BAREKENG: Journal of Mathematics and Its Application
December 2022 Volume 16 Issue 4 Page 1541–1550
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol16iss4pp1541-1550

SMALL AREA ESTIMATION OF MEAN YEARS SCHOOL IN BOGOR DISTRICT USING SEMIPARAMETRIC P-SPLINE

Christiana Anggraeni Putri [1*], Indahwati[2], Anang Kurnia[3]

[1] *Department of Statistics, STIS Polytechnic of Statistics*
Jl. Otto Iskandardinata No.64C, Jakarta 13330, Indonesia
[2,3] *Department of Statistics, Faculty of Mathematics and Natural Sciences, IPB University*
Jl. Raya Dramaga Kampus IPB Darmaga, Bogor 16680, Indonesia

Corresponding author's e-mail: [1] christianaaputri@gmail.com*

Abstract. *The Fay-Herriot model generally uses the EBLUP (Empirical Best Linear Unbiased Prediction) method, which is less flexible due to the assumption of linearity. The P-Spline semiparametric model is a modification of the Fay-Herriot model which can accommodate the presence of two components, linear and nonlinear predictors. This paper also deals with spatial dependence among the random area effects so that a model with spatially autocorrelated errors will be implemented, known as the SEBLUP (Spatial Empirical Best Linear Unbiased Prediction) method. Using data from SUSENAS, PODES, and some publications from BPS, the main objective of this study is to estimate the mean years school at the sub-district level in Bogor district using the EBLUP, Semiparametric P-Spline approach, and the SEBLUP method. The results show that based on the RRMSE value, the cubic P-Spline model with three knots predicts the mean years school better than EBLUP. Meanwhile, the addition of spatial effects into the small area estimation has not been able to improve the estimated value of the P-Spline semiparametric approach.*

Keywords: *mean years school, semiparametric penalized spline, small area estimation, spatial EBLUP*

Article info:
Submitted: 19th September 2022 *Accepted: 28th November 2022*

How to cite this article:
C. A. Putri, Indahwati and A Kurnia, "SMALL AREA ESTIMATION OF MEAN YEARS SCHOOL IN BOGOR DISTRICT USING SEMIPARAMETRIC P-SPLINE", *BAREKENG: J. Math. & App.*, vol. 16, iss. 4, pp. 1541-1550, Dec., 2022.

1. INTRODUCTION

One of the education indicators in the Sustainable Development Goals (SDGs) is the mean years school (MYS). In addition, the MYS also becomes the indicator of the Human Development Index (HDI) which can reflect the educational attainment of a region. The years of schooling in this indicator are defined as the number of years used by residents aged 25 years and over in taking formal education. The MYS in Indonesia officially which released by Badan Pusat Statistik (BPS) are sourced from the National Socioeconomic Survey (Susenas). The Susenas, conducted twice a year, is designed for direct assessment at the national, province and district/city levels. The problem occured when we want to make a direct estimation for a smaller area from the Susenas data. Although Susenas is a large-scale national survey, the number of samples in each small area is too small for direct estimation. When there are subpopulations with small sample sizes, direct estimation can produce large errors [1]

The estimation of a small area can be done by increasing the sample size. However, the size of the sample is directly proportional to the time, human resources, and costs required. It can be done without having to increase the sample size by indirect estimation using Small Area Estimation (SAE). SAE is an indirect estimation method that utilizes the auxiliary variables and the relationship between regions to minimize the standard error of the estimation results.

The development of an SAE model in research generally uses a parametric approach in relating small area statistics to its supporting variables. One of the models that are often used is the Fay-Herriot model with the Empirical Best Linear Unbiased Predictor (EBLUP) estimation method. This model is less flexible in adjusting the pattern of survey data due to the assumption that the direct estimator is a linear function of the covariates. Errors in the specification of this model can result in biased small-area parameter estimators. Opsomers, et al. [2] conducted SAE with the nonparametric Penalized Spline (P-Spline) approach as an alternative which is considered more flexible and more profitable than the parametric approach if the form of the relationship between the response variable and the covariates cannot be determined. However, the research was conducted under the assumption that data is available up to the unit level. Furthermore, Giusti et al. [3] conducted a similar study but under the assumption that data is only available at the area level. The approach used is semiparametric SAE which is a combination of the P-Spline model and the Fay-Herriot model.

The Fay-Herriot semiparametric model with the P-Spline approach is a modification of the Fay-Herriot model when the linearity assumption is violated. The semiparametric approach is capable of covering not only parametric components but also nonparametric components. Thus, this approach can be used as an alternative, especially if not all functional forms of the relationship between variables and their covariates can be specified or the relationship is not linear.

In addition to linearity, this paper also deals with spatial dependence among the random area effects. This assumption is often violated because in general the diversity of an area is influenced by its surrounding area, so that spatial effects can be included in random effects. This is based on Tobler's first law of geography "everything is related to everything else, but near things are more related than distant things"[4]. In its development, the random effect of spatial correlation is considered to be included in the small area estimation method. This method became known as the Spatial Empirical Best Linear Unbiased Predictor (SEBLUP) method. Pratesi and Salvati [5] then developed the SEBLUP method using a simultaneously autoregressive (SAR) process. The research showed that the SEBLUP method has better accuracy than the direct and EBLUP estimation methods.

The previous studies that applied small area estimation using P-Spline generally did not include spatial effects in the model [6][7][8]. In addition, small area estimation research for the MYS indicator is still rarely carried out. In fact, the availability of more detailed data is needed not only as an evaluating material of development results but also as a policy formulation basis that is right on target at the regional level. Based on the description above, the main objective of this research is to estimate MYS at the sub-district level in Bogor district using several methods, namely EBLUP FH, semiparametric P-Spline, and SEBLUP.

2. RESEARCH METHODS

2.1 Data

The data used in this study is secondary data from BPS. The response variable in this study is the MYS per sub-district in Bogor District from the direct estimation of Susenas Kor 2019. While the auxiliary variables used are the percentage of agricultural families (X_1), the number of residents participating in BPJS Kesehatan PBI (X_2), population density (X_3), number of villages that have schools (X_4), population growth rate (X_5).

2.2 Data Analysis Procedure

The steps of data analysis in this study are as follows:

1. Data exploration.
   a. Explore the distribution of data for each variable
   b. Detecting the pattern of MYS relationships with each auxiliary variable using a scatter plot
   c. Define parametric and nonparametric variables
2. Estimating MYS in sub-district level with the Fay-Herriot Model [9] :

$$\hat{\theta}_i = \boldsymbol{x_i}^T\boldsymbol{\beta} + d_i v_i + e_i$$

or in matrix form:

$$\hat{\boldsymbol{\theta}} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{v} + \boldsymbol{e} \quad (1)$$

i = $1,2,\cdots,m$ is a small area index (sub-district);
$\hat{\theta}_i$ = direct estimator of the i-th small area parameter (θ_i);
$\boldsymbol{x_i}$ = vector of auxiliary variables in the i-th small area;
$\boldsymbol{\beta}$ = vector of regression coefficient;
d_i = known positive constant ($d_i = 1$);
v_i = random effect of the i-th small area, assumed to be independent and identical, $v_i \sim (0, \hat{\sigma}_v^2)$;
e_i = sampling error, assumed to be independent $e_i \sim N(0, \Psi_i)$. e_i and v_i are assumed to be independent.

The EBLUP estimator for the Fay-Herriot model [10]:

$$\tilde{\theta}_i = \hat{\gamma}_i\hat{\theta}_i + (1-\hat{\gamma}_i)\boldsymbol{x_i^T}\hat{\boldsymbol{\beta}}$$

with $\hat{\gamma}_i = \frac{\hat{\sigma}_v^2}{\Psi_i + \hat{\sigma}_v^2}$

3. Estimate MYS with semiparametric P-Spline method
   a. Find the estimated value of the parametric function
   b. Model the residual value of the parametric function with nonparametric components to determine the number of knots and the appropriate P-Spline model, i.e. linear, quadratic, or cubic.
   P-Spline regression model with one covariate [11] :

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

   or in matrix notation:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{e} \quad (2)$$

   with $\mathbf{X} = [1\ x_i \ldots x_i{}^p]_{1\le i\le n}$, $\mathbf{Z} = [(x_i - k_1)_+^p \ \ldots \ (x_i - k_K)_+^p]_{1\le i\le n}$, p is the order of the spline, $(x_i - k_j)_+^p = max\{0, (x_i - k_j)^p\}$, $k_j, j = 1,\ldots,K$ is the set of knots used, $\boldsymbol{\beta} = (\beta_0 \ldots \beta_p)^T$ is the parametric coefficient vector of the unknown parameter, and $\boldsymbol{\gamma} = (\gamma_0 \ldots \gamma_k)^T$ is the spline coefficient vector.
   P-Spline parameter estimator [12]:

$$\hat{\boldsymbol{\Theta}} = \left(\mathbf{C}^{\mathrm{T}}\mathbf{C} + \lambda\mathbf{D}\right)^{-1}\mathbf{C}^{\mathrm{T}}\boldsymbol{Y}$$

with $\mathbf{C} = [\boldsymbol{X} \quad \boldsymbol{Z}]$, $\ddot{\boldsymbol{\Theta}} = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{bmatrix}$, $\mathbf{D} = \text{diag}(\mathbf{0}_{p+1}, \mathbf{1}_K)$ is the penalty matrix, and $\lambda \geq 0$ is the smoothing parameter

c. Determine the optimum number of knots using the Ruppert's fixed selection method [13]:

$$\kappa = \min\left(\frac{1}{4} \times \textit{the number of unique } x_i, 35\right)$$

d. Calculate the GCV values for each P-Spline model [14]

$$GCV(\lambda) = \frac{n^{-1}RSS(\lambda)}{(1 - n^{-1}df_\lambda)^2} = \frac{MSE(\lambda)}{(n^{-1}tr(I - S_\lambda))^2}$$

$$RSS(\lambda) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \quad S_\lambda = \mathbf{C}\left(\mathbf{C}^{\mathrm{T}}\mathbf{C} + \lambda\mathbf{D}\right)^{-1}\mathbf{C}^{\mathrm{T}}, \text{ and } df_\lambda = tr(S_\lambda)$$

e. Determine the best P-Spline model based on the optimum GCV value
f. Determine knot locations
g. Estimate the parameters of the semiparametric P-Spline model
The semiparametric P-Spline model is obtained by combining equations (1) and (2) as follows:

$$\widehat{\boldsymbol{\theta}} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{D}\boldsymbol{v} + \boldsymbol{e}$$

The limited number of observations will result in a model that tends to over-parameter. Because of that in this study $\boldsymbol{\gamma}$ is considered as a fixed effect. So that the resulting model will correspond to the Fay-Herriot model with modifications to the auxiliary variables which are considered as nonlinear components.

4. Perform spatial autocorrelation test using Moran's test
5. Prepare a spatial weighting matrix (W) using the queen contiguity method
6. Add spatial information to the model. The SEBLUP model and estimator are formulated as follows [15]:

$$\widehat{\boldsymbol{\theta}} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}(\boldsymbol{I} - \boldsymbol{\rho}\boldsymbol{W})^{-\mathbf{1}}\boldsymbol{u} + \boldsymbol{e}$$

$$\widetilde{\boldsymbol{\theta}}_i^{SEBLUP} = x_i\widehat{\boldsymbol{\beta}} + \boldsymbol{b}_i^T\left\{\hat{\sigma}_u^2\left[(I - \hat{\rho}\boldsymbol{W})\left(I - \hat{\rho}\boldsymbol{W}^{\boldsymbol{T}}\right)\right]^{-1}\right\}\boldsymbol{Z}^{\boldsymbol{T}}$$
$$\times\left\{\boldsymbol{diag}(\sigma_e^2) + \boldsymbol{Z}\hat{\sigma}_u^2\left[(I - \hat{\rho}\boldsymbol{W})\left(I - \hat{\rho}\boldsymbol{W}^{\boldsymbol{T}}\right)\right]^{-1}\boldsymbol{Z}^{\boldsymbol{T}}\right\}^{-\mathbf{1}}\left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)$$

7. Calibrate the results of small area estimation against BPS publications using the ratio benchmarking method [16]:

$$\hat{Y}_{\boldsymbol{i}}^{\boldsymbol{RB}} = \hat{Y}_{\boldsymbol{i}}^{\boldsymbol{B}}\left(\frac{\sum_{i=1}^{40} W_i\hat{Y}_i}{\sum_{i=1}^{40} W_i\hat{Y}_i^B}\right)$$

8. Estimate MSE of EBLUP and semiparametric P-Spline (Datta and Lahiri 2000 in [17]):

$$mse\left(\tilde{\theta}_i\right) \approx g_{1i}(\hat{\sigma}_v^2) + g_{2i}(\hat{\sigma}_v^2) + 2g_{3i}(\hat{\sigma}_v^2)$$

9. Estimate MSE of SEBLUP (Singh et al. 2005 in [17]):

$$mse[\tilde{\theta}_i(\hat{\sigma}_u^2, \hat{\rho})] \approx g_{1i}(\hat{\sigma}_u^2, \hat{\rho}) + g_{2i}(\hat{\sigma}_u^2, \hat{\rho}) + 2g_{3i}(\hat{\sigma}_u^2, \hat{\rho}) - g_{4i}(\hat{\sigma}_u^2, \hat{\rho})$$

3. RESULTS AND DISCUSSION

3.1 Data Exploration

There is no example less sub-district in Susenas Kor 2019, so the unit of analysis in this study includes 40 sub-district. The range of MYS in each sub-district in Bogor district is 4.35 to 11.44 years. The distribution of MYS is shown in Figure 1. Sub-district Babakan Madang has the highest MYS with the average population has attended formal education to upper secondary level (11.44 years). Meanwhile, Sub-district Sukajaya has the lowest MYS with the average population has not graduated from elementary school (4.35 years).

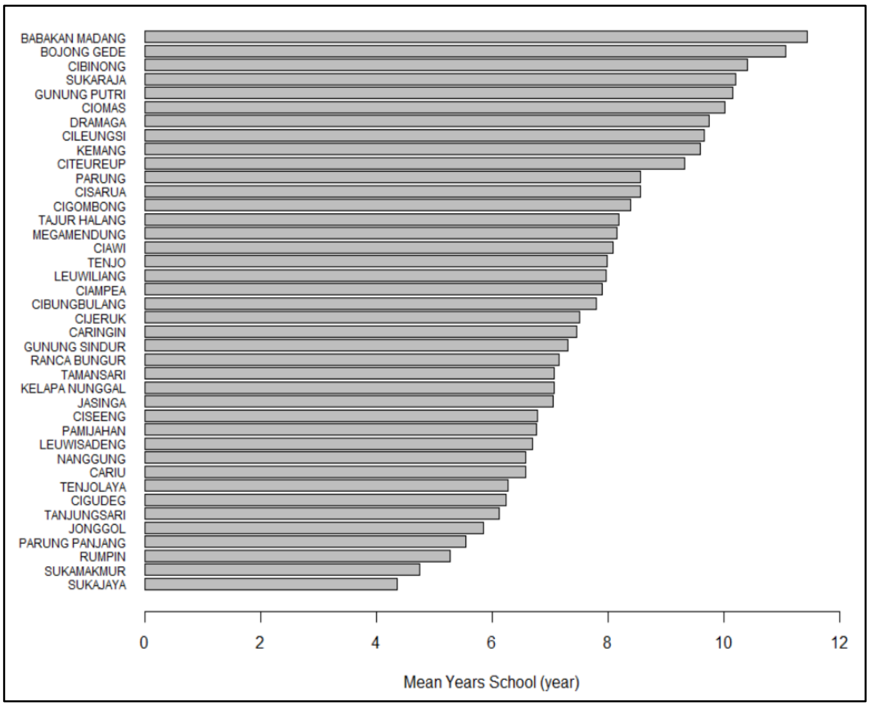
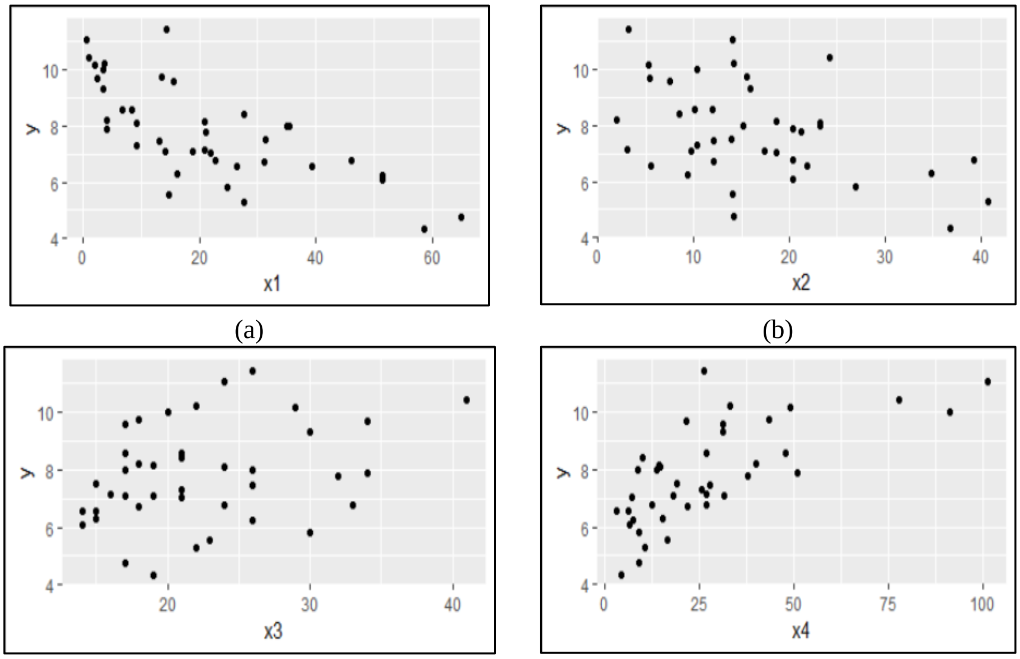


Figure 1. MYS direct estimator in each Sub-district, 2019

The relationship pattern of each auxiliary variable with MYS is visualized in Figure 2. It shows that X_1, X_2, X_3, dan X_4 have a linear relationship with MYS. Unlike the other auxiliary variables, the distribution pattern of X_5 tends to be irregular. Exploration of the pattern of relationships between variables was also carried out using statistical tests with the test results showing that only X_5 did not have a linear relationship with MYS (p-value=0.2004). Thus, a small area estimation model with p-spline semiparametric approach will be used to accommodate the MYS relationship with both linear and nonlinear auxiliary variables.



(c) (d)

Figure 2. Scatter plot of MYS with X_1 (a), X_2 (b), X_3 (c), X_4 (d), and X_5 (e)

3.2 EBLUP

Based on the results of the estimated model parameters in Table 2, the MYS small area model using the EBLUP method can be written as follows:

$$\hat{\theta}_i = 8.3899 - 0.0364x_{1i} - 0.0429x_{2i} + 0.0207x_{3i} + 0.0282x_{4i} - 0.4133x_{5i} + v_i$$

Table 2. The Estimated Value of the Fay-Herriot Model Parameters

Parameter	Estimator	Std. error	*p-value*
β_0	8.3899	0.9445	0.0000
β_1	-0.0364	0.0125	0.0037
β_2	-0.0429	0.0178	0.0159
β_3	0.0207	0.0302	0.4934
β_4	0.0282	0.0109	0.0094
β_5	-0.4133	0.3138	0.1879

3.3 Semiparametric P-Spline and SEBLUP

The number of knots and the P-Spline model to be used are based on the optimum GCV value. This process is carried out by modeling the residual value in a parametric model with a nonlinear variable (X_5). The optimum number of knots determined using the fixed selection method produces 9 knots, so the number of knots to be used is 1 to 9 knots. Table 3 shows the GCV values of the three P-Spline models, namely linear, quadratic, and cubic for each number of knots. The P-Spline cubic model with 3 knots produces a minimum GCV value of 0.731989. So that the next MYS small of area modeling will use a cubic P-Spline semiparametric model with 3 knots, each located at points 1.0625, 1.3300, and 1.6025.

Table 3. GCV Values of Linear, Quadratic, and Cubic P-Spline Models

Knot	Linear	Quadratic	Cubic
1	0.934182	0.971482	1.010658
2	0.934027	0.971482	1.010423
3	0.935994	0.971483	0.731989
4	0.934182	0.835501	0.760225
5	0.934182	0.764278	0.769294
6	0.934182	0.76167	0.775549
7	0.934182	0.763896	0.776571
8	0.913703	0.774477	0.772898
9	0.934182	0.778939	0.777953

Spatial autocorrelation test using Moran's test results that there is a spatial autocorrelation of MYS values in each sub-district (p-value = 0.000). So the next analysis is to include the spatial effects on the semiparametric P-Spline model. Table 4 shows the estimated fixed effect values of the P-Spline semiparametric model with and without spatial information, where the regression coefficient estimator in the P-Spline semiparametric model without spatial information is slightly smaller.

Table 4. Estimated Values of Model Parameters

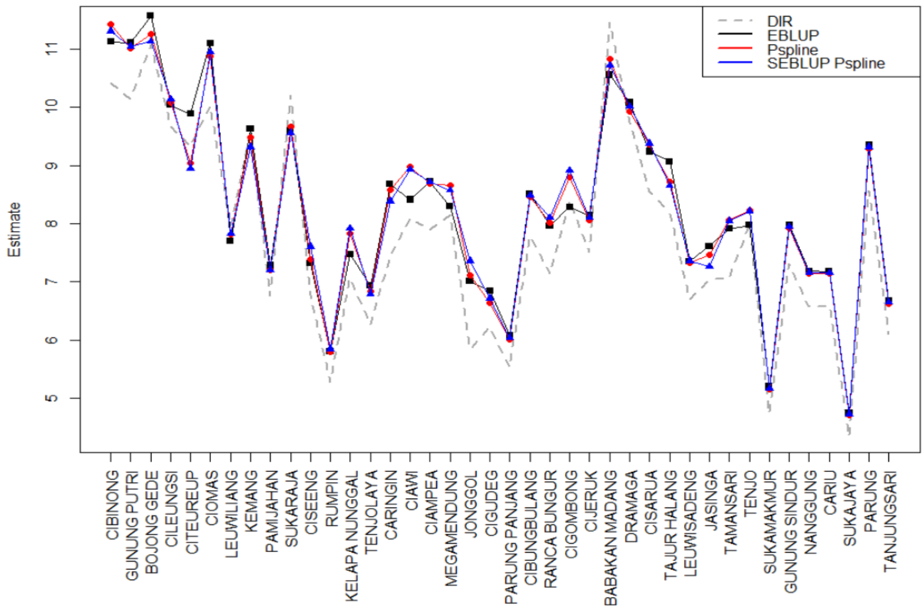
Parameter	**P-Spline**			**Spatial P-Spline**		
	Estimator	**Std. error**	***p-value***	**Estimator**	**Std. error**	***p-value***
β_0	-0.3687	2.3257	0.8740	-0.7539	2.3117	0.7443
β_1	-0.0430	0.0110	0.0001	-0.0401	0.0108	0.0002
β_2	-0.0485	0.0155	0.0018	-0.0541	0.0151	0.0004
β_3	0.0209	0.0261	0.4232	0.0251	0.0249	0.3129
β_4	0.0247	0.0097	0.0108	0.0261	0.0095	0.0059
β_{51}	-5.8571	1.6402	0.0004	-6.0757	1.6017	0.0001
β_{52}	36.2036	9.0843	0.0001	37.5263	8.9789	0.0000
β_{53}	-21.0531	5.3829	0.0001	-21.8888	5.3363	0.0000

Parameter	P-Spline			Spatial P-Spline		
	Estimator	Std. error	*p-value*	Estimator	Std. error	*p-value*
γ_1	86.9626	24.8268	0.0005	92.6276	24.6250	0.0002
γ_2	-102.4816	33.5074	0.0022	-112.3410	33.0047	0.0007
γ_3	41.5257	16.8972	0.0140	47.6007	16.3276	0.0036

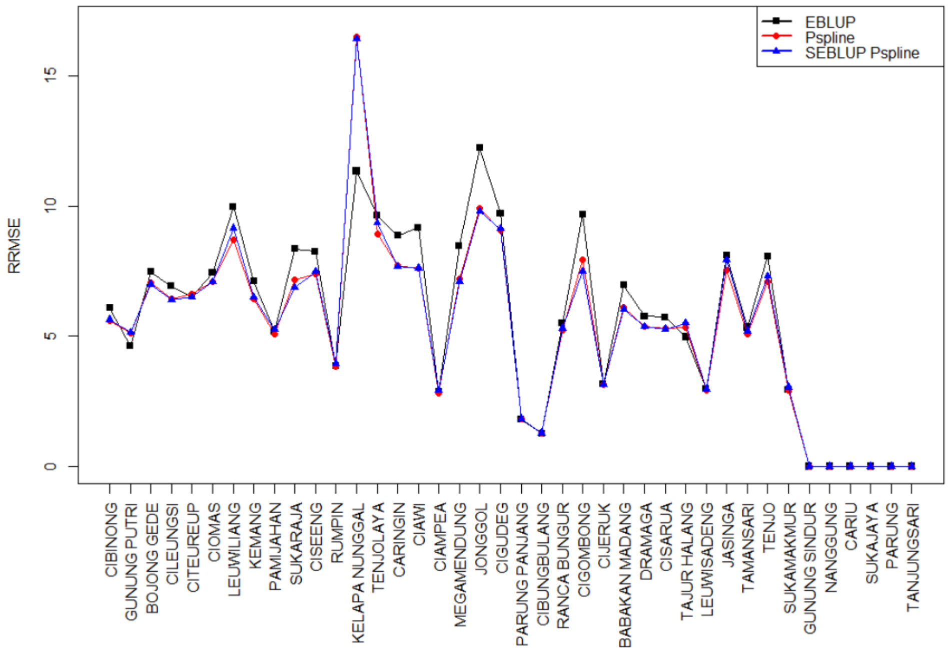
3.4 Comparison of EBLUP, Semiparametric P-Spline, and SEBLUP Estimation Results

A comparison of the estimated results of MYS in each sub-district based on direct estimation, EBLUP, semiparametric P-Spline, and SEBLUP P-Spline can be seen in Figure 4(a). In general, the estimated value of MYS using direct estimation and indirect estimation produces almost the same pattern. Meanwhile, in the P-Spline and SEBLUP P-Spline semiparametric models, there is no significant difference in the estimated MYS values.

The selection of the best model is done by comparing the RRMSE values of each model, where the best model is the model with the smallest RRMSE value. Figure 4(b) shows a comparison of the estimated RRMSE values with the EBLUP method and the P-Spline Semiparametric method. In general, the P-Spline semiparametric model produces a smaller RRMSE value than the EBLUP-FH model. The average value of RRMSE generated by the EBLUP-FH and semiparametric P-Spline models is 5.66% and 5.34%. Thus, there is a decrease in the value of RRMSE after the correction of the variables that are suspected to have no linear relationship with the observed variables, but the difference is not too large. This could be due to the auxiliary variable of the nonlinear component (X_5) which has less effect on the MYS either linearly or nonlinearly.



(a)



(b)

Figure 4. Comparison of Estimated Values (a) and RRMSE (b) Direct Estimation, EBLUP, P-Spline, and SEBLUP P-Spline

Based on Figure 4(b), the addition of spatial effects produces a slightly similar RRMSE value to the previous P-Spline semiparametric model. With the RRMSE average value of 5.36%, it can be concluded that the addition of spatial effects into the model has not been able to improve the estimated value obtained. This can be related to the spatial weighting matrix used. With an appropriate spatial weighting matrix, evaluation of the presence of spatial effects should be able to improve the estimated value. However, as the estimator method that pays attention to the random effect of spatial correlation, the determination of the spatial weighting matrix is a very sensitive element in obtaining optimum estimation results.

4. CONCLUSIONS

Modifications to the EBLUP method with the P-Spline semiparametric approach can be used to estimate MYS in the sub-district level of Bogor district, especially if not all of the auxiliary variables have a linear relationship with the direct estimator. The estimation results using a cubic spline model with 3 knots are able to produce a smaller average of RRMSE than the EBLUP method. This shows that the modification of the Fay-Herriot model with the P-Spline semiparametric approach is better in estimating the sub-district level of MYS in Bogor district. Meanwhile, the addition of spatial effects into the small area estimation has not been able to improve the estimated value of MYS from the P-Spline semiparametric approach.

REFERENCES

[1] J. N. K. Rao and I. Molina, *Small area estimation*, 2nd Ed. New York: Wiley and Sons, 2015.
[2] J. D. Opsomer, G. Claeskens, M. G. Ranalli, G. Kauermann, and F. J. Breidt, “Non-parametric small area estimation using penalized spline regression,” 2008.
[3] C. Giusti, S. Marchetti, M. Pratesi, and N. Salvati, “Semiparametric Fay-Herriot Model using Penalized Splines,” *Journal of The Indian Society of Agricultural Statistics*, vol. 66, no. 1, pp. 1–14, 2012, [Online]. Available: www.isas.org.in/jisas
[4] N. Waters, “Tobler’s First Law of Geography,” in *The International Encyclopedia of Geography*, John Wiley & Sons, Ltd, 2017, pp. 1–13. doi: 10.1002/9781118786352.wbieg1011.
[5] M. Pratesi and N. Salvati, “Small area estimation: The EBLUP estimator based on spatially correlated random area effects,” *Stat Methods Appt*, vol. 17, no. 1, pp. 113–141, Feb. 2008, doi: 10.1007/s10260-007-0061-9.

[6] I. Sriliana, D. Agustina, and E. Sunandi, “Pemetaan Kemiskinan di District Mukomuko Menggunakan Small Area Estimation Dengan Pendekatan Regresi Penalized Spline,” *Jurnal Matematika Integratif*, vol. 12, no. 2, pp. 125–133, Jul. 2016, doi: 10.24198/jmi.v12.n2.11929.59-67.

[7] F. Apriani, “Pemodelan Pengeluaran Per Kapita Menggunakan Small Area Estimation Dengan Pendekatan Semiparametrik Penalized Spline,” ITS, Surabaya, 2017.

[8] I. Sriliana, E. Sunandi, and U. Rafflesia, “The Poverty Modeling Using Small Area Estimation with Semiparametric P-spline (A case study: Poverty in Bengkulu Province),” *Asian Journal of Applied Sciences* , vol. 06, no. 04, pp. 166–174, 2018, [Online]. Available: www.ajouronline.com

[9] M. Yoshimori and P. Lahiri, “A new adjusted maximum likelihood method for the Fay-Herriot small area model,” *J Multivar Anal*, vol. 124, pp. 281–294, Feb. 2014, doi: 10.1016/j.jmva.2013.10.012.

[10] J. Luis´avila, L. Luis´avila-Valdez, M. Huerta, V. Leiva, M. Riquelme, and L. Trujillo, “The Fay-Herriot Model in Small Area Estimation: EM Algorithm and Application To Official Data,” *REVSTAT-Statistical Journal*, vol. 18, no. 5, pp. 613–635, 2020, [Online]. Available: www.victorleiva.cl

[11] N. Ghaida Zia, Suparti, and D. Safitri, “Pemodelan Regresi Spline Menggunakan Metode Penalized Spline pada Data Longitudinal (Studi Kasus: Harga Penutupan Saham LQ45 Sektor Keuangan dengan Kurs USD terhadap Rupiah Periode Januari 2011-Januari 2016),” *Jurnal Gaussian*, vol. 6, no. 2, pp. 221–230, 2017, [Online]. Available: http://ejournal-s1.undip.ac.id/index.php/gaussian

[12] A. Djuraidah and D. Aunuddin, “Pendugaan Regresi Spline Terpenalti dengan Pendekatan Model Linear Campuran,” 2006.

[13] A. W. Ndung’u, S. Mwalili, and L. Odongo, “Hierarchical Penalized Mixed Model,” *Open J Stat*, vol. 09, no. 06, pp. 657–663, 2019, doi: 10.4236/ojs.2019.96042.

[14] M. Maharani and D. R. S. Saputro, “Generalized Cross Validation (GCV) in Smoothing Spline Nonparametric Regression Models,” *J Phys Conf Ser*, vol. 1808, no. 1, Mar. 2021, doi: 10.1088/1742-6596/1808/1/012053.

[15] J. Kubacki and A. Jedrzejczak, “Small area estimation of income under spatial sar model,” *Statistics in Transition*, vol. 17, no. 3, pp. 365–390, Sep. 2016, doi: 10.21307/stattrans-2016-028.

[16] R. Oktavia, “Pendugaan Area Kecil Indikator Kemiskinan Tingkat Sub-district di District Bantul dengan Model Regresi Beta,” IPB, Bogor, 2019.

[17] I. Molina and Y. Marhuenda, “R package sae: Methodology,” 2015. Accessed: Oct. 04, 2021. [Online]. Available: https://cran.r-project.org/web/packages/sae/vignettes/sae_methodology.pdf