Nonstationary EBLUP on Prediction of Poverty Rate at Village Level in Lembata Regency

Riza Ghaniswati*[1], Asep Saefuddin[2], Anang Kurnia[2]

*[1]Statistics Indonesia, Jakarta, Indonesia

[2]Department of Statistic, Bogor Agricultural University, Bogor, Indonesia

ABSTRACT

The village development program requires accurate village level data, such as the poverty rate. However data poverty rate in Indonesia can only be obtained at the regency/municipality level. An analysis technique to overcome this problem is Small Area Estimation (SAE). SAE model related to poverty rate must be able to produce an estimated proportion that is in the interval of 0 and 1. One approach that can be done is to use logit transformation. The purpose of this study was to estimate the poverty rate at village level in Lembata Regency, Nusa Tenggara Timur Province. This estimation was done by comparing the Empirical Best Linear Unbiased Prediction (EBLUP), Spatial Empirical Best Linear Unbiased Prediction (SEBLUP), and Nonstationary Empirical Best Linear Unbiased Prediction (NSEBLUP). The results showed that logit transformations produced estimates between 0 and 1. The best method to estimate poverty rate at village level in Lembata Regency was NSEBLUP, which produced estimation that more precise than EBLUP and SEBLUP.

Keywords: Poverty, Small Area, Spatial, Village

I. INTRODUCTION

Village law makes the village as subject of development. The village development program requires accurate village level data, such as poverty rate. In this paper, poverty rate refers to the proportion of poor people. In Indonesia, poverty rate was obtained through the National Socio-Economic Survey (Susenas). Susenas has limitations because it is only able to produce poverty rates at regency/municipality level. Estimation at village level cannot be implemented because the sample size is small and even zero.

Rao and Molina [13] state that when there are sub-populations with small sample sizes, direct estimates can produce large standard errors. The analysis technique to overcome this problem is indirect estimation known as Small Area Estimation (SAE). The small area in this study refers to villages.

Research on the proportion of poverty using SAE in Indonesia has been done by several researchers. Suhartini [15] used EBLUP and principal component analysis to estimate the proportion of poor household at all regency/municipality in Jawa Barat. Sulistiyono [16] used boostrap SEBLUP for village poverty mapping in Pati Regency. Both studies assumed that proportions had normal distribution. The use of normal assumptions in proportion variable can produce the estimates that outside the interval 0 and 1. Another approach that can be done for modelling proportions is logit transformation on the proportion variable [2]. The logit transformation produces a new

variable with a value in the interval $(-\infty, \infty)$, therefore we can assume that the results of logit transformation are normally distributed.

Some studies on poverty with SAE showed that there was a spatial effect on poverty data [1,16]. The spatial effect in SAE model depends on the characteristics of the data. The development of SAE model that incorporates spatial effect was first carried out by Cressie (1991) as referenced in Rao and Molina [13] using a conditional autoregressive (CAR) model . Salvati [14] developed a Simultaneously Autoregressive (SAR) model. CAR and SAR assumed spatial stationary, i.e regression coefficient did not vary spatially. According to Fotheringham et al. [9], in the social process the measurement of relationships depends on where the measurement is done called spatial nonstationary. The existence of spatial nonstationarity causes the parameter regression to differ spatially [7]. SAE model with spatial nonstationary effect was developed by Chandra et al [5] for unit level model. Furthermore Chandra et al [7] developed Fay-Herriot model with spatial nonstationary effects, referred as Nonstationary Empirical Best Linear Unbiased Prediction (NSEBLUP).

Nusa Tenggara Timur (NTT) is province with the third largest poverty rate in Indonesia. One of regencies in NTT which has high poverty rate is Lembata. Poverty rate in Lembata on March 2017 reached 0.2648 [3]. That number was higher than poverty rate in NTT (0.2185%). The objective of this research are (1) comparing the SAE model (EBLUP, SEBLUP, and NSEBLUP) with logit transformation to obtain the estimates of proportion of poor people at village level in Lembata, (2) mapping the proportion of poor people at village level in Lembata.

II. MATERIAL AND METHOD

A. MATERIAL

The data used in this study were secondary data from Statistics Indonesia (BPS). The unit analysis in this study was villages in Lembata Regency. Data poverty rate at village level was calculated from national socio-economic survey (Susenas) March 2017. The auxiliary variables were obtained from the 2014 village potential data collection (Podes). Data for 2016 village population in Lembata was cited from 2017 Subdistrict In Figures (KCDA) publications.

In more detail, all variables used in this study are presented in Table 1. X1 to X4 are auxiliary variables. The auxiliary variable was chosen based on the research of Aminah [1], Jumiartanti [10], and Sulistiyono [16] and also availability the data. X6 and X7 are used to create spatial weighting matrices.

TABLE 1
Variable in Research

Variable		Source
Y	Poverty rate	Susenas March 2017
X1	proportion of family using electricity	Podes 2014
X2	The ratio of educational facilities per 1000 residents	Podes 2014
X3	Closest distance to vocational high school (km)	Podes 2014
X4	Proportion of agricultural families	Podes 2014
X5	Population in each village	KCDA 2017
X6	Village's Latitude	Village Shp
X7	Village's Longitude	Village Shp

B. METHOD

The steps of data analysis are as follows:

1. Exploring data.

2. Estimate the proportion of the poor people per sample village using direct estimation.

3. Calculate logit transformation for result from step 2. If there is proportion that has a value of 0 or 1, Cox and Sneil [8] make modification for logit transform as follow:

$$y_i = logit\,(p_i) = ln\left[\frac{r_i + ^1/_2}{n_i - r_i + ^1/_2}\right]$$

where :

n_i : the number of individu on i[th] small area

r_i : the number of poor people on i[th] small are

p_i : proportion of the poor, result from step 2

i : small area (i=1,2,...,m).

4. Making a spatial weighting matrix with rows standardization.

5. Test spatial dependency with Moran's I and Geary'C statistics. According to Cliff and Ord (1981) quoted from Prastesi and Salvati [12], two statistics that can be used to see the presence of global spatial patterns are Moran's I and Geary'C.

6. Seeing linear correlation between auxilary variables as an indication of multicollinearity.

7. Estimating the poverty rate in sample village using EBLUP, SEBLUP [12], and NSEBLUP [7] models.

8. Test for spatial nonstationary effect on the data.

9. Evaluate SAE model using model diagnosis and small areas estimate diagnosis [4,6]. Small areas estimate diagnose is done by comparing Coefficient of Variation (CV) between SAE Model .

$$CV(\tilde{\pi}_i) = \frac{\sqrt{MSE(\tilde{\pi}_i)}}{\tilde{\pi}_i} \times 100$$

where:

$\tilde{\pi}_i$: Estimate of poverty rate in i[th] village

$MSE(\tilde{\pi}_i)$: Mean Square error of povert rate estimate in i[th] village

i : village (i=1,2,...., m)

10. Select the best method for estimating poverty rate at village level in Lembata Regency based on the results of step 9.

11. Estimating poverty rate in nonsample villages based on the following stages:

a. clustering all villages in Lembata Regency based on auxlary variables.

b. In each cluster, calculate average of random effect from sample village.

c. Estimating logit proportion of nonsample villages uses random effect from step b.

d. Back transforming the result on step c to get poverty rate at nonsample villages in Lembata Regency.

12. Calculate the aggregation poverty rate at regency level. If the aggregation results differ greatly from poverty rate published by BPS, then do different calibration with the formula:

$$\tilde{\pi}_i^{DB} = \tilde{\pi}_i + \left(\sum_{l=1}^{M} W_l \hat{\pi}_l - \sum_{l=1}^{M} W_l \tilde{\pi}_l\right)$$

where M is the number of villages in Lembata Regency, $\sum_{l=1}^{M} W_l \hat{\pi}_l$ is direct estimation at regency level, W_l is the ratio between l[th] village population and total regency population, $\tilde{\pi}_i$ is the proportion of poor in the ith village, results from steps 10 and 11, and $\tilde{\pi}_i^{DB}$ is the proportion of poor estimate in the ith village as the result of different calibration.

13. Mapping the estimated value of poverty rate at village level in Lembata Regency.

III. RESULTS AND DISCUSSION

A. General description of Lembata Regency

Lembata Regency is one of the regency in East Nusa Tenggara Province. It consists of one island with an area of 1,266.40 km². The capital of Lembata Regency is Lewoleba, which is in the Subdistrict of Nubatukan. Administratively Lembata District in 2017 consists of 9 subdistricts, 151 villages.

Based on the results of the 2016 population registration, Lembata Regency consists of 124,231 residents. The majority of the population works in the agricultural sector. In March 2017, the percentage of poor people in Lembata Regency reached 26.48% with a poverty line of Rp335,693 per capita per

month. This poverty rate is above the provincial figure, which is 21.85%.

B. Direct Estimation on Poverty Rate

Direct estimate of poverty rate carried out on 39 sample villages. The direct estimation results show that there are three villages (7.7%) whose proportion of poor is 0, namely North Lewoleba and West Lewoleba which are located in the Nubatukan subdistrict, and Benihading located in Buyasari subdistrict. Even though the estimates of the three villages are 0, it doesn't means that in reality all residents in the village are not poor. This can be caused by a small sample size so that the poor population in the village is not selected as a sample. The highest poverty rate as a result of direct estimation is 0.886 located in Lerek village, Atadei subdistrict.

Through direct estimation, the poverty rate in nonsample village cannot be obtained. To produce a better estimate of the proportion of poor people at the village level in Lembata Regency, an indirect estimation method is used. It is expected that the estimation results with the indirect estimation method can increase the precision of estimate. Indirect estimation is done through a model. The model built must be in accordance with the characteristics of the data so that the efficiency of the estimation of small areas can be achieved.

C. Spatial Dependency Test for Poverty Rate

The spatial dependence between the interest variables needs to be considered in the formation of the model. This aims is to eliminate the bias that occurs due to spatial autocorrelation in the observed variables. Two statistics that can be used to see the existence of global spatial patterns are Moran's I and Geary'C [12]. Testing for the presence of spatial dependency with these two methods requires a spatial weighting matrix. The spatial weighting matrix used in this study was the inverse distance matrix with row standardization.

The results of testing the spatial dependency on proportions and logit proportions of poor people are shown in Table 2.

TABLE 2
Test results for spatial dependency

Statistik	Proporsi	Logit Proporsi
Morans'I	0.068*	0.113*
Geary'C	0.903*	0.873*

* significance level 0.05

Based on Table 2, it can be concluded that there is spatial dependency in the data on the proportion and logit proportions of poor people with a level of significance 5%.

D. Modeling Small Area Estimates

The choice of auxiliary variables influences the results of the model predictions. A good auxiliary variable will produce a good model predictions. The criteria for the auxiliary variables used were those allegely related to poverty based on previous studies and had linear correlation with the logit proportions variable. In addition, the auxiliary variables must be available for all villages in Lembata Regency. This study used four auxiliary variables, namely proportion of family using electricity (X1), The ratio of educational facilities per 1000 residents (X2), Closest distance to vocational high school (X3), and Proportion of agricultural families (X4). Educational facilities in this study included elementary school, junior high school, senior high school not only in private school but also in public school.

The results of back transformation in the logit transformation model (Table 3) show that there is no suspicion which the proportion is outside the interval (0-1). This means that the use of logit transformations is appropriate in this study. Based on Table 3, it is known that the estimated average of the three methods is relatively the same. The minimum proportion of the three methods is greater than zero,

meaning that there is no sample village that does not have a poor population. This condition is more likely than the results of direct estimates (there are villages with no poor people).

TABLE 3
Summary of indirect estimation on poverty rate for sample village

Statistics	EBLUP	SEBLUP	NSEBLUP
Averages	0.359	0.359	0.360
Standard Deviation	0.231	0.226	0.229
Minimum	0.032	0.010	0.019
Median	0.321	0.334	0.326
Maximum	0.867	0.845	0.868

Testing of the spatial nonstationary on the data produces a p-value of 0.008. It means that NSEBLUP is the better method compared to the EBLUP model for this data at level of significance 5%.

The best method in estimating poverty rate at village level in Kabupaten Lembata is obtained through a diagnosis of estimation of small areas. Small area estimation diagnosis consists of models diagnosis and small areas estimate diagnosis. Diagnosis of small area models checks whether the assumptions used in forming a small area model are appropriate. Shapiro Wilk test is applied to evaluate residual normality assumptions. Shapiro-Wilk W statistics produce a value of 0.98 for the EBLUP model, 0.962 for SEBLUP, and 0.978 for the NSEBLUP model. With the EBLUP, SEBLUP, and NSEBLUP p-value models of 0.704, 0.204, 0.634 then with level of significance 5% there is no evidence to reject the null hypothesis that the residuals are normally distributed.

The objective of small area estimate diagnosis is to see the estimated reliability. In this study we use coefficient of variation (CV) and MSE evalute the reliabilty of estimate. The estimated value of poverty rate at the village level is said to be more precise if the MSE produced is smaller than the MSE of direct estimate. Similar to MSE, a smaller CV indicates that the model used is more efficient.

The MSE of EBLUP, SEBLUP, and NSEBLUP models in this study used parametric bootsrap. The comparison of MSE estimates from direct estimates, EBLUP, SEBLUP, and NSEBLUP presented in Figure 1. It shows that the SEBLUP and NSEBLUP models generally produce MSEs which are smaller than the direct estimates and EBLUP. This means that the addition of spatial information in estimating the proportion of poor people at the village level in Lembata Regency can increase the precision of the estimated results compared to the direct estimation method and EBLUP. Compared with SEBLUP, NSEBLUP provides more precise results. This indicates that when a spatial nonstationary influence was detected, NSEBLUP was better used to estimate the proportion of the poor in Lembata Regency. Figure 1 shows that villages with direct estimates of poverty rate are 0 (North Lewoleba, West Lewoleba, and Benihading) have a large MSE of EBLUP, SEBLUP, and NSEBLUP.
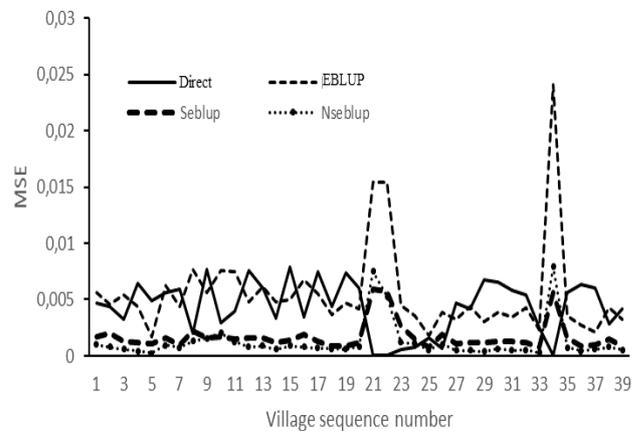

Figure 1: MSE of poverty rate estimates

The CV of poverty rate estimate with direct estimation, EBLUP, SEBLUP, and NSEBLUP models yielded a very large value for the village with a direct estimate of poverty rate of 0. Therefore, to compare the CV from direct and indirect estimates we used median rather than mean. Judging from the median, indirect estimation (EBLUP, SEBLUP, and NSEBLUP) produce a smaller CV than the direct estimate. This means that indirect estimates can increase the precision of the estimates. Similar to the results of MSE comparison, based on comparison of CVs in

Table 4 NSEBLUP models are the most efficient method used to produce poverty rate estimate at village level in Lembata compared with direct estimation, EBLUP, and SEBLUP methods.

TABLE 4

MSE and CV of poverty rate estimate for sample village

Model	Median	
	MSE	CV(%)
Direct	0.0047	21.30
EBLUP	0.0044	18.09
SEBLUP	0.0013	10.14
NSEBLUP	0.0008	7.50

E. Estimating the Poverty Rate

Before estimating poverty rate at nonsample village in Lembata, we clustered all village in Lembata. Determination of number of cluster was done subjectively based on the dendogram. There were nine cluster which was formed. Estimation of poverty rate at village level in Lembata Regency used the NSEBLUP method with cluster correction for nonsample villages. NSEBLUP is used because this method is the best method based on the diagnosis. After obtaining poverty rate estimate in sample villages and nonsample villages, then aggregation was carried out at the regency level to see whether the results obtained were appropriate for use. Aggregation was carried out using a weighted average with the number of population per village as weighting. The aggregation results at the regency level produced higher numbers (0.2945) compared to the figures released by BPS (0.2648). Therefore, calibration was implemented.

Calibration using a different calibration technique resulted in an estimate of poverty rate at regency level of 0.2652. This number is not much different from the one produced by BPS. Therefore, SAE with calibration is appropriate to estimate poverty rate at level in Lembata Regency.

The estimated value of poverty rate at village level in Lembata Regency using calibration method ranged from 0 to 0.905. To make it easier to see the pattern of the distribution of poverty rate in Lembata Regency, a mapping of poverty rate at village level was conducted. Figure 2 shows that villages in the Nubatukan subdistrict, the capital of Lembata Regency, have lower poverty rates compared to other villages. This can happen because Lewoleba village in Nubatukan is the entrance to Lembata Regency. In Nubatukan there are airports and ports which are the main entrances to Lembata Regency. Figure 2 also shows that there is further distance from the regency capital, the poverty rate tends to increase. This can be caused the farther away an area from the city center, the access to the center of the economy, health and education is also increasingly difficult.
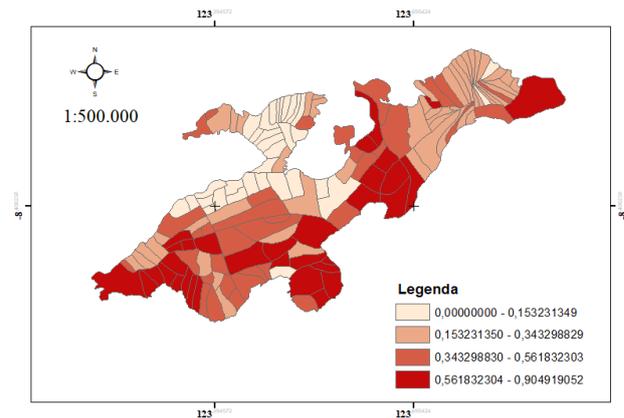

Figure 2: Distribution of proportion of the poor at village level in Lembata Regency

IV.CONCLUSION

The use of logit transformations in proportions was able to produce poverty rate estimate in the interval (0,1). The results of comparison between direct estimates, EBLUP, SEBLUP, and NSEBLUP showed that NSEBLUP was the best method for estimating poverty rate at village level in Lembata Regency. Estimation with NSEBLUP and calibration yields an estimate of poverty rate in Lembata Regency of 0.2652. These results were not much different from the BPS figures (0.2648) so that the estimated results

of the poverty rate at the village level with NSEBLUP and calibration were appropriate.

Mapping the proportion of poor people showed that villages in the regency capital have a smaller poverty rate. The farther away the village from city center, the poverty rate is tend to greater.

V. REFERENCES

[1] Aminah AS. 2017. Penerapan Metode Spatial Empirical Best Linear Unbiased Prediction Dengan Prosedur Restricted Maximum Likelihood Dan Bootstrap Untuk Estimasi Persentase Penduduk Miskin Tingkat Kecamatan Di Kabupaten Wonosobo. [Thesis]. Bandung (ID): Padjadjaran University.

[2] Baum CF. 2008. Modeling Proportion. The Stata Journal. 2: 299-303.

[3] [BPS] Statistics Indonesia. 2017. Data dan Informasi Kemiskinan Kabupaten/Kota Tahun 2017. Jakarta (ID): Statistics Indonesia.

[4] Chandra H, Salvati N, UC Sud. 2011. Disaggregate Level Estimates of Indebtedness in the State of Uttar Pradesh in India: an Application of Small-Area Estimation Technique. Journal of Applied Statistics. 38: 2413-2432.

[5] Chandra H, Salvati N, Chambers R, Tzavidis N. 2012. Small Area Estimation Under Nonstationary Spatial. Journal of Computational and Data Analysis. 56: 2875-2888.

[6] Chandra H. 2013. Exploring Spatial Dependence in Area-Level Random Effect Models for Disaggregate-Level Crop Yield Estimation. Journal of Applied Statistics. 40: 823-842.

[7] Chandra H, Salvati N, Chambers R. 2015. A Spatially N onstationary Fay-Herriot Model f or Small Area Estimation. Journal of Statistics and Methodology Survey. 3: 109-135.

[8] Cox DR, Snell EJ. 2002. Analysis of Binary Data Second Edition. USA: Chapman and Hall

[9] Fotheringham AS, Brunsdon C, Charlton M. 2002. Geographically Weighted Regression. England: John Wiley and Sons.

[10] Jumiartanti. 2018. Model Zero Inflated Beta Pada Pendugaan Area Kecil Untuk Menduga Proporsi Penduduk. [Thesis]. Bogor (ID): Bogor Agricultural University

[11] Kurnia A. 2009. Prediksi Terbaik Empirik untuk Model Transformasi Logaritma di dalam Pendugaan Area Kecil dengan Penerapan pada Data Susenas. [Dissertation]. Bogor (ID): Bogor Agricultural University.

[12] Pratesi M, Salvati N. 2008. Small Area Estimation: The EBLUP estimator based on spatially correlated random area effects. Stat Meth & Appl. 17: 113-141.

[13] Rao JNK, Molina I. 2015. Small Area Estimation Second Edition. New York (USA): John Wiley and Sons.

[14] Salvati N. 2004. Small Area Estimation by Spatial Models: The Spatial Empirical Best Linear Unbiased Prediction (Spatial EBLUP). [Working Paper]. Italy: University of Florence.

[15] Suhartini T. 201 5. Kajian Pendugaan Proporsi Rumah Tangga Miskin Melalui Metode Pendugaan Area Kecil (Studi Kasus Proporsi Rumah Tangga miskin di Jawa Barat. [Thesis]. Bogor (ID): Bogor Agricultural University.

[16] Sulistiyono D. 2016. Bootstrap Spatial Empirical Best Linear Unbiased Prediction untuk Pemetaan Kemiskinan Tingkat Desa di Kabupaten Pati. [Thesis]. Surabaya (ID): Sepuluh November Institute of Technology.