The restricted maximum likelihood method for variance estimation in a mixed model with additive penalized-spline

A N R Chytrasari¹,²*, S H Kartiko² and Danardono³

¹,²,³ Department of Mathematics, FMIPA, Universitas Gadjah Mada, Indonesia
¹ Department of Mathematics Education, FKIP, Universitas Pancasakti Tegal, Indonesia

*Corresponding author: angela.nina.r@mail.ugm.ac.id

Abstract. Estimation of variance components in the model often needed in the problem of model-based prediction. We present the concept of restricted maximum likelihood estimation for variance components in the mixed model with additive p-spline. We provide simulations as a result of applying the REML estimator concept to variance components in diverse models with p-spline additives. The model is then used to predict the means of sub-area of the population. The limited simulation shows the performance of the model using the REML estimator for the variance component to provide better predictive results, indicated by the average RMSE value and smaller absolute average values and the denser distribution of the MSPE with a lower median, than that produced by the model the same as the ML estimator for component variance.

1. Introduction

Estimation concept for variance components has found in many statistical kinds of literature. One of the estimation methods for variance components is restricted maximum likelihood (REML) method. The REML method maximizes the likelihood functions that do not contain fixed parts of the model and take into account the degree of freedom associated with the fixed effects of the model. This process is carried out by basing estimates on residues calculated after fitting a fixed effect part of the model using ordinary least square.

Regression problems may involve several covariates where the functional relationships between covariates and responses cannot be specified beforehand. The model used in this paper is limited to the two additive univariate covariates in the unit-level mixed model. Let N unit population is partitioned into m areas, with N states the number of population units in area i (i = 1,...,m). Samples of size n, is taken in each area, and the covariate variables x, and t, responses of the sample are observed for every i and j (j = 1,...,n). Generally, the model for this case stated as

\[ y_{ij} = f_1(x_{ij}) + f_2(t_{ij}) + u_i + e_{ij} \]  

for i = 1,...,m; j = 1,...,n, where \( u_i \) is a random effect of the area i, \( e_{ij} \) is an error of unit j in the area i, while \( f_1 \) and \( f_2 \) are smooth but not specified functions that require a nonparametric approach.
The nonparametric approach using spline in the context of mixed models has been studied. Maiti et al. used basis-spline to estimate the fixed effect and random effect of the model. They developed a functional regression theory and methodology in the context of small area estimation [4]. Meanwhile, the method based on penalized-spline (p-spline) for the functional mixed effect model with varying coefficients offer a flexible estimation of both the population and subject-level curves [3]. P-spline is a method of fitting a smoothing spline using knots and simple penalties to control the smoothing. P-spline approach in the mixed model framework has computational advantages [1]. Wood et al. applied likelihood maximization in p-spline to estimate smoothing parameters through Laplace approximate marginal likelihood. By construction, the method is numerically stable and convergent [10].

Model (1) can be used to estimate population parameters, such as population / sub-population averages. In estimates based on this model, information on variance in the model often needed. However, this variance value information is not always available, so it needs to be estimated. We provided a theoretical concept to obtain variance component estimators in an additive p-spline mixed model using the restricted maximum likelihood (REML) method.

2. Method
We used p-spline to approximate \( f_1(x_j) \) and \( f_2(t_j) \). Firstly, we estimated the coefficients in the model by solving Henderson’s mixed-model equation and obtain the estimators of variance component using the REML method. Simulations were carried out in the estimation context for the mean area. We compared the performance of the model, where the REML method and the maximum likelihood (ML) method were used to estimate the component variance model. Model performance was observed through mean square prediction error and absolute-bias values.

3. Results
3.1. Penalized-spline additive mixed model
In general, the p-spline additive model for the mixed model (1) can be written as

\[
y_{ij} = \alpha + \sum_{k=1}^{p} \beta_k x_{ijk} + \sum_{k=1}^{p} r_k (x_{ijk} - q_k)^2 + \sum_{k=1}^{p} w_k t_{ijk} + \sum_{k=1}^{p} d_k (t_{ij} - Q_k)^2 + v + e_{ij}
\]

For \( i = 1, ..., m \) and \( j = 1, ..., n_i \). Using all sample data, (2) can be written in matrix notation as

\[
y = Xa + Tw + Zh + e
\]

\[
y = \text{col} \left\{ [y_{i1}, y_{i2}, \ldots, y_{in_i}] \right\}, \quad X = \text{col} \left\{ \text{col} \left\{ [1 \ x_{i1} \ \cdots \ x_{in_i}] \right\} \right\}, \quad T = \text{col} \left\{ \text{col} \left\{ [t_{i1} \ \ t_{i2} \ \cdots \ t_{in_i}] \right\} \right\}. \]

Where he parametric part coefficients for \( f_1 \) and \( f_2 \) are respectively stated by vectors \( a = [a_0 \ a_1 \ \cdots \ a_p]^T \) \( w = [w_1 \ w_2 \ \cdots \ w_p]^T \), and the partitioned matrix \( Z \) is defined by \( Z = [Z_1 \ Z_2 \ Z_3] \), where \( Z_1 \) and \( Z_2 \) successively stated the spline section for \( f_1 \) and \( f_2 \), namely the column matrices

\[
Z_1 = \text{col} \left\{ \text{col} \left\{ [(x_{ij} - q_1)^2 \ \cdots \ (x_{ij} - q_k)^2] \right\} \right\}
\]

\[
Z_2 = \text{col} \left\{ \text{col} \left\{ [(t_{ij} - Q_1)^2 \ \cdots \ (t_{ij} - Q_k)^2] \right\} \right\}
\]

Meanwhile \( Z_3 = \text{diag} \left\{ [1_{n_i}] \right\} \). Matrix \( h = [h_1 \ \ h_2 \ \ \cdots \ \ h_{n_i}]^T \), where \( h_i = [r_1 \ r_2 \ \cdots \ r_{n_i}]^T \) and \( h_2 = [d_1 \ d_2 \ \cdots \ d_{n_i}]^T \), is the coefficients of the spline part for \( f_1 \) dan \( f_2 \). Meanwhile, \( h_3 = [v_1 \ v_2 \ \cdots \ v_{n_i}] \) is a vector that states the random area effect. Vector \( e = \text{col} \left\{ [e_{i1} \ e_{i2} \ \cdots \ e_{in_i}] \right\} \) states unit error. Model (3) is a p-spline mixed model that contains fixed effects \( X, T \) and random effects \( Z \). In this paper, the matrices \( X, T \) and are assumed to have the full column rank. Vektor \( h_1 \) and \( e \) are assumed to have a normal distribution with mean \( 0 \) and fixed variance, namely \( h_1 \overset{iid}{\sim} N(0, \sigma^2_{h_i} 1_{n_i}) \) and \( e \overset{iid}{\sim} N(0, \sigma^2 I_n) \) with \( n = \Sigma_{i=1}^{m} n_i \). The variance of \( y \) can be expressed as \( V = \sum_{k=1}^{p} Z'_k G_k Z_k \) for
\( G_k = \sigma^2_{h_k} I_{K_k} \) with \( K_j = m, h_2 = e \) and \( K_4 = n \). The application of p-spline in the context of a regression model is not a completely new concept. Model (3) is an extension of Rao’s p-spline mixed model [7]. The p-spline fitting criteria can be made equal to the prediction criteria in the mixed model by treating the coefficients of the spline section as random coefficients[8]. In line with these opinions, \( h \) and \( h_2 \)
are then seen as random variables. For simplification, it is assumed \( h_i \sim N(0, \sigma^2_{h_i} I_{K_{h_i}}) \), and the spline smoothing parameters for \( f_i \) and \( f_2 \) are expressed as \( \lambda_i^2 = \sigma^2_e / \sigma^2_{h_i} \) and \( \lambda_i^2 = \sigma^2_e / \sigma^2_{h_2} \).

Estimator for the coefficients of fixed effects as well as coefficients of random effect simultaneously obtained by solving the Henderson mixed model equation. By maximizing the joint density of \( y, h_1, h_2 \), and \( h_3 \) and by assuming the conditional density \( (y|h) \sim N(Xa + Tw + Zh, G_k) \), the Henderson mixed model for this problem is

\[
\begin{bmatrix}
XG_1\hat{X} & XG_1\hat{X} & XG_1\hat{X} & XG_1\hat{X} \\
T'G_1\hat{T} & T'G_1\hat{T} & T'G_1\hat{T} & T'G_1\hat{T} \\
Z_iG_1\hat{X} & Z_iG_1\hat{X} & Z_iG_1\hat{X} & Z_iG_1\hat{X} \\
Z_iG_1\hat{T} & Z_iG_1\hat{T} & Z_iG_1\hat{T} & Z_iG_1\hat{T}
\end{bmatrix}
\begin{bmatrix}
a \\
w \\
h_1 \\
h_2
\end{bmatrix}
= 
\begin{bmatrix}
XG_1\hat{y} \\
T'G_1\hat{y} \\
Z_iG_1\hat{y} \\
Z_iG_1\hat{y}
\end{bmatrix}
\tag{4}
\]

Equation (4) has a solution that can be written as follows,

\[
\left[ \begin{array}{c}
\hat{a} \\
\hat{w} \\
\hat{h}_1 \\
\hat{h}_2
\end{array} \right] = J \left[ (X'G_1\hat{X})' (T'G_1\hat{T})' (Z_iG_1\hat{X})' (Z_iG_1\hat{X})' \right]^{-1}
\left[ (X'G_1\hat{y})' (T'G_1\hat{y})' (Z_iG_1\hat{y})' (Z_iG_1\hat{y})' \right],
\]

where \( J \) is an inverse of the matrix coefficient in (4). This solution requires information on variance values, but often the component variances \( \sigma^2_1, \sigma^2_2, \sigma^2_3, \) and \( \sigma^2_4 \) are unknown so it must be estimated.

3.2. REML estimation of variance components model
The restricted maximum likelihood method estimates variance components based on linear combinations \( K'y \) which do not contain fixed effects \( a \) and \( w \). In other words, \( K'X = 0 \) and \( K'T = 0 \).

We obtained \( K' \) as a matrix with the line formed by matrices \( K' = c'(M - MX(MX)^{-1}XM) \) with \( M = I - (T'T)^{-1}T'c \) and is a random constant vector with the appropriate order. The \( K' \) matrix that satisfies both \( K'X = 0 \) and \( K'T = 0 \) is a new form that we offer for this case. We have also proven that \( K' \) is a symmetric idempotent matrix.

The likelihood function for \( K'y \) can be reduced by assuming \( y \sim N(Xa + Tw, V) \), so \( K'y \sim N(0, K'VK) \) that is

\[
L \approx |K'VK|^{-\frac{1}{2}} e^{-\frac{1}{2}y'[K'(K'VK)^{-1}]K'y}
\tag{5}
\]

By maximizing function log \( L \) on \( \sigma^2_k \) ( \( k = 1, 2, 3, 4 \)) and seeing that \( \frac{\partial V}{\partial \sigma^2_k} = Z_iZ_i' \) for every \( k \) we obtain the equation of REML, that is

\[
\text{trace}(P) = y'P'y, \quad \text{for} \ k = 4
\tag{6}
\]

and

\[
\text{trace}(PZ_kZ_k') = y'PZ_kZ_k'P_y, \quad \text{for} \ k = 1, 2, 3
\tag{7}
\]

Where \( P = (K'(K'VK)^{-1}K')' \). In line with [9], multiplying (7) by \( \sigma^2_k \) then adding to \( k = 1, 2, 3 \) and see the fact that \( \sum_{k=1}^{4} \sigma^2_k Z_kZ_k' = V - \sigma^2_kZ_kZ_k' \) and \( PVP = P \) we obtain \( \text{trace}(PV) = y'P_y \). because
\( K'X = 0, \ K'T = 0 \) with \( K' \) have full row rank and \( V \) is a definite positive matrix, it can be proved that
\[
P = M' - M'X(X'M'X)^{-1}X'M'
\] (8)
Where \( M' = V^{-1} - V^{-1}T(T'V^{-1}T)^{-1}T'V^{-1} \) and \( V^{-1} = G_i^{-1} - G_i^{-1}(Z'G_4^{-1}Z + D^{-1})^{-1}Z'G_4^{-1} \). Furthermore, the matrix \( \hat{\hat{h}} = (Z'G_4^{-1}Z + D^{-1})^{-1}Z'G_4^{-1}(y - \hat{Xa} - \hat{T}\hat{w}) \) \( D = \text{diag}(G_i) \) can be stated, which uses the third, fourth, and fifth rows of equation (4). Subsequently substituted \( \hat{h} \) in the first and second lines of (4) with \( \hat{\hat{h}} \) to get
\[
X'V^{-1}\hat{Xa} + X'V^{-1}\hat{T}\hat{w} = X'V^{-1}y
\] (9)
and
\[
T'V^{-1}\hat{Xa} + T'V^{-1}\hat{T}\hat{w} = T'V^{-1}y
\] (10)
The equation \( VP = y - X\hat{b} - T\hat{w} \) can be obtained using (8), (9), and (10). Together with facts \( PVP = P \) and \( \text{trace}(PV) = \text{rank}(K') \) can then be obtained \( \text{rank}(K') = y'V^{-1}(y - \hat{Xa} - \hat{T}\hat{w}) \). Estimator for error variance obtained by substituting \( V^{-1} \) and \( \hat{\hat{h}} \) to \( V^{-1}(y - \hat{Xa} - \hat{T}\hat{w}) \), that is
\[
\hat{\sigma}_e^2 = y'(y - \hat{Xa} - \hat{T}\hat{w} - Z\hat{h}) / (n - p - \text{trace}(MX'(MX)^{-1}X'M))
\] (11)
Estimator for \( \hat{\sigma}_h^2, \hat{\sigma}_k^2, \) and \( \hat{\sigma}^2 \) is obtained by first substituting \( V \) for \( Z = 0 \) into \( P \) so that it can be stated as \( S = M'' - M''X(X'M''X)^{-1}X'M'' \) to where \( M'' = G_4^{-1} - G_4^{-1}(T'G_4^{-1}T)^{-1}T'G_4^{-1} \). Matrix \( P \) can be restated as
\[
P = S - SZDTZ'S
\] (12)
Where \( T = (I + Z'SZD)^{-1} \). Let \( F_{ik} \) is defined as \( D \) with identity element in \( \hat{\sigma}_i^2 \) and 0 in \( \hat{\sigma}_j^2, j \neq k \), thus
\[
F_{ik} = \frac{DF_{ik}}{\hat{\sigma}_h^2}, \quad k = 1, 2, 3.
\] (13)
Using (12) and (13), the left-hand side of (7) becomes \( \text{trace}(PZ_kZ_k') = \text{trace}((F_{ik} - T\hat{F}_{ik}) / \hat{\sigma}_h^2) \). Because \( DZ'V^{-1} = (D^{-1} + Z'G_4^{-1}Z)^{-1}Z'G_4^{-1} \) we get
\[
DZ'Py = DZ'V^{-1}VPy = \hat{\hat{h}}.
\] (14)
Using (13) and (14), the right side of (7) becomes \( y'PZ_kZ_k'Py = \hat{\hat{h}}_k, \hat{\hat{h}}_k / \hat{\sigma}_h^2 \). Furthermore, from (7) and using \( \text{trace}(PZ_kZ_k') \) we can obtain estimators for variance components, that is
\[
\hat{\sigma}_{h_k}^2 = \hat{\hat{h}}_k / (\text{tr}(F_{ik} - T\hat{F}_{ik})); \quad k = 1, 2, 3.
\]
Together with (11), the estimated values of this variance are determined using the iteration formula
\[
\hat{\sigma}_e^{2(b+1)} = \frac{y'(y - \hat{Xa} - \hat{T}\hat{w} - \hat{Z}\hat{h})}{n - p - \text{trace}(MX'(MX)^{-1}X'M)}
\] (15)
\[
\hat{\sigma}_{h_k}^{2(b+1)} = \frac{\hat{\hat{h}}_k / \hat{\sigma}_h^2}{K_i - \text{tr}(T_{ii})}
\] (16)
\[
\hat{\sigma}_{h_k}^{2(b+1)} = \frac{\hat{\hat{h}}_k / \hat{\sigma}_h^2}{K_2 - \text{tr}(T_{22})}
\] (17)
\[
\sigma_{b_k}^{2(b+1)} = \frac{\hat{\hat{h}}_k / \hat{\sigma}_h^2}{m - \text{tr}(T_{33})}
\] (18)
4. Simulation

We report a simple simulation that used to show the variance components estimator's performance produced using REML method compared to those using a maximum likelihood (ML) method to predict the population area means. Consider a finite population consist of $m = 10$ non-overlapping areas with $N_i = 200$ units within each area $i, (i = 1, \cdots, 10)$. We generate the population $x_{ij}$ values from $x_{ij} \sim iid N(1, 1)$, $t_{ij} \sim iid N(1, 1)$, $u_{ij}$ values from $u_{ij} \sim iid N(0, 0.5^2)$ and $v_{ij}$ values from $v_{ij} \sim N(0, 1)$ for every $i$ and $j (j = 1, \cdots, 200)$. Furthermore, model (1) is used to build the population $y_{ij}$. In this simulation $f_1(x_{ij})$ and $f_2(t_{ij})$ are each approached by first and second orders of p-spline. Knots in $x$ and $t$ are determined using the quantile formula expressed by [8]. The sample is taken by various size $n_i = 15, 30, 40$ and 50 in every area with unequal inclusion probability $\pi_i = n_i b_{ij} / \sum_j b_{ij}$ where $b_{ij} = (\exp(1/3 (-u_{ij} + e_{ij}) / \sqrt{0.5}) + \delta_{ij} / 5)$ and with $\delta_{ij} \sim iid N(0, 1)$. The variance component is calculated using formulas (15)-(18) with iterations. The prediction value of mean areas was counted using $\hat{\mu}_i = \frac{1}{N_i} \left\{ \sum_{i \in \bar{s}_i} y_{ij} + \sum_{i \in \bar{s}_i} \hat{y}_{ij} \right\}$, where $\bar{s}_i$ is a non sampled unit set in the area $i$ and $\hat{y}_{ij}$ is a predicted value for $y_{ij}$ using the proposed approach. We also calculate the mean square prediction error (MSPE) and the absolute bias using bootstrap with 1000 replications.

4.1. Simulation Results and Discussion

Figure 1 and Figure 2 give a distribution of MSPE values generated by the model with the first and second order p-spline approach. In both images, the distribution of MSPE values generated by the model whose variance components are estimated using the ML method is more diverse than that produced by the REML method. The median value of MSPE produced by the REML approach for the variance component in both images smaller than that given by the ML method except for cases sample $n = 40$.

Figure 1. The distribution of MSPE values for each sample size uses a prediction model with a first-order p-spline approach

Figure 2. The distribution of MSPE values for each sample size uses a prediction model with a second-order p-spline approach
With a first-order p-spline approach that looks slightly higher. However, ML produces a further outer. This result shows that REML estimation for the model variance component tends to produce a better accuracy in estimating the mean area for the sample size being tested. The average of RMSPE and the average of absolute bias generated by the prediction model (3) with the first or second order of p-spline with the components of variance estimated to use REML tend to be smaller than those produced by ML. However, increasing the sample size in this simulation does not guarantee a decrease in the average of RMSE value or the average of absolute bias obtained. This can be seen in Table 1, that provides a comparison of the results of the RMSPE values and the average absolute bias values for each sample size.

Table 1. Average of RMSPE and average of absolute bias values for REML and ML method estimation of component variances for four sample sizes and each spline approach. For each sample, the first row for the first order spline approach and the second row for the second order.

| Sample | ML method | REML method |
|--------|-----------|-------------|
|        | Average of RMSPE | Average of absolute bias | Average of RMSPE | Average of absolute bias |
| 15     | 0.143348994 | 0.14325143 | 0.128248883 | 0.128121432 |
| 30     | 0.164623381 | 0.16449694 | 0.140997652 | 0.140833432 |
| 40     | 0.133208009 | 0.133136059 | 0.078665079 | 0.077580382 |
| 50     | 0.131419214 | 0.131279763 | 0.053068006 | 0.051640645 |

Studies are comparing the performance of REML and ML as well as the use of the p-spline approach in a mixed model framework is not new [2, 5, 6]. However, we have discovered the matrix $K'$ that meets the requirements to obtain the REML estimator and its properties so that it gets the REML estimator explicitly for the variance component in the mixed model involving two additive p-spline functions in the model.

5. Conclusion

We have obtained the REML estimator for the mixed-model component variances involving two p-spline additive functions and effect area random. Next, we use the model to estimate the average area of the population. The limited simulation results show the performance of the p-spline mixed model with the REML estimator for the variance component, which provides better performance in estimating the average area compared to the results obtained using the ML estimator.

References

[1] Aydin D and Memmedli M 2012 Optimization 61 459
[2] Cao Y, Lin H, Wu T Z and Yu Y 2010 Comput. Stat. Data Anal. 54 891
[3] Chen H and Wang Y 2011 Biometrics 67 861
[4] Maiti T, Sinha S and Zhong P S 2016 Scand. J. Stat. 43 886
[5] Nancy H, Richard L and Jason D N 2013 Electron J. Stat. 71 517
[6] Wu C, Marcia L G and Dennis D 2001 Am. Stat. 55 125
[7] Rao J N K, Sinha S K and Dumitrescu L 2014 Can. J. Stat. 42 126
[8] Ruppert D, Wand M P and Carroll, R J 2003 Semiparametric Regression (New York: Cambridge University Press)
[9] Searle S R, Casella G and McCulloch C E 1992 Variance Components (New York: Wiley)
[10] Wood S N, Pya Nand Säfken B 2016 J. Am. Stat. Assoc 111 1548