yusiong and Naval who have proposed an unsupervised framework for monocular depth estimation that trains a Siamese convolutional long short-term memory (Siamese convLSTM) network to jointly perform estimation and refinement of depth maps using rectified stereo image pairs and produce a depth map from a single RGB image at test time [6].

Based on the research results that have been mentioned, it is necessary to develop a method of estimating small areas in populations that are not symmetrical with the auxiliary variables obtained from the survey results. It is important to develop a small area estimation method to increase the effectiveness of small sample sizes in the September period so that the estimator value for per-capita expenditure/consumption is obtained by utilizing information available in the March survey period.

The remainder of the paper is arranged as follows. Section 2 discusses the methods used. Section 3 describes in detail our proposed model, lognormal small area estimation with measurement error. Section 4 reports the implementation with *Susenas* 2015 data. Section 5 describes the results of the simulation study obtained. Finally, Section 6 concludes the paper.

2. **Methods.** To resolve the problems previously mentioned, this research used a small area estimation approach method by considering the case of measurement error on the auxiliary variables and applying the lognormal transformation to the interest variable.

2.1. **Small area estimation.** The small area estimation model was first introduced by Fay and Herriot in 1979, popularly called the Fay-Herriot Model [7], in the form of a model:

$$y_i = \theta_i + e_i, \quad \theta_i = X_i^{\mathrm{T}}\beta + v_i \tag{1}$$

where $e_i$ and $v_i$ are mutually independent, $E(e_i) = E(v_i) = 0$, $Var(e_i) = \sigma_e^2$ and $Var(v_i) = \sigma_v^2$, $i = 1, 2, \ldots, m$. The $y_i$ component is a direct estimator for the $i$-th area and obtained from the corresponding survey data, $\theta_i$ is an interesting parameter, $e_i$ is a sample error, $X_i = (X_{i1}, X_{i2}, \ldots, X_{ip})^{\mathrm{T}}$ is the auxiliary variable and $v_i$ is an area random effect. The natural way to estimate $\theta_i$ in Equation (1) is to replace $\beta$ and $v_i$ with $\hat{\beta}$ and $\hat{v}_i$ respectively.

Two model approaches which are applied in small area estimation are basic area-level and the basic unit-level model [8]. The basic area-level model is a model based on the availability of supporting data that only exist for a certain area level:

$$y_i = X_i^{\mathrm{T}}\beta + v_i + e_i \tag{2}$$

Furthermore, the basic unit-level model is a model in which the supporting data available are compatible individually with response data, for example, $X_{ij} = (X_{ij1}, X_{ij2}, \ldots, X_{ijp})^{\mathrm{T}}$:

$$y_{ij} = X_{ij}^{\mathrm{T}}\beta + v_i + e_{ij} \tag{3}$$

with $i = 1, 2, \ldots, m$, and $j = 1, 2, \ldots, n_i$, also $e_{ij} \sim N(0, \sigma_e^2)$.

2.2. **Measurement error on the auxiliary variables.** The assumption in the small area estimation model is that the auxiliary variables are measured without errors [9]. Therefore, the auxiliary information commonly used is the data from the census and administrative or registry. However, in reality, the census and administrative data are often not fully available and up to date to be used as the auxiliary information. The solution is to use survey data as auxiliary information in a small area estimation model, although consequently the use of survey data may contain sample errors.

Carroll et al. explain that there are two widely used measurement error models [10]. The first model is the classic measurement error model, with the form

$$w_i = X_i + \eta_i \tag{4}$$

where $w$ is the auxiliary variable that contains an error, and $X$ is the actual variable value. The $\eta$ component is an additive error with $\eta_i \sim N\left(0, \sigma_\eta^2\right)$; it assumes that $\eta$ and $X$ are mutually independent, $w$ fluctuates around $X$, diversity of $w$ is more than $X$, and $E(w_i|X_i) = X_i$, then $w$ is an unbiased estimator for $X$. The second model is the Berkson measurement error model:

$$X_i = w_i + \eta_i \tag{5}$$

where $X$ is the actual variable value, and $w$ is the auxiliary variable that contains an error. The $\eta$ component is an additive error with $\eta_i \sim N\left(0, \sigma_\eta^2\right)$; it assumes that $w$ and $\eta$ are mutually independent, $X$ fluctuates around $w$, and $E(X_i|w_i) = w_i$, then $X$ is an unbiased estimator for $w$.

2.3. **Lognormal transformation model.** A random variable $X$ has a lognormal distribution if the transformation $Y = \ln(X)$ has a normal distribution, the pdf for $X$:

$$f\left(X|\mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi}\sigma X} e^{-\frac{(\ln(X)-\mu)^2}{2\sigma^2}} X > 0 \tag{6}$$

Expected value and variances value of $X \sim \text{lognormal}\left(\mu, \sigma^2\right)$:

$$E(X) = e^{\mu + \frac{1}{2}\sigma^2} \text{ and } Var(X) = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right) \tag{7}$$

The parameters $\mu$ and $\sigma$:

$$\mu = \ln\left(\frac{E(X)^2}{\sqrt{Var(X) + E(X)^2}}\right) \text{ and } \sigma^2 = \ln\left(1 + \frac{Var(X)}{E(X)^2}\right) \tag{8}$$

3. **Proposed Model.** The basic model in this paper refers to Torabi et al., with auxiliary variables at the area level, $w$, which contains measurement errors [5]:

$$y_{ij} = \beta_0 + w_i\alpha + v_i + e_{ij}, \quad W_{ij} = w_i + \eta_{ij}, \quad (i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n_i) \tag{9}$$

**Modification model.** The modified model hereinafter referred to the small area estimation model with measurement error on the auxiliary variable (SAE-ME):

$$y_{ij}^* = X_{ij}^{\mathrm{T}}\beta + w_i\alpha + v_i + e_{ij}, \quad W_i = w_i + \eta_i \tag{10}$$

where $y_{ij}^* = \log(y_{ij})$ [11], $y_{ij}$ is the value of the interesting variable for the $j$-th unit in the $i$-th area; $X_{ij}$ is an auxiliary variable at the unit level in the $i$-th area (fixed effect); $\beta$ is a fixed effect coefficient; $w_i$ is the actual value of the area-specific covariate, which is unknown related to $y_{ij}$, with $w_i \sim N\left(\mu_w, \sigma_w^2\right)$; $\alpha$ is a coefficient of the variable with measurement error; $v_i$ is the area random effect with $v_i \sim N\left(0, \sigma_v^2\right)$; $e_{ij}$ is a model error with $e_{ij} \sim N\left(0, \sigma_e^2\right)$; components $v_i$, $e_{ij}$ and $w_i$ are assumed to be mutually independent; $W_i$ is an auxiliary variable at area level with measurement errors; $\eta$ is the measurement error on the auxiliary variable with $\eta_{ij} \sim N\left(0, \sigma_\eta^2\right)$; $m$ is the number of areas.

Based on Equation (10) we get the expected values and variance values of $y_{ij}^*$:

$$E\left(y_{ij}^*\right) = X_{ij}^{\mathrm{T}}\beta + \alpha\mu_w \text{ and } Var\left(y_{ij}^*\right) = \sigma_v^2 + \sigma_e^2 + \alpha^2\sigma_w^2 \tag{11}$$

Equation (10) can be written in the form of a matrix:

$$\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{w}\alpha + \mathbf{Z}\mathbf{v} + \mathbf{e} \tag{12}$$

where $\mathbf{y}^* = (\mathbf{y}_1^*, \ldots, \mathbf{y}_m^*)^{\mathrm{T}}$ is the vector of the interesting variable with sized $n \times 1$; $\mathbf{y}_i^* = (\mathbf{y}_{i1}^*, \ldots, \mathbf{y}_{in_i}^*)^{\mathrm{T}}$ with the size $n_i \times 1$; $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_m)^{\mathrm{T}}$ is a fixed effect matrix of size $n \times p$, $p$ is number of fixed variables, with $\mathbf{X}_i = (X_{i1}, \ldots, X_{in_i})^{\mathrm{T}}$ matrix of size $n_i \times p$; $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ is a fixed effect coefficient vector with size $p \times 1$, $\alpha$ is scalar value for coefficient of auxiliary variable with measurement error, $\mathbf{w} = (w_1, \ldots, w_m)^{\mathrm{T}}$ is auxiliary variable vector with measurement error with the size $m \times 1$; $\mathbf{v} = (v_1, \ldots, v_m)^{\mathrm{T}}$ is the vector of area random effect with the size $m \times 1$; $\mathbf{e} = (e_1, \ldots, e_m)^{\mathrm{T}}$ with $e_i = (e_{i1}, \ldots, e_{in_i})^{\mathrm{T}}$ is the vector of error model with the size $n \times 1$; $\mathbf{Z} = \oplus_{i=1}^m \mathbf{Z}_i = diag\,(\mathbf{Z}_1, \ldots, \mathbf{Z}_m)$ with the size $n \times m$, $\mathbf{Z}_i = \mathbf{1}_{n_i}$ is a vector of size $n_i$ with all elements of 1.

Equation (12) produces the values of $E\,(\mathbf{y}^*|\mathbf{v})$ and $Var\,(\mathbf{y}^*|\mathbf{v})$:

$$E\,(\mathbf{y}^*|\mathbf{v}) = \mathbf{X}\boldsymbol{\beta} + \mu_w \alpha \mathbf{1}_n + \mathbf{Z}\mathbf{v} \text{ and } Var\,(\mathbf{y}^*|\mathbf{v}) = \alpha^2 \sigma_w^2 \mathbf{1}_n + \sigma_e^2 \mathbf{1}_n \tag{13}$$

where $\mathbf{1}_n$ is a vector of the size $n$ with all elements of value 1. Based on Equation (13), the value of $E(\mathbf{y}|\mathbf{v})$ or $\hat{\mathbf{y}}$ is obtained through back transformation, and then obtained in the form of an equation:

$$\hat{\mathbf{y}} = \exp\left[(\mathbf{X}\boldsymbol{\beta} + \mu_w \alpha \mathbf{1}_n + \mathbf{Z}\mathbf{v}) + 0.5\left(\alpha^2 \sigma_w^2 \mathbf{1}_n + \sigma_e^2 \mathbf{1}_n\right)\right] \tag{14}$$

By substituting the estimators of the model parameters with $\left(\hat{\beta}, \hat{\mu}_w, \hat{\alpha}, \hat{v}_i, \hat{\sigma}_e^2, \hat{\sigma}_w^2\right)$ based on the sample data, we produce

$$\hat{\mathbf{y}} = \exp\left[\left(\mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mu}_w \hat{\alpha} \mathbf{1}_n + \mathbf{Z}\hat{v}\right) + 0.5\left(\hat{\alpha}^2 \hat{\sigma}_w^2 \mathbf{1}_n + \hat{\sigma}_e^2 \mathbf{1}_n\right)\right] \tag{15}$$

The target parameters obtained from the sum of the observed values that are members of the sample, with index $(s)$, and unit values that are not sample members, with index $(r)$. The average value by area is estimated by

$$\hat{\bar{\mathbf{Y}}}_i = \frac{1}{N_i}\left(\sum_{(s)} \mathbf{y} + \sum_{(r)} \hat{\mathbf{y}}\right) \tag{16}$$

Equation (16) is obtained with the following stages:

- estimating the variance component $(\hat{\sigma}_v^2, \hat{\sigma}_e^2, \hat{\sigma}_w^2)$ with the restricted maximum likelihood (REML) method,
- estimating the coefficient of a random variable that has a measurement error $(\hat{\alpha})$, referred to Torabi et al. [5]:

$$\hat{\alpha} = \left(\frac{\hat{\sigma}_w^2}{\hat{\sigma}_w^2 + \hat{\sigma}_\eta^2}\right)^{-1} \frac{\sum_{i \in s} (y_i - \bar{y}_s)(w_i - \bar{w}_s)}{\sum_{i \in s} (w_i - \bar{w}_s)^2}, \text{ and } \hat{\sigma}_\eta^2 = \frac{\sum_{i=1}^m \sum_{j=1}^{n_i} (w_i - \bar{w}_s)^2}{\left(\sum_i n_i\right) - m} \tag{17}$$

- estimating the coefficient of a fixed effect variable $\left(\hat{\boldsymbol{\beta}}\right)$:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{y}^* - \mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\hat{\alpha}\hat{\mu}_w\right) \tag{18}$$

  with $\mathbf{V} = \hat{\sigma}_v^2 \mathbf{J}_{n_i} + \alpha^2 \hat{\sigma}_w^2 \mathbf{J}_{n_i} + \hat{\sigma}_e^2 \mathbf{J}_{n_i}$, $\mathbf{J}_{n_i}$ is a square matrix with size $n_i \times n_i$,
- estimating the area random effect value $(\hat{v}_i)$ [12]:

$$\hat{v}_i = \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + \hat{\alpha}^2 \hat{\sigma}_w^2 + \hat{\sigma}_e^2 / n_i}\left(\bar{\mathbf{y}}_i^* - \bar{\mathbf{X}}_i \hat{\boldsymbol{\beta}} - w_i \hat{\alpha}\right) \tag{19}$$

  $\bar{\mathbf{y}}_i^*$ is vector of the average value of $y_{ij}^*$ at area level, $\bar{\mathbf{X}}_i$ is matrix of average value of fixed variables of $X_{ij}$ at area level, and $w_i$ is vector of the auxiliary variable with measurement error at area level.

After obtaining Equation (16), the estimator value for provincial average is obtained from

$$\hat{\hat{\mathbf{Y}}} = \frac{1}{N}\sum_{i=1}^{m}\hat{\hat{\mathbf{Y}}}_i \times N_i \tag{20}$$

The next stage is to estimate the value of mean squared error, denoted $mse\left(\hat{\hat{\mathbf{Y}}}_i\right)$ by using the bootstrap method [3], through equation:

$$mse\left(\hat{\hat{\mathbf{Y}}}_i^*\right) = B^{-1}\sum_{b=1}^{B}\left(\hat{\hat{\mathbf{Y}}}_{i^*}^H(b) - \hat{\hat{\mathbf{Y}}}_{i^*}(b)\right)^2 \tag{21}$$

Assume $\bar{\mathbf{Y}}_i^* = \bar{\mathbf{X}}_i^T\hat{\boldsymbol{\beta}} + \mathbf{w}_i^*\hat{\alpha} + \mathbf{v}_i^*$ is the bootstrap version of the target parameter $\bar{\mathbf{Y}}_i = \bar{\mathbf{X}}_i^T\boldsymbol{\beta} + \mathbf{w}_i\alpha + \mathbf{v}_i$. Then, by using bootstrap data, the bootstrap version $\hat{\hat{\mathbf{Y}}}_{i^*}^H$ from the EBLUP estimator $\hat{\hat{\mathbf{Y}}}_i^H$ which is generated from $\hat{\hat{\mathbf{Y}}}_{i^*}^H = \bar{\mathbf{X}}_i^T\hat{\boldsymbol{\beta}}^* + \mathbf{w}_i^*\hat{\alpha}^* + \hat{\mathbf{v}}_i^*$ with $\hat{\boldsymbol{\beta}}^*$, $\hat{\alpha}^*$, $\hat{\mathbf{v}}_i^*$ is obtained in the same way to obtain $\hat{\boldsymbol{\beta}}$, $\hat{\alpha}$ and $\hat{\mathbf{v}}_i$, but by using bootstrap sample data. The bootstrap estimation theory of $mse\left(\hat{\hat{\mathbf{Y}}}_i^H\right)$ is determined to refer to $mse_B\left(\hat{\hat{\mathbf{Y}}}_i^H\right) = E\left(\hat{\hat{\mathbf{Y}}}_{i^*}^H - \bar{\mathbf{Y}}_i^*\right)^2$. This bootstrap estimate is an approximation of Monte Carlo simulation, where each stage is repeated as many large numbers, $B$. At this stage a number of $B$ values of $\bar{\mathbf{Y}}_i^*$ are obtained, i.e., $\bar{\mathbf{Y}}_i^*(1), \ldots, \bar{\mathbf{Y}}_i^*(B)$ of the actual value of the bootstrap is $\bar{\mathbf{Y}}_i^*$, along with a number of $B$ values from $\hat{\hat{\mathbf{Y}}}_{i^*}^H$.

4. **Implementation with *Susenas* 2015 Data.** The 2015 *Susenas* data from West Java Province were applied using the modified model, according to Equation (10). The interesting variable was the log of an average of per capita household consumption at the unit level, which was obtained from the September period. The unit in this study was a sub-district, while the area level in this study was the district/city. When using sub-districts as a unit level, the weaknesses of application to the model emerge. In the concept of small area estimation, the unit level auxiliary variable that is used must be without error. The average information of household per-capita consumption in September is used as an interesting variable ($y$), which is the aggregate of households in the sub-district, so correction needs to be done first.

One of the auxiliary variables related to the household consumption and population composition in each district was the availability of facilities to meet household consumption needs, which was obtained from PODES (Village Potential) data in 2014. Information on the proportion of the total restaurants and food and drink stalls is adjusted to the composition of the population in each district, used as fixed variables. Variable $X_1$ is the proportion of the number of restaurants to the total population in each district. Variable $X_2$ is the proportion of the number of food and drink stalls with a total population in each district. The $W$ variable is the log of the average per capita household consumption at the district level resulting from the March *Susenas*.

Table 1 shows the district/city level estimates obtained based on Equation (16), by adding recorded unit information and units obtained from the estimated SAE-ME or SAE methods. Units that were not recorded in September were previously estimated by the SAE-ME method by using the information in March as an auxiliary variable with measurement errors, and by the SAE method with an auxiliary variable without measurement errors.

Table 1 presents the goodness value model of the SAE-ME estimation model with the SAE estimation. Table 1 shows the root mean squared error (rmse) values for the two types of models. the root mean square error (rmse) is a frequently used measure of the differences between values (sample or population values) predicted by a model

TABLE 1. Estimation value and goodness of model by area and estimation method in West Java Province, 2015

| Number of districts/ cities | SAE-ME | | | | SAE | | | |
|---|---|---|---|---|---|---|---|---|
| | Estimator | Relative bias | rmse | CV | Estimator | Relative bias | rmse | CV |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| 1 | 1053733 | −0.049 | 30388 | 11.57 | 1077328 | −0.249 | 72818 | 27.71 |
| 2 | 978637 | 0.011 | 26476 | 9.61 | 937416 | −0.142 | 47530 | 17.25 |
| 3 | 766866 | 0.001 | 26440 | 9.04 | 680416 | 0.155 | 45770 | 15.65 |
| 4 | 890268 | 0.014 | 26700 | 10.50 | 873180 | −0.256 | 74239 | 29.20 |
| 5 | 898527 | −0.038 | 29237 | 13.39 | 759267 | −0.140 | 47426 | 21.72 |
| 6 | 708168 | 0.014 | 26154 | 9.40 | 593710 | −0.129 | 44403 | 15.96 |
| 7 | 774911 | 0.017 | 26895 | 11.48 | 720901 | 0.090 | 33942 | 14.49 |
| 8 | 883285 | −0.006 | 26657 | 9.63 | 835468 | −0.202 | 61424 | 22.19 |
| 9 | 762396 | 0.036 | 26986 | 11.70 | 678524 | 0.087 | 32929 | 14.28 |
| 10 | 877221 | 0.015 | 26450 | 10.85 | 848665 | 0.431 | 112892 | 46.33 |
| 11 | 924221 | 0.021 | 26990 | 8.06 | 913130 | 0.410 | 106950 | 31.95 |
| 12 | 778274 | 0.021 | 26688 | 9.97 | 734721 | −0.024 | 28009 | 10.46 |
| 13 | 859237 | 0.053 | 28799 | 11.31 | 826356 | −0.066 | 33075 | 12.99 |
| 14 | 905313 | 0.023 | 26282 | 10.08 | 890117 | 0.198 | 55652 | 21.34 |
| 15 | 887475 | −0.030 | 28200 | 8.32 | 876411 | −0.145 | 48021 | 14.17 |
| 16 | 1246485 | 0.027 | 26589 | 9.53 | 1305776 | −0.006 | 26574 | 9.52 |
| 17 | 650530 | 0.005 | 26510 | 8.81 | 618905 | 0.029 | 26941 | 8.96 |
| 18 | 695975 | 0.030 | 26671 | 10.62 | 671424 | −0.032 | 28370 | 11.30 |
| 19 | 1600269 | −0.047 | 30578 | 12.10 | 1612832 | −0.080 | 35779 | 14.16 |
| 20 | 1104629 | −0.001 | 25947 | 9.37 | 1104629 | −0.095 | 37467 | 13.53 |
| 21 | 1502913 | 0.015 | 26160 | 10.56 | 1613311 | 0.384 | 100611 | 40.63 |
| 22 | 927319 | 0.011 | 26775 | 11.74 | 927319 | 0.119 | 38947 | 17.08 |
| 23 | 1635938 | 0.047 | 27682 | 9.79 | 1635938 | 0.386 | 101352 | 35.84 |
| 24 | 2033990 | −0.020 | 27825 | 11.31 | 2083311 | 0.078 | 31987 | 13.00 |
| 25 | 1332503 | 0.031 | 27403 | 10.89 | 1332503 | 0.142 | 43645 | 17.34 |
| 26 | 876442 | 0.025 | 26549 | 10.75 | 876442 | −0.014 | 26991 | 10.93 |
| 27 | 890235 | −0.004 | 26695 | 13.33 | 890235 | 0.027 | 26850 | 13.41 |
| Province | 1053671 | 0.008 | 27212 | 10.51 | 1037430 | 0.035 | 50763 | 19.31 |

or an estimator and the values observed. It is used to estimate positional accuracy, to measure how well your model performed. It does this by measuring the difference between predicted values and the actual values.

From Table 1, it can be seen that the small area estimation model with measurement errors in the auxiliary variable (SAE-ME) produces a smaller rmse value. This shows that the distribution of estimator data produced by the small area estimation method with the error measurement method (SAE-ME) is relatively more reliable. The difference in the predicted value with the SAE-ME method is smaller, and this model is better at estimating. The same case can be seen from the resulting coefficient of variation, which shows that the estimation results generated by the SAE-ME approach are more homogeneous. The SAE-ME model produces a more homogeneous estimation value, and the resulting estimate is closer to the average value. So using the SAE-ME method will produce estimates with a value distribution that is closer to the average value. The SAE-ME method provides good and accurate prediction results.

5. **Simulation Study.** We conducted simulation studies on the proposed prediction model's (SAE-ME) goodness test and its basic prediction model (SAE), and each was obtained by considering the presence of an element of measurement error in the auxiliary variable. It is assumed that $y_{ij}$ for the population unit is generated from model (10) with $\beta_0 = -1$, $\beta_1 = 2$, $\beta_2 = -1$, $\alpha = 1$, $\mu_w = 14$, $\sigma^2_{v\_small} = 0.01$, $\sigma^2_{v\_medium} = 0.05$, $\sigma^2_{v\_large} = 0.1$, $\sigma^2_e = 0.01$, $\sigma^2_{w\_small} = 0.01$, $\sigma^2_{w\_medium} = 0.05$, $\sigma^2_{w\_large} = 0.1$. The population has $N = 61100$ units spread over 27 areas $(m)$ of each size $(N_i)$: 4000, 4400, 3100, 3100, 3600, 3900, 2600, 3100, 4000, 2600, 2600, 3100, 2900, 1600, 2900, 2300, 1600, 1000, 600, 700, 2900, 500, 1200, 1100, 300, 1000, and 400. Sample size $(n_i)$ in the specified area is respectively: 23, 18, 17, 20, 19, 20, 17, 16, 22, 17, 14, 20, 16, 12, 20, 15, 13, 8, 5, 7, 18, 5, 12, 9, 3, 10, and 4.

For this purpose, we produce $B = 5000$ independent sets with normal distribution $\{v_i(b); i = 1, \ldots, m\}$, $\{e_{ij}(b); j = 1, \ldots, N_i; i = 1, \ldots, m\}$, $(b = 1, \ldots, B)$, with a mean value of zero and the variance specified for $\sigma^2_v$, $\sigma^2_w$ and $\sigma^2_e$. We also generated $\{w_i; i=1, \ldots, m\}$ normally distributed with mean value $\mu_w$ and variance $\sigma^2_w$. Using $\{v_i(b), e_{ij}(b), w_i\}$, a total of $B$, $\{y_{ij}(b); j = 1, \ldots, N_i; i = 1, \ldots, m\}$ set of populations is obtained with Equation (10). In each data generated, parameters are obtained according to the data equation that has been generated and then applied to Equations (11)-(21), so that the simulation results are obtained from each of the average value of relative bias $(\overline{RB}_i)$, average value of root mean squared error $(\overline{RMSE}_i)$ and average value of coefficient of variation $(\overline{CV}_i)$, the results can be seen in Table 2. The values of $RB_i$, $RMSE_i$ and $CV_i$ are obtained from: $RB_i = \frac{1}{B}\sum_{b=1}^{B} \frac{\bar{y}_{i\_s\_indirect} - \bar{y}_{i\_population}}{\bar{y}_{i\_population}}$, $RMSE_i = \sqrt{\frac{1}{B}\sum_{b=1}^{B}\left(\bar{y}_{i\_s\_indirect} - \bar{y}_{i\_population}\right)^2}$, and $CV_i = \frac{RMSE_i}{\bar{y}_{i\_population}} \times 100\%$.

TABLE 2. The value of the goodness of the model according to the estimation method, the variance value of area random effect $(\sigma^2_v)$ and the variance value of the auxiliary variable with measurement error $(\sigma^2_w)$

| Goodness of model value | Variance value of area random effect $(\sigma^2_v)$ | Method | | Method | | Method | |
|---|---|---|---|---|---|---|---|
| | | SAE-ME (Small $\sigma^2_w$) | SAE | SAE-ME (Medium $\sigma^2_w$) | SAE | SAE-ME (Large $\sigma^2_w$) | SAE |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Average value of relative bias | Small | 0.1099 | 0.1100 | 0.1064 | 0.1066 | 0.1201 | 0.1215 |
| | Medium | 0.1221 | 0.1223 | 0.1092 | 0.1100 | 0.1007 | 0.1016 |
| | Large | 0.2554 | 0.2566 | 0.0918 | 0.0928 | 0.1213 | 0.1253 |
| Average value of RMSE | Small | 176921 | 177332 | 175336 | 176286 | 181737 | 185409 |
| | Medium | 187017 | 187856 | 179999 | 181851 | 177393 | 182118 |
| | Large | 296988 | 300885 | 183717 | 187406 | 194783 | 202571 |
| Average value of CV | Small | 24.28 | 24.34 | 24.05 | 24.18 | 24.96 | 25.47 |
| | Medium | 25.43 | 25.54 | 24.72 | 24.97 | 24.25 | 24.96 |
| | Large | 41.51 | 42.07 | 25.15 | 25.63 | 26.96 | 28.10 |

6. **Conclusion.** The use of information with measurement errors as an auxiliary variable in the form of random variables can result in estimating small areas at the provincial and district/city level by first estimating unregistered units (sub-districts). Estimates are made for the average per capita consumption in September by using the results of the March period survey as an auxiliary variable at each unit level.

The application of the 2015 *Susenas* data to the small area estimation model with measurement errors has resulted in a better estimation than without regard to measurement errors. The small area estimation model with measurement error on the auxiliary variable produces a smaller standard deviation and coefficient of variation value compared to the small area estimation model without regard to measurement errors on the auxiliary variable.

Further research related to this problem is the development of a small area estimation model for repeated data in a certain time unit. The model that is formed includes the time random effect, in addition to the area random effect and the element of measurement error on the auxiliary variables.

## REFERENCES

[1] [BPS] Badan Pusat Statistik, *Buku 1 Susenas September 2015*, Jakarta, 2015.

[2] F. Karlberg, Population total prediction under a lognormal superpopulation model, *Metron*, vol.8, pp.53-80, 2000.

[3] J. N. K. Rao and Molina, *Small Area Estimation*, 2nd Edition, Wiley, New York, 2010.

[4] E. Tanur, A. Kurnia, K. A. Notodiputro and A. M. Soleh, Estimation of small area means for subsample repeated measurement data, *IOP Conf. Ser.: Earth Environ. Sci.*, 2018.

[5] 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, *Scandinavian Journal of Statistics*, DOI: 10.1111/j.1467-9469.2008.00623.x, 2009.

[6] J. P. T. Yusiong and P. C. Naval, Jr., Unsupervised monocular depth estimation of driving scenes using siamese convolutional LSTM networks, *International Journal of Innovative Computing, Information and Control*, vol.16, no.1, pp.91-106, 2020.

[7] R. E. Fay and R. A. Herriot, Estimates of income for small places an application of James-Stein procedures to census data, *Journal of the American Statistical Association*, vol.74, pp.269-277, 1979.

[8] J. N. K. Rao, *Small Area Estimation*, John Wiley and Sons, New York, 2003.

[9] L. M. R. Ybarra and S. L. Lohr, Small area estimation when auxiliary information is measured with error, *Biometrika*, vol.94, no.4, pp.919-931, 2008.

[10] R. J. Carroll, D. Ruppert, L. A. Stefanski and C. Crainiceanu, *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd Edition, Chapman & Hall, New York, 2006.

[11] D. Thorburn, Model-based estimation in survey sampling of lognormal distribution, *A Spectrum of Statistical Thought, Essays in Statistical Theory, Economics and Population Genetics in Honor of Johan Fellman*, pp.228-243, The Swedish School of Economics and Business Administration Helsinki, 1991.

[12] C. E. McCulloch and S. R. Searle, *Generalized, Linear, and Mixed Models*, John Wiley & Sons, New York, 2001.