© 2018 IJSRSET | Volume 4 | Issue 9 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology

Estimating Poverty Indicator with Small Area Estimation in Simulation Study of Different Population and Sample Size

Fera Kuraysia, Kusman Sadik, Anang Kurnia
Department of Statistics, Bogor Agricultural University, Bogor, Indonesia

ABSTRACT

The estimation of poverty indicators of the sub-district or village level can be calculated by small area estimation using direct estimation, empirical Bayes and fast empirical Bayes method. These three methods are evaluated through a simulation study. The usual simulation uses the same population size and sample for each area. This study compares three SAE methods with four population size scenarios with different samples for each area. Based on Bias and MSE values, direct predictions are well used in small populations. The EB method is capable of generating estimation with small bias and MSEs for all scenarios but take longer computation time. While the FEB method produces estimations with bias and MSE are small in large population conditions with faster computational time.

Keywords: direct estimation, empirical Bayes, fast empirical Bayes, poverty indicator, small area estimation

I. INTRODUCTION

Small Area Estimation (SAE) became popular in the past few decades. One application that used this method is an estimation of poverty indicator FGT (Foster, Greer and Thorbecke) and Multidimensional Poverty Indicator in an area that has a small sample size such as subdistrics or villages. Poverty is a crucial topic, especially in Indonesia that is populated by more than 250 million peoples. Indonesia through Central Bureau of Statistics (BPS) conducted a survey called The National Socioeconomic Survey (SUSENAS) of which the sample size were designed for estimation at district and provincial level[1]. However, this sample cannot be used for estimating at the smaller level, such as sub-districts or villages because the sample size is too small. SAE method than can be used to estimate this area.[5]

Molina and Rao in [3] used the Bayes empirical method to estimate FGT poverty indicator. They conducted a simulation study that used same population and sample size for each area with population size, sample size and the number of areas. Vinny *et al.* in [6] also conducted a similar study with modifications to areas with no sample. Their simulation study used population size ranging from 6 to 16withsample size from 0 to 9. But this simulation focus to see the effect of clustering to estimate area that has no sample. In 2012 Ferretti and Molina in [2] developed fast empirical Bayes (FEB) method and conducted a simulation study with the same population and sample size as the research in 2010. Their research specially focuses on large population size and multidimensional indicators poverty.

The problem is whether all of these methods can be used for all population and samples sizes and which method will give the best results in terms ofbias, MSE and computation time. This study would like to evaluate the differences between several SAE method such as direct estimation, EB and FEB to estimate FGT poverty indicators especially for head count index (HCI-P0) for different population and sample sizes. This simulation study can be used as a reference for SAE users who would like to estimate

the poverty indicator . Through this study, it is hoped that the condition under which the three SAE methods are good to use can be identified.

II. DATA AND METHODS

A. Data

This research used simulation data that were generated with four scenarios of population and sample size. This simulation used the same parameter as Molina and Rao in [3]. There were $D = 40$ areas that were estimated. The population Y_{ij} was generated from nested linear regression model using 2 covariates variables. The first was generated from a binomial distribution $x_{ij1} \sim Binom(1,0.2)$. The second was generated from the binomial distribution $x_{ij2} \sim Binom(1, p_d)$ with $p_d = 0.3 + \frac{0.5d}{D}$. Randomareas effect were generated from a normal distribution $u_i \sim iid\ N(0,0.15)$ and the sampling error was also generated from the normal distribution $e_{ij} \sim iid\ N(0,0.5)$. The intercept and slope of model were set as $\beta = (3,0.03,-0.04)$. The walfare variables used is $E_{ij} = Exp(Y_{ij})$ while the poverty line is $z = 0.6 * median\ (E_{ij})$ [2]. The nested linear regression model used is

$$y_{ij} = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + u_i + e_{ij} \quad (1)$$

Table 1. The Four Scenarios Of Population And Sample Size

Scenarios	Population Size		Sample Size	
	Max	Min	Max	Min
Small	96	10	89	8
Medium	300	102	89	8
Large 1	4982	1011	89	8
Large 2	4982	1011	50	22

B. Method of Simulation

For this simulation study, the steps that were used are:

1. Calculate the value of poverty indicator parameter P0 for each area of population data that has been generated with the formula P0 as follows:

$$P_{0i} = \frac{1}{N_i}\sum_{j=1}^{N_i} P_{0ij} j = 1, \dots, N_i, \qquad i = 1, \dots, D \quad (2)$$

$$P_{0ij} = \left(\frac{z - E_{ij}}{z}\right)^0 I\left(E_{ij} < z\right), j = 1, \dots, N_i, i = 1, \dots, d \quad (3)$$

2. Set a random sample of each area with a simple randomized design without replication.
3. Calculate the expected poverty indicator P0 using the method:

a. Direct estimation

$$\hat{P}_{0i} = \frac{1}{n_i}\sum_{j=1}^{n_i} P_{0ij} j = 1, \dots, N_i, i = 1, \dots, D \quad (4)$$

b. Empirical Bayes

$$\hat{P}_{0ij}^{EB} = E_{y_r}\left[h_0\left(Y_{ij}\right)\middle|\boldsymbol{y_s}\right] \approx \frac{1}{L}\sum_{l=1}^{L} h_0\left(Y_{ij}^{(l)}\right), j \in r_i \quad (5)$$

with $\mathrm{h}_0\left(Y_{ij}\right) = P_{0ij}$

c. Fast empirical Bayes

$$\hat{P}_{0i}^{FEB} = \frac{1}{L}\sum_{l=1}^{L} \hat{P}_{0i}^{DB}(l), \quad i = 1, \dots, D \quad (6)$$

with

$$\hat{P}_{0i}^{DB}(l) = \frac{1}{N_i}\sum_{\mathrm{j}\in \mathrm{s}_{\mathrm{i(l)}}} P_{0ij}(l) \quad (7)$$

with

$$P_{0ij}(l) = \left(\frac{z - T^{-1}\left(Y_{ij}{}^{(l)}\right)}{z}\right)^0 I\left(T^{-1}\left(Y_{ij}{}^{(l)}\right) < z\right) \quad (8)$$

The prediction model used to estimate the variables of Y_{ij}'s welfare is as follows:

$$\hat{y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 x_{ij1} + \hat{\beta}_2 x_{ij2} + \hat{u}_i \quad (9)$$

4. Repeat steps 2 to 3 as many as 100 times.
5. Calculate the value of Bias and Mean Squares Error (MSE) from the result of parameter estimation of each area
6. Evaluate methods based on Bias values and Mean

Squares Error (MSE)

III. RESULTS AND DISCUSSION

Figures 1 and 2 show the bias and MSE values of the predicted poverty indicators P0 for each area. From Figure 1, it can be seen that for areas with small population size, the direct estimation method results very small bias, whereas the bias of FEB and EB is not much different but has a greater bias than bias of direct predictions. Meanwhile, for the large population size, the bias differences for the three methods tend not to be much different. Although in some areas, the biases of EB and FEB become larger when compared with direct estimates.

The value of MSE shown in Figure 2 (a) shows that, in small population, data of the FEB method and direct predictions tend to have unstable MSE values in all areas. In other hand, the EB method has a more stable MSE in all areas. For data with large populations as in figure 2 (b), it can be seen that the direct prediction method has the largest MSE, while EB and FEB have almost same MSE values.

Evaluating of the bias and MSE values in these two data conditions concluded that the EB method can be used in small and large population conditions. However, in large populations, EB requires considerable computational time compared to FEB. Thus, the FEB method is best used with a large population size.

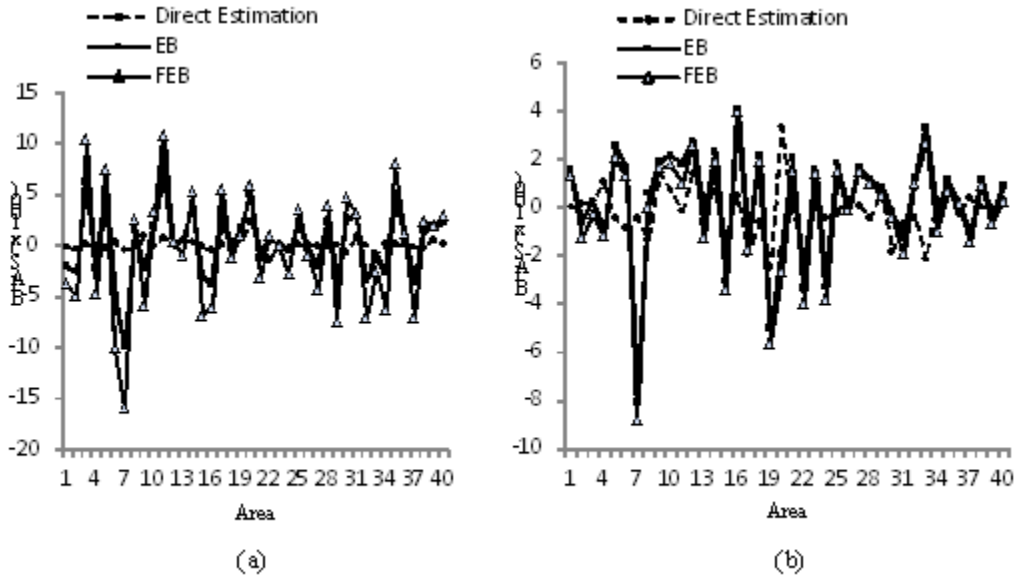

Figure 1. Graph of Bias (x 100): (a) small population; (b) large population

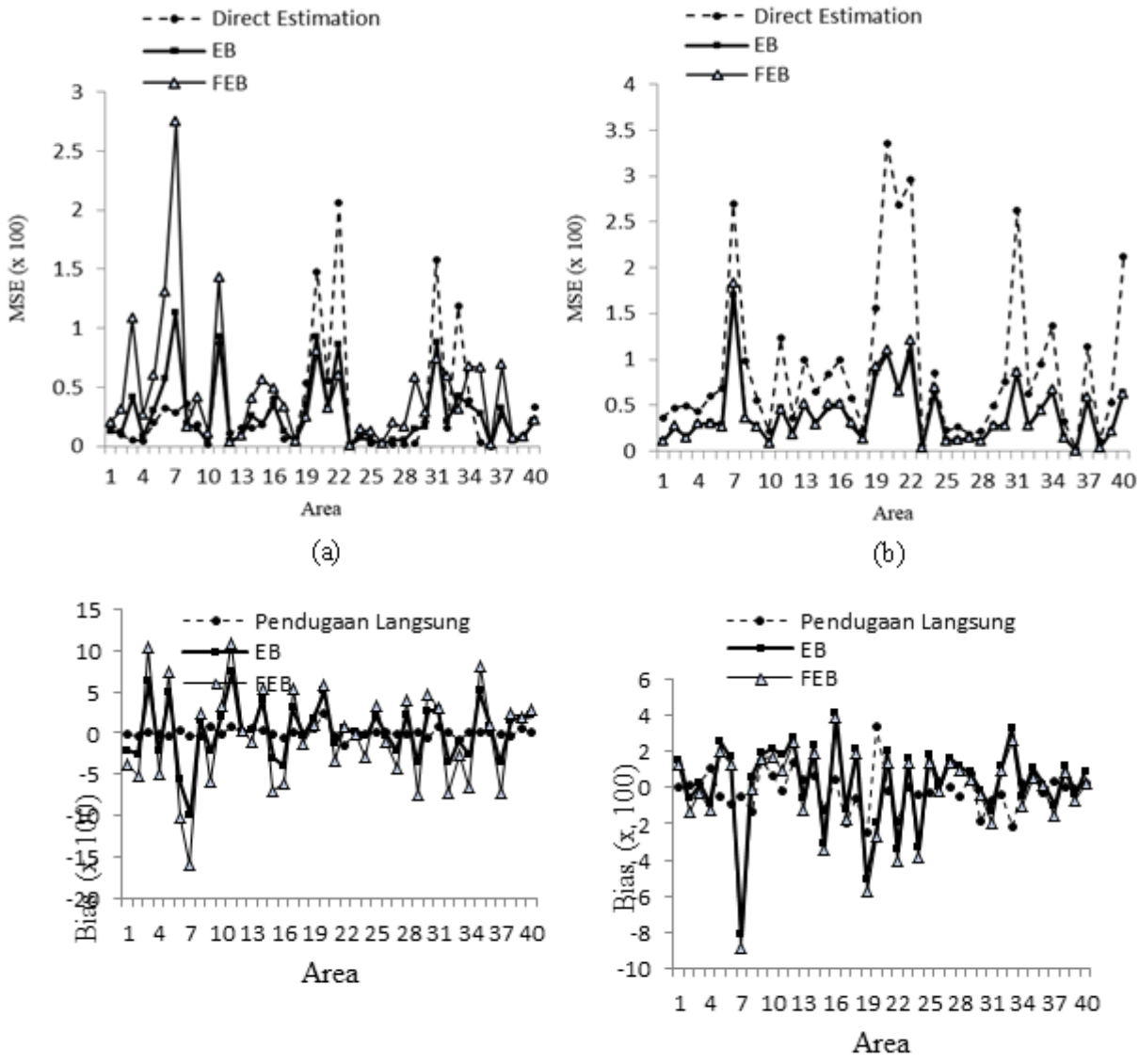

Figure 2. Graph of MSE (x 100): (a) small population; (b) large population

Table 2 shows the average of biases and MSEs in various data sizes in four scenarios using several methods. Estimation using direct estimates indicates that the resulting bias becomes larger if the population size difference and the sample are larger. Then, for small population size, direct predictive bias becomes very small. Based on average of data, it can be seen that the bias of FEB results a bias that is closer to zero when compared with EB, even under conditions of small population size, the EB bias is less than FEB. Meanwhile, for the average MSE score, the direct prediction method has the largest MSE overall of 0.643, while MSE of EB and FEB value is not much different. However, for a small population, MSE of EB is smaller than MSE of FEB that are 0.331 and 0.384 successively. So, it can be concluded that the EB method is more suitable for estimating in small population conditions, whereas FEB is more suited to be applied in large population conditions. However, FEB method can speed up computational estimation times, but it still produces good estimations despite loss a slight efficiency compared to EB.

Table 2. Average Of Biases And Mses In Various Data Sizes In Four Scenarios Using Several Methods

Scenarios	Population Size		Sample Size		Bias (x 100)			MSE (x 100)		
	Max	Min	Max	Min	Direct	EB	FEB	Direct	EB	FEB
Small	96	10	89	8	0.013	0.159	-0.494	0.324	0.276	0.455
Medium	300	102	89	8	0.121	0.57	0.084	0.827	0.398	0.413
Large 1	4982	1011	89	8	-0.14	0.227	-0.212	0.918	0.397	0.408
Large 2	4982	1011	50	22	0.047	0.34	-0.015	0.504	0.254	0.261
Average	2570	533	80	12	0.01	0.324	-0.159	0.643	0.331	0.384

IV. CONCLUSION

Simulation studies can provide an overview of the use of small area estimation methods in any data conditions. Simulation results show that the direct estimation method can be used in small population conditions, although EB can be used in various data conditions with a lower level of predicted efficiency. However, in large population conditions, the FEB is better to use, as it is capable of generating unreliable and small alterations as well as has faster computing time.

V. REFERENCES

[1]. [BPS] Badan Pusat Statistik. 2014. Perhitungan dan analisis kemiskinan makro Indonesia tahun 2014. Jakarta (ID): BPS.

[2]. Ferretti C, Molina I. 2012.Fast EB for estimating complex poverty indicators in large populations. Journal of The Indian Society of Agricultural Statistics. 66(1): 105-120.

[3]. Molina I, Rao JNK. 2010. Small area estimation of poverty indicators. The Canadian Journal of Statistics. 38(3): 369-385.

[4]. Rao JNK. 2003. Small Area Estimation. New York (US): John Wiley and Sons.

[5]. Sadik K. 2009. Metodeprediksitak-bias linier terbaikdanbayesberhirarkiuntukpendugaan area kecilberdasarkan model state space disertasi]. Bogor (ID): InstitutPertanian Bogor.

[6]. Vinny. 2017. PengembanganMetodePendugaan Area Kecil MelaluiPenggunaanInformasiGerombolPada Area Yang TidakTerdapatContoh (StudiKasusIndikatorKemiskinan di Wilayah Bogor) tesis]. Bogor (ID): InstitutPertanian Bogor.