Thailand Statistician
October 2025; 23(4): 839-849
http://statassoc.or.th
Contributed paper

Parameter Estimation for Generalized Random Coefficient in the Linear Mixed Models

Yassine Ou Larbi [a]*, Rachid El Halimi [a], Abdelhadi Akharif [a], and Amal Mellouk [b]

[a] Laboratoire de Mathmatiques et Applications, Facult des Sciences et Techniques, Abdelmalek Essadi University, Tangier, Morocco

[b] Centre Rgional des Mtiers de lEducation et de la Formation, Tangier, Morocco

*Corresponding author; e-mail: yassineoularbi2@gmail.com

Received: 28 November 2022
Revised: 18 October 2023
Accepted: 10 November 2023

Abstract

The analysis of longitudinal data, comprising repeated measurements of the same individuals over time, requires models with a random effects because traditional linear regression is not suitable and makes the strong assumption that the measurements are independent, which is often unrealistic for longitudinal data. However, values repeatedly measured in the same individual are usually correlated, and ignoring the correlation between repeated measurements may lead to biased estimates as well as invalid P-values and confidence intervals. Therefore, careful consideration is needed to enable valid inference of covariate effects on longitudinal responses. In this regard, Residual Maximum Likelihood (REML) analysis is the most widely used method to estimate parameters. This method is based on the assumption that there is no correlation between the random effects and the error term (or residual effects). Hence, it is unclear if the failure to meet this assumption will affect the conclusions when conducting a real datasets study. In the present article, we propose Conditional Least Squares (CLS) and Weighted Conditional Least-Squares (WCLS) methods for estimating the model parameters, considering the presence of the correlation between the random effects and the error term. The choice of these methods is motivated by the fact that it is not necessary to specify the distributions of random effects and/or errors. These methods are illustrated via simulation studies that were performed with different sample sizes and different parameter values. In addition, we compared these methods with traditional estimators. This comparison is made by utilizing the ratio of their mean square error. Our results highlight that the weighted conditional least-squares estimator is efficient and attractive compared to the restricted maximum likelihood when random effects are permitted to be correlated with the error term. Also, real data analysis is conducted to confirm the advantages of the improved method.

Keywords: Linear mixed model, inference for linear model, conditional least squares, weighted conditional least squares, mean-squared errors.

1. Introduction

The analysis of repeated-measures response is often complicated by variation among individuals with respect to the times or conditions under which observations are taken. The resulting data are used to estimate an average linear response function. A model often used for modeling and analyzing longitudinal (popular measures) and hierarchical data is the mixed model described by Laird and Ware (1982), Foulley et al (2002), and Hartley and J. N. K. Rao (1967). They allow for the study of multiple sources of heterogeneity and/or correlation in data through the inclusion of random effects and explanative variables in the model. A crucial problem when adjusting such a model to data is to identify the fixed and/or random effects. In the literature several tests have been proposed to overcome this difficulty, however, all of these tests rely on regularity conditions. Recently, Ou Larbi, Y. et al. (2021) propose an optimal test based on the so-called ULAN (Uniform Local Asymptotic Normality) to detect the possible presence of random effects in linear mixed models. Concerning the problem of estimation of the parameters of interest, the standard method used is the restricted maximum likelihood (REML, the default in lme4), however, this method assumes that the random coefficient and the error term are independent, in practice, this assumption is not always valid. Diggle et al. (2002) is a fairly comprehensive reference in this area. However, it is unclear whether the failure of this hypothesis will have an impact on conclusions when conducting a real datasets study. Then, it is reasonable to pose the question of the performance of these methods when the assumed correlation structures, specify the relationships between random effects and error terms, are inappropriate. To avoid this problem and to attempt to obtain more efficient parameter estimates, several approximations have been proposed. Hwang et al. (1998) showed that Parameter estimation for generalized random coefficient autoregressive processes could be used to perform a more complete analysis when the random coefficients are permitted to be correlated with the error process.

Our work aims to extend the Hwang procedure to include ”the correlation measure to quantify the association between random effects and residuals terms” in the simple linear mixed model, we utilized the weighted conditional least-squares methods of estimating the fixed parameters. The simulation study compares, on the one hand, the weighted and the unweighted conditional least-squares estimators, and, on the other hand, the Maximum Likelihood (REML) and least-squares weighted estimators. The results are then applied to clinical data designed to study the distance between the pituitary and the pterygomaxillary of children between 8 and 14 years.

Traditionally, the term ”linear regression model” has been applied to models for a continuous response as it relates to one or more continuous covariates. An example of the simple linear regression model with a random effect for the slope, as described in the work of Ou Larbi, Y. et al. (2021), would be

$$y_{ij} = \beta_0 + (\beta_1 + \theta_i)x_{ij} + \varepsilon_{ij}, \qquad i = 1, \dots, n \quad and \quad j = 1, \dots, m. \tag{1}$$

The two-dimensional fixed-effects vector $\beta = (\beta_0, \beta_1)'$ consists of the mean intercept β_0 for the population and the common slope or growth rate β_1; the one-dimensional random-effects vectors θ_i, describe a shift in the slope for each individual. Here we have assumed that there are no random effects in the intercept of the model (1). Note that, the random effects θ_i are assumed to be independent for different i with mean 0, the within-individual errors ε_{ij} are assumed to be independent for different i or j with mean 0 and to be independent of the random effects, and, the random effects θ_i and the within-individual errors ε_{ij} are assumed to be independent for different individuals, it is completely characterized by its variance-covariance matrix:

$$Var\begin{pmatrix}\theta_i\\ \varepsilon_{ij}\end{pmatrix} = \begin{pmatrix}\sigma_\theta^2 & 0\\ 0 & \sigma_\varepsilon^2\end{pmatrix} \tag{2}$$

The parameters of this model are β, σ_{θ^2}, and σ_{ε^2}. Thus, it may be of interest to relax this assumption by allowing the dependence between the random effects θ_i and the errors ε_{ij} to be expressed as:

$$\mathrm{E}\begin{pmatrix}\theta_i\\ \varepsilon_{ij}\end{pmatrix} = \begin{pmatrix}0\\ 0\end{pmatrix} \text{ and } Var\begin{pmatrix}\theta_i\\ \varepsilon_{ij}\end{pmatrix} = \begin{pmatrix}\sigma_\theta^2 & \sigma_{\theta\varepsilon}\\ \sigma_{\theta\varepsilon} & \sigma_\varepsilon^2\end{pmatrix}. \tag{3}$$

The parameters of the statistical model created by combining (1) and (3) are β, σ_{θ^2}, σ_{ε^2}, and $\sigma_{\theta\varepsilon}$.

For matrix writing of the model (1), we can write it in the following matrix form, for each ith individual

$$Y_i = X_i\beta_i + \varepsilon_i,$$

where $\beta_i = (\beta_0, \beta_1 + \theta_i)'$, X_i is a matrix of "regressors" with dimensions $m \times 2$, and ε_i is a $m \times 1$ vector of unobservable random errors. Let $\beta = (\beta_0, \beta_1)'$, our objective is to estimate the unknown parameters $(\beta, \sigma^2_\theta, \sigma_{\theta\varepsilon}, \sigma^2_\varepsilon)$, from a given sample of observations.

Once the model has been formulated, methods are needed to estimate the unknown parameters, β, σ_{θ^2}, σ_{ε^2}, and $\sigma_{\theta\varepsilon}$, from a given sample of observations. Notice that the parameter β is considered the main parameter of interest, and the others are treated as unknown nuisance parameters that must also be estimated.

The paper is organized as follows. Section 2 relates the proposed conditional least squares estimator of the mean of the random coefficient vector. In Section 3, we estimate the variance-covariance parameters which are treated as nuisance parameters. The weighted conditional least-squares estimator of the mean of the random coefficient vector is studied in Section 4. Section 5 presents some simulation results for efficiency comparisons between the weighted and the unweighted conditional least-squares estimators, and between Restricted Maximum Likelihood (REML) and least-squares weighted estimators. We apply our estimation procedure to the real famous dental growth dataset from Potthoff and Roy (1964).

2. Proposed Conditional Least-Squares Estimation of β

In this section, we propose an estimation procedure for dependent observations based on the minimization of a sum of squared deviations on conditional expectations. This approach, known as "conditional least-squares" (CLS), offers a unified treatment of estimation problems for widely used regression model classes.

Based on the sample $(Y_1, Y_2, ..., Y_n)$ with size $N = n{*}m$, the conditional least-squares estimator $\widehat{\beta}$ of β, is obtained by minimizing:

$$Q = \sum_{i=1}^{n} \big[Y_i - \mathrm{E}(Y_i|X_i)\big]'\big[Y_i - \mathrm{E}(Y_i|X_i)\big],$$

with respect to β. Substituting $\mathrm{E}(Y_i|X_i) = X_i\beta$ in Q, and solving $dQ/d\beta = 0$ for β, we obtain

$$\widehat{\beta}_{(2\times 1)} = \left(\sum_{i=1}^{n} X_i'X_i\right)^{-1}\left(\sum_{i=1}^{n} X_i'Y_i\right). \tag{4}$$

The following theorem shows the distribution of the limit of the conditional least squares estimator of $\widehat{\beta}$.

Theorem 1 *Let $\widehat{\beta}$ the estimator of β, we have*

$$\sqrt{N}(\widehat{\beta} - \beta)\overset{d}{\longrightarrow}\mathcal{N}_2(0, T^{-1}UT^{-1}) \;\; as \;\; N \longrightarrow \infty,$$

where $T^{(N)}_{(2\times 2)} = \frac{1}{N}\sum\limits_{i=1}^{n} X_i'X_i \underset{N\longrightarrow\infty}{\longrightarrow} T = \begin{pmatrix} 1 & \mu^x_1 \\ \mu^x_1 & \mu^x_2 \end{pmatrix}$, *and* $U^{(N)}_{(2\times 2)} = \frac{1}{N}\sum\limits_{i=1}^{n} X_i'\alpha_i X_i \underset{N\longrightarrow\infty}{\longrightarrow} U$, *with* $\frac{1}{N}\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} x^k_{ij}$ *converge to* μ^x_k, $k = 1, ..., 4$, *and* $\alpha_i = Var(Y_i|X_i)$.

Proof: Note that

$$\begin{aligned}
\widehat{\beta} - \beta &= \left(\sum_{i=1}^{n} X_i'X_i\right)^{-1}\left(\sum_{i=1}^{n} X_i'Y_i\right) - \beta \\
&= \left[\sum_{i=1}^{n} X_i'X_i\right]^{-1}\left[\sum_{i=1}^{n} X_i'Y_i - \sum_{i=1}^{n} X_i'X_i\beta\right]
\end{aligned}$$

$$= \left[\sum_{i=1}^{n} X_i' X_i\right]^{-1} \left[\sum_{i=1}^{n} X_i' \left(Y_i - X_i\beta\right)\right].$$

Denote $\tilde{\beta}_i = \beta_i - \beta$. We consider

$$\sum_{i=1}^{n} X_i' \left(Y_i - X_i\beta\right) = \sum_{i=1}^{n} X_i' \left(\varepsilon_i + X_i\tilde{\beta}_i\right) = \sum_{i=1}^{n} G_i,$$

where $G_i = X_i' \left(\varepsilon_i + X_i\tilde{\beta}_i\right)$.

Then, we have

$$\sqrt{N}\left(\widehat{\beta} - \beta\right) = \left[\frac{1}{N}\sum_{i=1}^{n} X_i' X_i\right]^{-1} \left[N^{-1/2}\sum_{i=1}^{n} G_i\right].$$

Applying the central limit theorem, we find that

$$N^{-1/2}\sum_{i=1}^{n} G_i \xrightarrow{d} \mathcal{N}(0, U).$$

3. Conditional Least-Squares Estimation of η

Now, define $\eta = (\sigma_\varepsilon^2, \sigma_\theta^2, \sigma_{\theta\varepsilon})'$, we are considering the η-estimation problem. Let $R_{ij}(\beta) = y_{ij} - \mathrm{E}(y_{ij}|x_{ij})$, $\alpha_{ij} = Var(y_{ij}|x_{ij})$, and, denote α_{ij} by $\alpha_{ij}(\eta)$. Then

$$\begin{aligned} \alpha_{ij}(\eta) &= \mathrm{E}(R_{ij}^2(\beta)|x_{ij}) \\ &= \sigma_\varepsilon^2 + x_{ij}^2\sigma_\theta^2 + 2x_{ij}\sigma_{\theta\varepsilon} \end{aligned} \tag{5}$$

A conditional least-squares estimator $\widehat{\eta}(\beta)$ of η can be obtained when β is known by minimizing

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \left(R_{ij}^2(\beta) - \alpha_{ij}(\eta)\right)^2.$$

Therefore, a solution to the equation (6) gives $\widehat{\eta}(\beta)$,

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \left(R_{ij}^2(\beta) - \alpha_{ij}(\eta)\right) \frac{\partial \alpha_{ij}(\eta)}{\partial \eta} = 0. \tag{6}$$

When β is known, the $\widehat{\eta}(\beta)$ estimate is then considered to be

$$\widehat{\eta}(\beta) = (L'L)^{-1}L'Z(\beta), \tag{7}$$

where

$$Z(\beta) = \left(R_{11}^2(\beta), ..., R_{nm}^2(\beta)\right)',$$

and

$$L_{(N\times 3)} = \begin{pmatrix} 1 & x_{11}^2 & 2x_{11} \\ \vdots & \vdots & \vdots \\ 1 & x_{1m}^2 & 2x_{1m} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1}^2 & 2x_{n1} \\ \vdots & \vdots & \vdots \\ 1 & x_{nm}^2 & 2x_{nm} \end{pmatrix}. \tag{8}$$

The following theorem gives the distribution limit for $\widehat{\eta}$.

Theorem 2 *Let $\widehat{\eta}=\widehat{\eta}(\widehat{\beta})$, where $\widehat{\beta}$ is indicated by* (4).

$$\sqrt{N}(\widehat{\eta}-\eta)\xrightarrow{d}\mathcal{N}(0,M^{-1}\Sigma M^{-1})\ \ as\ \ N\longrightarrow\infty,$$

where $M^{(N)}_{(3\times3)}=\frac{1}{N}\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m}\left[\left(\frac{\partial\alpha_{ij}(\eta)}{\partial\eta}\right)\left(\frac{\partial\alpha_{ij}(\eta)}{\partial\eta}\right)'\right]\underset{N\longrightarrow\infty}{\longrightarrow}M=\begin{pmatrix}1 & \mu_2^x & 2\mu_1^x\\ \mu_2^x & \mu_4^x & 2\mu_3^x\\ 2\mu_1^x & 2\mu_3^x & 4\mu_2^x\end{pmatrix}$, and

$$\Sigma^{(N)}_{(3\times3)}=\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{m}\left[E\Big(R^2_{ij}(\beta)-\alpha_{ij}(\eta)\Big)^2\left(\frac{\partial\alpha_{ij}(\eta)}{\partial\eta}\right)\left(\frac{\partial\alpha_{ij}(\eta)}{\partial\eta}\right)'\right]\underset{N\longrightarrow\infty}{\longrightarrow}\Sigma. \tag{9}$$

Proof: We have

$$\begin{pmatrix}\mathrm{E}\big(R^2_{11}(\beta)|x_{11}\big)\\ \vdots\\ \mathrm{E}\big(R^2_{nm}(\beta)|x_{nm}\big)\end{pmatrix}=L\eta. \tag{10}$$

Within the appendix, the following technical lemma is proved.

Lemma 1 As $N\longrightarrow\infty$, we have,

(i) $N^{-1}L'L\xrightarrow{a.s}M$,

(ii) $N^{-1}L'(Z(\beta)-L\eta)\xrightarrow{a.s}0$,

(iii) $N^{-1/2}L'(Z(\widehat{\beta})-Z(\beta))\xrightarrow{P}0$,

(iv) $N^{-1/2}L'(Z(\beta)-L\eta)\xrightarrow{d}\mathcal{N}(0,\Sigma)$.

Now, consider

$$\sqrt{N}(\widehat{\eta}-\eta)=\sqrt{N}(\widehat{\eta}-\widehat{\eta}(\beta))+\sqrt{N}(\widehat{\eta}(\beta)-\eta).$$

Not that, by (i) and (iii) of Lemma 4.1 we have,

$$\begin{aligned}\sqrt{N}(\widehat{\eta}-\widehat{\eta}(\beta)) &= \sqrt{N}(L'L)^{-1}L'(Z(\widehat{\beta})-Z(\beta))\\ &= M^{-1}N^{-1/2}L'(Z(\widehat{\beta})-Z(\beta))+o_p(1)\\ &= o_p(1).\end{aligned} \tag{11}$$

Moreover, by (i) and (iv) of Lemma 4.1 we have,

$$\begin{aligned}\sqrt{N}(\widehat{\eta}(\beta)-\eta) &= \sqrt{N}(L'L)^{-1}L'(Z(\beta)-L\eta)\\ &= M^{-1}N^{-1/2}L'(Z(\beta)-L\eta)+o_p(1)\\ &\xrightarrow{d} \mathcal{N}(0,M^{-1}\Sigma M^{-1}),\end{aligned} \tag{12}$$

the theorem is proved by (11) and (12).

For the problem of estimating the correlation $\rho_{\theta\varepsilon}=Corr(\theta_i,\varepsilon_{ij})=\frac{\sigma_{\theta\varepsilon}}{\sigma_\theta\sigma_\varepsilon}$, a natural estimator of $\rho_{\theta\varepsilon}$ is given by $\hat{\rho}_{\theta\varepsilon}=\frac{\hat{\sigma}_{\theta\varepsilon}}{\hat{\sigma}_\theta\hat{\sigma}_\varepsilon}$.

4. Weighted Conditional Least-Squares Estimation of β

The method of ordinary least squares assumes that there is a constant variance in the errors (which is called homoscedasticity). But in many practical problems, the OLS assumption of constant variance in the errors is violated; in other words, the error variances are unequal (heteroscedastic). In this case, the weighted least squares method can be used to handle regression situations in which the data points are of varying quality. If the standard deviation of the random errors in the data is not constant across all levels of the explanatory variables, using weighted least squares with weights that are inversely proportional to the variance at each level of the explanatory variables yields the most precise parameter estimates possible.

Since $Var(Y_i|X_i)$ is based on X_i, one might find a conditional weighted least-squares estimator of β for improving efficiency. We derive such an estimator in this section, and we study its properties. Recall that

$$\mathrm{E}(Y_i|X_i) = X_i\beta \text{ and } Var(Y_i|Xi) = \alpha_i(\eta).$$

First assume the nuisance parameter η is defined, we can obtain the weighted conditional least-squares estimator $\beta_w(\eta)$ of β by minimizing:

$$\sum_{i=1}^{n} \big[Y_i - \mathrm{E}(Y_i|X_i)\big]'(\alpha_i(\eta))^{-1}\big[Y_i - \mathrm{E}(Y_i|X_i)\big],$$

that means to minimize:

$$\sum_{i=1}^{n} \big[Y_i - X_i\beta\big]'(\alpha_i(\eta))^{-1}\big[Y_i - X_i\beta\big],$$

where $\widehat{\eta}(\beta)$ in (7) is defined, therefore $\alpha_i(\eta)$. It is easily verified that

$$\widehat{\beta}_w(\eta) = \left(\sum_{i=1}^{n} X_i'(\alpha_i(\eta))^{-1}X_i\right)^{-1}\left(\sum_{i=1}^{n} X_i'(\alpha_i(\eta))^{-1}Y_i\right). \tag{13}$$

We replace η in $\widehat{\beta}_w(\eta)$ by $\widehat{\eta}$ given in (7) when η is unknown. Denote $\widehat{\beta}_w = \widehat{\beta}_w(\widehat{\eta})$. The following theorem gives the limit distribution of $\widehat{\beta}_w$.

Theorem 3 *Let $\widehat{\beta}_w$ the estimator of β, we have*

$$\sqrt{N}(\widehat{\beta}_w - \beta)\xrightarrow{d}\mathcal{N}_2(0, T_\alpha^{-1}) \;\; as \;\; N \longrightarrow \infty,$$

where $T_\alpha = \frac{1}{N}\sum_{i=1}^{n} X_i'(\alpha_i(\widehat{\eta}))^{-1}X_i$

Proof: Note that

$$\begin{aligned}
\widehat{\beta}_w - \beta &= \left(\sum_{i=1}^{n} X_i'(\alpha_i(\widehat{\eta}))^{-1}X_i\right)^{-1}\left(\sum_{i=1}^{n} X_i'(\alpha_i(\widehat{\eta}))^{-1}Y_i\right) - \beta \\
&= \left[\sum_{i=1}^{n} X_i'(\alpha_i(\widehat{\eta}))^{-1}X_i\right]^{-1}\left[\sum_{i=1}^{n} X_i'(\alpha_i(\widehat{\eta}))^{-1}Y_i - \sum_{i=1}^{n} X_i'(\alpha_i(\widehat{\eta}))^{-1}X_i\beta\right] \\
&= \left[\sum_{i=1}^{n} X_i'(\alpha_i(\widehat{\eta}))^{-1}X_i\right]^{-1}\left[\sum_{i=1}^{n} X_i'(\alpha_i(\widehat{\eta}))^{-1}\left(Y_i - X_i\beta\right)\right] \qquad (14)
\end{aligned}$$

On the right of (14) the second factor converges in distribution to a normal vector with mean zero, and covariance matrix $\frac{1}{N}\sum_{i=1}^{n} X_i'(\alpha_i(\widehat{\eta}))^{-1}X_i$.

5. Simulation

In this section, we present a simulation experiment (using R-programming) to compare, on the one hand, $\widehat{\beta}$ and $\widehat{\beta}_w$ (Table 1), on the other hand, $\widehat{\beta}_{REML}$ and $\widehat{\beta}_w$ (Table 2), to improve the interpretation and validity of the theoretical results of the previous sections, we set $m = 10$, we consider the model:

$$Y_{ij} = \beta_0 + (\beta_1 + \theta_i)X_{ij} + \varepsilon_{ij}, \qquad i = 1, \ldots, 50, \;\; and, \qquad j = 1, \ldots, 10,$$

where

- $\beta_0 = -1$ and $\beta_1 = 0.5$;
- the X_{ij}'s are $i.i.d.$ normal $(0, 1)$;

and

$$\begin{pmatrix} \theta_i \\ \varepsilon_{ij} \end{pmatrix} are\ \ i.i.d\ \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\theta^2 & \sigma_{\theta\varepsilon} \\ \sigma_{\theta\varepsilon} & 1 \end{pmatrix}\right).$$

We generated realizations with $n = 50$, $n = 75$, and, $n = 100$, for $\sigma_\theta^2 = 0.1, 0.2, 0.3$, and, 0.4, and $\rho_{\theta\varepsilon} = -0.8, -0.4, 0, 0.4$, and, 0.8.

We used the ratio of mean-squared errors (MSE)(simulation version)
$\mathrm{E}\Big((\widehat{\beta} - \beta)'(\widehat{\beta} - \beta)\Big)/\mathrm{E}\Big((\widehat{\beta}_w - \beta)'(\widehat{\beta}_w - \beta)\Big)$, and,
$\mathrm{E}\Big((\widehat{\beta}_{REML} - \beta)'(\widehat{\beta}_{REML} - \beta)\Big)/\mathrm{E}\Big((\widehat{\beta}_w - \beta)'(\widehat{\beta}_w - \beta)\Big)$.

Table 1 Ratio of mean-squared errors of $\widehat{\beta}\ and\ \widehat{\beta}_w$

n	$\rho_{\theta\varepsilon}$	σ_θ^2			
		0.1	0.2	0.3	0.4
	0.8	1.359359	1.395495	1.532641	1.869584
	0.4	1.197054	1.207617	1.373875	1.602843
50	0	1.117028	1.128982	1.268104	1.437612
	-0.4	1.185248	1.232476	1.310724	1.694612
	-0.8	1.415381	1.451137	1.607392	1.921483
	0.8	1.424608	1.617739	1.692227	2.160999
	0.4	1.374462	1.458082	1.736580	2.049992
75	0	1.198832	1.203693	1.416753	1.591268
	-0.4	1.359088	1.445395	1.654181	1.994612
	-0.8	1.397835	1.636729	1.728250	2.057458
	0.8	1.626075	1.966686	2.626474	2.908092
	0.4	1.536073	1.737017	1.935537	2.361636
100	0	1.245761	1.387348	1.633601	1.827943
	-0.4	1.372188	1.829915	2.085198	2.226514
	-0.8	1.777537	1.836029	2.527386	2.911533

Inspection of Tables: Each of the above simulations was replicated 100 times and the values of $\widehat{\beta}$, $\widehat{\beta}_{REML}$, and $\widehat{\beta}_w$ were calculated. Tables 1 and 2 summarize the results for for the ratio of mean-squared errors of ($\widehat{\beta}_w$ and $\widehat{\beta}_w$) and ($\widehat{\beta}_{REML}$ and $\widehat{\beta}_w$) respectively. $\widehat{\beta}_w$ is seen as being always more efficient than $\widehat{\beta}$ and $\widehat{\beta}_{REML}$ as σ_θ^2, n, and $|\rho_{\theta\varepsilon}|$ increase.

Table 2 Ratio of mean-squared errors of $\widehat{\beta}_{REML}$ and $\widehat{\beta}_w$

n	$\rho_{\theta\varepsilon}$	σ^2_θ			
		0.1	0.2	0.3	0.4
	0.8	1.252802	1.301108	1.423658	1.606676
	0.4	1.115304	1.192158	1.361992	1.425782
50	0	1.001793	1.020279	1.098203	1.138350
	-0.4	1.180884	1.389044	1.388832	1.512429
	-0.8	1.315381	1.202226	1.569503	1.712257
	0.8	1.376726	1.440681	1.594710	1.717377
	0.4	1.256880	1.316841	1.527196	1.533357
75	0	1.009582	1.055467	1.100187	1.156502
	-0.4	1.377173	1.434246	1.434246	1.367092
	-0.8	1.418113	1.421580	1.621584	1.693885
	0.8	1.409952	1.502766	1.716366	2.034322
	0.4	1.307171	1.419290	1.551238	1.681656
100	0	1.014806	1.024055	1.104276	1.200266
	-0.4	1.284842	1.347044	1.537082	1.706824
	-0.8	1.520679	1.489360	1.637833	1.907261

6. Real Data Application

In this section, we consider the well-known Orthodont dataset (package nlme of R) introduced by Potthoff and Roy (1964). Researchers at the University of North Carolina School of Dentistry followed the growth of 27 children (16 boys and 11 girls) aged 8 to 14 years. Every two years, they measured the distance between the pituitary and the pterygomaxillary fissure (”distance” noted by y) by the explanatory variable (”age” noted by x). The study objectives were to determine if the distances were longer for boys than girls and if the rate of change of distance differed between boys and girls. This data set was analyzed by Fearn (1975), Drikvandi et al. (2012), and Verbeke and Molenberghs (2000).

In this example, we are interested in estimating the parameter for effects fixed in the simple linear mixed model. Consider here:

$$y_{ij} = \beta_0 + (\beta_1 + \theta_i)x_{ij} + \varepsilon_{ij} \ \ for \ \ i = 1, \ldots, 27 \ \ and \ \ j = 1, \ldots, 4 \tag{15}$$

with β_0 and β_1 are, respectively, the fixed effects for the intercept and the slope (parameter of interest), θ_i is a random effects with zero mean, and with var(θ_i) = σ^2_θ, and ε_{ij} is a random error with zero mean, and with var(ε_{ij}) = σ^2_ε.

When we apply, on the one hand, the conditional least-squares estimation to the model (15), we find: $\hat{\beta}_0 = 16.7611132$ and $\hat{\beta}_1 = 0.6601326$. On the other hand, the weighted conditional least-squares estimation gives: $\hat{\beta}_{0w} = 16.7611219$ and $\hat{\beta}_{1w} = 0.6602256$.

Then, we generate the model 200 times, where we calculate the output variable y_{ij}, by the estimator $\hat{\beta}$ and by the estimator $\hat{\beta}_w$, which we denote, respectively, $y(\hat{\beta})$ and $y(\hat{\beta}_w)$. In the end, we calculate the average of the errors of each output variable with the true values of Orthodont (distance), and we find $MSE(y(\hat{\beta}))/MSE(y(\hat{\beta}_w)) = 1.040632$.

7. Conclusion

During this work, we were interested in statistical inference, more specifically in the theory of vector β estimation of the fixed parameters in the mixed linear model, in which we considered that the random coefficient and the random errors are not always independent. On the one hand, we have established the estimator by the ordinary least squares method, and on the other hand, we have established it the weighted least squares method.

The simulation by the Monte Carlo study guarantees the good performance of the weighted estimator $\widehat{\beta}_w$, which dominates both the unweighted (ordinary) least squares $\widehat{\beta}$ and restricted maximum likelihood (REML) $\widehat{\beta}_{REML}$ estimators.

Acknowledgements

The authors would like to express their gratitude to the referees and the editor for their constructive comments and suggestions, which improved the presentation of the manuscript.

References

Anderson RL. Bancroft TA. Statistical Theory in Research. New York: McGrawHill; 1952.

Bates DM. Pinheiro JC. Linear and nonlinear mixed-effects models. Proceeding of the 10th annual conference on applied statistics in agriculture; 1998. pp. 1-22.

Breslow NE. Clayton DG. Approximate inference in generalized linear mixed models. J Am Stat Assoc. 1993; 88(421): 9-25.

Cornillon PA. Matzner-Lober E. Rgression: thorie et applications. Springer; 2007.

Davezies L. Modles effets fixes, effets alatoires, modles mixtes ou multi-niveaux: proprits et mises en oeuvre des modlisations de l'htrognit dans le cas de donnes groupes; 2011.

Demidenko E. Mixed models: theory and applications with R. John Wiley and Sons; 2013.

Diggle P, Diggle PJ, Heagerty P, Liang KY, Zeger S. Analysis of longitudinal data. Oxford university press; 2002.

Drikvandi R, Khodadadi A, Verbeke G. Testing variance components in balanced linear growth curve models. J Appl Stat. 2012; 39(3): 563-572.

Foulley JL, Delmas C, Robert-Grani C. Mthodes du maximum de vraisemblance en modle linaire mixte. Journal de la socit franaise de statistique. 2002; 143(1-2): 5-52.

Hartley HO, Rao JN. Maximum-likelihood estimation for the mixed analysis of variance model. Biometrika. 1967; 54(1-2): 93-108.

Harville DA. Maximum likelihood approaches to variance component estimation and to related problems. J Am Stat Assoc. 1977; 72(358): 320-338.

Hwang SY, Basawa IV. Parameter estimation for generalized random coefficient autoregressive processes. J Stat Plan Infer. 1998; 68(2): 323-337.

Ibragimov IA, Hasminskii RZ. Statistical Estimation: Asymptotic Theory. New York: Springer-Verlag; 1981.

Jiang J. Linear and generalized linear mixed models and their applications. Springer Science and Business Media; 2007.

Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics. 1982; 38(4): 963-974.

Ou Larbi Y, Halimi RE, Akharif A, Mellouk A. Optimal tests for random effects in linear mixed models . Hacet J Math Stat. 2021; 50(4): 1185-1211.

Pinheiro J, Bates D. Mixed-effects models in S and S-PLUS. Springer Science and Business Media; 2006.

Regis M, Brini A, Nooraee N, Haakma R, van den Heuvel ER. The t linear mixed model: model formulation, identifiability and estimation. Commun Stat Simulat. 2019; 51(5): 2318-2342.

Swamy PAVB. Statistical inference in random coefficient regression models (Vol. 55). Springer Science and Business Media; 2012.

Shao J. Asymptotic distribution of the weighted least squares estimator. Ann I Stat Math. 1989; 41(2): 365-382.

Thompson Jr. WA. The problem of negative estimates of variance components. Ann Math Stat. 1962; 33(1): 273-289.

Verbeke G. Linear mixed models for longitudinal data. In Linear mixed models in practice. New York: Springer; 1997.

Appendix: Proof of Lemma 4.1

In order to prove (ii) and (iv), we consider

$$L'(Z(\beta)-L\eta) = \begin{pmatrix} 1 & \cdots & 1 \\ x_{11}^2 & \cdots & x_{nm}^2 \\ 2x_{11} & \cdots & 2x_{nm} \end{pmatrix} \begin{pmatrix} R_{11}^2(\beta)-\mathrm{E}\big(R_{11}^2(\beta)|x_{11}\big) \\ \vdots \\ R_{nm}^2(\beta)-\mathrm{E}\big(R_{nm}^2(\beta)|x_{nm}\big) \end{pmatrix} \quad (16)$$

$$= \begin{pmatrix} \sum_{i=1}^{n}\sum_{j=1}^{m}\left[R_{ij}^2(\beta)-\mathrm{E}\big(R_{ij}^2(\beta)|x_{ij}\big)\right] \\ \sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij}^2\left[R_{ij}^2(\beta)-\mathrm{E}\big(R_{ij}^2(\beta)|x_{ij}\big)\right] \\ \sum_{i=1}^{n}\sum_{j=1}^{m} 2x_{ij}\left[R_{ij}^2(\beta)-\mathrm{E}\big(R_{ij}^2(\beta)|x_{ij}\big)\right] \end{pmatrix} \quad (17)$$

$$= \left(\sum_{i=1}^{n}\sum_{j=1}^{m} D_{1,ij}, \sum_{i=1}^{n}\sum_{j=1}^{m} D_{2,ij}, \sum_{i=1}^{n}\sum_{j=1}^{m} D_{3,ij}\right)', \quad (18)$$

where

$$D_{1,ij} = \left[R_{ij}^2(\beta)-\mathrm{E}\big(R_{ij}^2(\beta)|x_{ij}\big)\right], D_{2,ij} = x_{ij}^2\left[R_{ij}^2(\beta)-\mathrm{E}\big(R_{ij}^2(\beta)|x_{ij}\big)\right], and$$

$$D_{3,ij} = 2x_{ij}\left[R_{ij}^2(\beta)-\mathrm{E}\big(R_{ij}^2(\beta)|x_{ij}\big)\right].$$

Note that each $D_{k,ij}, k=1,2,3$ is with mean zero. Part (ii) can be easily obtained from (A.3). Also, part (iv) follows from the central limit theorem.

Notice that it is possible to write Σ in (9) in terms of D_{ij} as

$$\Sigma = \mathrm{E}\big(D_{ij}D_{ij}'\big), \ \ with \ \ D_{ij} = (D_{1,ij}, D_{2,ij}, D_{3,ij})'.$$

To establish (iii), we consider

$$Z(\hat{\beta})-Z(\beta) = \begin{pmatrix} [y_{11}-(\hat{\beta}_0+\hat{\beta}_1 x_{11})]^2-[y_{11}-(\beta_0+\beta_1 x_{11})]^2 \\ \vdots \\ [y_{nm}-(\hat{\beta}_0+\hat{\beta}_1 x_{nm})]^2-[y_{nm}-(\beta_0+\beta_1 x_{nm})]^2 \end{pmatrix}.$$

Let $X_{ij} = \begin{pmatrix}1 & x_{ij}\end{pmatrix}$ and not that for the ith individual, the jth element of $Z(\hat{\beta})-Z(\beta)$, $i=1,...,n$ and $j=1,...,m$ can be written as

$$\begin{aligned} [y_{ij}-(\hat{\beta}_0+\hat{\beta}_1 x_{ij})]^2-[y_{ij}-(\beta_0+\beta_1 x_{ij})]^2 &= [y_{ij}-X_{ij}\hat{\beta}]^2-[y_{ij}-X_{ij}\beta]^2 \\ &= -2y_{ij}X_{ij}(\hat{\beta}-\beta)+X_{ij}(\hat{\beta}-\beta)X_{ij}(\hat{\beta}+\beta) \\ &= \begin{pmatrix}-2y_{ij}+X_{ij}(\hat{\beta}+\beta) & x_{ij}[-2y_{ij}+X_{ij}(\hat{\beta}+\beta)]\end{pmatrix} \\ &\quad \times \begin{pmatrix}\hat{\beta}_0-\beta_0 \\ \hat{\beta}_1-\beta_1\end{pmatrix}. \end{aligned}$$

Then, we have

$$\begin{aligned} N^{-1/2}L'(Z(\widehat{\beta})-Z(\beta)) &= N^{-1}L'\begin{pmatrix} -2y_{11}+X_{11}(\hat{\beta}+\beta) & x_{11}[-2y_{11}+X_{11}(\hat{\beta}+\beta)] \\ \vdots & \vdots \\ -2y_{nm}+X_{nm}(\hat{\beta}+\beta) & x_{nm}[-2y_{nm}+X_{nm}(\hat{\beta}+\beta)] \end{pmatrix} \\ &\quad \times\sqrt{N}\begin{pmatrix}\hat{\beta}_0-\beta_0 \\ \hat{\beta}_1-\beta_1\end{pmatrix} \\ &= W_N\sqrt{N}(\hat{\beta}-\beta), \end{aligned}$$

where

$$W_N = N^{-1} L' \begin{pmatrix} -2y_{11} + X_{11}(\hat{\beta} + \beta) & x_{11}[-2y_{11} + X_{11}(\hat{\beta} + \beta)] \\ \vdots & \vdots \\ -2y_{nm} + X_{nm}(\hat{\beta} + \beta) & x_{nm}[-2y_{nm} + X_{nm}(\hat{\beta} + \beta)] \end{pmatrix}_{(N \times 2)}.$$

Since $\hat{\beta} \xrightarrow{a.s} \beta$, $\sqrt{N}(\hat{\beta} - \beta)$ is bounded in probability, and $W_N \xrightarrow{a.s} 0$, the result in part (iii) finally follows.