Estimating the Poverty level in the Coastal Areas of Mukomuko District Using Small Area Estimation: Empirical Best Linear Unbiased Prediction Method

Etis Sunandi[1], Dian Agustina[2], Herlin Fransiska[3]
{esunandi@unib.ac.id, dianagustina@unib.ac.id, hfransiska@unib.ac.id}

Department of Mathematics, Faculty of Mathematics and Natural Sciences, The University of Bengkulu, Bengkulu, 38125, Indonesia[1,2,3]

Abstract. This research aims to estimate the poverty level in the Coastal Areas of Mukomuko District using small area estimation. One of the estimation methods on small area estimation is Empirical Best Linear Unbiased Prediction (EBLUP). using the method, the poverty estimator in the coastal area of Mukomuko District is obtained. The evaluation of parameter estimator is calculated by the value of MSE (Mean Square Error) using Bootstrap resampling method. Based on the result of the study is seen that the MSE value of EBLUP estimators is smaller than the MSE value of the direct estimator in each village. The MSE value of the EBLUP estimators is smaller than the MSE value from the direct estimator for each village. This indicates that the estimation with the EBLUP method can improve the estimation of parameters.

Keywords: Coastal Area, EBLUP, Poverty Data, Small Area Estimation

1   Introduction

Poverty is a problem faced by the Indonesian people. According to Bureau of Statistic (BPS) from the Bengkulu Central Statistics Bureau, in March 2016, the number of poor people in Bengkulu Province reached 328.610 people (17.32%). The percentage of poor people in the urban areas has decreased. However, in the Village, there was an increasing. The percentage of poor people in the urban areas in March 2016 was 16.19% and 17.85% in the rural areas. Some rural areas in Bengkulu province are in the coastal areas. Sadly, the level of poverty in the coastal communities in Indonesia is still very worrying by 32.4% [1].

Fishermen's poverty problem is a multidimensional problem so; government needs a complete solution to solve it. There are several aspects that caused poverty of fishermen or coastal communities, including: Government policies that do not take a side to the poor, many policies related to the poverty reduction are top down and always make the community as an object, not a subject. Another condition which derioted the level of fishermen prosperous is the lack of poverty data. As a result, the statistics data are neglected and inaccurate.

Statistical methods that are often used as an effort to get accurate information about poverty are Small Area Estimation (SAE). The SAE method is a statistical method for estimating parameters in a subpopulation where the number of samples is small or even absent. This method utilizes data from large domains to predict parameters that concern a smaller domain. One parameter estimation method used in SAE is Empirical Best Linear Unbiased Prediction (EBLUP).

SAE research using the EBLUP method has been carried out; including [2] Ningtyas R, Rahmawati R, Wilandari Y (2015) uses the EBLUP method for estimating the per capita expenditure of Brebes District. Another study was carried out by Malik NFJ, Hoyyi A, Ispriyanti D (2014) [3] who used EBLUP for estimating dropout rates in Semarang District. In the end, this research aims to estimate the poverty level in the Coastal Areas of Mukomuko District using small area estimation through Empirical Best Linear Unbiased Prediction (EBLUP) method.

2 Materials

This study uses secondary data from the Agency for the Implementation of Food Counseling and Food Security, and the Office of Marine and Fisheries of the Mukomuko District and Village Potential (recorded by BPS) in 2014. The research variables used are the percentage of poor families (Y), Number of malnutrition Sufferers in the Last 3 Years (X1), Number of maternal mortality (X2), Number of Families living in slums (X3), Number of Health Facilities (X4), Number of Recipients of Public Health Insurance (Jamkesmas) (X5), Number of Not Affordable Certificates (SKTM) Receipts (X6), Number of BPJS Receipts (X7), Number of Family without Electricity (X8), and Number of people with disabilities (X9). These data have also been used in previous study. The study focused on Poverty Mapping Coastal Region Using the Best Linear Empirical Spatial Prediction Method [4].

3 Methods

A. Poverty
The poor are people who have an average per capita expenditure per month below the poverty line (GK). Technically, GK is built from two components, namely the Food Poverty Line (GKM) and the Non-Food Poverty Line (GKNM). GKM is the expenditure value of minimum food requirements which is equal to 2,100 kilo calories per capita per day; meanwhile, GKNM is a minimum requirement for housing, clothing, education and health [5].

In addition, BPS defines household outcome a month as all household expenses for a month to meet consumption needs for all household members. Outcome Per capita data is obtained from the total monthly household outcome divided by the number of household members. Based on the assumption, the sampling is based on drawing a simple random sample, the average of outcome per capita of a village is obtained by the formula:

$$\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$$

where:
$\bar{y}_i$ = the average of outcome village per capita, $i = 1, 2, \ldots, m$
y_{ij} = outcome of the household-j per capita in village to-i and to- $j = 1, 2, \ldots, n_i$

n_i = number of households in the village to-i
m = number of villages

B. Small Area Estimation (SAE)

Small area estimation or SAE is a statistical technique to estimate the parameters of subpopulations with small sample sizes [6]. In Indonesia, these subpopulations can be in the form of provinces, districts/cities, sub-districts or villages. In general, there are three approaches to get SAE parameter estimators, namely direct estimation, indirect estimation, and composite estimation.

SAE technique is an indirect estimation that combines survey data with other supporting data, for example from previous census data that contains variables with the same characteristics as survey data so that they can be used to estimate smaller areas and provide a better level of accuracy. The indirect estimation process is estimation in a domain by connecting information in that area with other areas through a right model. This means that these allegations include data from other domains.

In SAE, there are used two types of basic models, namely area-based models and unit-based models [6]. In the area-based small area estimation model, the supporting data available is only up to the area level. The area level model connects the estimator directly to a small area with supporting data from another domain for each area.

The small area parameter that to observe is θ_i, index i shows area index. The linear model that explains the relationship is:

$$\theta_i = x_i^T \beta + z_i v_i \tag{1}$$

where :

$\beta = (\beta_l, \dots, \beta_p)^T$ is the regression coefficient $p \times l$

z_i = known positive constants to-i
v_i = random area effect, assumed $v_i \sim iid\ N(0, \sigma^2),\ i = 1, 2, \dots, m$

In making conclusions about the population, it is assumed that the estimated value is direct $\hat{\theta}_i$ is known then it can be stated as follows:

$$\hat{\theta}_i = \theta_i + e_i \tag{2}$$

Random variable e_i is sampling error, assumed $e_i \sim iid\ N(0, \psi_i)$ and $i = 1, 2, \dots, m$.

A small area estimation model for the area level consists of two model component levels, namely the indirect estimation model component according to equation (1) and the direct estimation model component according to equation (2). The models in equations (1) and (2) if combined form the following equation:

$$\hat{\theta}_i = x_i^T \beta + z_i v_i + e_i \tag{3}$$

C. Empirical Best Linear Unbiased Predictor (EBLUP) Method

One method used in building the SAE model is based on the mixed linear model. The EBLUP method is a parameter estimation that minimizes the Mean Square Error (MSE) among classes of estimation of other non-linear linear parameters by adding spatial influence. EBLUP is generated with the assumption that the sampling and area random effect components are unknown. EBLUP estimators obtained by the Maximum Likelihood (ML) and Restricted Maximum Likelihood (REML) methods are unbiased estimators. variables v_i and e_i are Normal Distribution with null mean value.

The EBLUP estimator of θ_i $\left(\tilde{\theta}_i^{\,EBLUP}\right)$, based on equation (3) is:

$$\tilde{\theta}_i^{\,EBLUP} = \hat{\gamma}_i \hat{\theta}_i + (1 - \hat{\gamma}_i) x_i^T \tilde{\beta} \tag{4}$$

Where $\hat{\gamma}_i = \dfrac{\hat{\sigma}_v^2 z_i^2}{(\hat{\sigma}_v^2 z_i^2 + \hat{\sigma}_i^2)}$, estimator of $\tilde{\beta}$ is regression coefficient which is estimated by *Generalized Least Square* (GLS), namely $\tilde{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} \hat{\theta}$, x_i^T is auxiliary data of area to-i, and $\hat{\theta}_i$ is direct estimator of θ_i.

The formulation of MSE from EBLUP (Rao, 2015) is:

$$MSE\left[\tilde{\theta}_i\right] \approx g_{1i}(\sigma_v^2) + g_{2i}(\sigma_v^2) + g_{3i}(\sigma_v^2)$$

where $g_{3i}(\sigma_v^2) = \sigma_v^4 z_i^4 (\sigma_i^2 + \sigma_v^2 z_i^2)^{-3} \bar{V}(\hat{\sigma}_v^2)$. $\bar{V}(\hat{\sigma}_v^2)$ is asymptotic variance of $\hat{\sigma}_v^2$ with the formula:

$$\bar{V}(\hat{\sigma}_v^2) = [\mathfrak{I}(\hat{\sigma}_v^2)]^{-1} = 2 \left[\sum_{i=1}^{m} z_i^4 / (\sigma_i^2 + \sigma_v^2 z_i^2)^2 \right]^{-1}$$

Formula of $MSE\left[\tilde{\theta}_i\right]$ estimator:

$$mse_1\left(\tilde{\theta}_i\right) = g_{1i}(\sigma_v^2) + g_{2i}(\sigma_v^2) + 2g_{3i}(\sigma_v^2) \tag{5}$$

where $mse_1\left(\tilde{\theta}_i\right) = g_{1i}$ is estimator of $MSE\left[\tilde{\theta}_i\right]$. In an area-based model, there are two choices of items $mse\left(\tilde{\theta}_i\right)$ namely:

$$mse_1\left(\tilde{\theta}_i\right) = g_{1i}(\sigma_v^2) + g_{2i}(\sigma_v^2) + 2g_{3i}^*\left(\sigma_v^2, \hat{\theta}_i\right)$$

And

$$mse_2\left(\tilde{\theta}_i\right) = g_{1i}(\sigma_v^2) + g_{2i}(\sigma_v^2) + g_{3i}(\sigma_v^2) + g_{3i}^*\left(\sigma_v^2, \hat{\theta}_i\right)$$

Where

$$g_{3i}^*\left(\sigma_v^2, \hat{\theta}_i\right) = [z_i^4 \sigma_v^4 / (\sigma_i^2 + \sigma_v^2 z_i^2)^4]\left(\hat{\theta}_i - x_i^T \tilde{\beta}\right)^2 \bar{V}(\hat{\sigma}_v^2)$$

Several studies have been done using the EBLUP method, namely: Aminah AS, et al (2017) [7], Petrucci A, Salvati N (2004) [8], Pratesi M, Salvati N (2008) [9], Minasny (2007) [10], and Pereira and Coelho (2012) [11].

D. Estimating MSE with Bootstrap Resampling

The Bootstrap method was first introduced by Bradley Efron in 1979. The Bootstrap method is a nonparametric approach to estimating various quantities of statistics such as the mean, standard error, and bias of an estimator or to establish confidence intervals by utilizing computer technology. The Bootstrap method can also be used to estimate the distribution of statistics. With the Bootstrap method there is no need to make distribution assumptions and initial assumptions to estimate the form of distribution and statistical tests [12].

MSE estimator with bootstrap is given by:

$$mse^*\left(\hat{\theta}_i\right)=\frac{1}{J}\sum_{j=1}^{B}\left(\hat{\theta}_i^{*(j)}-\theta_i^{*(j)}\right)^2$$

where J is the number of bootstrap populations, $\hat{\theta}_i^{*(j)}$ is the estimator of the i-small small area of the j-bootstrap population, $\theta_i^{*(j)}$ is the actual value of the estimator of the small i-small area of the j-bootstrap population.

4 Results and Discussion

A. Statistics Description of Research Variables

Mukomuko District consists of 152 villages. 20 villages (13.16%) are villages in the coastal area [13], [14], [15], [16], [17], [18], [19], [20], with the number of households for each village selected as an example ranging from 9 to 10 households. According to BPS, the poor are residents who have an average per capita outcome below the poverty line. Based on Table 1, the highest percentage of poor families is 35.36%, namely Dusun Baru Pelokan Village. While Lubuk Sanai Village, Pasar Bantal Village, and Mandi Angin Village there were no families belonging to poor families (0%). Overall, the average poor family in Mukomuko district was 8.56%. While the median is 8.24% and the standard deviation is 8.71.

Table 1. Description of the variable statistical research

	Mean	Median	Standard Deviation	Minimum	Maximum
Percentage of poor families (Y)	8.56	8.24	8.71	0.00	35.36
Number of malnutrition Sufferers in the Last 3 Years (X1)	0.75	0.00	1.89	0.00	8.00
Number of maternal mortality (X2)	0.50	0.00	0.83	0.00	2.00
The number of families living in slums (X3)	1.00	0.00	4.47	0.00	20.00
Number of health facilities (X4)	3.95	3.00	3.03	1.00	13.00
The number of Public Health Insurance (Jamkesmas) recipients (X5)	218.80	204.50	115.45	7.00	420.00
Number of Not Affordable Certificates (SKTM) Reciepts (X6)	50.30	39.00	44.52	0.00	200.00
Number of BPJS participants (X7)	72.15	29.00	129.00	3.00	567.00
Number of families without electricity (X8)	13.70	2.00	27.64	0.00	120.00
Number of people with disabilities (X9)	6.15	4.00	6.20	0.00	21.00

On the variable number of malnourished people in the last 3 years (X1), Rawa Bangun Village is the village with the most cases of malnutrition, namely 8 cases of malnutrition in the last 3 years. Then, New Hamlet Pelokan occupies the second position

where there are 3 cases of malnutrition. Meanwhile, 75% of other villages are free from malnutrition.

In the variable Number of maternal mortality (X2), Dusun Baru Pelokan Village, Teramang Jaya Village, Air Buluh Village, and Air Dikit Village, each of these have two cases. Furthermore, in the Rawa Bangun Village and Mandi Angin Jaya Village have only one case each. Whereas in 70% of other villages cases maternal mortality can be reduced to zero.

In general, in twenty sample villages, there were no families living in slums. However, there is still one village that has residents who live in slums. The village is Air Rami Village (20 families).

Distribution of health facilities in Mukomuko District has been evenly distributed. The indication is that in every village there is at least one health facility. The highest number of health facilities is Bandar Ratu Village with 13 health facilities. The opposite is Koto Jaya Village. This village has only one health facility.

In addition, the number of Jamkesmas recipients is the least in Teramang Jaya village, which is only 7, and the largest in Rawa Bangun Village is 420. Of the total 20 coastal villages there are 4,376 number of Jamkesmas recipients. With an average of 218 rural Jamkesmas recipients.

Of the 20 coastal villages in Mukomuko District, as many as 1006 families received SKTM, with an average of 50.3. The most SKTM recipients is issued by the Pasar Sebelah Village, which is 200 SKTM. Meanwhile Mandi Angin Jaya Village did not issue one SKTM during the year.

Based on Table 1 above, it can be seen that coastal villages with the least number of recipients of BPJS are only 3 participants in Mandi Angin Jaya and Sinar Laut villages and the largest in Pasar Mukomuko Village is 567. Of the total 20 coastal villages there are 1443 BPJS participants, with a village average of 72 BPJS participants.

Of the 20 coastal villages in Mukomuko District, there were 274 families; an average of 24 families still did not have access to electricity. Most families without electricity are in Dusun Baru Pelokan Village, which is as many as 120 families. While some villages have accessed electricity.

In addition, of the 20 coastal villages in Mukomuko district, there are 123 people with disabilities, with an average of 6 people per village. The most disabled people are in Koto Jaya Village, which is 21 people. While Pulau Baru Village and Pasar Sebelah Village have no disabilities.

B. Modeling the Poverty Level in the Coastal Area of Mukomuko Regency by Using the Empirical Best Linear Unbiased Prediction Method

Mukomuko District consists of 152 villages. 13.16% of these or 20 villages are villages in the coastal area, with the number of households for each village selected as examples ranging from 9 to 10 households. The number of samples for each village is very

small compared to the number of households in each village, which is only between 0.1% and 1.26%. The small number of samples used directly causes the MSE variable Y (referred to θ^{DE}) that as the direct estimator produced is very large. To improve the estimation of the percentage of poor families, the SAE model is then used with the EBLUP method through Maximum Likelihood (ML) and Restricted Maximum Likelihood (REML). The results of the estimation of percentage of Poor Families can be seen in Table 2.

Table 2. Statistics of Direct Estimator and EBLUP Estimator

	$\widehat{\theta}_i^{DE}$	$\widehat{\theta}_i^{ML}$	$\widehat{\theta}_i^{REML}$
Mean	8.56	7.90	8.22
Standard Error	1.95	1.83	1.88
Median	8.24	7.95	8.13
Standard Deviation	8.71	8.20	8.41
Sample Variance	75.89	67.29	70.80
Minimum	0.00	0.00	0.00
Maximum	35.36	34.50	34.88

Comparison of observation data and estimation results of the percentage of poor families through EBLUP in each village in Mukomuko District can be seen in Figure 1. From Figure 1 it can be seen that the estimation results using SAE with EBLUP method have trends that are almost same as the observation data. The model obtained has good flexibility. It can be seen from the plot of the alleged results that are able to follow the pattern of distribution of observation data.

Table 3. Comparison of MSE of several Estimation Methods

	$MSE(\widehat{\theta}_i^{DE})$	$MSE(\widehat{\theta}_i^{ML})$	$MSE(\widehat{\theta}_i^{REML})$
Mean	7.17	7.28	6.65
Standard Error	1.05	1.11	0.99
Median	8.69	8.54	8.02
Mode	0.00	0.00	0.00
Standard Deviation	4.71	4.95	4.42
Sample Variance	22.18	24.49	19.55
Minimum	0.00	0.00	0.00
Maximum	15.94	18.27	16.06

From Figure 2, it can be seen that majority of MSE values of $\widehat{\theta}_i^{ML}$ and $\widehat{\theta}_i^{REML}$ are smaller than the MSE values of $\widehat{\theta}_i^{DE}$ for each village. From Table 3, mean of MSE value of $\widehat{\theta}_i^{REML}$ is smallest. This indicates that the estimation with the EBLUP method using REML can improve the parameter estimation obtained by using the direct estimation method.

Poverty mapping of the coastal areas of Mukomuko District can be seen in the Figure 3. The aggregation data for each sub-district used is the maximum statistic percentage of poor families in each subdistrict. On this map, the poverty level is divided based on the data quintile. Then, given the coloring on the subdistrict polygon with the color gradation according to the poverty level. Subdistricts that have a high percentage of poor families, have the brightest colored polygon. Whereas subdistricts which have a low percentage of poor families, have dark blue polygons.

It can be seen that the spread of poverty in Mukomuko District spread randomly. The highest poverty rates are in Air Rami and Pondok Suguh Subdistricts. Whereas subdistricts which have 75% of poor families are Penarik Subdistrict and XIV Koto District. Teramang Jaya sub-district has 50% of poor families. Finally, subdistricts that have low poverty rates ($\leq$ 25%) are Mukomuko and Air Dikit Subdistrict.
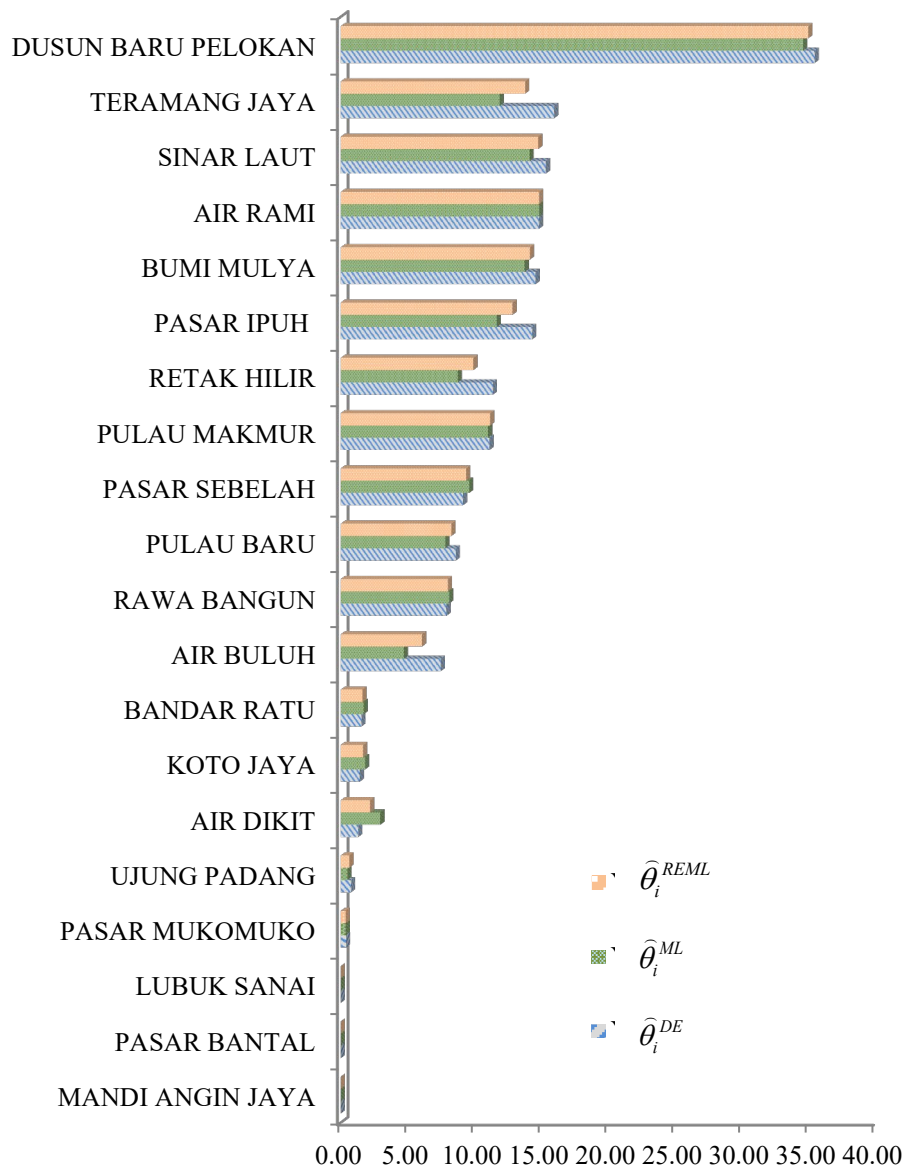
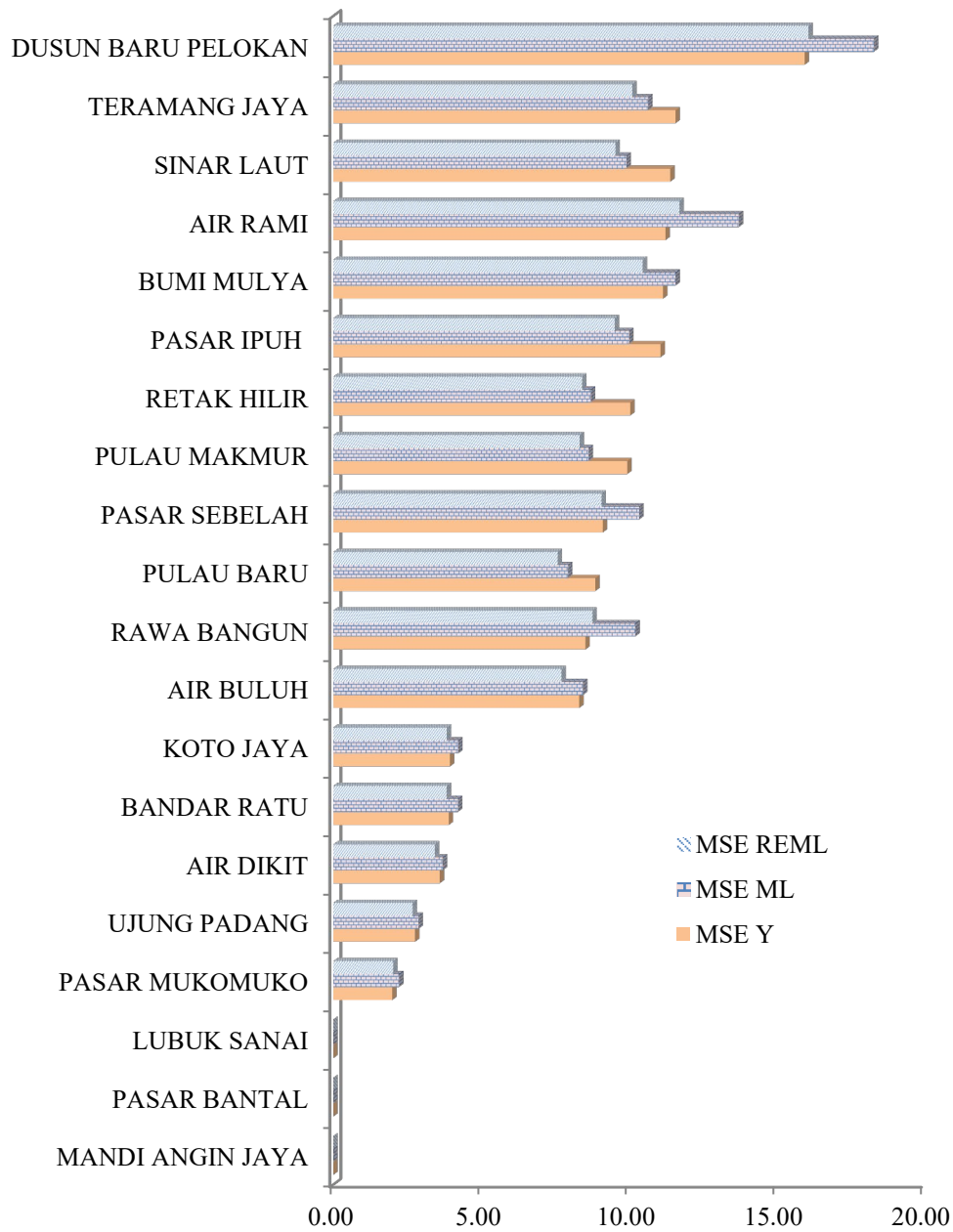**Figure 1.** Distribution Comparison of Several Percentage of Poor Families Estimates

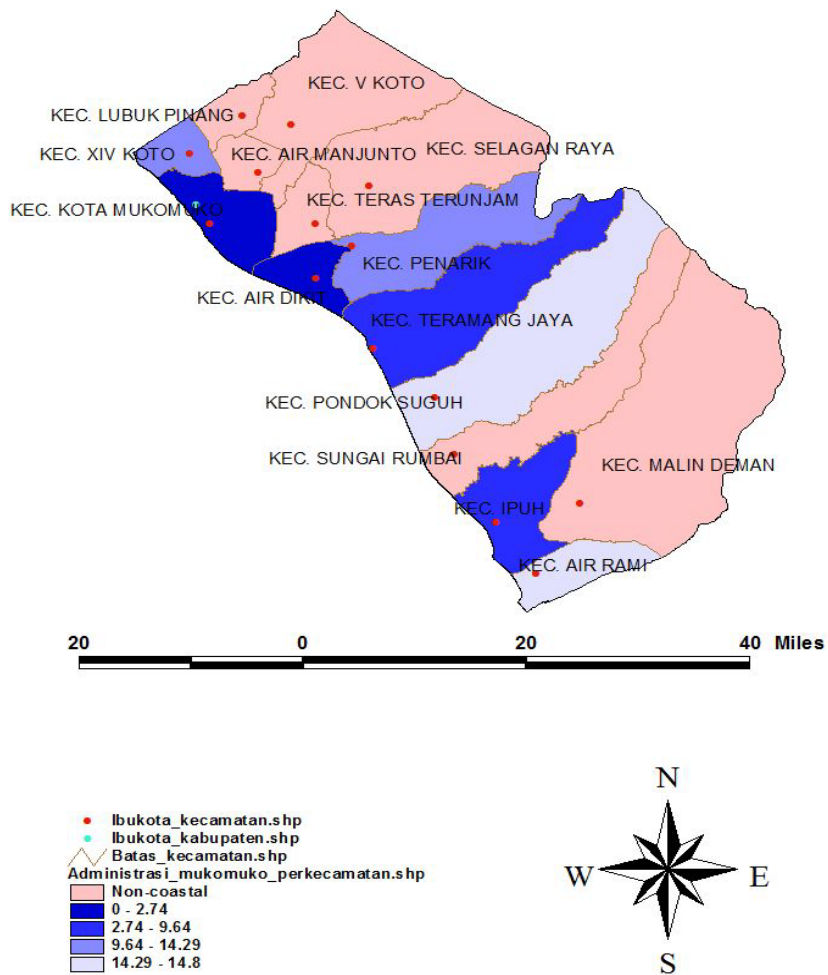**Figure 2.** Comparison of Distribution of Several Estimates of MSE

Figure 3. Poverty Mapping of the Coastal Areas of Mukomuko District

5   Conclusion

Mukomuko District consists of 152 villages. 13.16% of these or 20 villages are villages in the coastal area, with the number of households for each village selected as examples ranging from 9 to 10 households. The number of samples for each village is very small compared to the number of households in each village, which is only between 0.1% and 1.26%. The small number of samples used directly causes the MSE variable Y (referred to θ^{DE}) that as the direct estimator produced is very large. To improve the estimation of the percentage of poor families, the SAE model is then used with the EBLUP.

The estimation results using SAE with the EBLUP method have the same trend as the observation data. The model obtained has good flexibility. It can be seen from the plot of the alleged results that are able to follow the pattern of distribution of observation data. In addition, the MSE value of EBLUP REML estimators is smaller than the MSE value from the direct estimator (Y) for each village. This indicates that the estimation with the EBLUP method can improve the estimation of parameters obtained using the direct method.

References

[1] Kristiyanti M: Pemberdayaan Masyarakat Pesisir Pantai Melalui Pendekatan ICZM (*Integrated Coastal Zone Management*). Proceeding: *Seminar Nasional Multi Disiplin Ilmu Unisbank (Sendi_U) Ke-2*, ISBN: 978-979-3649-96-2 (2016)

[2] Ningtyas R, Rahmawati R, Wilandari Y: Penerapan Metode *Empirical Best Linear Unbiased Prediction* (Eblup) pada Model Penduga Area Kecil dalam Pendugaan Pengeluaran per Kapita di Kabupaten Brebes. *J GAUSSIAN, Volume 4, Nomor 4*, 977-986 (2015)

[3] Malik NFJ, Hoyyi A, Ispriyanti D: Pendugaan Angka Putus Sekolah Di Kabupaten Semarang Dengan Metode Prediksi Tak Bias Linier Terbaik Empirik Pada Model Pendugaan Area Kecil. *J GAUSSIAN, Volume 3, Nomor 1, pp101-110.* (2014)

[4] Sunandi E, Agustina D, Fransiska H. Poverty Mapping of the Coastal Areas Using Spatial Empirical Best Linear Unbiased Prediction Method. *Journal of Physics: Conference Series, Volume 1188, conference 1*. (2018)

[5] BPS: Indicator Kesejahteraan Rakyat Provinsi Bengkulu. *Badan Pusat Statistika, Bengkulu* (2015)

[6] J. N. K Rao and I. Molina: *Small Area Estimation.* John Wiley & Sons, New York. pp. 1-442. (2015)

[7] Aminah AS, Pawitan G, Tantular B: Empirical best linear unbiased prediction method for small areas with restricted maximum likelihood and bootstrap procedure to estimate the average of household expenditure per capita in Banjar Regency. Citation: *AIP Conference Proceedings* 1827, 020010 (2017); doi: 10.1063/1.4979426 (2017)

[8] Petrucci A, Salvati N: Small area estimation using spatial information. The rathbun lake watershed case study. *Dipartimento di Statistica"G. Parenti" viale morgagni,* 59-50134 (2004)

[9] Pratesi M, Salvati N: Small area estimation: the EBLUP estimator based on spatially correlated random area effects, *Statistical methods and applications, Stat. Meth. & Appl.* 17:113-141 (2008)

[10] Minasny B, McBratney AB: Spatial Prediction of Soil Properties Using Eblup With The Matérn Covariance Function, *Elsivier*: *Geoderma*, 140 (2007), 324–336 (2007)

[11] Pereira LN, Coelho PS: Small Area Estimation Using a Spatio-Temporal Linear Mixed Model. *REVSTAT – Statistical Journal,Volume 10, Number 3, 285–308* (2012)

[12] Efron B, Tibshirani R: An Introduction to the Bootstrap. *Chapman and Hall, London* (1993)

[13] BPS Kabupaten Mukomuko (a): Kecamatan Air Dikit dalam Angka 2017. BPS Kab Mukomuko (2017)

[14] BPS Kabupaten Mukomuko (b): Kecamatan Air Rami dalam Angka 2017. BPS Kab Mukomuko. (2017)

[15] BPS Kabupaten Mukomuko (c): Kecamatan Ipuh dalam Angka 2017. BPS Kab Mukomuko (2017)

[16] Kabupaten Mukomuko (d): Kecamatan Kota Mukomuko dalam Angka 2017. BPS Kab Mukomuko (2017)

[17] Kabupaten Mukomuko (e): Kecamatan Penarik dalam Angka 2017. BPS Kab Mukomuko (2017)

[18] Kabupaten Mukomuko (f): Kecamatan Pondok Suguh dalam Angka 2017. BPS Kab Mukomuko (2017)

[19] Kabupaten Mukomuko (g): Kecamatan Teramang Jaya dalam Angka 2017. BPS Kab Mukomuko (2017)

[20] Kabupaten Mukomuko (h): Kecamatan XIV Koto dalam Angka 2017. BPS Kab Mukomuko (2017)