An Alternative Procedure to Produce a P-Spline Small Area Estimation Model Based on Partial Residual Plot and Significance Test of Spline Term

Taly Purwa*
BPS Provinsi Bali, Jl. Raya Puputan (Renon) No. 1, Denpasar, Bali, Indonesia

*E-mail: taly@bps.go.id

Abstract. This study provides an alternative procedure to produce the penalized spline small area estimation (P-Spline SAE) model by plotting each of auxiliary variable and variable interest as a simple nonlinearity identification and performing the iterative procedure: by estimating the model, producing partial residual plots to check model adequacy and also identify the nonlinearity, and testing the significance of spline term using Restricted Likelihood Ratio Test (RLRT). These procedures are applied to estimate the monthly average per-capita expenditure of district level in the Province of Bali, 2014 using direct survey estimates from the National Socio-Economic Survey (Susenas 2014) and auxiliary variables from the administrative record of village data (PODES 2014). The Fay-Herriot (FH) model (M1) as a benchmark and four P-Spline SAE model, i.e. the P-Spline SAE with spline term: x2 and x3 (M2), x3 and x4 (M3), x2, x3 and x4 (M4), and x1, x2, x3 and x4 (M5), are obtained. From RLRT, the spline term of M3, M4, and M5 are statistically significant. According to the parametric bootstrap of mean squared prediction error (MSPE) and coefficient of variation (CV), the M4 and M5 show a significant improvement with the CV values range from about 1%-6% compared to M1 with a range from about 4%-17%, shows that these two models more efficient. The M3 model shows the opposite performance even though has the smallest AIC value. More detail, the MSPE and CV produced by M4 are slightly better than M5 makes the M4 is the best model in this study.

1. Introduction
Small area estimation (SAE) has been becoming the most widely used model based method when dealing with inadequate efficiency results from direct survey estimates with small samples for a particular domain or area by borrowing strength from auxiliary variable and random effect area. The auxiliary variables used are usually obtained from the results of registration and census. The choice of these auxiliary variables as good predictors is crucial for small area estimation performance [1]. While random area effect accounts for between area variations beyond the auxiliary variables.

There are two types of Small Area models, i.e. the Fay-Herriot (FH) area level model [2] and unit level model or nested error regression model [3], with linearity assumption between auxiliary variables and direct survey estimator or variable interest. Whereas in some cases, the linearity assumption does not always apply or functional form between auxiliary variables and variable interest cannot be specified a priori and lead to a misspecified SAE model so that the estimators produced will be biased [4][5].

Based on that problem, [6] proposed a non-parametric SAE model for unit level case by taking advantage of Penalized Spline (P-Spline) regression as a mixed model so that provides a more robust
model alternative even when a specific functional form appears reasonable. The Empirical Best Linear Unbiased Estimator (EBLUP), analytical approximation and nonparametric bootstrap prediction of mean squared error (MSE) for P-Spline SAE unit level model are constructed. That study also proposed restricted likelihood ratio test (RLRT) for random effect area and spline term based on nonparametric bootstrap. The estimation of mean Acid Neutralizing Capacity (ANC) from survey of lakes in the northeastern USA using the P-Spline SAE model was performed with the elevation as fixed/linear term, geographical coordinates of the centroid lake as spline term and indicator of hydrologic unit code (HUC) as random effect area. The models with at least one random effect, from spline term or random effect area, outperform the model with only fixed/linear term according to the AIC value.

The P-Spline SAE for area level model, as an extension of the FH model, and also the prediction of MSE, i.e. analytical approximation, nonparametric bootstrap based and combination of them, was proposed by [5]. By using the same real dataset as [6], these study estimated mean ANC and also calcium (CA), both with the same spline term but different auxiliary variables compared to [6] and without random effect area. Through a simulation study, the P-Spline SAE model still has a good performance even the linear assumption of the FH model holds. These results show that the P-Spline SAE model can produce more robust estimators. [4] reviewed the proposed P-Spline SAE model by [6] and use this model and standard unit level model to estimate average household per-capita consumption expenditures at district level of Albania using the Living Standards Measurements Study (LSMS) data. The estimated MSE, using analytical approximation by [6] and conditional approach by [7], of the P-Spline SAE model can produce smaller value than the estimated MSE of the FH model.

[8] used P-Spline SAE unit level model to forecast dwelling prices for the next five years in nine neighborhoods of Victoria, Spain. This study also presented the RLRT for spline term and random effect area by [6]. The different setting performed by treating random effect area as nonparametric part, i.e. unknown function of the area-indicative variable [9]. From these previous studies, there was no explicitly and comprehensively explained procedure from identifying nonlinearity relationship between auxiliary variables and response variable through how to get the best model. [10] and [11] performed preliminary visual identification and nonlinear test, respectively, to determine spline term but there was no selection of the best model conducted.

This study provides an alternative procedure to produce the P-Spline SAE model. First, simple nonlinearity identification is conducted by plotting each auxiliary variable and response variable to inspect nonlinearity subjectively. Then proceed iterative procedure that consists of modeling, inspect the partial residual plot of each auxiliary variable to check fit adequacy and an indication of nonlinearity, testing the significance of nonlinear term using RLRT [6][8] and choose the best model based on parametric bootstrap of MSE [12]. These procedures are applied to estimate monthly average per-capita expenditure of district level in the Province of Bali, 2014.

2. Literature Review

2.1. Fay-Herriot (FH) Area Level Model
The Fay-Herriot (FH) area level model [2] is a combination of sampling model and linking model. At the first stage, there is a sampling model,

\[ \hat{\theta}_i = \theta_i + e_i \]  

(1)

where \( \hat{\theta}_i \) is the direct estimation of the parameter of interest \( \theta_i \) at the \( i^{th} \) area, \( i = 1, 2, ..., m \) and \( e_i \) is sampling error, \( e_i \overset{iid}{\sim} (0, \sigma^2_i) \) with \( \sigma^2_i \) assumed to be known. Then at the second stage, unobservable parameters of interest \( \theta_i \) at (1) are assumed has a linear relationship with a set of auxiliary variables, \( x_i = [x_{i1} \ x_{i2} \ ... \ x_{ip}]^T \), from a census or registration results:

\[ \theta_i = x_i^T \beta + u_i \]  

(2)

Model above is called the linking model since it relates all areas through the common regression coefficients \( \beta = [\beta_1 \ \beta_2 \ ... \ \beta_p]^T \), allowing to borrow strength from all areas [13] and the random
area effect, \( u_i \sim \mathcal{N}(0, \sigma_u^2) \), accounts for between area variation beyond the auxiliary variables [1]. The substitution of the equation (2) to the equation (1) is called the Fay-Herriot model as follow:
\[
\hat{\theta}_i = x_i^T \hat{\beta} + u_i + \epsilon_i
\]
where \( u_i \) and \( \epsilon_i \) are assumed independent for all \( i \). The representation of that equation in the matrix form is:
\[
\hat{\theta} = X\hat{\beta} + Du + \epsilon
\]
\( D = I_m, u \sim (0, \Sigma_u) \) and \( \epsilon \sim (0, \Sigma_e) \), with \( \Sigma_u = \sigma_u^2 I_m \) and \( \Sigma_e = \text{diag}(\sigma_e^2, \sigma_e^2, ..., \sigma_e^2) \). The FH model above is also called the Linear Mixed Model, i.e. the linear model that contains fixed effect part \( x_i^T \beta \) and random effect part \( u_i \) [14].

According to [15], the best linear unbiased estimator (BLUP) estimator of \( \theta_i \) is as follow:
\[
\hat{\theta}^{BLUP} = X\hat{\beta} + D\hat{u}
\]
where
\[
\hat{\beta} = (X^TV^{-1}X)^{-1}X^TV^{-1}\hat{\theta}
\]
\[\hat{u} = \Sigma_uD^TV^{-1}(\hat{\theta} - X\hat{\beta})\]
with \( \text{var}(\hat{\theta}) = D\Sigma_uD^T + \Sigma_e \). The EBLUP estimator is obtained by substitute \( \Sigma_u \) and \( \Sigma_e \) by its estimators, i.e. \( \hat{\Sigma}_u \) and \( \hat{\Sigma}_e \).

### 2.2. P-Spline SAE Model

The connection between P-Spline regression and mixed model is done by treating the spline coefficient as an additional random effect [16] hence the incorporation of semiparametric term, i.e. P-Spline, become popular for SAE modeling started by [6] for unit level model and [5] for are level model. The latter will be used in this study and has general form as follow,
\[
\hat{\theta} = X\hat{\beta} + Z\gamma + Du + \epsilon
\]
[5] also called these model as semiparametric FH model since containing both of parametric term and nonparametric term. For simple illustration, the model with only has a univariate auxiliary variable is
\[
\hat{\theta}_i = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_q x^q + \sum_{k=1}^{K} \gamma_k (x - \kappa_k)^q + u_i + \epsilon_i
\]
where \( q \) is degree of the spline, \( \gamma \sim (0, \Sigma_y) \) with \( \Sigma_y = \sigma_y^2 I_K \) is vector of spline coefficient random effect, \( \kappa_k \) is \( k \)-th knot and \((x - \kappa_k)^q = (x - \kappa_k)\) if \( x > \kappa_k \) and 0 otherwise, \( k = 1, 2, ..., K \), with \( K \) is number of knots. The setting of \( D, u \) and \( \epsilon \) are equal as described in equation (4).

The simple default choice for number of knots by [17] is
\[
K = \max(5, \min(\text{number of unique } x_i's/4, 35))
\]
and by [16] is
\[
K = \min(\text{number of unique } x_i's/4, 35)
\]
with the rule for knots location is
\[
\kappa_k = (\frac{k+1}{K+2})\text{th sample quantile of the unique } x_i's
\]
The BLUP estimator of \( \theta_i \) according to [5] is as follow
\[
\hat{\theta}^{BLUP} = X\hat{\beta} + Z\hat{\gamma} + D\hat{u}
\]
where
\[
\hat{\beta} = (X^TV^{-1}X)^{-1}X^TV^{-1}\hat{\theta}
\]
\[\hat{\gamma} = \Sigma_yZ^TV^{-1}(\hat{\theta} - X\hat{\beta})
\]
\[\hat{u} = \Sigma_uD^TV^{-1}(\hat{\theta} - X\hat{\beta})\]
with \( \text{var}(\hat{\gamma}) = Z\Sigma_yZ^T + D\Sigma_uD^T + \Sigma_e \). Similar to the FH model, the EBLUP estimator of the P-Spline SAE model also can be obtained by substituting \( \Sigma_y, \Sigma_u \) and \( \Sigma_e \) by its estimators, i.e. \( \hat{\Sigma}_y, \hat{\Sigma}_u \) and \( \hat{\Sigma}_e \).
2.3. Partial Residuals Plot

According to [18], partial residual can detect nonlinearity. Supposed there are two auxiliary variables, i.e. $x_1$ and $x_2$, in the model then the partial residuals for $x_1$, is $e_{i1} = \hat{\beta}_1 x_{i1} + e_i$, with $e_i$ is residual. The unmodelled nonlinear component from the relationship of $x_1$ and $y$ should appears in the residuals, so plotting $x_{i1}$ against $e_{i(1)}$ will reveal the partial relationship between $x_1$ and $y$.

This means that the partial residual plots can make allowance for the effect of other auxiliary variables, through the addition of the residual, while only focusing on the relationship between certain auxiliary variable to variable of interest [19]. The partial residuals are usually plotted around the line of fitted values. This plot can also provide a useful diagnostic for model adequacy [16]. If the partial residuals are scattered randomly around the line of fitted values then there is an indication that the model is adequate. While there are certain patterns of partial residuals, then there is an indication that the model is not adequate or there is a nonlinear relationship.

2.4. Hypothesis Testing of Spline Term

[6] proposed restricted likelihood ratio test (RLRT) for testing the significance of random effect area and spline coefficient random effect using the following hypothesis:

$H_0 : \sigma^2_u = 0$ and $\sigma^2_r = 0$

$H_1 : \sigma^2_u > 0$ and $\sigma^2_r > 0$

Under the null hypothesis, the random effect area or spline coefficient random effect is not incorporated in the model. As an example, the procedure to test the significance of the spline coefficient random effect is as follows:

1. Fit model and get the log-likelihood under the null and alternative hypothesis, i.e. $L_0$ and $L_1$, respectively. Compute the restricted likelihood ratio statistic, $L_{obs} = 2(L_1 - L_0)$.

2. Generate $R$ bootstrap replicates of $\tilde{\theta}^r = X\tilde{\beta} + Du^r + e^r$ with $\tilde{\beta}$ is obtained under the null hypothesis, $u^r$ and $e^r$ obtained by sampling with replacement from $\tilde{u}$ and $e$ under the null hypothesis.

   Fit model using $\tilde{\theta}^r$ and $X$ under the null and alternative hypothesis for each bootstrap replication $r$, $r = 1, 2, ..., R$. Compute the restricted likelihood ratio statistic, $L^{r'} = 2(L_{1r}^{r'} - L_{0r}^{r'})$.

3. Compute the bootstrap $p$-value,

$$ P_{boot} = \frac{1 + \#\{L^{r'} \geq L_{obs}\}}{R + 1} \quad (14) $$

with $\#\{L^{r'} \geq L_{obs}\}$ is number of $L^{r'} \geq L_{obs}$, $r = 1, 2, ..., R$. If $P_{boot} < \alpha$ then the null hypothesis is rejected hence the spline coefficient random effect is significant in the model.

2.5. Parametric Bootstrap of MSPE

This study adopted the parametric bootstrap of MSPE by [12] with the following procedures:

1. Generate $R$ bootstrap replicates of $\tilde{\theta}^{r'} = X\tilde{\beta} + Z\gamma^{r'} + Du^{r'} + e^{r'}$, $r = 1, 2, ..., R$, with $\tilde{\beta}$ is obtained from model estimation, $\gamma^{r'}$, $u^{r'}$, and $e^{r'}$ generated from $N(0, \Sigma_\gamma)$, $N(0, \Sigma_u)$, and $N(0, \Sigma_e)$, respectively.

2. Fit $\tilde{\theta}^{r'}$ and $X$ and get the parameter estimate of $\hat{\beta}^{r'}$, $\hat{\gamma}^{r'}$, and $\hat{u}^{r'}$. Then use these parameters estimate to get the prediction value, $\hat{\theta}^{r'}$.

3. The parametric bootstrap estimate of MSPE for $i$th small area estimate is

$$ MSPE_i = \frac{1}{R} \sum_{r=1}^{R} (\hat{\theta}_{i}^{r'} - \hat{\theta}_{i}^{r})^2 \quad (15) $$

3. Methodology

3.1. Sources of Data

This study uses the same data from the previous study [20]. Instead of using food and non-food expenditure, this study uses total expenditure as a response variable from National Socio-Economic...
Survey (Susenas 2014) and only use four auxiliary variables, i.e. proportion of family as electricity user (x1), ratio of elementary school per 1.000 population (x2), proportion of residents that receive health insurance card (x3) and ratio of poor certificate per 1.000 population (x4), from the administrative record of village data (PODES 2014) of Province Bali.

3.2. Analytical Procedures

The analytical procedures applied in this study are as follows:

1. Calculate the direct estimate of average per-capita expenditure of district level in the Province of Bali from Susenas 2014 as a response variable with samples in each district are assumed obtained from simple random sampling (SRS) process. Then aggregate the village level PODES 2014 data of four auxiliary variables at district level.

2. Plotting each of auxiliary variable and response variable that is useful to identify, visually, whether there is a linear or nonlinear relationship.

3. If there is a visually linear relationship then fit the FH SAE model. In this study, the FH SAE model is still estimated since used as a benchmark model.

4. Produce partial residual plot for each variable to conduct visual diagnostic as model adequacy indication.

5. If there is an indication of a nonlinear relationship of $x_j$ and $y$, with $j = 1, 2, ..., p$, from step 2 and 4 then estimate the P-Spline SAE model with $x_j$ as a spline term.

6. Repeat step 4 and 5 until the adequate model based on partial residual plot is obtained. The several alternative P-Spline SAE models would be produced from this process.

7. Testing the significance of spline term presence in the P-Spline SAE models compared to the FH SAE model using RLRT.

8. Choose the best model by comparing the parametric bootstrap of MSPE and Coefficient of Variation (CV) of FH SAE model as benchmark model and P-Spline SAE model with significant spline term from step 7.

![Flowchart](image)

**Figure 1.** The analytical procedures
The estimation of the FH and P-Spline SAE model above is obtained using lme() function from nlme package [21] in R software.

4. Results and Discussion
The scatter plots of each auxiliary variable vs. response variable are presented in Figure 2. The indication of a nonlinear relationship appears for all auxiliary variables when the scatter plot is compared to the fitted line from linear regression (red solid line), subjectively.

![Figure 2. Scatter Plot of Each Auxiliary Variable and Response Variable](image)

Next, the data is fitted using the FH SAE model (M1) as in equation (4) and the partial residual plots are obtained to identify the P-Spline SAE model (Figure 3). The partial residual plot in Figure 3(a) and 3(d) scattered closer to the red line of fitted values than the plot in Figure 3(b) and 3(c) so $x_2$ and $x_3$ might have a nonlinear relationship with the response variable, respectively. Therefore, the P-Spline SAE model with two spline terms of $x_2$ and $x_3$ (M2) is fitted and the partial residual plots are identified to observe whether the model is already adequate or need more spline term for other auxiliary variables. The degree of the spline term $q$ used in this study is 1 since according to [4] if the number of knot, $\kappa$, is sufficiently large and cover all the range of $x$ then the model will produce high accuracy even the degree of the spline is small.

![Figure 3. Partial Residual Plot of Each Auxiliary Variable Produced by the M1 Model](image)
From the identification process, four possible P-Spline SAE models are obtained, i.e. the P-Spline SAE with x2 and x3 spline term (M2) as mentioned above, x3 and x4 spline term (M3), x2, x3 and x4 spline term (M4), and x1, x2, x3 and x4 spline term (M5). Using the specification as (9), the number of knots used in the spline term for x1, x2, x3, and x4 are 7, 14, 14, and 14 knots, respectively. Then the significance of the spline term for each model are tested using RLRT (R =500 replicates) with the FH SAE model under the null hypothesis. The parameter estimation result, AIC value and p-value of RLRT are presented in Table 1. The significant parameters in the M1 and M2 model are x1 and x4. When the spline term of x4 is incorporated in M3, M4, and M5 model, the x4 variable become insignificant.

| Parameter | M1         | M2         | M3         | M4         | M5         |
|-----------|------------|------------|------------|------------|------------|
| $\beta_0$ | 16.8080*   | 16.7115*   | 17.2920*   | 17.2920*   | 17.4379*   |
|           | (0.0000)   | (0.0000)   | (0.0000)   | (0.0000)   | (0.0000)   |
| $\beta_1$ | -5.8079*   | -6.3903*   | -6.0238*   | -6.0238*   | -8.4935*   |
|           | (0.0019)   | (0.0007)   | (0.0010)   | (0.0010)   | (0.0212)   |
| $\beta_2$ | -1.6203    | 0.0541     | 1.1201     | 1.1200     | 0.7301     |
|           | (0.6346)   | (0.9888)   | (0.7394)   | (0.7394)   | (0.8258)   |
| $\beta_3$ | -0.6334    | -6.3968    | -4.5162    | -4.5161    | -3.0868    |
|           | (0.7295)   | (0.4133)   | (0.5319)   | (0.5319)   | (0.6403)   |
| $\beta_4$ | -0.0398*   | -0.0360*   | -0.1029    | -0.1029    | -0.0785    |
|           | (0.0035)   | (0.0070)   | (0.0735)   | (0.0735)   | (0.1436)   |
| p-value of RLRT | 0.0798 | 0.0140* | 0.0040* | 0.0040* |
| AIC       | 268.6802   | 270.6943   | 268.1798** | 270.1798   | 271.6726   |

*) significant at $\alpha$ =0.05; **) the smallest value

From the RLRT results, the P-Spline SAE models with significant spline term are M3, M4, and M5 model with an equal p-value for M4 and M5. Next, the parametric bootstrap of MSPE and CV of these models need to be compared to choose the best model with the benchmark model, M1, also included.

![Figure 4. Comparison of MSPE (a) and CV (b)](image)

According to the MSPE, the best model is the M4 model that produces the smallest MSPE. This value is slightly smaller than MSPE of M5 model. Using the CV to make comparison more objective, Figure 4(b) also confirms the previous result that the M4 model has the smallest CV values. The boxplot range of M4 and M5 is similar and shorter compared to the boxplot range of M1 and M2. These two results indicated that the estimation results produced by M4 and M5 are more reliable. Contrary, the
smallest AIC value obtained by the M3 model, the model with the worst performance according to the MSPE. While the best model, the M4 model, only produces the third smallest AIC value.

![Partial Residual Plot](image)

**Figure 5.** Partial Residual Plot of Each Auxiliary Variable Produced by the M4 Model

To see the comparison of model adequacy of the M4 and M5 models, the partial residual plots from these two models are presented in Figure 5 and 6, respectively. When the spline term of all auxiliary variables are included in the M5 model, the nonlinear pattern of fitted value for \(x_1\) and \(x_2\) do not strongly appear (Figure 6(a) and 6(b)). The partial residual plot of \(x_2\) is not close enough to the fitted line. The different result obtained when the spline term of \(x_1\) is excluded in the M4 model makes the nonlinear pattern of fitted value for \(x_2\) strongly appears and its partial residual is closer to the fitted line (Figure 5).

![Partial Residual Plot](image)

**Figure 6.** Partial Residual Plot of Each Auxiliary Variable Produced by the M5 Model

5. Conclusions

This study gives a simple procedure to produce the P-Spline SAE model. First, the scatter plot of each auxiliary variable and response variable is used to visually detect nonlinearity relationship and determine possible P-Spline SAE models. The only drawback, this step is not practical when the number of
auxiliary variables is large enough. Next, all possible models will be estimated and the partial residual plot will be produced to diagnose model adequacy. From this diagnostic step, the specification of the model, the spline term included in the model can be further adjusted. The trial and error are needed in this step.

The significance of the spline term is confirmed by the RLRT. In the case of the M2 model, when the partial residual plot is not closed enough to the fitted line then the RLRT tends to give an insignificant result, vice versa. According to the parametric bootstrap of MSPE and CV, the M4 model is slightly better than the M5 model shows that a more complex model does not always outperform the simple one. This makes the M4 is the best model in this study. The M4 and M5 models have a significant improvement with the CV value ranges of about 1%-6% compared to M1 as a benchmark model ranges about 4%-17% shows that these two models more efficient.

References
[1] I. Rao, J. N. K & Molina, Small area estimation, Second Edi. Hoboken, New Jersey, 2015.
[2] R. Fay and R. Herriot, “Estimation of Income from Small Places: an Application of James-Stein Procedure to Census Data,” J. Am. Stat. Assoc., vol. 74, pp. 269–277, 1979.
[3] G. Battese, R. Harter, and W. Fuller, “An Error Component Model for Prediction of County Crop Areas Using Survey and Satellite Data,” J. Am. Stat. Assoc., vol. 83, pp. 28–36, 1988.
[4] M. G. Ranalli, F. J. Breidt, and J. D. Opsomer, “Nonparametric Regression Methods for Small Area Estimation,” in Analysis of Poverty Data by Small Area Estimation, M. Pratesi, Ed. West Sussex, England: John Wiley and Sons Ltd, 2016, pp. 187–204.
[5] C. Giusti, S. Marchetti, M. Pratesi, and N. Salvati, “Semiparametric Fay-Herriot model using Penalized Splines,” J. Indian Soc. Agric. Stat., vol. 66, no. 1, pp. 1–14, 2012.
[6] J. D. Opsomer, G. Claeskens, M. G. Ranalli, G. Kauermann, and F. J. Breidt, “Non-Parametric Small Area Estimation Using Penalized Spline Regression,” J. R. Stat. Soc. Ser. B (Statistical Methodol., vol. 70, no. 1, pp. 265–286, 2008.
[7] N. Salvati, H. Chandra, M. G. Ranalli, and R. Chambers, “Small area estimation using a nonparametric model-based direct estimator,” Comput. Stat. Data Anal., vol. 54, no. 9, pp. 2159–2171, 2010.
[8] M. D. Ugarte, T. Goicoa, A. F. Militino, and M. Durbán, “Spline smoothing in small area trend estimation and forecasting,” Comput. Stat. Data Anal., vol. 53, no. 10, pp. 3616–3629, 2009.
[9] R. Zhu, G. Zou, H. Xie, and Y. Hu, “Semiparametric Small Area Estimation of Crop Acreage under Partially Linear Model,” arXiv, vol. 1212.0270, pp. 1–17, 2012.
[10] F. APRIANI, “Pemodelan Pengeluaran Per Kapita Menggunakan Small Area Estimation Dengan Pendekatan Semiparametrik Penalized Spline,” Institut Teknologi Sepuluh Nopember (ITS), 2017.
[11] 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 J. Appl. Sci., vol. 06, no. 04, pp. 166–174, 2018.
[12] S. K. Sinha and J. N. K. Rao, “Robust small area estimation,” Can. J. Stat., vol. 37, no. 3, pp. 381–399, 2009.
[13] I. Molina and Y. Marhuenda, “sa: An R Package for Small Area Estimation,” R J., vol. 7, no. 1, pp. 81–98, 2015.
[14] J. C. Pinheiro and D. M. Bates, Mixed-Effects Models in S and S-PLUS. New York: Springer, 2000.
[15] C. R. Henderson, “Best linear unbiased estimation and prediction under a selection model,” Biometrics, vol. 31, no. 2, pp. 423–447, 1975.
[16] D. Ruppert, M. P. Wand, and R. J. Carroll, Semiparametric Regression. New York: Cambridge University Press, 2003.
[17] L. Ngo and M. P. Wand, “Smoothing with mixed model software,” J. Stat. Softw., vol. 9, no. 1, pp. 1–54, 2004.
[18] J. Fox, Nonparametric Simple Regression: Smoothing Scatterplots. California: Sage Publication, Inc., 2000.

[19] J. J. Faraway, Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models. Boca Raton: Taylor & Francis Group, 2016.

[20] T. Purwa, A. T. Rumiati, and I. Zain, “Small Area Estimation with Bivariate Hierarchical Bayes (HB) Approach to Estimate Monthly Average per Capita Expenditure of Food and Non-Food Commodities in Province of Bali,” IOP Conf. Ser. Mater. Sci. Eng., vol. 546, pp. 1–11, 2019.

[21] J. Pinheiro, D. Bates, S. DebRoy, and D. Sarkar, “nlme: Linear and Nonlinear Mixed Effects Models.” p. 83, 2013.