ANADOLU ÜNİVERSİTESİ

Bilim ve Teknoloji Dergisi B-Teorik Bilimler
Cilt: 2 Sayı: 2 2013
Sayfa:91-102

ARAŞTIRMA MAKALESİ / **RESEARCH ARTICLE**

Özlem AKSOY [1], Dursun AYDIN [2]

ESTIMATION AND INFERENCES IN LINEAR MIXED EFFECTS MODELS: A COMPARATIVE STUDY

ABSTRACT

Smoothing methods that use basis functions with penalization can be formulated as fits in form linear mixed effects models. This allows such models to be fitted using standard mixed models structures. In this paper we provide an estimation and inference for linear mixed models using restricted maximum likelihood and penalized spline smoothing, and describe the connection between the two. To this end, a real data example is considered and model is fitted in R using different package. We see that penalized spline smoothing expressed in form of linear mixed model gives the better results than standard mixed effects model.

Keywords: Mixed effect model, Semi-parametric regression, Penalized spline, Smoothing parameter, Generalized cross validation

DOĞRUSAL KARMA ETKİLİ MODELLERDE TAHMİN VE ÇIKARSAMALAR: BİR KARŞLAŞTIRMALI ÇALIŞMA

ÖZ

Cezalandırma ile taban fonksiyonları kullanan düzeltme yöntemleri doğrusal karma etkili modellerde uyumlar olarak formüle edilebilir. Bu durum standart karma model yapılarını kullanarak bu tür modellerin tahmin edilmesine olanak sağlar. Bu çalışmada doğrusal karma etkili modeller için sınırlı maksimum olabilirlik ve cezalı splayn düzeltme yöntemi kullanarak tahmin ve çıkarsama sağlanmakta ve bu iki yöntem arasında bağlantı tanımlanmaktadır. Bu amaçla bir gerçek örneklem farklı paketler kullanarak R ortamında tahmin edilmiştir. Görülüyor ki, doğrusal karma etkili model şeklinde formüle edilen cezalı splayn düzeltme standart karma etkili modelden daha iyi sonuçlar vermektedir.

Anahtar Kelimeler: Karma etkili model, Yarı-parametrik regresyon, Cezalı splayn, düzeltme parametresi, Genelleştirilmiş çapraz geçerlilik

[1,] Faculty of Medicine, Department of Biostatistics, University of Dumlupınar, 43100, Kütahya, Turkey.
Tel: 0506 473 18 73, E-mail: ozlemaksoy@mu.edu.tr

[2,] Faculty of Science, Department of Statistics, University of Muğla, 48000, Muğla Sıtkı Koçman, Turkey.
Tel:, 0533 023 04 65, E-mail: duaydin@mu.edu.tr

Recieved: 28 November 2012 **Revised:** 21 March 2013 **Accepted:** 17 June 2013

1. INTRODUCTION

Many common statistical models can be expressed as linear models that incorporate both fixed effects, which are parameters associated with an entire population or with certain repeatable levels of experimental factors and random effects, which are associated with individual experimental units drawn at random from a population. A model with both fixed effects and random effects is called a mixed-effects model. In order to fit linear mixed-effects models, such as maximum likelihood (ML) and restricted maximum likelihood (REML) method, standart methods, can be used (Pinheiro and M.Bates, 2000).

Linear mixed effects models are powerful and useful approaches to many applications. They have received considerable attention from both the theoretical and applied points of view. Much of work on linear mixed effects models was motivated by the analysis of data animal breeding experiments and driven by the need to incorporate heritability and generic correlations in a parsimonious fashion. They have also played an important role in establishing quality control procedures and determination of sampling, among other applications (Davidian and Giltinan, 1996, McCulloch and Searle, 2001, Verbeke and Molenberghs, 2000, Vonesh and Chinchili, 1997). Another note worthy reference is (Diggle, Heagerty, Liang and Zeger, 2002) in which the commonalities between longitudinal data analysis and spatial statistics are observed. Several authors have pointed out a close relationship between the mixed model and penalized splines (P-splines).

Eilers and Marx (Eilers and Marx, 1996) introduced smoothing with P-splines, extending an original idea of O'Sullivan (O'Sullivan, 1986). This powerful and applicable technique based on the minimization problem of penalized residuals sum of squares have gained much popularity as a flexible tool for smoothing and nonparametric models. Moreover, P-spline smoothing using truncated power basis functions can be easily extended to a linear mixed effects model by treating the basis functions as random variables. Here we refer to (Ruppert, Wand and Carroll, 2003) for P-splines using truncated power functions, knots based on quantiles of the independent variable and a penalty parameter. The ability to combine nonparametric regression and mixed model regression with P-splines has recently been used in other contexts. Parise et al. (Parise, Wand, Ruppert and Ryan ,2001) Coull et al. Provide examples of using P-splines in the construction of mixed effect regression models for the analysis of data containing random effects. In a recent book, Ruppert et al. present an excellent overview of theory and applications of semi-parametric models based on P-splines. However, they used different ingredients: truncated power functions in the basis, knots at quantiles of the independent variable, and a ridge penalty. This might be confusing to potential users: which type of splines should be used, how should knots be spaced and how should a smoothing parameter be chosen.

This paper is organized as follows. In section 2, we give a brief summary of the estimation based on REML for linear mixed effect models. Estimation procedur in linear mixed effect models are given in Section 3. Section 4 compares these methods via a real example, and finally, the conclusion and recommendations are presented in section 5.

2. LINEAR MIXED EFFECT MODELS

Many common statistical models can be expressed as linear models that incorporate both *fixed effects*, which are parameters associated with an entire population or with certain repeatable levels of experimental factors, and *random effects*, which are associated with individual experimental units drawn at random from a population. A model with both fixed effects ana random effects is called a *mixed-effects* model (see, West et al., 2007; Brown ana Prescott, 2006).

Using the hierarchal notation of Laird and Ware (1982), we can express the linear mixed effects (LME) model as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{u}_i + \boldsymbol{\varepsilon}_i, \;\; i = 1, 2, ..., m \;, \tag{1}$$

where;

- $\mathbf{y}_i$ is the $n_i \times 1$ response vector for observations in the i th group.
- $\mathbf{X}_i$ is the $n_i \times p$ model matrix for the fixed effects for observations in group i.
- $\boldsymbol{\beta}$ is the $p \times 1$ vector of fixed-effect coefficients.
- $\mathbf{Z}_\mathrm{i}$ is the $n_i \times q$ model matrix for the random effects for observations in group i.
- $\boldsymbol{u}_i$ is the $q \times 1$ vector of random-effect coefficients for group i.
- $\boldsymbol{\varepsilon}_i$ is the $n_i \times 1$ vector of errors for observations in group i.
- $\mathbf{D}$ is the $q \times q$ covariance matrix for the random effects.
- $\mathbf{R}_i$ is the $n_i \times n_i$ covariance matrix for the errors in group i.

We assume taht the q random effects in the $\boldsymbol{u}_i$ vector follow a multivariate normal distribution, with mean vector **0** and a variance –covariance matrix indicated by $\mathbf{D}$:

$$\boldsymbol{u}_i \sim \mathrm{N}(0, \mathbf{D}). \tag{2}$$

The main diagonal elements of the $\mathbf{D}$ matrix represent the variances of each random effect in $\boldsymbol{u}_i$, and the off-diagonal elements represent the covariances between two corresponding random effects. Elements of symmetric and positive definite matrix $\mathbf{D}$ are defined as following way (West et al., 2007):

$$\mathbf{D} = Var(\boldsymbol{u}_i) = \begin{pmatrix} Var(u_{1i}) & cov(u_{1i}, u_{2i}) & \cdots & cov(u_{1i}, u_{qi}) \\ cov(u_{1i}, u_{2i}) & Var(u_{2i}) & & cov(u_{2i}, u_{qi}) \\ \vdots & \vdots & \ddots & \vdots \\ cov(u_{1i}, u_{qi}) & cov(u_{2i}, u_{qi}) & \cdots & Var(u_{qi}) \end{pmatrix}_{q \times q}.$$

The variance and covariances elements of the $\mathbf{D}$ matrix are defined as functions of a small set of covariance parameters stored in a vector denoted by $\boldsymbol{\theta_D}$. Note that the vector $\boldsymbol{\theta_D}$ imposes structure (or constraints) on the elements of the $\mathbf{D}$ matrix. In this case, the vector $\boldsymbol{\theta_D}$ contains two parameters:

$$\boldsymbol{\theta}_\mathrm{D}{}' = \begin{pmatrix} \sigma^2_{\mathrm{u}1} & \sigma^2_{\mathrm{u}2} \end{pmatrix}.$$

Finally, the $\boldsymbol{\varepsilon}_i = \begin{pmatrix} \varepsilon_{1i} & \varepsilon_{2i} & \cdots & \varepsilon_{n_i i} \end{pmatrix}'$ vector in Eq. (1) is a vector of n_i residuals. We assume that the $\boldsymbol{\varepsilon}_i$ vector is random variables that follow a multivariate normal distribution with a mean vector **0** and a positive definite symmetric covariance matrix $\mathbf{R}_i$:

$$\boldsymbol{\varepsilon}_i \sim \mathrm{N}(0, \mathbf{R}_i). \tag{3}$$

We also assume that residuals associated with different subjects are independent of each other. Further, we assume that the vectors of residuals, $\varepsilon_1, ..., \varepsilon_\mathrm{m}$, and random effects, $u_1, ..., u_\mathrm{m}$ are independent of each other.

We discuss some of the more commonly used covariance structures for the $\mathbf{R}_i$ matrix.The simplest covariance matrix for $\mathbf{R}_i$ is the diagonal structure,in which the residuals associated with observations on the same subject are assumed to be uncorrelated and to have equal variance. The diagonal $\mathbf{R}_i$ matrix for each subject i has the following structure.

$$\mathbf{R}_i = Var(\varepsilon_i) = \begin{pmatrix} Var(\varepsilon_{1i}) & cov(\varepsilon_{1i},\varepsilon_{2i}) & \cdots & cov(\varepsilon_{1i},\varepsilon_{n_i i}) \\ cov(\varepsilon_{1i},\varepsilon_{2i}) & Var(\varepsilon_{2i}) & & cov(\varepsilon_{2i},\varepsilon_{n_i i}) \\ \vdots & \vdots & \ddots & \vdots \\ cov(\varepsilon_{1i},\varepsilon_{n_i i}) & cov(\varepsilon_{2i},\varepsilon_{n_i i}) & \cdots & Var(\varepsilon_{n_i i}) \end{pmatrix} = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix}.$$

The diagonal structure requires one parameter in $\boldsymbol{\theta}_{\mathrm{R}}$, which defines the constant variance at each time point: $\boldsymbol{\theta}_{\mathrm{R}} = (\sigma^2)$ All software procedures that we discuss use the diagonal structure as the default structure for the $\mathbf{R}_i$ matrix.

The compound symmetry structure is frequently used for the $\mathbf{R}_i$ matrix. In the compound symmetry covariance structure, there are two parameters in the $\boldsymbol{\theta}_{\mathrm{R}}$ vector that define the variances and covariances in the R_i matrix:

$$\boldsymbol{\theta}_{\mathrm{R}}' = \begin{pmatrix} \sigma^2 & \sigma_1 \end{pmatrix}.$$

The first-order autoregressice structure, denoted by AR(1), is another commonly used covariance structure for the $\mathbf{R}_i$ matrix. The AR(1) structure has only two parameters in the $\boldsymbol{\theta}_{\mathrm{R}}$ vector that define all the variances and covariances in the $\mathbf{R}_i$ matrix: a variance parameter, σ^2, and a correlation parameter, ρ.

$$\boldsymbol{\theta}_{\mathrm{R}}' = \begin{pmatrix} \sigma^2 & \rho \end{pmatrix}.$$

Note that σ^2 must be positive, whereas ρ can range from -1 to 1.The AR(1) structure is often used to fit models to data sets with equally spaced longitudinal observations on the same units of analysis. This structure implies that observations closer to each other in time exhibit higher correlation than observations farther apart in time (West et al., 2007).

3. ESTIMATION IN LINEAR MIXED EFFECT MODEL

In the LME, we estimate the fixed parameters vector, $\boldsymbol{\beta}$, and the covariance parameters, $\boldsymbol{\theta}$ (i.e., $\boldsymbol{\theta}_{\mathrm{D}}$ and $\boldsymbol{\theta}_{\mathrm{R}}$ for the **D** and $\mathbf{R}_{\mathrm{i}}$ matrices, respectively). In this section, we discuss restricted maximum likelihood (REML) estimations and penalized smoothing splines, which are methods commonly used to estimate these parameters.

3.1. Restricted Maximum Likelihood (REML) Estimation

REML estimation is an alternative way of estimating the covariance parameters in $\boldsymbol{\theta}$. REML estimation, which is also sometimes called as called residual maximum likelihood estimation, was introduced by Patterson and Thompson (1971). Alternative and more general derivations of REML are given by Harville (1977), Cooper and Thompson (1977), and Verbyla (1990).

The REML estimates of $\boldsymbol{\theta}$ are based on optimization of the following REML log-likelihood function:

$$\begin{aligned} \mathrm{L}_{REML}(\boldsymbol{\theta}) &= -0.5(n-p)ln(2\pi) - 0.5\sum_i \ln\left[\det\left(\boldsymbol{V}_i\right)\right] \\ &- 0.5\sum_i\left(\boldsymbol{r}_i'\boldsymbol{V}_i^{-1}\boldsymbol{r}_i\right) - 0.5\sum_i \ln\left[\det\left(\boldsymbol{X}_i'\boldsymbol{V}_i^{-1}\boldsymbol{X}_i\right)\right] \end{aligned} \quad (4)$$

In Eq. (4) $\boldsymbol{r}_i$ and $\hat{\boldsymbol{V}}_i$ (an estimate of $\boldsymbol{V}_i$) can be defined as follows, respectively;

$$\boldsymbol{r}_i = \boldsymbol{y}_i - \boldsymbol{X}_i\left[\left(\sum_i \boldsymbol{X}_i'\boldsymbol{V}_i^{-1}\boldsymbol{X}_i\right)^{-1}\sum_i \boldsymbol{X}_i'\boldsymbol{V}_i^{-1}\boldsymbol{Y}_i\right], \quad (5)$$

$$\hat{\boldsymbol{V}}_i = \boldsymbol{Z}_i\hat{\boldsymbol{D}}\boldsymbol{Z}_i' + \hat{\boldsymbol{R}}_i' \quad . \quad (6)$$

Once the $\hat{V}_i$ matrix has been obtained, REML based on the $\hat{\beta}$ fixed-effects parameter esmations and $Var\left(\hat{\beta}\right)$ can be computed as following way:

$$\hat{\beta} = \left(\sum_i \boldsymbol{X}_i'\hat{\boldsymbol{V}}_i^{-1}\boldsymbol{X}_i\right)^{-1}\sum_i \boldsymbol{X}_i'\hat{\boldsymbol{V}}_i^{-1}\boldsymbol{y}_i , \quad (7)$$

$$Var\left(\hat{\beta}\right) = \left(\sum_i \boldsymbol{X}_i'\hat{\boldsymbol{V}}_i^{-1}\boldsymbol{X}_i\right)^{-1} . \quad (8)$$

3.2. Penalized Smoothing Spline

Suppose observed are n pairs of measurements, $(x_i, y_i), i = 1, 2, \ldots, n$ satisfying the model

$$y_i = f\left(x_i\right) + \varepsilon_i , \quad (9)$$

where f is unknown regression function and $\varepsilon_1, \ldots, \varepsilon_n$ are independent errors random variables with common mean zero and variance σ_S^2.Let $t_1 < \ldots < t_K$ be a set of fixed knots $min(x_i) \leq t_1 < \ldots < t_K \leq \max(x_i)$ and let $(x)_+ = \max\left(0, x\right)$ It is assumed that $f(x)$ can be well approximated by a pth-degree P-spline with truncated polynomial basis

$$f(x) = \beta_0 + \beta_1 x + \ldots + \beta_p x^p + \sum_{\mathrm{j}=1}^{\mathrm{K}} \mathrm{u}_{\mathrm{j}}(\mathrm{x} - \mathrm{t}_{\mathrm{j}})_+^{\mathrm{p}} , \quad (10)$$

where p≥1 is an integer, $\boldsymbol{\beta} = \left[\beta_0, \beta_1, \ldots, \beta_p\right]^T$ and $u = \left[u_1, \ldots, u_K\right]^T$ are vectors of regression coefficients fort the parametric and spline portions of the model, respectively. Note that $\hat{f}(x)$ is a linear combination of the set of functions $1, x, \ldots, x^p, (x - t_1)_+^p, \ldots, (x - t_K)_+^p$, known as the truncated power basis of degree p with K knots $t_1, t_2, \ldots, t_K$. The number of knots, K must be selected in

implementing the regression spline. A reasonable default rule for the knot locations is $t_j = \left\{(j+1)/(K+2)\right\}$ th sample quantile of the unique x_i s for j=1,…, K

To give an explicit expression for $f(x)$ in matrix notation, let denote the design matrices

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & \dots & x_1^p \\ 1 & x_2 & \dots & x_2^p \\ . & . & . & . \\ 1 & x_n & \dots & x_n^p \end{bmatrix}, \text{ and } \mathbf{Z} = \begin{bmatrix} (x_1-\kappa_1)_+^p & \dots & (x_1-\kappa_K)_+^p \\ (x_2-\kappa_1)_+^p & & (x_2-\kappa_K)_+^p \\ . & \dots & . \\ (x_n-\kappa_1)_+^p & \dots & (x_n-\kappa_K)_+^p \end{bmatrix},$$

then model (10) can be expressed as a linear mixed effects model

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \boldsymbol{\varepsilon}, \quad \text{where } \begin{bmatrix} \mathbf{u} \\ \boldsymbol{\varepsilon} \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 \mathbf{I} & 0 \\ 0 & \sigma_\varepsilon^2 \mathbf{I} \end{bmatrix} \right). \tag{11}$$

where $\mathbf{y}$ is a vector of observed responses, and $\mathbf{X}$ and $\mathbf{Z}$ are design matrices associated with a vector of fixed effects $\boldsymbol{\beta}$ and a vector of random effects $\mathbf{u}$, respectively. The connection between penalized regression and linear mixed effect models can be determined by considering that the estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{u}}$ minimize the penalized least squares

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \arg\min_{\boldsymbol{\beta},\mathbf{u}} \left(\left\| \mathbf{y} - \mathbf{X\beta} - \mathbf{Zu} \right\|^2 + \lambda \left\| \mathbf{u} \right\|^2 \right), \tag{12}$$

where $\|\mathbf{u}\|$ is the Euclidean norm of the vector $\mathbf{u}$, $\lambda > 0$ denotes smoothing parameter. In this paper, the smoothing parameter λ is selected by minimizing the function generalized cross-validation (GCV) (Eubank, 1999; Green and Silverman, 1994).

The likelihood approaches such as maximum likelihood (ML) and restricted maximum likelihood (REML) based smoothing parameter selection methods depend on the linear mixed model representation of penalized spline.The minimization of the residual sum of the squares $\|\mathbf{y} - \mathbf{X\beta} - \mathbf{Zu}\|^2$ in Eq. (12) is subject to the penalty $\lambda \|\mathbf{u}\|^2$. The solution for $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{u}}$ that minimizes the penalized least squares in (12) can be defined as (Wand, 2003)

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \left(\mathbf{C}^{\mathrm{T}}\mathbf{C} + \lambda \mathbf{D} \right)^{-1} \mathbf{C}^{\mathrm{T}}\mathbf{y}, \tag{13}$$

where $\mathbf{C} = [\mathbf{X} \ \ \mathbf{Z}]$, $\mathbf{D} = diag(0_m, 1, \dots, 1)$, the vector 0_m denoting the $p+1$-dimensional zero vector where m is the dimension of the vector $\boldsymbol{\beta}$ of fixed regression coefficients.

Equation (11) is also recognizable as linear mixed effect models and the best linear unbiased estimators (BLUEs) of y [27]:

$$\hat{\mathbf{f}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}}, \tag{14}$$

where the estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{u}}$ can be treated as an estimator of $\mathbf{f} = \left[f(x_1), \dots, f(x_n) \right]^{\mathrm{T}}$. Note that the fitted values $\hat{f}$ can be expressed as

$$\hat{\mathbf{f}}_\lambda = \mathbf{C}\left(\mathbf{C}^{\mathrm{T}}\mathbf{C} + \lambda \mathbf{D} \right)^{\mathrm{T}} \mathbf{C}^{\mathrm{T}}\mathbf{y}. \tag{15}$$

4. CASE STUDY

In this sutudy, the sample data are analyzed with normal linear mixed effects models and penalized spline model separately. Within the mixed models implementation of penalized splines the amount of smoothing is controlled by the relative magnitude of the relevant variance component and the residual error variance. Typically this parameter is estimated from the data, and in a real example in the text the amount of smoothing was determined by the restricted maximum likelihood estimation (REML) estimates of the variance parameters.

Data set used in this case study is taken from Organisation for Economic Co-operation and Development (OECD)'s, and given in Table 1. Data metioned here covers nominal gross domestic product (ngdp) of 34 countries for different years, from 2006 to 2010. All coputations are calculated by using R.2.13 software.

Table 1: OECD Data set for case study

Country	Year	NGDP	t
Australia	2006	3,5	1
⋮	⋮	⋮	⋮
Australia	2010	2,9	5
⋮	⋮	⋮	⋮
United States	2006	2,9	166
⋮	⋮	⋮	⋮
United States	2010	1,6	170

Countries: Australia, Austria, Belgium, Canada, Chile, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Kore, Luxemburg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom, United States

NGDP: Nominal Gross Domestic Product .

4.1 Empirical Results

Model (1) is easily fitted by REML method for data set called as "Özlem". This estimation method for the parameters in LME models is described in detail in section 3. We obtain the REML fit of the model given (1) as following way, Outcomes obtained by mixed effects model using REML are given following way:

```
Data: Özlem
       AIC      BIC    logLik
  725.3883 744.1321 -356.6942


Random effects:
             StdDev       Corr
(Intercept)  1.658442762  Intr)
 time        0.001295467  0.002
 Residual    1.638233543
```

```
 Fixed effects: ngdp ~ time
                Value Std.Error  DF   t-value p-value
(Intercept)  3.524118 0.4095414 135  8.605033  0.0000
time        -0.238235 0.0888460 135 -2.681442  0.0082

Correlation:
     (Intr)
time -0.651
Number of Observations: 170
Number of Groups: 34
```

We see that the REML estimates for the parameters have been calculated as

β_0=3.524118, β_1=-0.238235 and σ_u=1.658453, σ_e = 1.638235
corresponding to a log-restricted-likelihood of -356.6942.

The output of the summary function includes the values of the Akaike Information Criterion (AIC) (Sakamoto, Ishiguro and Kitagawa, 1986) and Bayesian Information Criterion (BIC) (Schwarz, 1978), which is also sometime called as Schwarz's Bayesian Criterion (SBC). These are model comparison criteria evaluated as

$$\text{AIC} = -2 \log \text{Lik} + 2\text{npar},$$
$$\text{BIC} = -2 \log \text{Lik} + \text{npar} \log(N).$$

where npar indicates the number of parameter in the model and N total number of observations used to fit the model. Under these definitions, "smaller is better." That is, if we are using AIC to compare two or more models for the same data, we prefer the model with the lowest AIC. Similarly, when using BIC we prefer the model with the lowest BIC.

We should examine the fitted model both graphically and numerically. The 95% confidence intervals provides an indication of the precision of the estimates of the variance components

```
Approximate 95% confidence intervals

Fixed effects:
                 lower       est.       upper
(Intercept)  2.7141678   3.5241176   4.33406746
     t      -0.4139449  -0.2382353  -0.06252569
attr(,"label")
[1] "Fixed effects:"

Random Effects:
Level: goverment
                   lower     est.    upper
sd((Intercept)) 1.241933 1.658453 2.214667

Within-group standard error:
   lower     est.     upper
1.454038  1.638235  1.845766
```

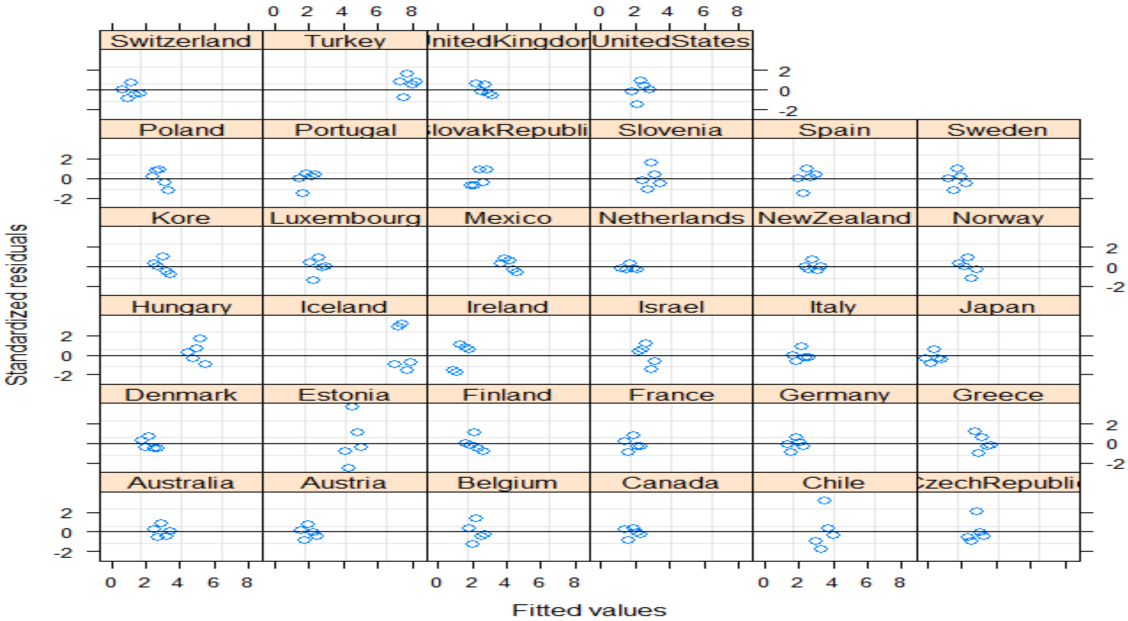

Figure 1: Plots of standardized residuals versus fitted values for each country

We see that σ is estimated relatively precisely. Further more, the plot of the standardized residuals versus the fitted values, shown in Figure 1, does not indicate a violation of the assumption of constant variance for the ε_i error terms.

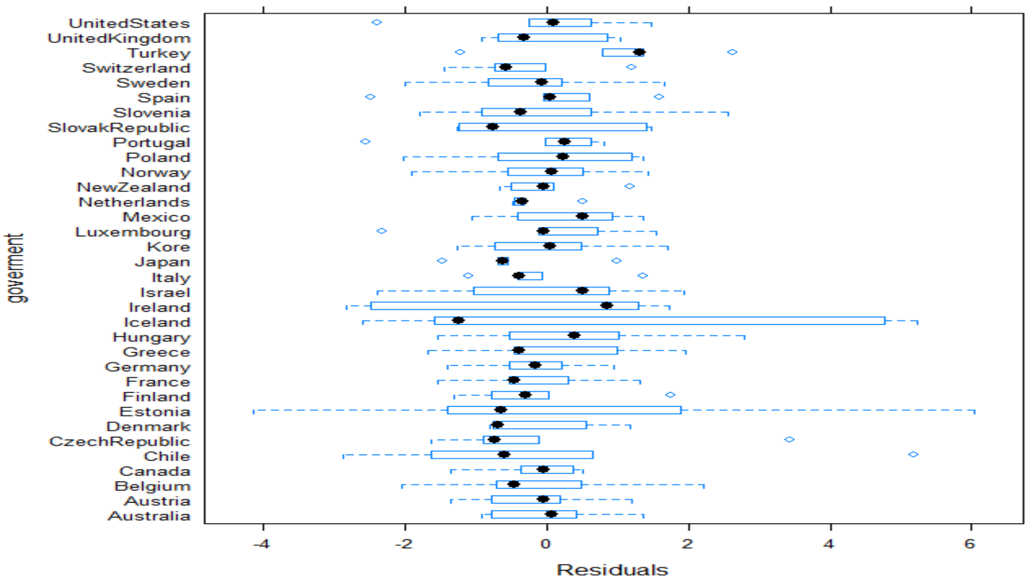

Figure 2: Boxplots of the residuals for each country

Figure 2 shows the boxplots of the residuals for each country. According to Fig. 2, it can be said that model is adequate for the data set called as "Özlem". A boxplots of the country against residuals may also reveal one or more unusually large residuals. These points maybe potential outliers. In other words, the values of some residuals depict an unusual structure. However, since the residuals are now centered around zero, this plot do not exhibit any strong unusual pattern, although the large residuals e_{Chile}, e_{Turkey}, and $e_{\text{CzechRepublic}}$ show up clearly.

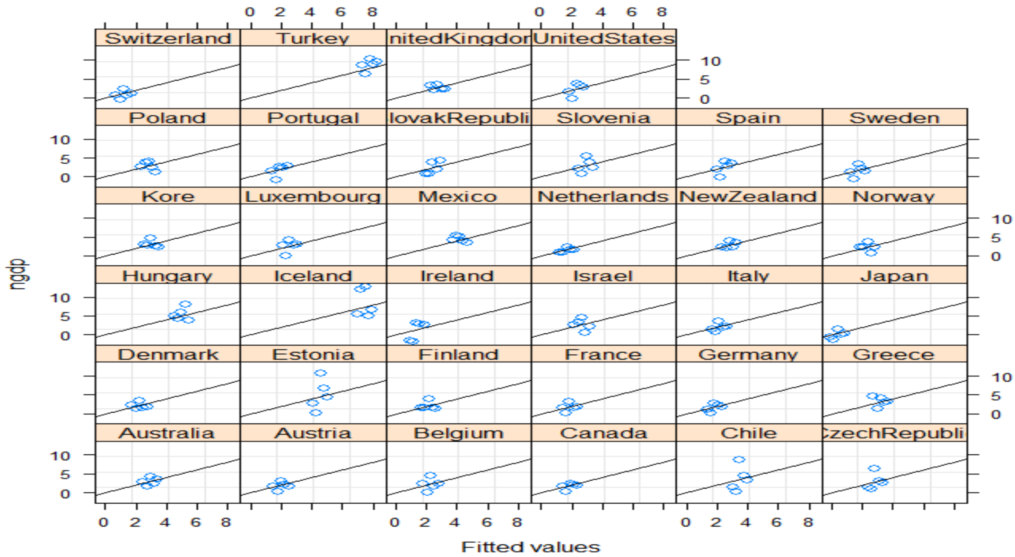

Figure 3: Plots of ngdp versus fitted values for each country

Figure 3 contains plots of ngdp against time for each country with the straight line fits from model(1) included. Once again these plots have been ordered from bottom left to top right in terms of increasing average value of ngdp. As can be seen from Fig.3 estimated random intercept for Japan is lower than others.

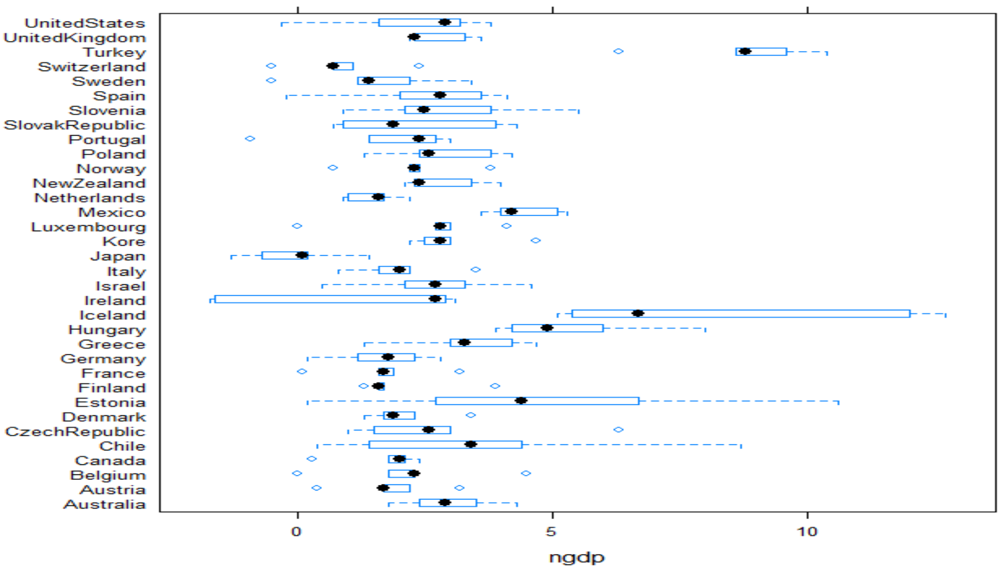

Figure 4: Boxplots of the ngdp for each country

Boxplots of the ngdp for each country are indicated in Figure 4. As shown Figure 4, it can be said that the ngdp values of Turkey is demonstrated much more different behaviour than other countries. We would examine residual plots such as Figure 2 for deficiencies in the model. There are no alarming patterns in this figure.

Model (1) is also fitted by penalized least squares in form of linear mixed effect models for the same data set. This estimation method for the parameters in LME models is described in detail in section 3. The penalizd spline fits are obtained by using R Software. Outcomes obtained by mixed effects model using penalized splines are given following way:

```
Model(1): ngdp ~ time + s(goverment), correlation = corAR1 (form=~1|time),
method="REML")
Parametric coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.5373     0.4137   8.551 1.73e-14
time         -0.2429     0.0792  -3.067  0.00259
Approximate significance of smooth terms:
                edf Ref.df     F  p-value
s(goverment) 25.54  25.54 2.737 8.84e-05
R-sq.(adj) = 0.518  Scale est. = 2.6753    n = 170

Approximate 95% confidence intervals for variance-covariance

Random Effects:
Level: goverment
                lower     est.    upper
sd(Xr - 1) 8.079208 14.18957 24.88740
Correlation structure:
        lower      est.     upper
Phi 0.1767087 0.3418008 0.4881868
Within-group standard error:
   lower     est.    upper
1.432647 1.635636 1.867387
```

We see that the penalized spline estimates for the parameters have been calculated as.

$$\beta_0 = 3.5373, \beta_1 = -0.2429$$

It can be seen that fitted values carried out from the model (1) are well enough and significant. Moreover, 51.8 % of variability in the ngdp explained by the regressor $\mathbf{X}$. Besides, 95 % confidence interval of variance parameters is shown from outputs of the summary. The confidence interval for correlation parameter, ρ is easily picked out, and provides strong evidence the AR1 model is preferable to and independence model (ρ=0), while interval for σ is ($1.432647 \le \sigma \le 1.867387$). Note that, for smoothing parameter is also available. Under the random effects heading the interval for government relates to the smoothing parameter for the smooth term.

5. CONCLUSION AND RECOMMENDATIONS

Notice that the restricted maximum likelihood estimate of σ is 1.638, the same as the penalized spline estimate. Equality of the restricted maximum likelihood and penalized spline estimates of σ occurs for this simple model, but will not occur in general. The penalized spline estimate of σ_u, 1.635, is smaller than the restricted maximum likelihood estimate, 1.638. Finally the restricted maximum likelihood estimate of, β_0 and β_1, is the same as the penalized estimate. Again, exact equality of the restricted maximum likelihood and penalized estimates of the fixed effects need not occur in more complex models, but it is commonplace for them to be nearly identical. However, it seems that penalized spline has provided an improvement in variances of subjects, and that model has given better fits than standard mixed effects model.

REFERENCES

Coull, B. A., Ruppert, D. and Wand, M.P. (2001). Simple Incorporation of Interactions into Additive Models. *Biometrics* 57, 539–545.

Coull, B.A. , Schwartz, J. and Wand, M.P. (2001). Respiratory Health and Air Pollution: Additive Mixed Model Analysis. *Biostatistics 2* , 337–349.

Davidian, M and Giltinan, D.M. (1996). *Nonlinear Models for Repeated Measurement Data,* Chapman and Hall, London.

Diggle, P., Heagerty, P., Liang, K.-L., and Zeger, S. (2002). *Analysis of Longitudinal Data. 2nd ed*, Oxford University Press.

Eilers, P.H.C and Marx, B.D. (1996). Flexible Smoothing with B-Splines and Penalties. *Statistical Science.* 11.

Green P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Model*, Chapman &Hall, London

McCulloch, C.E and Searle, S.R. (2001). *Generalized, Linear, and Mixed Models. Willey*, New York.

O'Sullivan, F. (1986). A Statistical Perspective on Ill-Posed Inverse Problems (C/R: P519-527). *Statistical Science* 1 , 502–518.

Parise, H., Wand, M.P. , Ruppert, D. and Ryan, L. (2001). Incorporation of Historical Controls using Semiparametric Mixed Models. *Journal of the Royal Statistical Society: Series C Applied Statistics, 50,* 31–42.

Pinheiro, J.C. and Bates, D.M. (2000). *Mixed Effects Models in S and S-Plus,* Springer-Verlag, New York.

Ruppert, D., Wand, M. and Carroll, R. (2003). *Semiparametric Regression*, Cambridge University Press, Cambridge.

Shively, T. S., Kohn, R. And Wood, S. (1999). Variable Selection and Function Estimation in Additive Nonparametric Regression using A Data-Based Prior (with discussion). *Journal of the American Statistical Association*, 94, 777–806.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer-Verlag, New York.

Vonesh, E.F. and Chinchili, V.M. (1997). *Linear and Nonlinear Models for The Analysis of Repeated Measurements*. Marcel Dekker, New York.

Wand, M. P. (2003). Smoothing and Mixed Models. *Computational Statistics*, 18, 223–249.

Wood, S.N. (2011). Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models. *Journal of the Royal Statistical Society (B),* 73 (1):3-36.

Zeger, S.L. and Diggle, P.L. (1994). Semiparametric Models for Longitudinal Data with Application to CD4 Cell Numbers in HIV Seroconverters. *Biometricis* 50, 689–699.

Zhang, D. , Lin, X., Raz, J. and Sowers, M.(1998). Semi-Parametric Stochastic Mixed Models for Longitudinal Data. *Journal of the American Statistical Association*, 93, 710–719.