DOI: 10.15849/IJASCA.240730.07

Int. J. Advance Soft Compu. Appl, Vol. 16, No. 2, July 2024
Print ISSN: 2710-1274, Online ISSN: 2074-8523

Spatial Empirical Best Predictor of Small Area Poverty Indicator

Dian Handayani[1], Khairil Anwar Notodiputro[2], Asep Saefuddin[2], I Wayan Mangku[3], Anang Kurnia[2#]

[1]Department of Statistics, Universitas Negeri Jakarta, Indonesia
[2]Department of Statistics, IPB University, Indonesia
[3]Department of Mathematics, IPB University, Indonesia

[#]Corresponding author e-mail: anangk@apps.ipb.ac.id

Abstract

Information about some poverty indicators is important not only for the large administrative level but also for lower administrative level. This information can be obtained from many surveys. Unfortunately, many surveys are usually designed to satisfy accuracy for large populations. As a result, it is often encountered that the sample size from some sub-populations which can be obtained from a survey is too small to produce a reliable direct estimator. The sub-population which the selected sample from it is not large enough to produce a reliable direct estimator is also called a small area. In this paper, we propose the spatial empirical best predictor (SEBP) for some poverty indicators in some small areas. The SEBP is derived under a unit-level spatial lognormal mixed model which incorporates spatial dependence into the covariance structure. The mean square prediction error (MSPE) of the SEBP is estimated by the parametric bootstrap method. A simulation study was conducted to evaluate the performance of the SEBP compared to the direct estimates as well as the empirical best predictor (EBP). Further, the SEBP was also applied to obtain the estimates of some poverty indicators for some sub-districts in Bogor, Indonesia. The results showed that there is a substantial reduction in MSPE of the SEBP over the direct estimates and the EBP for almost all sub-districts.

Keywords: *spatial dependence, spatial lognormal mixed model, small area parameter, small area estimation, Monte Carlo approximation.*

1 Introduction

Surveys are usually designed to estimate some parameters for large areas or large populations. The sample size of some subgroups of a population which are not planned in the survey is often very small [1,2]. In this case, direct estimation will produce large standard error and hence is unreliable. A subgroup of a population from which the selected sample is not large enough to produce direct estimates with adequate precision is known as a small area [1,3]. The estimation techniques related to produce the small area parameters are called small area estimation (SAE) [1]. It is important to note that small

Received 15 March 2024; Accepted 20 June 2024

areas do not only refer to geographical context. They can be a subdivision of a large population which is defined by cross-classification of geographical areas and social-economic or demographic characteristics such as age, sex, race [1,4].

An empirical best linear unbiased prediction (EBLUP) is designed to estimate linear small area parameters, such as means or totals of variable of interest [1,5-7]. Some studies which extend the EBLUP to estimate nonlinear small area parameter, namely poverty indicators, have also been conducted [8-18]. These studies estimate some poverty indicators by assuming independence among small areas. Some other SAE studies have been conducted by considering spatial dependence among some small areas [2,19-27]. These spatial SAE studies focus on estimating linear small area parameter, namely small area means.

In this paper, we propose a spatial empirical best prediction (SEBP) which extends the empirical best prediction (EBP) in Molina and Rao [15] as well as Handayani et al. [28] by considering spatial dependence among small areas. The SEBP is proposed to estimate the FGT (stands for Foster, Greer, and Thorbecke) poverty indicators [29]. The SEBP is derived under a unit-level spatial lognormal mixed model in Handayani et al. [19], which extends the nested error unit level regression model [15] by incorporating spatial information into the random small area effect. The unit-level spatial lognormal mixed model assumes normality of the logarithm transformation of the variable of interest.

The rest of this paper is organized as follows. Section 2 introduces a unit-level spatial lognormal mixed model. The SEBP of FGT poverty indicators, that our proposed, are also provided in this section. Section 3 discusses the uncertainty measure of the SEBP which is estimated by using parametric bootstrap techniques. Simulation studies to evaluate the performance of the SEBP are presented in Section 4. Application of the SEBP to predict the FGT poverty indicators for some sub-districts in Bogor Regency and Bogor Municipality in Indonesia is provided in Section 5. The conclusion follows in Section 6.

2 The Spatial Empirical Best Prediction (SEBP)

Consider a finite population U with N elements, divided into M sub-populations called small areas. The population size of the i^{th} small area is N_i, $i = 1.2 \dots M$, and the population size of U is $N = \sum_{i=1}^{M} N_i$. Suppose a random sample of size n_i $(n_i < N_i)$ is selected from i^{th} small area. Hence, the size of non-sampled observations from the i^{th} small area is $r_i = (N_i - n_i)$. Furthermore, the total sample size selected from U is $n = \sum_{i=1}^{M} n_i$ and the total non-sampled size is $r = (N - n) = \sum_{i=1}^{M} (N_i - n_i)$. The sampled elements of U are denoted by s and the non-sampled elements are denoted by $\bar{s}$. We also denote s_i be the sample from area i and $\bar{s}_i$ be the sample complement from area i, $i = 1.2 \dots M$.

The FGT poverty indicators for the i^{th} small area is [15,29]:

$$P_{i\alpha} = \frac{1}{N_i} \sum_{j=1}^{N_i} P_{ij\alpha} \; ; \; P_{ij\alpha} = \left(\frac{z - y_{ij}}{z}\right)^{\alpha} I\left(y_{ij} < z\right) \tag{1}$$

where $i = 1,2, \dots M$; $\alpha = 0,1,2$; y_{ij} is the variable of interest which underlies the poverty indicators $P_{i\alpha}$, such as income or expenditure, for unit j within small area i, z is the poverty line such that $I\left(y_{ij} < z\right) = 1$ if $y_{ij} < z$ (a person is under poverty) and $I\left(y_{ij} < z\right) = 0$ if $y_{ij} > z$ (a person is not under poverty). For $\alpha = 0$ the FGT poverty indicators called poverty incidence which represents the proportion of people under poverty. For $\alpha = 1$ and for $\alpha = 2$, the FGT poverty indicators are called poverty gap and poverty severity respectively. The poverty gap measures the intensity of poverty whereas poverty severity measures the severity of deprivation of people living in absolute poverty. The FGT poverty indicators $P_{i\alpha}$ (1) also can be decomposed as follows:

$$P_{i\alpha} = \frac{1}{N_i}\left\{\sum_{j\in s_i} P_{ij\alpha} + \sum_{j\in \bar{s}_i} P_{ij\alpha}\right\}, i = 1,2, \dots M. \tag{2}$$

The second term of the right-hand side (2), the values of $P_{ij\alpha}$ for non-sampled units, will be predicted under the unit-level spatial lognormal mixed model [19]:

$$y^{\#} = \log(y) = X\beta + Zv + e = X\beta + Z(I_M - \rho W)^{-1}u + e \tag{3}$$

where $y^{\#} = \left(logy_{11}, \dots, logy_{1N_1}, \dots, \log y_{M1} \dots \log y_{MN_M}\right)^T$ is the $(N \times 1)$ vector of logarithm transformation of variable of interest y, which is assumed to follow normal distribution, $X = \left(x_{ij1}, x_{ij2}, \dots x_{ijp}\right)^T$, $i = 1,2, \dots M$; $j = 1,2, \dots N_i$, is the $(N \times p)$ matrix of values of p auxiliary variables which is assumed to be fixed , β the $(p \times 1)$ the vector of model coefficients, $v = (I_M - \rho W)^{-1}u$ the $(M \times 1)$ vector of correlated random area effects which is assumed to follow an autoregressive process, namely simultaneous autoregressive process (SAR), I_M is $(M \times M)$ identity matrix, W the $(M \times M)$ weights or proximity matrix which measures the proximity among small areas, ρ the spatial correlation coefficient which measure the strength of correlation among small areas, u the $(M \times 1)$ vector of random area effects, e the $(N \times 1)$ vector of sampling errors and Z the $(N \times M)$ design matrix of v . There are many choices for proximity matrix W [30]. In our research, we define the elements of proximity matrix W as follows: w_{ij} is 1 if area i shares an edge with area j and 0 otherwise. On the other hand, we specify design matrix Z as follows:

$$Z = \begin{bmatrix} 1_{N_1} & 0 & \Lambda & 0 \\ 0 & 1_{N_2} & M & 0 \\ M & M & M & M \\ 0 & 0 & 0 & 1_{N_M} \end{bmatrix}.$$

where 1_{N_i}; $i = 1,2, \dots M$ column vector of ones which its size is N_i.

The assumption for the vector of independent error terms u is $u \sim iid\ N(0, G = \sigma_u^2 I_M)$ and for the vector of independent sampling errors is given by $e \sim iid\ N(0, R = \sigma_e^2 I_N)$. Thus, $v = (I_M - \rho W)^{-1}u \sim N(0, D)$, $D = \sigma_u^2[(I_M - \rho W^T)(I_M - \rho W)]^{-1}$, $y^{\#} \sim N\left(\mu_{y^{\#}}, \Sigma_{y^{\#}}\right)$, $\mu_{y^{\#}} = X\beta, \Sigma_{y^{\#}} = ZDZ^T + R$. I_k is the $(k \times k)$ identity matrix. Note that the vector of sampling errors e and the vector of error terms u are mutually independent.

Consider $y^{\#} = \log(y)$ can be partitioned into $y^{\#} = \left(y_s^{\#T}, y_{\bar{s}}^{\#T}\right)^T$; $y_s^{\#} = \log(y_s)$, $y_{\bar{s}}^{\#} = \log(y_{\bar{s}})$ and $X = (X_s^T, X_{\bar{s}}^T)^T$. In a similar vein, the mean vector $\mu_{y^{\#}}$ is partitioned as $\mu_{y^{\#}} = \left(\mu_{y_s^{\#}}^T, \mu_{y_{\bar{s}}^{\#}}^T\right)^T$ and the covariance matrix $\Sigma_{y^{\#}}$ as $\Sigma_{y^{\#}} = \begin{bmatrix} \Sigma_{ss} & \Sigma_{s\bar{s}} \\ \Sigma_{\bar{s}s} & \Sigma_{\bar{s}\bar{s}} \end{bmatrix}$. Since $y^{\#}$ is normally distributed with mean vector $\mu_{y^{\#}}$ and covariance matrix $\Sigma_{y^{\#}}$, the conditional distribution of $y_{\bar{s}}^{\#}$ given the sample data $y_s^{\#}$ is given by Molina and Rao [15]:

$$y_{\bar{s}}^{\#}|y_s^{\#} \sim N(\mu_{\bar{s}|s}, \Sigma_{\bar{s}|s}) \tag{4}$$

$$\mu_{\bar{s}|s} = \mu_{\bar{s}} + \Sigma_{\bar{s}s}\Sigma_{ss}^{-1}(y_s^{\#} - \mu_s) \tag{5}$$

$$\Sigma_{\bar{s}|s} = \Sigma_{\bar{s}\bar{s}} - \Sigma_{\bar{s}s}\Sigma_{ss}^{-1}\Sigma_{s\bar{s}} \tag{6}$$

Note that: $\mu_{\bar{s}} = X_{\bar{s}}\beta$, $\mu_s = X_s\beta$, $\Sigma_{\bar{s}s} = Z_{\bar{s}}DZ_s^T$, $\Sigma_{ss} = Z_sDZ_s^T + \sigma_e^2 I_n$, $\Sigma_{\bar{s}\bar{s}} = Z_{\bar{s}}DZ_{\bar{s}}^T + \sigma_e^2 I_r, \Sigma_{s\bar{s}} = Z_sDZ_{\bar{s}}^T$, $D = \sigma_u^2[(I_M - \rho W^T)(I_M - \rho W)]^{-1}$ is the $(M \times M)$ covariance matrix of v.

Based on (3), we can obtain

$$y = \left(y_{11}. \dots y_{1N_1}, y_{21}. \dots y_{2N_2}, \dots . y_{M_1}, y_{M_2} \dots y_{MN_M}\right)^T = \exp(y^{\#})$$
$$= exp\left(y_{11}^{\#}. \dots y_{1N_1}^{\#}, y_{21}^{\#}. \dots y_{2N_2}^{\#}, \dots . y_{M1}^{\#}. \dots y_{MN_M}^{\#}\right) \text{ or}$$

$y_{ij} = \exp(y_{ij}^{\#})$, $i = 1.2 \dots M, j = 1.2 \dots N_i$, so that $P_{ij\alpha}$ in (1) also can be written

$$P_{ij\alpha} = \left(\frac{z - exp\left(y_{ij}^{\#}\right)}{z}\right)^{\alpha} I\left(exp\left(y_{ij}^{\#}\right) < z\right) = h_\alpha\left(y_{ij}^{\#}\right). \tag{7}$$

Based on (7), it can be seen that $P_{ij\alpha}$ is a function of $y_{ij}^{\#}$. Thus, we can write $P_{ij\alpha} = h_\alpha(y_{ij}^{\#})$.

The spatial best predictor (SBP) for FGT poverty indicators $P_{i\alpha}$, denoted by $\hat{P}_{i\alpha}^{SBP}$ is obtained by minimizing the mean square error $MSE\left(\hat{P}_{i\alpha}^{SBP}\right) = E\left(\hat{P}_{i\alpha}^{SBP} - P_{i\alpha}\right)^2$. It will be given by the conditional expectation:

$$\hat{P}_{i\alpha}^{SBP} = E_{y_{\bar{s}}}[P_{i\alpha}|y_s]. \tag{8}$$

The SBP of $P_{i\alpha}$ based on the decomposition (2) is:

$$\hat{P}_{i\alpha}^{SBP} = \frac{1}{N_i}\left\{\sum_{j\in s_i} P_{ij\alpha} + \sum_{j\in \bar{s}_i} \hat{P}_{ij\alpha}^{SBP}\right\}. \tag{9}$$

The second term of the right-hand side (9), $\hat{P}_{ij\alpha}^{SBP}$, is the SBP of $P_{ij\alpha} = h_\alpha\left(y_{ij}^{\#}\right)$. It is given by:

$$\begin{aligned}\hat{P}_{ij\alpha}^{SBP} &= E_{y_{\bar{s}}}\left[h_\alpha\left(y_{ij}^{\#}\right)|y_s^{\#}\right] \\ &= \int h_\alpha\left(y_{ij}^{\#}\right) f_{y_{ij}^{\#}}\left(y_{ij}^{\#}|y_s^{\#}\right)dy_{ij}^{\#}; j \in \bar{s}_i\end{aligned} \tag{10}$$

where $f_{y_{ij}^{\#}}\left(y_{ij}^{\#}|y_s^{\#}\right)$ is the conditional density of $y_{ij}^{\#}$ given vector $y_s^{\#}$.

The conditional expectation (10) cannot be calculated explicitly because of the complexity of $h_\alpha(y_{ij}^{\#})$. We can approximate it empirically by Monte Carlo simulation. For this purpose, we can generate a large number L of non-sampled vectors $y_{\bar{s}}^{\#}$ from (4).

Note that $y_{\bar{s}}^{\#} = \left(y_{11}^{\#}. \dots y_{1r_1}^{\#}, y_{21}^{\#}. \dots y_{2r_2}^{\#}, \dots . y_{M1}^{\#}, y_{M2}^{\#} \dots y_{Mr_M}^{\#}\right)^T$. Suppose $y_{ij}^{\#(l)}; i = 1.2 \dots M;\ j = 1.2 \dots \bar{s}_i$ be the value of non-sampled observation y_{ij} which is obtained from the l^{th} simulation. The Monte Carlo approximation to the SBP of $P_{ij\alpha}$ is given by:

$$\hat{P}_{ij\alpha}^{SBP} = E_{y_{\bar{s}}}\left[h_\alpha\left(y_{ij}^{\#}\right)|y_s^{\#}\right] = \frac{1}{L}\sum_{l=1}^{L} h_\alpha\left(y_{ij}^{\#(l)}\right). \tag{11}$$

$i = 1,2, \dots M;\ j = 1,2, \dots r_i$

where $h_\alpha\left(y_{ij}^{\#(l)}\right)$ is calculated from (7) using values of $y_{ij}^{\#(l)}$.

In practice, the mean vector μ and covariance matrix Σ usually depend on vector unknown parameter $\theta = (\sigma_u^2, \sigma_e^2, \rho)^T$. The $\hat{\theta} = (\hat{\sigma}_u^2, \hat{\sigma}_e^2, \hat{\rho})^T$ which is an estimator of θ can be obtained by using maximum likelihood (ML) or restricted ML (REML) method. Then, we will approximate (11) by generating $y_{ij}^{\#(l)}$ from (4) but $\theta = (\sigma_u^2, \sigma_e^2, \rho)^T$ is replaced by $\hat{\theta} = (\hat{\sigma}_u^2, \hat{\sigma}_e^2, \hat{\rho})^T$. The predictor of $P_{ij\alpha}$ which is obtained by using the values of estimator $\hat{\theta} = (\hat{\sigma}_u^2, \hat{\sigma}_e^2, \hat{\rho})^T$ is called spatial empirical best predictor (SEBP), denoted by $\hat{P}_{ij\alpha}^{SEBP}$. Finally, the SEBP of the FGT poverty indicators $P_{i\alpha}$ is given by:

$$\hat{P}_{i\alpha}^{SEBP} = \frac{1}{N_i}\left\{\sum_{j\in s_i} P_{ij\alpha} + \sum_{j\in \bar{s}_i} \hat{P}_{ij\alpha}^{SEBP}\right\}. \tag{12}$$

Generating M multivariate normal vector $y_{i\bar{s}}$ of size $r_i = (N_i - n_i)$ and repeating L times may be unfeasible. Alternatively, at l^{th} simulation, the non-sampled values of $y_{ij}^{\#}$ in i^{th} small area, denoted by $y_{i\bar{s}_i}^{\#(l)}$, can be obtained by generating univariate normal $u_i^{*(l)}$ and $e_{i\bar{s}}^{*(l)}$ independently, and then calculating the corresponding $y_{i\bar{s}}^{(l)}$ from the model:

$$y_{i\bar{s}}^{\#(l)} = \mu_{i\bar{s}|s} + u_i^{*(l)} + e_{i\bar{s}}^{*(l)} \tag{13}$$

with $\mu_{i\bar{s}|s}$ obtained from $\mu_{\bar{s}|s}$ in (5) for specific i^{th} area, $u_i^{*(l)} \sim N(0, \hat{a}_1)$; $\hat{a}_1 = Z_{i\bar{s}}\left[\hat{D} - \hat{D}Z_{is}^T\left(Z_{is}\hat{D}Z_{is}^T + \hat{\sigma}_e^2 I_{n_i}\right)^{-1} Z_{is}\hat{D}\right] Z_{i\bar{s}}^T$ and $e_{i\bar{s}}^{*(l)} \sim N\left(0_{r_i}, \hat{a}_2\right)$; $\hat{a}_2 = \hat{\sigma}_e^2 I_{r_i}$. where 0_{r_i} is $(r_i \times r_i)$ null vector and I_{r_i} is $(r_i \times r_i)$ identity matrix.

The covariance matrix of $y_{\bar{s}}|y_s$, denoted by $\Sigma_{\bar{s}^{\#}|s^{\#}}$ (6), under the unit-level spatial lognormal model (3), corresponds to the covariance matrix of $y_{i\bar{s}}$ which is obtained by generating u_i^* and $e_{i\bar{s}}^*$ independently under model (13).

3 The Parametric Bootstrap Estimator for the MSE of the SEBP

The mean square error (MSE) of $\hat{P}_{i\alpha}^{SEBP}$ is given by:

$$MSE\left(\hat{P}_{i\alpha}^{SEBP}\right) = E\left(\hat{P}_{i\alpha}^{SEBP} - P_{i\alpha}\right)^2. \tag{14}$$

Because the FGT poverty indicators $P_{i\alpha}$ is a nonlinear parameter, it is difficult to derive the $MSE\left(\hat{P}_{i\alpha}^{SEBP}\right)$ analytically in [15]. Thus, by using parametric bootstrap, we estimate the $MSE\left(\hat{P}_{i\alpha}^{SEBP}\right)$. The steps to obtain the estimates of $MSE\left(\hat{P}_{i\alpha}^{SEBP}\right)$ are described below:

1) Fit model (3) to the sample data $(y_s^{\#}, X_s)$ and obtain the model parameter estimates $\hat{\beta}$, $\hat{\sigma}_u^2, \hat{\sigma}_e^2, \hat{\rho}$.
2) Generate bootstrap random area effect $u^*\sim N(0, \hat{\sigma}_u^2 I_M)$ and sampling error $e^*\sim iid\ N(0, \hat{\sigma}_e^2 I_N)$ using $\hat{\sigma}_u^2$ and $\hat{\sigma}_e^2$ obtained at Step 1.
3) Construct the bootstrap population model using $\hat{\beta}, \hat{\rho}, u^*, e^*$, the known proximity matrix W and the population values of X :

$$y^{\#*} = X\hat{\beta} + Z(I - \hat{\rho}W)^{-1}u^* + e^*. \tag{15}$$

4) For the bootstrap model (15), generate B bootstrap replications $(y^{\#*(1)}, y^{\#*(2)} \dots . y^{\#*(B)})$ and calculate the bootstrap population parameter $P_{i\alpha}^{*(b)} = \frac{1}{N_i}\sum_{j=1}^{N_i} P_{ij\alpha}^{*(b)}$; where $P_{ij\alpha}^{*(b)} = h_\alpha\left(y_{ij}^{\#*(b)}\right); b = 1,2, \dots B$.
5) From each bootstrap replication b in step 4, take a sample with its size is similar to the original sample which is available at the survey data, then calculate the bootstrap SEBP, $\hat{P}_{i\alpha}^{SEBP*(b)}, b = 1,2, \dots B.$ as described in Section 2 using the bootstrap sample data $y_s^{\#*}$, W and X.
6) The bootstrap estimator $MSE\left(\hat{P}_{i\alpha}^{SEBP*}\right) = E\left\{\left(\hat{P}_{i\alpha}^{SEBP*} - P_{i\alpha}^*\right)\right\}^2$ is approximated by

$$mspe\left(\hat{P}_{i\alpha}^{SEBP*}\right) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{P}_{i\alpha}^{SEBP*(b)} - P_{i\alpha}^{*(b)}\right)^2. \tag{16}$$

Finally, the estimator (16) is used to estimate $MSE\left(\hat{P}_{i\alpha}^{SEBP}\right)$ given in (14).

4 Simulation Studies

To study the performance of the SEBP, we conducted a simulation study. We assume that all small areas in population U are selected in the sample and that the sample model represents the population model, i.e. that there is no sample selection bias of small areas.

The auxiliary variables that are used in the small area estimation do not always have an explicit causal relationship with the variable of interest (the dependent variable in the model). In the small area estimation context, auxiliary variables play the role of giving additional information that can explain changes in the variable of interest. In this research, simulation studies are designed by constructing a population which variable of interest as a function of auxiliary variables with random error.

The number of small areas is set at 45 ($M = 45$) and the range of sample size for each of the small areas is set up from 3 to 47. The small areas are arranged ascending based on its sample size ($n_1 = 3, n_2 = 4, n_3 = 5$ up to $n_{45} = 47$). Sample sizes less than or equal to 15 are considered "small", between 15 and 30 "medium" and larger than or equal to 30 as "large". Moreover, the sample sizes are set approximately 3% of the corresponding populations. The area population size thus ranges between 100 and 1567, the total population is $N = 37500$ and the total sample size $n = 1125$.

The variable of interest y_{ij} is assumed to follow a log normal distribution or $y_{ij}^{\#} = \log y_{ij}$ follow normal distribution. The values $y_{ij}^{\#}$ are generated from model (3) using an intercept and one auxiliary information, that is $x_{ij} = \left(1, x_{ij1}\right)^T$ where the values of auxiliary variable are generated from $x_{ij} \sim N(8,4)$ as well as from $x_{ij} \sim N(6,9)$. The expectation of $y_{ij}^{\#}$ with $x_{ij} \sim N(8,4)$ will be the same as the expectation of $y_{ij}^{\#}$ with $x_{ij} \sim N(6,9)$. However, the distribution of y_{ij} based on $x_{ij} \sim N(6,9)$ has a longer tail (heavily skewed) than $x_{ij} \sim N(8.4)$. The intercept and the regression coefficient of the auxiliary variable are $\beta = (2,1)^T$. For the cut off value of z_{ij}, we follow [15] and is fixed at $z_{ij} = 0.3 * \text{median}(y_{ij})$.

The matrix W is the contiguity matrix which is kept fixed for all simulations. The elements of W is specified by $w_{ij} = 1$ if small area i shares an edge with small area j and $w_{ij} = 0$ otherwise. The maximum number of neighbours for each area is restricted and equal to 5, and the W matrix was standardized by row. The elements of W which its dimension (45×45) is given by:

$$W = \begin{bmatrix} 0 & 1/5 & 1/5 & 1/5 & \dots & 0 \\ 1/3 & 0 & 1/3 & 0 & \dots & 0 \\ 1/4 & 1/4 & 0 & 1/4 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1/4 \\ 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

The random area effects v and sampling errors e are generated independently from $v \sim MVN(0,\ \sigma_u^2[(I_M - \rho W)(I_M - \rho W^T)]^{-1})$ and $e \sim N(0,\ \sigma_e^2)$ with $\sigma_u^2 = 0.09$ and $\sigma_e^2 = 0.25$. The spatial correlation coefficients ρ are $\rho = 0,\ 0.25, 0.50, 0.75$ and 0.90. Each of the five values of ρ is combined with the two sets of X variables: $X \sim N(8,4)$ and $X \sim N(6,9)$. As a result, there will be ten synthetic (simulated) populations. The parameters that are set up in our scenario simulation are like to our previous research [19].

A set of sample indices s_i with sample size n_i is proportional to its population is drawn independently in each area i using simple random sampling without replacement. The values of the auxiliary variable for the population units and the sample indices s_i were kept fixed over all Monte Carlo simulations. Then, we generate $K = 500$ Monte Carlo population vectors $y_{ij}^{\#}$ from the true model. For each Monte Carlo population k, for $k = 1,2, \dots K$, the following quantities were computed:

1) The true parameter for each area

   $P_{i\alpha}^{(k)} = \frac{1}{N_i}\sum_{j=1}^{N_i} P_{ij\alpha}^{(k)}$; $i = 1,2, \dots M$; $\alpha = 0,1,2$

   where $P_{ij\alpha}^{(k)}$ is:

$$P_{ij\alpha}^{(k)} = \left(\frac{z-y_{ij}^{(k)}}{z}\right)^{\alpha} I\left(y_{ij}^{(k)} < z\right),\ y_{ij}^{(k)} = exp\left(y_{ij}^{\#(k)}\right),\ j = 1, 2, \dots N_i.$$

2) Direct estimates of $P_{i\alpha}$ for each area i using the sampled part $y_s^{\#(k)}$ of the k^{th} vector population $y^{\#(k)}$:

$$\hat{P}_{i\alpha}^{DIR(k)} = \frac{1}{n_i}\sum_{j=1}^{n_i} P_{ij\alpha}^{(k)}\ ; i = 1, 2, \dots M$$

3) Using the sampled part $y_s^{\#(k)}$, Monte Carlo approximation of the EBP of $P_{i\alpha}$ were computed with $L = 50$ replicates.
4) Model (3) was fitted to the sample data $\left(y_s^{\#(k)}, X_s\right)$ for each population k. Then, substituting the estimated model parameters in (5) and (6), $L = 50$ out-of-sample vectors $y_{\bar{s}}^{\#(kl)}, l = 1.2 \dots L$ were generated from the conditional distribution (4) using (13). The sample data $y_s^{\#(k)}$ is attached to the generated out-of-sample $y_{\bar{s}}^{\#(kl)}$ to form a population vector $y^{\#(kl)}$. The small area parameter $P_{i\alpha}$ is calculated from each population $y^{\#(kl)}$ as follows:

$$P_{i\alpha}^{(kl)} = \frac{1}{N_i}\left\{\sum_{j\in s_i} P_{ij\alpha}^{(k)} + \sum_{j\in \bar{s}_i} P_{ij\alpha}^{(kl)}\right\}, i = 1, 2, \dots M$$

where for $j \in s_i$. $P_{ij\alpha}^{(k)}$ was already obtained in (1), while for non-sampled units $j \in \bar{s}_i$, $P_{ij\alpha}^{(kl)}$ is calculated as:

$$P_{ij\alpha}^{(kl)} = \left(\frac{z - y_{ij}^{(kl)}}{z}\right) I\left(y_{ij}^{(kl)} < z\right),\ y_{ij}^{(kl)} = exp\left(y_{ij}^{\#(kl)}\right), j \in \bar{s}_i.$$

Then, the SEBP of FGT poverty measure $P_{i\alpha}$ was calculated for each small area i as:

$$\hat{P}_{i\alpha}^{SEBP(k)} = \frac{1}{L}\sum_{l=1}^{L} P_{i\alpha}^{(kl)}$$

5) Calculate the means over the Monte Carlo populations $k = 1,2, \dots K$ of true values, design-based, EBP and SEBP outcomes as:

$$E(P_{i\alpha}) = \frac{1}{K}\sum_{k=1}^{K} P_{i\alpha}^{(k)}\ , i = 1,2, \dots M$$

6) Similarly, bias over Monte Carlo populations of DIR (direct estimates), EBP and SEBP along with the corresponding MSE are calculated using the formula that is given in the Table 1.

Table 1. The Formula of Bias and MSE for Direct, EBP and SEBP

	Bias	MSE
Direct	$Bias\left(\hat{P}_{i\alpha}^{DIR}\right) = E\left(\hat{P}_{i\alpha}^{DIR}\right) - E(P_{i\alpha})$	$MSE\left(\hat{P}_{i\alpha}^{DIR}\right) = E\left(\hat{P}_{i\alpha}^{DIR} - P_{\alpha_i}\right)^2$
EBP	$Bias\left(\hat{P}_{i\alpha}^{EBP}\right) = E\left(\hat{P}_{i\alpha}^{EBP}\right) - E(P_{i\alpha})$	$MSE\left(\hat{P}_{i\alpha}^{EBP}\right) = E\left(\hat{P}_{i\alpha}^{EBP} - P_{i\alpha}\right)^2$
SEBP	$Bias\left(\hat{P}_{i\alpha}^{SEBP}\right) = E\left(\hat{P}_{i\alpha}^{SEBP}\right) - E(P_{i\alpha})$	$MSE\left(\hat{P}_{i\alpha}^{SEBP}\right) = E\left(\hat{P}_{i\alpha}^{SEBP} - P_{i\alpha}\right)^2$

The following Table 2, Table 3 and Table 4 report the mean (and median) of relative bias and the mean (and median) of relative RMSE of FGT poverty indictors P0, P1, and P2 respectively which are estimated by Direct, EBP and SEBP method. It can be seen from the Table 2, Table 3 and Table 4 that the mean (and median) of relative bias of P0, P1, and P2 which is estimated by Direct, EBP as well as SEBP are approximate to zero for all of ten simulated populations. Subject to the relative RMSE, the mean (and median) of relative RMSE of P0, P1, and P2 which is estimated by SEBP is the smallest, relative to the Direct and EBP method for all of ten simulated populations.

Table 2. Mean (Median) of Relative Bias and Mean (Median) of Relative RMSE of P0 over area for ten simulated populations for Direct, EBP and SEBP

	Data Sets	ρ	Mean of Relative Bias (Median of Relative Bias)			Mean of Relative RMSE (Median of Relative RMSE)		
			Direct	EBP	SEBP	Direct	EBP	SEBP
Small variance	1	0	0.17 (0.29)	0.18 (0.17)	0.34 (0.34)	38.64 (33.68)	15.26 (14.21)	14.08 (13.52)
	2	0.25	0.32 (0.16)	0.22 (0.22)	0.36 (0.37)	38.13 (31.01)	16.13 (14.72)	14.47 (14.06)
	3	0.50	-0.31 (-0.45)	0.12 (0.05)	0.29 (0.28)	39.98 (33.06)	14.60 (13.41)	13.26 (12.14)
	4	0.75	0.45 (0.25)	0.20 (0.23)	0.34 (0.36)	37.60 (32.97)	16.61 (14.56)	14.62 (15.19)
	5	0.90	0.61 (0.33)	0.14 (0.15)	0.29 (0.34)	40.71 (31.96)	16.86 (15.12)	14.47 (13.69)
Large variance	1	0	-0.12 (-0.09)	0.01 (0.02)	0.12 (0.13)	32.35 28.22	13.07 (11.02)	10.02 (8.20)
	2	0.25	-0.30 (0.00)	0.02 (0.04)	0.12 (0.14)	31.83 (28.18)	13.29 (12.65)	10.07 (9.17)
	3	0.50	0.16 (0.14)	-0.06 (-0.05)	0.07 (0.05)	32.15 (28.59)	11.96 (10.90)	8.46 (6.76)
	4	0.75	-0.04 (0.01)	0.02 (-0.05)	0.10 (0.07)	32.25 (24.90)	14.98 (13.22)	11.06 (9.20)
	5	0.90	-0.12 (0.24)	0.07 (0.14)	0.11 (0.21)	33.59 (29.83)	22.55 (25.22)	18.17 (18.36)

Table 3. Mean (Median) of Relative Bias and Mean (Median) of Relative RMSE of P1 over area for ten simulated populations for Direct, EBP and SEBP

	Data Sets	ρ	Mean of Relative Bias (Median of Relative Bias)			Mean of Relative RMSE (Median of Relative RMSE)		
			Direct	EBP	SEBP	Direct	EBP	SEBP
Small variance	1	0	0.52 (0.51)	-0.41 (-0.45)	-0.17 (-0.20)	44.75 (37.60)	15.80 (15.50)	9.72 (8.51)
	2	0.25	-0.06 (-0.07)	-0.39 (-0.43)	-0.16 (-0.18)	43.39 (36.57)	15.56 (16.20)	9.49 (8.97)
	3	0.50	-0.23 (-0.28)	-0.46 (-0.52)	-0.22 (-0.25)	46.17 (38.25)	17.30 (18.51)	10.61 (10.28)
	4	0.75	0.43 (0.71)	-0.39 (-0.41)	-0.15 (-0.15)	42.95 (38.28)	15.89 (15.23)	9.91 (9.03)
	5	0.90	0.79 (0.26)	-0.43 (-0.48)	-0.20 (-0.19)	46.58 (36.91)	16.62 (17.03)	10.56 (9.04)
Large variance	1	0	-0.17 (-0.34)	-0.56 (-0.60)	-0.40 (-0.42)	35.98 (30.80)	19.68 (20.51)	14.53 (14.52)
	2	0.25	-0.39 (-0.21)	-0.56 (-0.59)	-0.40 (-0.39)	35.30 (31.26)	19.20 (19.72)	13.90 (13.54)
	3	0.50	0.10 (-0.07)	-0.61 (-0.63)	-0.44 (-0.48)	35.65 (31.83)	20.63 (21.26)	15.20 (16.40)
	4	0.75	-0.09 (-0.05)	-0.54 (-0.62)	-0.39 (-0.46)	35.78 (27.17)	19.25 (21.14)	14.58 (15.63)
	5	0.90	-0.26 (0.17)	-0.40 (-0.50)	-0.31 (-0.30)	37.45 (32.93)	19.24 (18.50)	15.88 (15.52)

Table 4. Mean (Median) of Relative Bias and Mean (Median) of Relative RMSE of P2 over area for ten simulated populations for Direct, EBP and SEBP

	Data Sets	ρ	Mean of Relative Bias (Median of Relative Bias)			Mean of Relative RMSE (Median of Relative RMSE)		
			Direct	EBP	SEBP	Direct	EBP	SEBP
Small variance	1	0	0.72 (0.39)	-0.66 (-0.69)	-0.42 (-0.45)	51.52 (43.13)	22.25 (23.17)	14.87 (15.53)
	2	0.25	-0.27 (0.03)	-0.65 (-0.68)	-0.42 (-0.45)	49.59 (41.85)	21.98 (23.12)	14.80 (15.47)
	3	0.50	-0.12 (-0.11)	-0.69 (-0.73)	-0.47 (-0.52)	52.76 (43.60)	23.23 (24.80)	16.36 (17.54)
	4	0.75	0.44 (0.99)	-0.63 (-0.67)	-0.40 (-0.42)	49.23 (43.90)	21.71 (22.76)	14.59 (14.51)
	5	0.90	0.91 (0.26)	-0.67 (-0.71)	-0.45 (-0.46)	53.61 (41.54)	22.63 (24.02)	15.66 (15.61)
Large variance	1	0	-0.15 (-0.34)	-0.76 (-0.79)	-0.62 (-0.64)	39.57 (33.47)	25.60 (26.61)	20.99 (21.66)
	2	0.25	-0.37 (-0.21)	-0.76 (-0.79)	-0.62 (-0.61)	38.99 (34.62)	25.53 (26.31)	20.77 (20.62)
	3	0.50	0.07 (-0.18)	-0.79 (-0.82)	-0.65 (-0.68)	39.36 (35.11)	26.426 (27.26)	21.76 (22.89)
	4	0.75	-0.09 (-0.07)	-0.74 (-0.80)	-0.61 (-0.67)	39.59 (29.94)	24.78 (26.74)	20.47 (22.40)
	5	0.90	-0.34 (0.11)	-0.61 (-0.73)	-0.51 (-0.55)	41.45 (36.63)	22.06 (24.62)	18.87 (18.63)

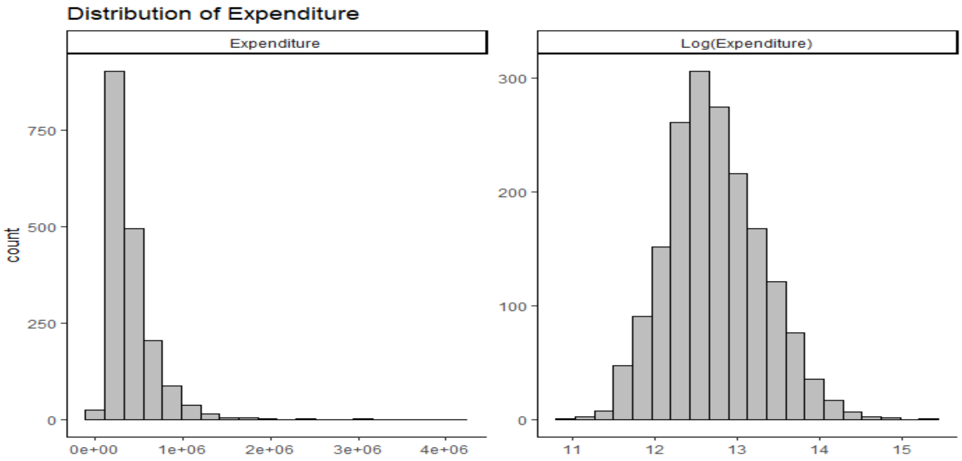
The graphs of relative bias and relative RMSE of P0, P1, and P2 which are estimated by Direct, EBP and SEBP using simulated data with small variance are provided in Appendix 1. On the other hand, Appendix 2 presents the estimates of P0, P1, and P2 based on simulated data with large variance. The relative bias of P0, P1, and P2 which are estimated by Direct, EBP and SEBP method are quite well (close to zero). However, the fluctuation of relative bias of P0, P1, and P2 which is estimated by EBP and SEBP tend to be more stable than Direct estimation method. The relative RMSE for Direct estimates of P0, P1, and P2 is the largest, compared to EBP and SEBP for all the conditions (small, medium, and large) spatial dependence. The RMSE of Direct estimates of P0, P1, and P2 will decrease as sample size increases where the RMSE of EBP and SEBP tend to be stable (it is not affected by sample size of area). The RMSE of Direct, EBP and SEBP coincided (similar) for the area with large sample, namely area 29th until area 45th.

5 Application

In this section we apply the methodology outlined above to estimate the FGT poverty indicators for some sub-districts in Bogor, Indonesia. Particularly. we apply the SEBP and evaluate it to the EBP and direct estimates (DIR) to estimate the FGT poverty indicators.

Bogor is one of the regions in the Province of West Java, Indonesia. Bogor is divided into two administrative areas: Bogor Regency and Bogor Municipality. Each of the two areas consists of several sub-districts. In the data that we analyze, there are 39 sub-districts in Bogor Regency and six sub-districts in Bogor Municipality. A sub-district consists of some villages. The status of a village could be urban or rural. Information of the status of a village is available from the Potensi Desa/PODES (2008 Village Potential). In addition to PODES data, we also utilize data from Survey Sosial Ekonomi Nasional/SUSENAS (2007 National Socioeconomic Survey) to obtain information about the monthly household per capita expenditure. The SUSENAS data has been designed to allow accurate estimation of poverty indicators at regencies or municipalities level but not at the lower level.

The variable of interest y_{ij} is per capita monthly expenditure of a household-j in sub-districts-i on basic needs (food and nonfood). Figure 1 (i) shows that y_{ij} does not follow a normal distribution. However, $y_{ij}^{\#} = \log(y_{ij})$ approximately follow normal distribution, as shown in Figure 1 (ii).



(i) (ii)

Figure 1. (i) Histogram of y_{ij} and (ii) histogram of $y_{ij}^{\#} = \log(y_{ij})$

According to [31], households in urban villages tend to have higher expenditures on basic needs than households in rural villages. Hence, in our research, the status of a village (rural or urban) where the household is located, is used as auxiliary information to predict the poverty measures.

The average number of households for 45 sub-districts in Bogor is 28,480 and the average number of households that was selected as a sample is approximately 40, which is only 0.14 % of its population size. The sample size of households for the 45 sub-districts in Bogor ranges from 16 (2.22) to 144 (33%).

As a measure of spatial dependence among the districts, we use the first-order contiguity, given by the (45×45) matrix W with $w_{ij} = 1$, if sub-district i shares an edge with sub-district j and 0 otherwise. Furthermore, the weights are row-normalized such that the sum of row elements is 1. The poverty line for the Province of West Java, where Bogor Regency and Bogor Municipalities are located, for the period January-December 2007 is IDR 180,821 for households in urban villages and IDR 144,204 for households in rural villages.

To obtain the EBP and SEBP of the FGT poverty measures, we generated 500 Monte Carlo $r_i = (N_i - n_i)$ non-sampled y_{ij} values. For the spatial dependence case, the y_{ij} values are generated according to model (3), otherwise according to nested error regression model in [15]. Based on the generated values and the sampled values that we have from survey data, we calculate the direct (DIR), EBP and SEBP of P0, P1, and P2 for each sub-district. The results of estimates are presented in Table 4. Furthermore, the boxplots in Figure 2 also depict the results.

Table 4. DIR. EBP and SEBP Estimates of P0, P1, and P2 for the districts in Bogor Regency and Bogor Municipality

No	Sub-District	N	n	P0 Estimates			P1 Estimates			P2 Estimates		
				DIR	EBP	SEBP	DIR	EBP	SEBP	DIR	EBP	SEBP
1	Nanggung	20,382	32	3.13	7.31	13.42	0.23	1.32	2.70	0.02	0.38	0.84
2	Leuwiliang	26,047	32	40.63	31.07	15.33	8.00	7.67	3.17	2.77	2.76	1.00
3	Leuwisadeng	16,225	16	12.50	19.96	14.97	1.15	4.42	3.12	0.17	1.47	0.99
4	Pamijahan	30,860	32	53.13	36.41	14.34	10.22	9.46	2.92	3.02	3.53	0.92
5	Cibungbulang	30,311	32	31.25	26.50	14.98	4.77	6.24	3.08	1.01	2.17	0.97
6	Ciampea	32,373	32	18.75	10.97	14.07	1.96	2.12	2.86	0.33	0.64	0.89
7	Tenjolaya	13,514	16	25.00	18.44	14.40	3.07	4.01	2.97	0.56	1.32	0.94
8	Dramaga	23,138	32	12.50	9.58	14.60	2.25	1.81	2.99	0.53	0.54	0.94
9	Ciomas	35,115	32	6.25	7.83	13.90	1.94	1.44	2.82	1.13	0.42	0.88
10	Tamansari	19,715	32	3.13	9.99	14.78	0.30	1.91	3.03	0.03	0.57	0.95
11	Cijeruk	23,717	16	0.00	16.62	14.59	0.00	3.53	3.02	0.00	1.14	0.96
12	Cigombong	20,703	16	6.25	16.87	14.13	0.95	3.59	2.90	0.15	1.16	0.91
13	Caringin	25,027	48	12.50	7.24	14.54	1.44	1.30	2.97	0.25	0.37	0.93
14	Ciawi	22,309	16	0.00	3.70	15.48	0.00	0.62	3.24	0.00	0.16	1.04
15	Cisarua	25,161	32	0.00	1.90	13.60	0.00	0.29	2.75	0.00	0.07	0.86
16	Megamendung	22,929	16	0.00	7.49	15.43	0.00	1.37	3.23	0.00	0.40	1.03
17	Sukaraja	38,812	48	4.17	4.65	14.52	0.36	0.79	2.96	0.03	0.21	0.93
18	Babakan Madang	21,848	32	43.75	27.57	14.14	7.39	6.56	2.88	2.43	2.30	0.90
19	Sukamakmur	17,945	16	0.00	10.56	17.39	0.00	2.06	3.74	0.00	0.62	1.22
20	Cariu	13,977	16	25.00	24.59	14.47	2.21	5.73	2.99	0.37	1.98	0.94
21	Tanjungsari	13,020	16	18.75	20.97	13.73	4.24	4.70	2.81	1.04	1.58	0.88
22	Jonggol	34,138	32	0.00	10.22	14.97	0.00	1.96	3.08	0.00	0.58	0.97
23	Cileungsi	51,349	48	8.33	6.75	14.06	1.98	1.20	2.85	0.49	0.34	0.89
24	Kelapa Nunggal	22,110	48	39.58	37.49	14.61	10.33	9.81	2.98	3.95	3.68	0.94
25	Gunung Putri	64,980	64	0.00	2.85	14.40	0.00	0.45	2.92	0.00	0.12	0.91
26	Citeureup	46,101	32	6.25	8.93	15.37	2.65	1.67	3.19	1.17	0.49	1.01
27	Cibinong	64,654	80	0.00	0.60	8.18	0.00	0.08	1.50	0.00	0.02	0.43
28	Bojong Gede	50,423	48	0.00	0.19	14.27	0.00	0.02	2.90	0.00	0.01	0.90
29	Tajur Halang	23,121	32	3.13	6.68	13.39	0.95	1.19	2.70	0.29	0.34	0.84
30	Kemang	18,991	16	0.00	1.63	15.99	0.00	0.25	3.38	0.00	0.06	1.09
31	Ranca Bungur	12,047	16	0.00	4.40	14.21	0.00	0.75	2.93	0.00	0.20	0.92
32	Parung	26,472	16	0.00	14.91	14.72	0.00	3.10	3.05	0.00	0.98	0.97
33	Ciseeng	22,744	32	0.00	8.15	14.67	0.00	1.50	3.01	0.00	0.44	0.95
34	Gunung Sindur	21,184	16	0.00	12.07	13.89	0.00	2.40	2.85	0.00	0.74	0.90
35	Rumpin	27,973	48	2.08	9.13	16.07	0.02	1.71	3.34	0.00	0.50	1.06
36	Cigudeg	24,592	32	6.25	16.99	15.96	0.73	3.59	3.33	0.10	1.15	1.06
37	Jasinga	21,975	16	18.75	28.04	14.40	2.91	6.77	2.97	0.62	2.40	0.94
38	Tenjo	15,933	16	50.00	35.01	14.45	7.51	9.06	2.98	1.24	3.37	0.94
39	Parung Panjang	24,684	32	46.88	38.69	13.99	9.05	10.27	2.84	2.40	3.89	0.89
40	Bogor Selatan	41,058	144	8.33	3.09	8.73	1.35	0.49	1.61	0.33	0.13	0.47
41	Bogor Timur	20,506	48	4.17	0.90	8.65	0.62	0.13	1.60	0.16	0.03	0.47
42	Bogor Utara	34,009	112	2.68	0.74	7.76	0.45	0.10	1.41	0.11	0.02	0.40
43	Bogor Tengah	31,498	64	4.69	0.54	8.35	0.25	0.07	1.54	0.02	0.02	0.44
44	Bogor Barat	47,122	128	0.00	0.88	8.30	0.00	0.12	1.52	0.00	0.03	0.44
45	Tanah Sereal	40,802	112	4.46	1.08	7.78	0.48	0.15	1.41	0.07	0.04	0.40

Note: Observations 1-39 belong to Bogor Regency and observations 40-45 belong to Bogor Municipality.

Figure 2 shows that the length of the boxplots for Bogor Regency are longer than for Bogor Municipality indicating that the variability of the poverty measures in the former is larger than in the latter. Furthermore, Figure 2 shows that the median for all poverty indicators P0, P1, and P2 obtained by all three estimators (DIR, EBP and SEBP) for Bogor Regency are larger than for Bogor Municipality. Table 4. However, shows that sub-district Cibinong (observation no 27), one of the sub-districts in Bogor Regency has the estimates of all three poverty measures are close to the corresponding poverty indicators of Bogor Municipality.

Moreover, all the estimates of P0, P1, and P2 in Cibinong are very small compared to those for the other sub-districts in Bogor Regency. This is because of most of status of villages in sub-district Cibinong are urban. Moreover, Cibinong is also the capital of Bogor Regency.

From Table 4 it follows that there are 15 sub-districts that have zero P0, P1, and P2 DIR estimates indicating that there is no household under poverty in these sub-districts. However, given their socioeconomic conditions, the 0 % estimates are highly unlikely. The EBP and SEBP estimates for these sub-districts are far more realistic than the DIR estimates because they are larger than zero. Note that the P0, P1, and P2 DIR estimates are zero in the 15 sub-districts because no household sampled in these sub-districts has income below the poverty line.

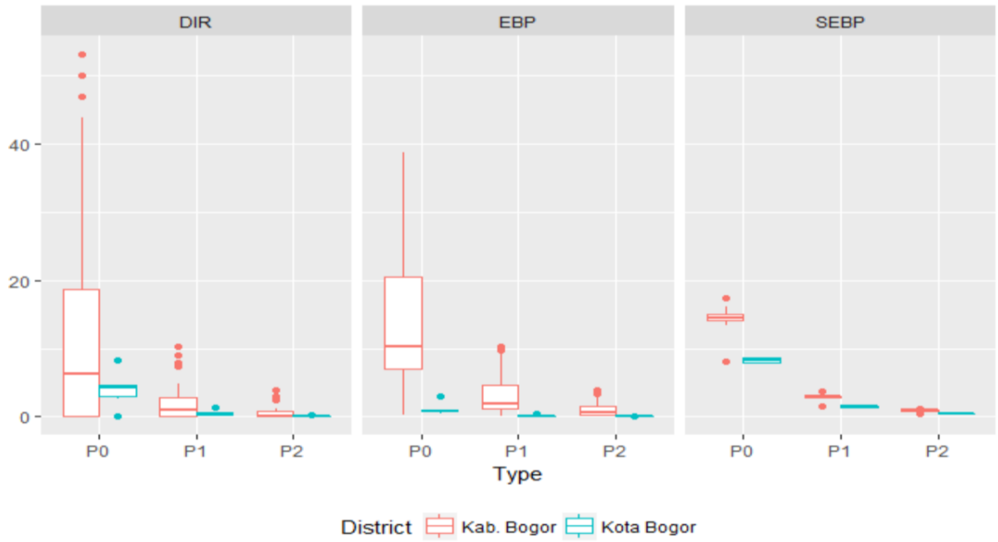
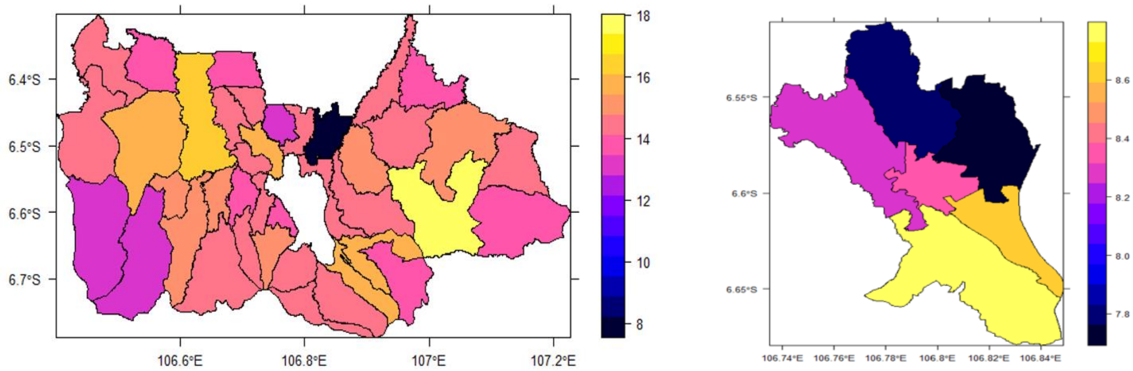


Figure 2. Boxplot of Direct Estimator (DIR), EBP and SEBP of P0 (i). P1(ii) and P2 (iii) in Bogor Regency (Kab. Bogor) and Bogor Municipality (Kota Bogor)

Figure 3(i) and Figure 3(ii) shows a geographical pattern of P0_SEBP for some sub-districts in Bogor Regency and Bogor Municipality respectively. The sub-district in Bogor Regency which has the smallest P0_SEBP is Cibinong (around 8%). It is located in the north ("black area") in Figure 3(i). Sukamakmur, the sub-district in Bogor Regency, which has the largest P0_SEBP (around 17%), is located in the south-east ("yellow area"). On the other hand, the sub-district in Bogor Municipality which has the largest P0_SEBP is South Bogor, and the smallest is North Bogor. The P0_SEBP among sub-districts in Bogor municipality are similar (around 8%).

The direct estimates of P0 for some sub-districts in Bogor Regency, particularly Leuwiliang (40.63%), Leuwisadeng (12.5%), Pamijahan (53.13%), Cibungbulang (31.25%), Ciampea (18.75%), Tenjolaya (25%) and Dramaga (12.5%) are different widely from the others, although they are quite similar in terms of socio-economic characteristics. These sub-districts form a cluster of first-order contiguous sub-districts, and they are "far away" from the central business district. The range of EBP estimates of P0 for the sub-districts in Bogor Regency mentioned above is also large but its range is narrower than for the DIR estimates. Particularly, it ranges between 10 % and 36 %. Finally, the SEBP estimates of P0 for these sub-districts are relatively similar, around 14 %. This outcome is in line with Statistics Indonesia (2008) which shows that in 2007 the average P0 estimate for Bogor Regency was 13.10% and for Bogor municipality 9.47%. Similar results are held for P1 and P2 estimates, which is shown in Table 4. To conclude, the SEBP is an improvement on EBP which in its turn is an improvement on direct estimates (DIR).



(i) (ii)

Figure 3. Spatial pattern of the P0_SEBP for Sub-Districts in Bogor Regency and Bogor Municipality

6 Concluding Remarks

In this study, we propose spatial empirical best predictor (SEBP) to estimate small area nonlinear parameter, namely poverty indicators. We use a parametric bootstrap method to estimate the MSE of the SEBP. Our simulation results indicate that the SEBP, in comparison with direct estimates and empirical best predictor (EBP), shows outstanding performance in term of bias and mean square error for all of the simulation scenarios.

For all the spatial correlation coefficient that we investigate whether in small variance condition as well as large variance condition, the SEBP has consistency of bias for small sample size as well as for large enough sample size. The RRMSE of SEBP is always smaller than direct estimates as well as EBP for all the simulation scenarios.

We apply the SEBP to estimate poverty indicators (poverty incidence, poverty gap and poverty severity) for 39 sub-districts in Bogor Regency and six sub-districts in Bogor Municipality, Indonesia. The results show that the SEBP performs well while it is compared to the EBP and the direct estimates.

Although we have developed the SEBP to estimate small area nonlinear parameters, specifically poverty measures, its application is not restricted to this area. The SEBP also could be applied to produce small area linear parameters such as small area means of monthly household per capita expenditure.

In this study, we derive the SEBP of poverty indicators by assuming that the variable of interest which underlies the poverty indicators has positively skew distribution and it will follow normal distribution after applying logarithm transformation. In other words, the variable of interest follows log normal distribution. In many cases, however, it is possible to encounter that the variable of interest has positively skewed distribution, but it does not follow log normal distribution. Regarding to this condition, for the future research, it is interesting for specifying other skewed distribution, for example generalized beta distribution of the second kind (GB2), which is more flexible than log normal distribution [32]. Otherwise, whenever variable of interest has skewed distribution, it is also interesting to find other transformation besides logarithm transformation, such that the result of transformation will have symmetric (normal) distribution.

We have estimated the MSE of SEBP of nonlinear small area parameter, specifically poverty indicators, by bootstrap parametric. Because of the intensive computing problem, we still have not evaluated the performance of the MSE estimates of the SEBP for nonlinear small area parameter. For the future research, it is necessary to make an efficient program simulation so that the performance of MSE estimates can be assessed.

References

[1] Rao, J.N.K. & Molina, I. (2015). Small Area Estimation (2^{nd} ed.). New Jersey: John Wiley and Sons.

[2] Coelho, P. & Pereira, L. (2011). A Spatial Unit Level Model for Small Area Estimation. *Revstat – Statistical Journal*, 9(2), 155–180.

[3] Congdon, P. (2014). Estimating life expectancies for US small areas: a regression framework. Journal of Geographical Systems, 16(1), 1-18.

[4] Moretti, A., Shlomo, N. & Sakshaug, J.W. (2019). Small Area Estimation of Well-Being. Sociological Methods & Research, 1-34.

[5] Bellow, M.E. & Lahiri, P.S. (2011). An Empirical Best Linear Unbiased Prediction Approach to Small Area Estimation of Crop Parameters. Section on Survey Research Methods, 3976-3986.

[6] Fay, R.E. & Herriot, R.A. (1979). Estimation of Income from Small Places: An Application of James–Stein Procedures to Census Data. *Journal of the American Statistical Association*, 74, 269–277.

[7] Battese, G.E., Harter, R.M. & Fuller. (1988). An error components model for prediction of county crops areas using survey and satellite data. *Journal of the American Statistical Association*, 80, 28-36.

[8] Das, S. & Haslett, S. (2019). A Comparison of Methods for Poverty Estimation in Developing Countries. International Statistical Review, 1-25. doi:10.1111/insr.12314.

[9] Guadarrama, M., Molina, I. & Rao, J.NK. (2018). Small Area Estimation of General Parameters Under Complex Sampling Designs. *Computational Statistics and Data Analysis*, 121, 20-40.

[10] Diallo, M.S. &d Rao, J.N.K. (2018). Small area estimation of complex parameters under unit-level models with skew-normal errors. *Scandinavian Journal of Statistics*, 45(4), 1092-1116.

[11] Molina, I., Nandram, B. & Rao, J.NK. (2014). Small Area Estimation of General Parameters with Application To Poverty Indicators: A Hierarchical Bayes Approach. *The Annals of Statistics*, 8(2), 852-885.

[12] Marhuenda, Y., Molina, I. & Morales, D. (2013). Small area estimation with spatio-temporal Fay–Herriot models. *Computational Statistics and Data Analysis*. 58, 308-325.

[13] Feretti, C. & Molina, I. (2012). Fast EB Method for Estimating Complex Poverty Indicators in Large Populations. *Journal of The Indian Society of Agricultural Statistics*, 66(1), 105-120.

[14] Esteban, M.D., Morales, D., Perez, A. & Santamaria, L. (2012). Small area estimation of poverty proportions under area-level time models. *Computational Statistics and Data Analysis*, 56, 2840–2855.

[15] Molina, I. & Rao, J.NK. (2010). Small Area Estimation of Poverty Indicators. *The Canadian Journal of Statistics*, 38(3), 369-385.

[16] Haslett, S.J., Isidro, M.C. & Jones, G. (2010). Comparison of survey regression techniques in the context of small area estimation of poverty. *Survey Methodology*, 36(2), 157-170.

[17] Srivastava, A.K. (2009). Some Aspects of Estimating Poverty at Small Area Level. *J. Ind. Soc. Agril. Statist*, 63(1), 1-23.

[18] Fabrizi, E. & Ferrante, M.R. (2008). Measuring Sub-National Income Poverty by Using a Small Area Multivariate Approach. *Review of Income and Wealth*, 54(4), 597-615.

[19] Handayani, D., Folmer, H., Kurnia, A. & Notodiputro, K.A. (2018). The Spatial Empirical Bayes Predictor of The Small Area Mean for a Lognormal Variable of Interest and Spatially Correlated Random Effects. *Empirical Economics*, 55, 147-167.

[20] Petrucci, A. & Pratesi, M. (2014). Spatial Models in Small Area Estimation in The Context of Official Statistics. *Statistica Applicata-Italian Journal of Applied Statistics*, 24(1), 9-27.

[21] Porter, A.T., Holan, S.H., Wikle, C.K. & Noel. C. (2014). Spatial Fay–Herriot models for small area estimation with functional covariates. *Spatial Statistics*, 10, 27–42.

[22] Pratesi, M. & Salvati, N. (2008). Small Area Estimation: the EBLUP Estimator Based on Spatially Correlated Random Effects. *Stat.Meth. & Appl*, 17, 113-141.

[23] Chandra, H., Salvati, N. & Chambers, R. (2007). Small area estimation for Spatially Correlated Populations - A Comparison of Direct and Indirect Model-Based Method". *Statistics in Transition* 8(20, 331-350.

[24] Singh, B.B., Shukla, G.K. & Kundu, D. (2005). Spatio-temporal models in small area estimation. *Survey Methodology*, 31(2), 183–195.

[25] Petrucci, A. & Salvati, N. (2004a). Small Area Estimation Using Spatial Information, The Rathbun Lake Watershed Case Study. Working Paper, No 2004/02. "G. Parenti" Department of Statistics. University of Florence.

[26] Petrucci, A. & Salvati, N. (2004b). Small Area Estimation Considering Spatially Correlated Errors: The Unit Level Random Effects Model. Working Paper, No 2004/10. "G. Parenti" Department of Statistics. University of Florence.

[27] Salvati, N. (2004). Small Area Estimation by Spatial Models: The Spatial Empirical Best Linear Unbiased Prediction (Spatial EBLUP). Working Paper, No 2004/03. "G. Parenti" Department of Statistics. University of Florence.

[28] Handayani, D., Notodiputro, K.A., Saefuddin, A., Mangku, I. W. & Kurnia, A. (2018). Empirical Best Predictor for Nested Error Regression Small Area Models.

IOP Conf. Ser.: Earth Environ. Sci. 187 012036. DOI 10.1088/1755-1315/187/1/012036

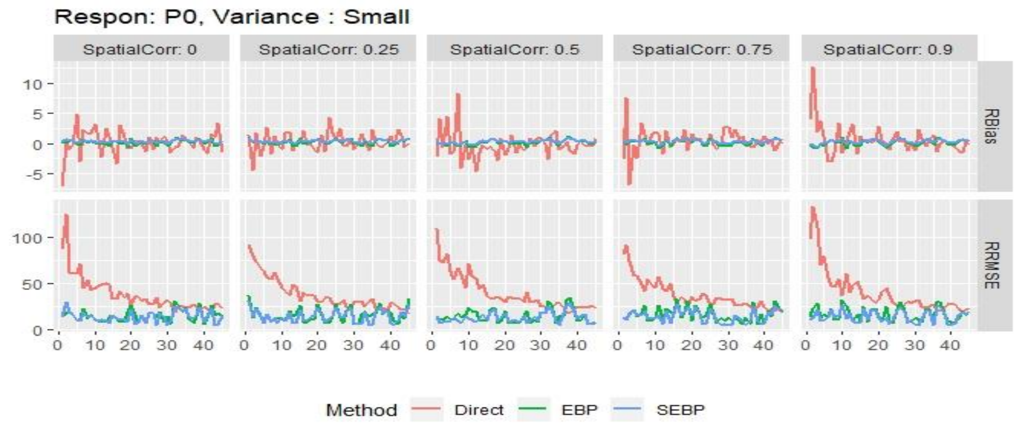
[29] Foster, J., Greer, J. & Thorbecke, E. (1984). A Class of Decomposable Poverty Measures. *Econometrica*, 52(3), 761-766.

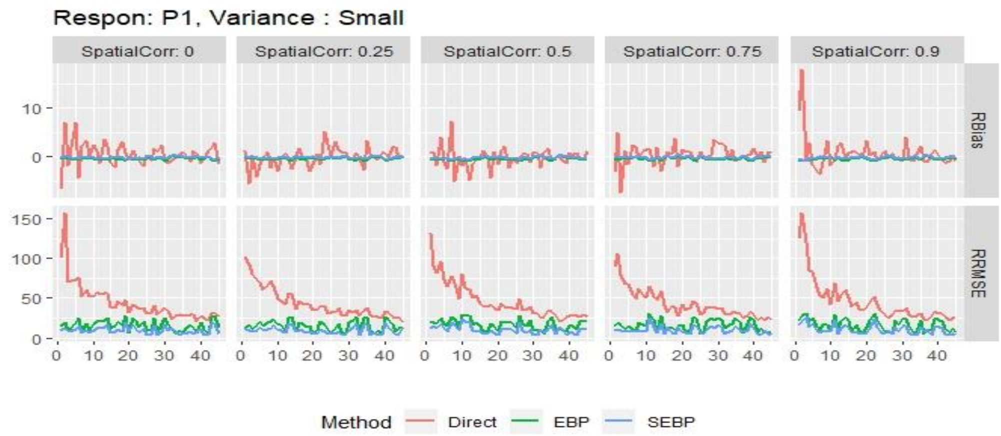
[30] Asfar, Kurnia, A. & Sadik, K. (2016) Optimum spatial weighted in small area estimation. *Global Journal of Pure and Applied Mathematics*, 12(5), 3977–3989

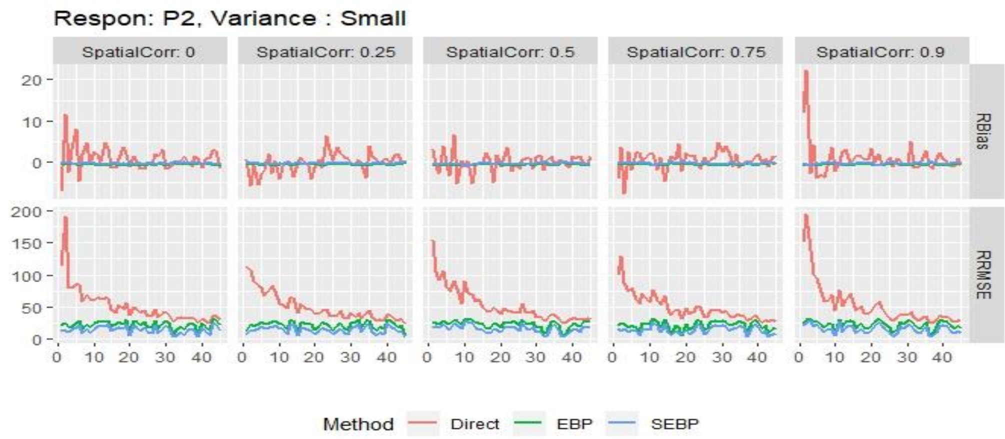
[31] Avenzora, A. & Karyono, Y. (2008). Analisis dan Penghitungan Tingkat Kemiskinan Tahun 2008. Jakarta: Badan Pusat Statistik.

[32] Graf, M., Marin, J.M & Molina, I. (2019). A generalized mixed model for skewed distributions applied to small area estimation. *TEST*, 28:565-597.

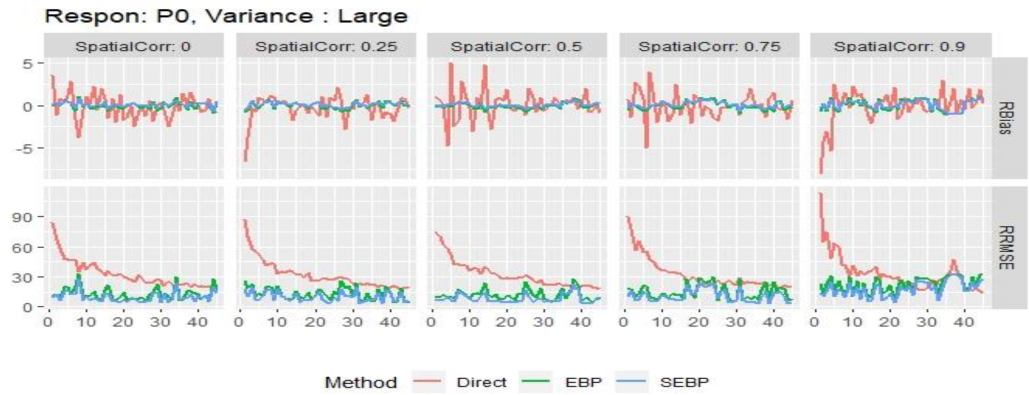
Appendix 1. The Graph of Relative Bias and Relative MSE for FGT Poverty Measures for Skewed Data with Small Variance

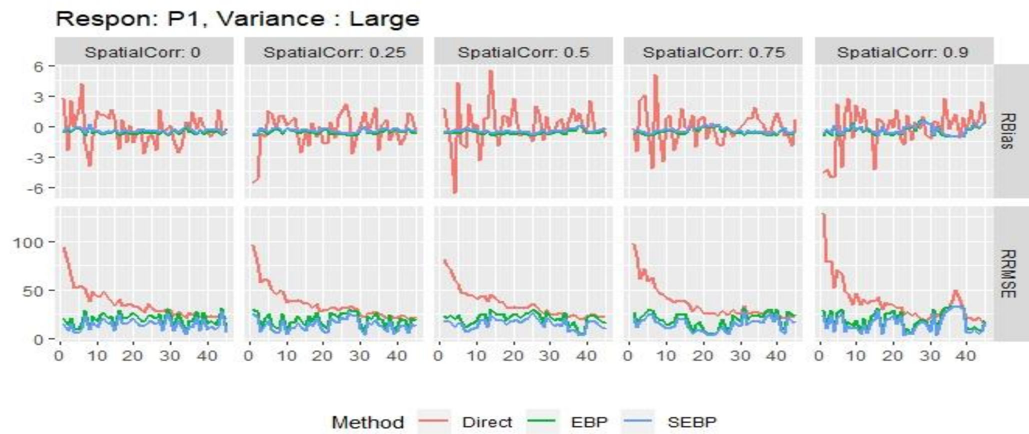


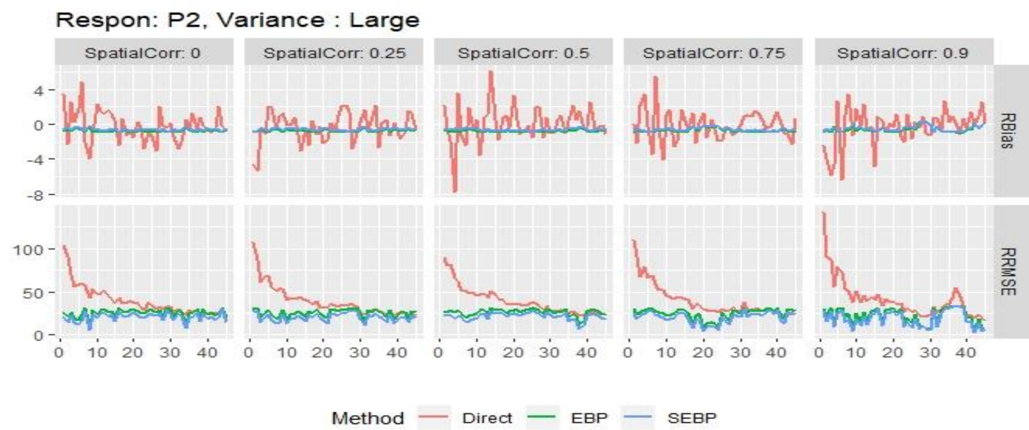




Appendix 2. The Graph of Relative Bias and Relative MSE for FGT Poverty Measures for Skewed Data with Large Variance







Notes on contributors

Dian Handayani is an Associate Professor and Head of Bachelor Program in Statistics at the Faculty of Mathematics and Natural Sciences, Universitas Negeri Jakarta, Indonesia. Her main teaching and research interests include statistical inference, generalized linear mixed model, item response theory, and small area estimation. She has published several research articles in international journals of statistics and its applications.

Khairil Anwar Notodiputro is is a Professor at the Department of Statistics, IPB University, Indonesia. His research interest include statistics and data science, statistical modeling, and small area estimation. He has published some books and research article in area of statsitics and data science.

Asep Saefuddin is a Professor at the Department of Statistics, IPB University, Indonesia. He also served as Rector at the University of Al Azhar Indonesia. His research interest include geoinformatics, spatial statistics analysis, and statistical modeling. He has published any journal and conference papers and also coordinated some research project in statsitics and data science.

I Wayan Mangku is a Professor and Head of the Mathematical Modeling Division at the Department of Mathematics, IPB University, Indonesia. His research interest include stochastic process and mathematical statistics. He has published two books and more than 90 articles in scientific journals.

Anang Kurnia is an Associate Professor and Head of the PhD Program in Statistics and Data Science at the Department of Statistics, IPB University, Indonesia. His main teaching and research interests include data science, statistical machine learning, statistical inference, generalized linear mixed model, and small area estimation. He has published several research articles in international journals in the area of statistics and data science.