The Poverty Modeling Using Small Area Estimation with Semiparametric P-spline

(A case study: Poverty in Bengkulu Province)

Idhia Sriliana[1,*], Etis Sunandi[2] and Ulfasari Rafflesia[3]

[1] Department of Statistics, Faculty of Mathematics and Natural Science, University of Bengkulu
Bengkulu, Indonesia

[2] Department of Statistics, Faculty of Mathematics and Natural Science, University of Bengkulu
Bengkulu, Indonesia

[3] Department of Mathematics, Faculty of Mathematics and Natural Science, University of Bengkulu
Bengkulu, Indonesia

*Corresponding author's email: idhiasriliana [AT] unib.ac.id

ABSTRACT—*The main objective of this research is to model poverty in Bengkulu Province using small area estimation (SAE) with semiparametric penalized spline (P-Spline). Small area estimation is a statistical method that is often used to obtain an accurate information about poverty. When the linearity assumption on the basic SAE model is violated, a nonparametric approach is used as an alternative. One is the semiparametric penalized spline. The small area method with semiparametric approach has a more flexible model because it accommodates the relationship between response with linear and nonlinear predictors. In this study, poverty modeling in Bengkulu Province was based on average per capita expenditure through the estimation of SAE model parameters using semiparametric P-Spline to obtain a mixed-effect model regression equation as a poverty model. Based on the analysis result, the poverty model in Bengkulu Province is P-Spline linear model with one knot. This model has a GCV value of 148928361265.95. Poverty mapping in Bengkulu Province based on sample villages indicates the estimation of poverty using SAE model with P-Spline having the same trend with the direct estimator.*

Keywords— Bengkulu Province, Poverty, Semiparametric Penalized Spline, Small Area Estimation

1. INTRODUCTION

A poverty is a social problem in Indonesia which until now has not been solved either by central government or by local government. Economic growth occurring in Indonesia is not in line with poverty reduction and economic imbalances. Based on data from the Central Bureau of Statistics (BPS) March 2017, the number of poor people in Indonesia increased by 6,900 people from September 2016 to March 2017. One of the provinces in Indonesia that has a complex poverty problem and needs special attention from the government is Bengkulu Province. Bengkulu is a province in Indonesia with a high poverty rate. The poverty rate of Bengkulu province is almost twice the national poverty rate. BPS Social Economic Data (March 2017) shows that Bengkulu occupies the first rank of the poorest province in Sumatra with a percentage of poor people of 16.45%. While in Indonesia, Bengkulu ranks sixth poorest province in Indonesia. The number of poor people in Bengkulu Province in March 2017 reached 316,980 people, decreased by 11,630 people compared with the poor in March 2016 which amounted to 328,610 people (17.23%) [1].

In general, a poverty is defined as the condition in which a person or group of people is unable to fulfill their basic rights to maintain and develop a dignified life. A Poverty is a multi-dimensional problem, so it is not easy to measure it and need to agree on the measurement approach used. To measure a poverty, BPS uses the concept of basic needs approach. With this approach, a poverty is seen as an economic inability to meet the basic needs of food and not food as measured by expenditure. The poor are people whose average monthly per capita expenditure is below the poverty line. The poverty line is the sum of the food poverty line and the non-food poverty line. The food poverty line is the minimum food expenditure value equivalent to 2100 kilocalories per day. The non-food poverty line is the minimum requirement for housing, clothing, education and health. In 2014, the poverty line of Bengkulu Province is Rp. 378.881,00 [2].

The method that can be used to estimate the average per capita expenditure as an indicator of poverty measurement is the Small Area Estimation (SAE). SAE is a statistical method for estimating parameters in a subpopulation where the number of samples is small or nonexistent. This estimation technique utilizes data from large domains to predict parameters in smaller domains that can be village, sub-district, district, ethnic group, or age group. The SAE method has a concept in the estimation of parameters indirectly in a relatively small area in the sampling survey, which direct predictions are not capable of providing sufficient accuracy when the sample size is small in small areas, so the resulting statistics will have a large variance or even estimation that can not be done because it is not represented in the survey [3].

In general, SAE uses parametric modeling to relate small area statistics with its supporting variables. The prediction of SAE basic model parameters generally uses EBLUP (Empirical Best Linear Unbiased Prediction) method which develops a mixed linear model. This modeling is less flexible in adjusting survey data patterns that do not resemble the existing formal distribution. So approach nonparametrik become alternative of choice, one of them by using semiparametric approach of penalized spline. The semiparametric penalized spline has a more flexible model because the presence of two components in the model accommodates that are the relationship between the response with the linear predictors and the relation between the responses and the nonlinear predictors.

Various studies have been conducted using small area estimation with nonparametric approaches such as: [4] SAE using Penalized Spline Regression to map poverty level in Mukomuko District, [5] SAE using Semiparametric Penalized Spline approach for modeling per capita expenditure in Sleman District, [6] SAE using P-Spline approach to estimate per capita expenditure in Sumenep regency, [7] SAE Kernel-Bootstrap to estimate poverty rate in Indonesia, [9] SAE using a nonparametric model based direct estimator, and [10] Development of SAE with a penalized spline regression approach.

In this research, authors perform an analysis of small area estimation using semiparametric penalized spline approach. The estimation of model parameters using semiparametric P-Spline is then used to model the village level per capita expenditure in Bengkulu Province based on several variables of poverty indicator. Evaluation of the estimation result is done by looking at the GCV value in the model. The result will be presented in poverty map.

2. THE PROPOSED METHOD/ ALGORITHM

2.1 Small Area Estimation

Small Area Estimation (SAE) is a statistical technique for predicting subpopulation parameters of small sample size or even non sampling areas. In Indonesia, such subpopulations could be provinces, districts, subdistricts or villages. SAE is an indirect estimation that combines survey data with other support data e.g. from previous Census data containing variables with similar characteristics to survey data so that it can be used to estimate smaller areas and provide better accuracy [3].

There are two basic model in small area estimation, i.e. area-based model and unit-based model [10]. In the area-based SAE model, supporting data are available only to the area level. The area-level model connects the direct estimator of a small area with supporting data from another domain for each area.

The small area parameter to be observed is θ_i . The linear model that describes the relationship is:

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

with $\beta = (\beta_l, \ldots, \beta_p)^T$ are the regression coefficients of $p \times l$, z_i = a known positive constants, v_i is the small area random effect, assumed $v_i \sim iid\, N(0, \sigma^2)$ where $i = 1, 2, \ldots, m$, and x_i^T is the supporting data of the i-th area.

In making the conclusions about the population, it is assumed that the value of direct estimation $\hat{\theta}_i$ is known. Then, it can be defined as follows:

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

where e_i is the sampling error, assumed $e_i \sim iid\, N(0, \psi_i)$ and $i = 1, 2, \ldots, m$

The SAE model for the area level consists of two levels of the model component i.e. the indirect estimation model component according to (1) and the direct model component of estimation according to (2). The model of (1) and (2) if combined form the following equation:

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

where $i = 1, 2, \ldots, m$.

2.2 Penalized Spline Regression

Penalized Spline or *P-spline* regression is a very interesting smoothing method because it has a simple properties [11]. Given model:

$$y_i = m(x_i) + \varepsilon_i \tag{4}$$

where ε_i are independent random variables with mean *npl* and variance σ_ε^2. Function $m(x_i)$ is an unknown function and is assumed to be approximated by P-splines:

$$m(x_i) = \beta_0 + \beta_1 x + ... + \beta_p x^p + \sum_{j=1}^{K} \gamma_j \left(x_i - k_j\right)_+^p \tag{5}$$

where p is the spline degree (*fixed*), $(x_i - k_j)_+ = maks\left\{0, \left(x_i - k_j\right)\right\}$, $k_j, j = 1,...,K$ are the set of knots. $\boldsymbol{\beta} = \left(\beta_0,...,\beta_p\right)^T$ is a parametric coefficients vector of unknown parameters, $\boldsymbol{\gamma} = \left(\gamma_1, ..., \gamma_k\right)^T$ is a vector of spline coefficients.

Given $X = \left[1\ x_i\ ...\ x_i^p\right]_{1 \le i \le n}$, $Z = \left[(x_i - k_1)\ ...(x_i - k_K)_+^p\right]_{1 \le i \le n}$ with $(x_i - k_j)_+^p = \begin{cases} (x_i - k_j)_+^p & untuk\ x_i \ge k_j \\ 0 & untuk\ x_i < k_j \end{cases}$

so the model in (4) can be written in the form of:

$$y_i = \beta_0 + \beta_1 x + ... + \beta_p x^p + \sum_{j=1}^{K} \gamma_j \left(x_i - k_j\right)_+^p + e_i \tag{6}$$

or it can be written in matrix form as follows:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{e} \quad \text{where} \quad \mathbf{Y} = (y_1 ... y_n)^T \tag{7}$$

Generalized Cross Validation (GCV). The GCV definition can be written as follows:

$$\text{GCV}(\mathbf{K}) = \frac{MSE(\mathbf{K})}{\left[n^{-1} trace\left(1 - A(\mathbf{K})\right)\right]^2} \tag{8}$$

where $MSE(\mathbf{K}) = n^{-1}\mathbf{Y}^T \left(\mathbf{I} - A(\mathbf{K})\right)^T \left(\mathbf{I} - A(\mathbf{K})\right)\mathbf{Y}$, $K = \left(K_1, K_2, ..., K_N\right)$ are knots and matrix $\mathbf{A}(\mathbf{K})$ is obtained from $\hat{\mathbf{Y}} = A(\mathbf{K})\mathbf{Y}$

2.3 Small Area Estimation Using Semiparametric Penalized Spline

In the small area estimation model using semiparametric penalized spline approach, the penalized spline model is a random effect that can be combined with the area-based SAE model to obtain a semiparametric small area estimate based on a mixed linear model. Based on (3) and (7), the Fay-Herriot semiparametric model can be written as follows:

$$\hat{\theta} = \begin{bmatrix} \mathbf{X} \\ \mathbf{X_1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} & \boldsymbol{\beta} \end{bmatrix} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{Du} + e \tag{9}$$

According to Giusti, Pratesi, and Salvati [12], if there are other variables that need to be included in the model, then the variable is added to the matrix X as a fixed effect.

Suppose there are T small areas, $U_1, U_{2,} ..., U_T$ are parameters to be estimated. Define that $d_{it} = I_{\{i \in U_T\}}$ and for each observation defined by $\mathbf{d}_i = \left(d_{i1}, d_{i2} ..., d_{iT}\right)$, $\mathbf{Y} = \left(y_1, y_2, ..., y_n\right)^T$,

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & \cdots & x_1^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^p \end{bmatrix}, \ Z = \begin{bmatrix} (x_1 - k_1)_+^p & \cdots & (x_1 - k_K)_+^p \\ \vdots & \ddots & \vdots \\ (x_n - k_1)_+^p & \cdots & (x_n - k_K)_+^p \end{bmatrix},$$

where $(x_i - k_j)_+^p = \begin{cases} (x_i - k_j)_+^p & \text{for } x_i \geq k_j \\ 0 & \text{for } x_i < k_j \end{cases}$

Opsomer et al in [9] using semiparametric *P-spline* to estimate a small area by adding the small area random effects in (7), thus obtained :

$$\mathbf{Y} = \mathbf{X\beta} + \mathbf{Z}\gamma + \mathbf{Du} + e \tag{10}$$

where semiparametric function $\mathbf{X\beta} + \mathbf{Z}\gamma$ is a spline function that contains linear and nonlinear components, **Du** are the small area random effects, $\mathbf{D} = (\mathbf{d}_1, ..., \mathbf{d}_n)^T$ are covariance matrix, $\mathbf{d}_i = (d_{i1}, ..., d_{iT})^T$ and **u** is a small area effect vector, each random component is assumed to be independent, and

$\gamma \sim (\mathbf{0}, \Sigma_\gamma)$, $\Sigma_\gamma \equiv \sigma_\gamma^2 I_K$

$\gamma \sim (\mathbf{0}, \Sigma_u)$, $\Sigma_u \equiv \sigma_u^2 I_T$

$\gamma \sim (\mathbf{0}, \Sigma_\varepsilon)$, $\Sigma_\varepsilon \equiv \sigma_\varepsilon^2 I_n$

If the variances of the random components are known, the estimation of fixed effect parameter $\boldsymbol{\beta}$ can be done by *Maximum Likelihood Estimation* (MLE) method by assuming γ and u as random effects. Equation (10) can be written in matrix form:

$$\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}^* \tag{11}$$

where $\boldsymbol{\varepsilon}^* = \mathbf{Z}\gamma + \mathbf{Du} + e$

The estimator of parameter $\boldsymbol{\beta}$ can be obtained by maximizing the likelihood function so that:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X^T V^{-1} X}\right)^{-1} \mathbf{X^T V^{-1} Y} \tag{12}$$

where $\mathbf{V} = \mathbf{Z}\Sigma_\gamma \mathbf{Z^T} + \mathbf{D}\Sigma_\mathbf{u} \mathbf{D^T} + \Sigma_\mathbf{e}$ are the variance covariance matrix of **Y**.

The best estimator of predictor variables γ and u are obtained by minimizing the mean squared error (MSE) of predictor variables for both γ and u. Thus, the best unbiased linear predictor (BLUP) is obtained for γ and u as follows:

$$\hat{\gamma} = \sum_\gamma \mathbf{Z}^T \mathbf{V}^{-1} \left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right) \tag{13}$$

$$\hat{u} = \sum_u \mathbf{Z}^T \mathbf{V}^{-1} \left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right) \tag{14}$$

For a given small area t, we are interested in predicting:

$$\bar{y}_t = \bar{x}_t \beta + \bar{z}_t \gamma + u_t \tag{15}$$

where $\bar{x}_t$ and $\bar{z}_t$ are the true means of the powers of x_i and of the spline basis function over the small area, and u_t is the small area random effect. Both $\bar{x}_t$ and $\bar{z}_t$ are assumed to be known. Clearly, $u_t = \mathbf{d_t u} = \mathbf{e_t u}$, where $\mathbf{e_t}$ is a vector with 1 in the *t*th position and 0s everywhere else. So the estimator for $\bar{y}_t$ that is:

$$\hat{y}_t = \bar{x}_t \hat{\beta} + \bar{z}_t \hat{\gamma} + \mathbf{e_t} \hat{u}_t \tag{16}$$

which is a linear combination of generalized least squares estimator (GLS) estimators at (10) and BLUPs in (11) and (12), so $\hat{y}_t$ is itself the BLUP for $\bar{y}_t$.

The BLUP estimator depends on the components of variance that are usually unknown. The estimation of variance components using MLE will result in the bias estimator, so the estimation is done using REML (*Restricted Maximum Likelihood*) based on the residual calculated after β is calculated. To estimate the components of variance using REML we can use several methods, namely Newton Raphson method and EM algorithm (expectation and maximization).

3. METHOD

The research data was secondary data obtained from Central Bureau of Statistics of Bengkulu Province (Susenas Data and Podes Data in 2014). The object of this research is the sampled villages in the SUSENAS survey in Bengkulu Province. Based on the SUSENAS survey, there are 502 villages in 113 sub-districts that are sampled for estimating average per capita expenditure at the village level in Bengkulu Province. The variables used in the study include: average per capita expenditure as the response variable, the number of Jamkesmas (Public Health Insurance) recipients as a linear component predictor variable, and the number of SKTM (Not Capable Certificate) as a nonlinear component predictor variable. Research begins with exploring the data to obtain a general description of factors that affect poverty in Bengkulu Province. Furthermore, data analysis is done, namely the formation of poverty model using SAE with semiparametric P-Spline approach. The model was constructed based on the data samples to obtain an equation of mixed effects linear regression as SAE model. Estimated parameters and prediction random effect variables in the model semiparametric P-spline can be obtained by REML, which is implemented in procedure lme ( ) in R, or by using programs that have been specifically written for penalized spline regression such as the SemiPar package in R. The mapping is done by Arcviews GIS.

4. RESULTS AND DISCUSSION

4.1 The Estimation Model of Poverty in Bengkulu Province

In this research, from the SUSENAS data, there are 502 villages in 113 sub-districts that are sampled in the estimation of per capita expenditure at the village level in Bengkulu Province. From 502 samples, there are 15 invalid sample data. So in this study, the small area model estimated by 487 sample villages using semiparametric penalized spline. Exploration of per capita expenditure data in Bengkulu Province can be seen in **Figure 1**.

Figure 1 present the distribution of sample villages in each sub-district that have an average per capita expenditure. Based on the sample data obtained from the SUSENAS survey of Bengkulu Province in 2014, about 75% of subdistricts in Bengkulu Province have village with the minimum average distribution of expenditure per capita of Rp 585,060.23 and 25% of Rp 411,487.53. The highest average spread of per capita expenditure in Bengkulu Province in 2014 is located in Lais Sub-district of North Bengkulu Regency which is Rp. 1,113,431.56. While the village with the lowest average per capita expenditure is in Kaur Selatan District, Kaur Regency is Rp. 248,376.69.

The pattern of relationship between the average per capita expenditure as a response variable and each predictor variable can be seen through the scatterplot in **Figure 2**.

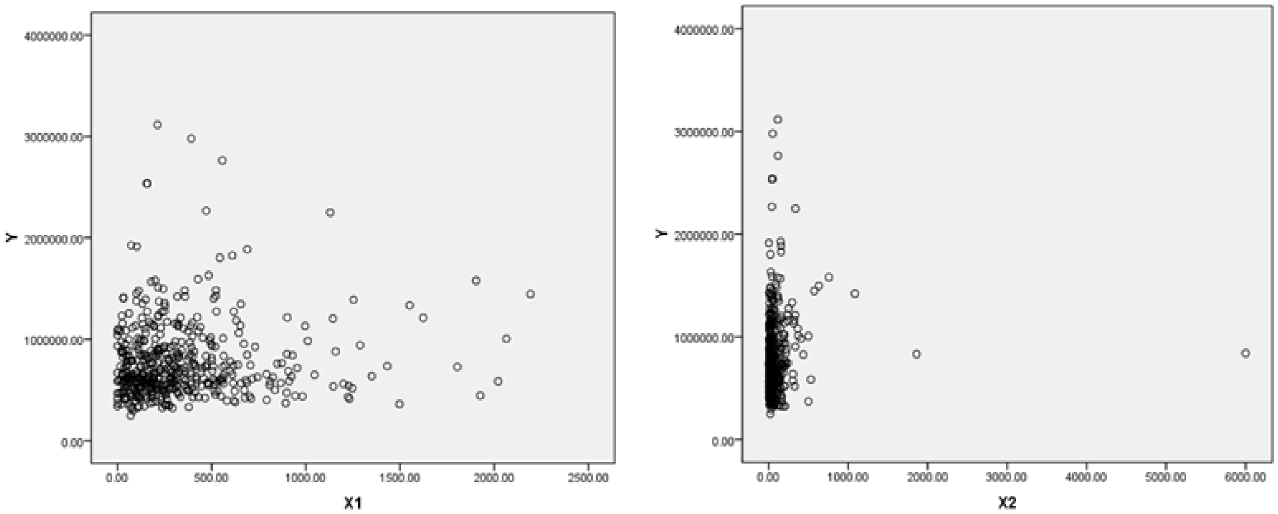

Figure 2: Scatterplot between Response and Predictors

The results of scatterplot in **Figure 2** shows The number of Jamkesmas recipients (X_1) has a linear relationship pattern with the response variable, whereas the predictor variable The number of SKTM (X_2) has a non-linear relationship pattern with the response variable. The data distribution of is clustered so that the assumption of SAE model is violated. Exploration of the relationship patterns among variables also done by linearity test. Based on the linearity test, the variable number of Jamkesmas receiver has a linear relationship to the response with significance of 0.057, while the variable of SKTM has non-linear relationship with significance level of 0.048. Thus, the formation of a fit SAE model is performed by a semiparametric penalized spline that will accommodate the relationship between the response and the linear and nonlinear predictors.

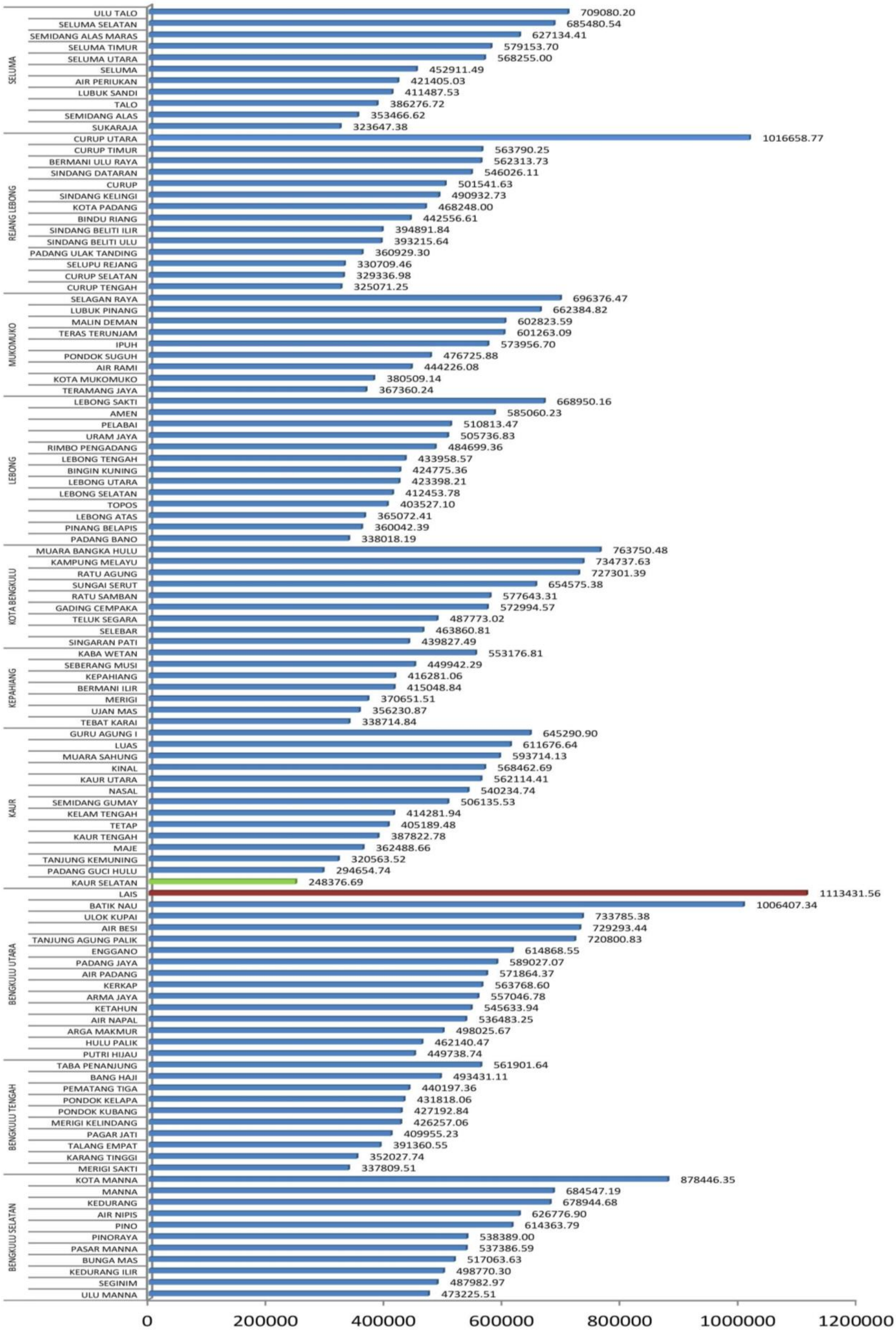

Figure 1: The Distribution of Bengkulu Average Per Capita Expenditure in 2014

Based on the result of exploration of relationship pattern between response variable and each predictor variable, there are two predictor variables used in model estimation that are the number of Jamkesmas recipients (X_1) which is linear pattern and the number of SKTM (X_2) which is non-linear pattern. So, this research is done by estimating SAE model using semiparametric *P-spline* approach. The estimation of SAE model with semiparametric approach of *P-spline* is done through two stages using Generalized Additive Model, which is looking for parametric function $f(x_1)$ and nonparametric function of *P-spline* $f(x_2)$. The process of estimating the model in the first stage to find the parametric value by using the Generalized Additive Model can produce the following model:

$$\mathbf{Y} = 745826 + 90{,}94\mathbf{X}_1 + f\left(x_2\right) + u \quad (17)$$

The second step is estimating nonparametric SAE model using *P-spline* regression. This process is done by modeling residual value on parametric model with non-linear predictor variable $\left(X_2\right)$. Based on the results of data analysis using R, the minimum GCV value for the three models of P-splines is obtained from GCV with one knot. **Table 1** presents the GCV values for linear, quadratic, and cubic *P-spline* models with a single knot point. Based on **Table 1**, it can be seen that the optimum GCV is obtained from the linear *P-spline* model with one knot. So it can be concluded that the SAE model using semiparametric *P-spline* used to model poverty based on the average of per capita expenditure on sample villages in Bengkulu Province is obtained from linear *P-spline* model with one knot.

Table 1: The GCV Values of Linear, Quadratic, and Cubic *P-spline* Models

The Number of Knots	GCV Value		
	Linear P- Spline	**Quadratic *P-spline***	**Cubic *P-spline***
1	148928361265,95*	149546301175,71	Singular
2	150793464357,71	151431105071,33	Singular
3	151532444408,51	152692315419,77	Singular
4	152426221524,96	154364244917,04	Singular
5	153734645823,07	155135245424,28	157347375809,58

Having known the location of the knot point and the *P-spline* model that has GCV Optimum value, the next step is to estimate the fixed effect and random effects. The estimate of influence remains by maximizing the likelihood function or log likelihood function, and predicting for $\hat{\gamma}$ and $\hat{u}$ which is the Empirical Best Linear Unbiased Predictors (EBLUP) of γ and u as random effects. The prediction value for the best semiparametric *P-spline* model (linear spline with one knot) can be seen in **Table 2**.

Table 2: Fix Effect Estimator

Parameter	Penduga
β_0	707678.80
β_1	90.94
β_2	541.21

So that the estimator model is obtained as follows:

$$\mathbf{Y} = 707678{,}80 + 90{,}94\mathbf{X}_1 + 541{,}21\mathbf{X}_2 + \gamma_1\left(\mathbf{X}_2 - 756{,}19\right)_+^1 + u \quad (18)$$

Where γ_i and u are random effect factors with the value of the estimator $\hat{\gamma}_i$ depends on the knot point and $\hat{u}$ depends on each area. The model in (18) is a linear *P-spline* semiparametric model with a knot point at 756.19 and has a penalty or smoothing parameter (λ) of 288504.70.

The model poverty in (18) shows there are differences in the estimation results for each village. The location of the knot at 756.19 point means that if the value and the other predictor variable (X_1) are assumed to be constant, then each increase of one unit affects $(707678.80 + 541.21\mathbf{X}_2 + \gamma_1)$ units against the response variable (*Y*)

4.2 Comparison between The Model and The Actual Data

Based on the model, the estimation of average per capita expenditure at the village level in Bengkulu Province in 2014 is Rp. 521.878,49. About 75% of subdistricts in Bengkulu Province have village with the minimum average distribution of expenditure per capita of Rp 590.254,91 and 25% of Rp 413.864,22. The exploration of the estimated results is similar to the actual data, the highest average spread of per capita expenditure in Bengkulu Province in 2014 is located in Lais Sub-district of North Bengkulu Regency which is Rp. 1.113.537,34. While the village with the lowest average per capita expenditure is in Kaur Selatan District, Kaur Regency is Rp. 248.400,28. **Table 3** shows the statistical comparison between the actual data and the estimated results.

Table 3: The Statistical Comparison between The actual data and the estimated results

Statistic	Average Per Capita Expenditure (Rp)	Estimation of Average Per Capita Expenditure (Rp)
Average	521.828,91	521.878,49
First Quartile	413.824,90	413.864,22
Third Quartile	590.198,84	590.254,91
Minimum	248.376,69	248.400,28
Maximum	1.113.431,56	1.113.537,34

4.3 The Poverty Map of Bengkulu

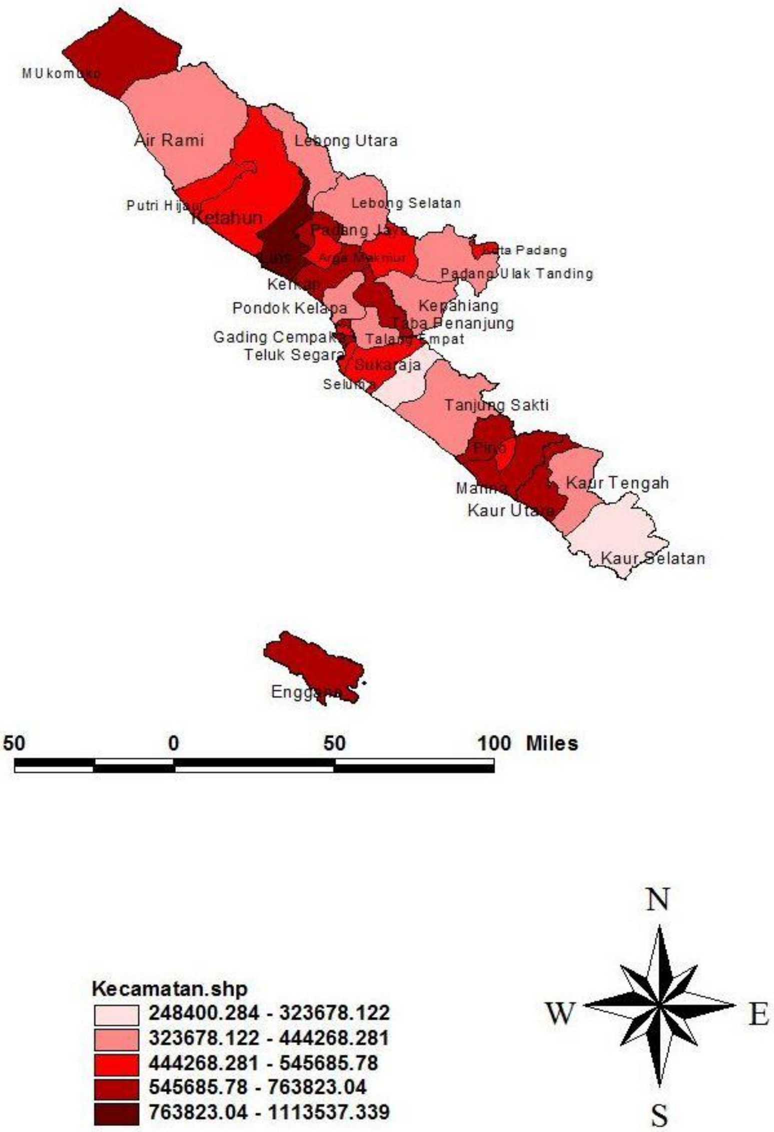

Figure 3: The Poverty Map of Bengkulu Based on Samples

Figure 3 shows the poverty map of Bengkulu Province based on the estimation of the average per capita expenditure for 487 sample villages. On the map, the color of the sub-district polygon being sampled is red gradation which gradation indicates the poverty level. Sub-districts with low average per capita expenditure, indicating high levels of poverty, is represented by pink polygons.

Based on **Figure 3**, it can be seen that sub-districts with high poverty rates are scattered in several sub-districts in southern and southeastern Bengkulu Province, namely Sukaraja in Seluma District. South Kaur in Kaur District has the highest average poverty rate. Most of the sub-districts that have low poverty rates spread in North Bengkulu, Bengkulu, and South Bengkulu. Sub-district with the lowest level of poverty is Lais, which located in North Bengkulu

5. CONCLUSION

Based on the analysis result, it can be concluded that small area estimation using semiparametric penalized spline can be applied to model poverty at village level in Bengkulu Province. The result of data analysis shows the best semiparametric *P-spline* model for the estimation of small area is linear *P-spline* model with one knot. This model has a GCV value of 148928361265.95. The model poverty in Bengkulu Province shows there are differences in the estimation results for each sample village. Poverty mapping in Bengkulu Province based on sample villages shows that the poverty estimation using SAE model with P-Spline has the same trend with the actual data. This method indicates that the estimation produces a consistent estimator. However, some samples have unequal trend that is suspected of being the outliers.

This research does not take into account spatial effect on the estimation of poverty level ini Bengkulu Province. Therefore, for future research, it is required to investigate this effect.

6. ACKNOWLEDGEMENT

Enormous thanks are due to those who have contributed in this research, especially for the Directorate of Research and Community Service, the Directorate General for Research and Development of Kemenristekdikti who has funded this research and LPPM of Bengkulu University as research institution.

7. REFERENCES

[1] Central Bureau of Statistics of Bengkulu Province, "Berita Resmi Statisti -Tingkat Kemiskinan di Provinsi Bengkulu Maret 2017". No. 42/07/17/XI, 17 July 2017, 2017.

[2] Central Bureau of Statistics (BPS), Data Strategis BPS, Katalog BPS 1103003, No. 03220.1202. ISSN, 2087-2011, 2012.

[3] N.G.N. Prasad, and J.N.K. Rao, "The estimation of the mean squared error of the small area estimators", Journal of American Statistical Association, vol 85, pp. 163-171, 1990.

[4] I. Sriliana, D. Agustina, and E. Sunandi. "Pemetaan kemiskinan di Kabupaten Mukomuko menggunakan small area estimation dengan pendekatan regresi penalized spline", Jurnal Matematika Integratif, vol 12, No. 2, pp. 125-133, 2016.

[5] F. Apriani, "Pemodelan pengeluaran per kapita menggunakan small area estimation dengan pendekatan semiparametrik penalized spline", Program Pascasarjana, Institut Teknologi Sepuluh Nopember, Surabaya, 2017. Unpublished.

[6] Z.W. Baskara, "Pendugaan area kecil menggunakan pendekatan penalized spline". Program Pascasarjana, Institut Teknologi Sepuluh Nopember, Surabaya, 2014. Unpublished.

[7] M.Y. Darsyah, and S. Iriyanto, "Analysis of poverty in Indonesia with small area estimation: case in Demak District", South East Asia Journal of Contemporary Business, Economics and Law, vol. 5, Issue 3, pp 18–23, 2014.

[8] N. Salvati, H. Chandra, M.G. Ranalli, and R. Chambers, "Small area estimation using a nonparametric model based direct estimator", Centre for Statistical & Survey Methodology, University of Wollongong, Wollongong NSW, 2008.

[9] D.J. Opsomer, G. Claeskens, M.G. Ranalli, G. Kauermann , and F.J. Breidt, "Non-parametric small area estimation using penalized spline regression", Royal Statistical Society Journal, vol.70, Part 1, pp 265–286, 2008.

[10] J.N.K. Rao, Small area estimation, Wiley, London. 2003.

[11] R.L. Eubank, Spline smoothing and nonparametric regression, Marcel Decker, New York, 1988.

[12] C. Giusti, M. Pratesi, and N. Salvati, "A Semiparametric Fay-Herriot model using penalized spline", Journal of The Indian Society Of Agricultural Statistics, 2012.