



## Henderson's method approach to Kernel prediction in partially linear mixed models

Özge KURAN<sup>1\*</sup> , Seçil YALAZ<sup>1</sup> 

<sup>1</sup>Dicle University, Faculty of Science, Department of Statistics, 21280, Diyarbakır / TURKEY

### Abstract

In this article, we propose Kernel prediction in partially linear mixed models by using Henderson's method approach. We derive the Kernel estimator and the Kernel predictor via the mixed model equations (MMEs) of Henderson's that they give the best linear unbiased estimation (BLUE) of the fixed effects parameters and the nonparametric function computationally easier and the best linear unbiased prediction (BLUP) of the random effects parameters as by-products. Additionally, asymptotic property of the Kernel estimator is investigated. A Monte Carlo simulation study is supported to illustrate the performance of Kernel prediction in partially linear mixed models and then, we finalize the article with the help of conclusion and discussion part to summarize the findings.

### Article info

*History:*

Received: 07.01.2020

Accepted: 15.06.2020

*Keywords:*

Henderson's method, Kernel estimator, Kernel predictor, partially linear mixed model, semiparametric model.

## 1. Introduction

The partially linear mixed models (PLMMs) can be viewed as a combination of the linear mixed models (LMMs) [1] and the partially linear models (PLMs) [2]. The PLMMs are popular in the analysis of correlated data including longitudinal and repeated measurement or clustered data over time by incorporating the between-subject and within-subject variations in many clinical and biomedical studies in recent years.

Let us consider the PLMMs

$$y_i = X_i\beta + g(T_i) + Z_iu_i + \varepsilon_i \quad i = 1, \dots, m \quad (1)$$

where  $y_i$  is an  $n_i \times 1$  vector of response variables measured on subject  $i$ ,  $\beta$  is a  $p \times 1$  parameter vector of fixed effects,  $T_i$  is a random variable defined on  $[0,1]$ , the function  $g(\cdot)$  is unknown function from  $\mathbb{R}^d$  to  $\mathbb{R}^1$ ,  $X_i = (x_{i1}, \dots, x_{ip})^T$ ,  $T_i = (t_{i1}, \dots, t_{id})^T$  and  $Z_i = (z_{i1}, \dots, z_{iq})^T$  are  $n_i \times p$ ,  $n_i \times d$  and  $n_i \times q$  known fixed and random effects design matrices, respectively,  $u_i$  random vector that the components of which are told random effects and  $\varepsilon_i$  is an  $n_i \times 1$  random errors vector.

The posits,  $u_i \stackrel{iid}{\sim} N_q(0, D)$  and  $\varepsilon_i \stackrel{iid}{\sim} N_{n_i}(0, W_i)$ ,  $i = 1, \dots, m$ , where  $u_i$  and  $\varepsilon_i$  are independent,  $D$  and  $W_i$  are  $q \times q$  and  $n_i \times n_i$  known positive definite (pd) matrices are usually used.

When  $X_i$ 's are observable and under the assumptions of model (1), the conditional distribution of  $y_i$  given  $u_i$  is  $y_i|u_i \sim N(X_i\beta + g(T_i) + Z_iu_i, W_i)$ . Defining the conditional expectations which are also known as the kernel regressions with bandwidth  $h$  of  $y$ ,  $X$  and  $Z$  as

$$\omega_y(T_i) = E(y_i|T_i) = \sum_{i=1}^n \omega_{ij}(t)y_i$$

\*Corresponding author. Email address: ozge.kuran@dicle.edu.tr

$$\omega_x(T_i) = E(X_i|T_i) = \sum_{i=1}^n \omega_{ij}(t)X_i$$

$$\omega_z(T_i) = E(Z_i|T_i) = \sum_{i=1}^n \omega_{ij}(t)Z_i$$

where  $\omega_{ij}(t) = K_h(t_{ij} - t) / \sum_{k=1}^n \sum_{l=1}^{n_k} K_h(t_{kl} - t)$ ,  $K_h(\cdot) = K(\cdot/h)$  and  $K(\cdot)$  is a Kernel function, the last expression could be written as

$$\omega_y(T_i) = \omega_x(T_i)\beta + \omega_z(T_i)u_i + g(T_i). \tag{2}$$

Subtracting equation (2) from equation (1), it is obtained that

$$y_i - \omega_y(T_i) = [X_i - \omega_x(T_i)]\beta + [Z_i - \omega_z(T_i)]u_i + \varepsilon_i$$

$$\tilde{y}_i = \tilde{X}_i\beta + \tilde{Z}_i u_i + \varepsilon_i. \tag{3}$$

Let  $\tilde{y} = (\tilde{y}_1^T, \dots, \tilde{y}_m^T)^T$ ,  $\tilde{X} = (\tilde{X}_1^T, \dots, \tilde{X}_m^T)^T$ ,  $\tilde{Z} = \bigoplus_{i=1}^m \tilde{Z}_i$ , where  $\bigoplus$  describes the direct sum,  $u = (u_1^T, \dots, u_m^T)^T$  and  $\varepsilon = (\varepsilon_1^T, \dots, \varepsilon_m^T)^T$ . Then, equation (3) is obtained more compactly as

$$\tilde{y} = \tilde{X}\beta + \tilde{Z}u + \varepsilon \tag{4}$$

which implies that

$$\begin{bmatrix} u \\ \varepsilon \end{bmatrix} \sim N_{qm+n} \left( \begin{bmatrix} 0_{qm} \\ 0_n \end{bmatrix}, \begin{bmatrix} \mathcal{D} & 0 \\ 0 & W \end{bmatrix} \right)$$

where  $n = \sum_{i=1}^m n_i$ ,  $\mathcal{D} = I_m \otimes D$  and  $W = \bigoplus_{i=1}^m W_i$  by  $\otimes$  indicating the Kronecker product. So, we derive  $\tilde{y} \sim N(\tilde{X}\beta, V)$  where  $V = \tilde{Z}\mathcal{D}\tilde{Z}^T + W$  in model (4).

There are the profile-kernel, backfitting, smoothing spline, penalized spline and local linear regression methodologies to estimate the nonparametric function in the PLMMs. Some of these methodologies were widespread for independent data, and some of them commonly used for correlated data. The asymptotic properties of profile-kernel estimators for the independent data were provided by [3], [4] and [5]. The bias problem of backfitting estimation was firstly discerned by [6]; see also [7] and [8]. Meanwhile, [8] demonstrated that the backfitting and kernel estimators share the same asymptotic behavior. [9] employed a semiparametric random intercept model (an extension of the PLMs) to examine the CD4 cell numbers in HIV seroconverters, where the nonparametric function is estimated by the backfitting method. [10] and [11] investigated PLMMs for longitudinal data and employed smoothing spline, while [12] and [13] employed the penalized spline to fit PLMMs. [14] characterized local linear regression in the framework of generalized PLMMs for longitudinal data.

[15]'s study is an extension of [9]'s model. [15] think a more general class of LMMs that the nonparametric component is estimated by the profile-kernel and backfitting methodologies. They work a natural extension of the linear mixed and semiparametric models called semiparametric mixed effect (or semiparametric linear mixed) model (SMEM) that uses parametric fixed effects to present the covariate effects and an arbitrary smooth function to model the time effect to account for the within subject correlation using random effects and its asymptotic behavior. To further highlighting the superiority of the methodology upon the backfitting, a comparison is also accomplished. They bring to an end that the kernel methodology rakes to have smaller bias and variance than the backfitting, asymptotically. Additionally, they demonstrate their theoretical results with the analysis of CD4 data in HIV disease and a small simulation study. They indicate that the SMEM is more stable and efficient than the linear mixed and semiparametric models. However, since [15] are obtained the profile-kernel and backfitting estimators under marginal model, they exclude the effect of the random effects in SMEM.

The principal goal of this article is to obtain Kernel predictors in the PLMMs by using [16]'s MMEs products. The plan of the article as follows. In Section 2, we recommend the Kernel prediction with the help of Henderson's MMEs and then, its asymptotic behavior is derived. In Section 3, a Monte Carlo simulation study is ensured to designate the theoretical outcomes. The article is finalized some summary and conclusions in Section 4.

## 2. Kernel Prediction in Partially Linear Mixed Models

In this section, we suggest the Kernel prediction in PLMMs via [16]'s MMEs different from [15]'s marginal model. Thus, we produce not only the estimation of the fixed effects and the nonparametric function but also the prediction of the random effects.

By following model (3) assumptions,  $u$  and  $\tilde{y}$  are jointly Gaussian distributed as

$$\begin{bmatrix} u \\ \tilde{y} \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ \tilde{X}\beta \end{bmatrix}, \begin{bmatrix} \mathfrak{D} & \mathfrak{D}\tilde{Z}^T \\ \tilde{Z}\mathfrak{D} & V \end{bmatrix} \right) \tag{5}$$

and then by using equation (5), the conditional distribution of  $\tilde{y}$  given  $u$  is  $\tilde{y}|u \sim N(\tilde{X}\beta + \tilde{Z}u, W)$ .

Following [16], we obtain the joint density of  $\tilde{y}$  and  $u$  given by

$$\begin{aligned} f(\tilde{y}, u) &= f(\tilde{y}|u)f(u) \\ &= (2\pi)^{-(n+qm)/2} |W|^{-1/2} |\mathfrak{D}|^{-1/2} \exp \left\{ -\frac{1}{2} \left[ (\tilde{y} - \tilde{X}\beta - \tilde{Z}u)^T W^{-1} (\tilde{y} - \tilde{X}\beta - \tilde{Z}u) \right. \right. \\ &\quad \left. \left. + u^T \mathfrak{D}^{-1} u \right] \right\} \end{aligned} \tag{6}$$

where  $|\cdot|$  designates a matrix determinate and the equations are similar to Henderson's MMEs.

Equation (6) is rewritten by taking the log-joint distribution of  $f(\tilde{y}, u)$

$$\begin{aligned} \log f(\tilde{y}, u) &= \log f(\tilde{y}|u) + \log f(u) \\ &= -\frac{1}{2} \left\{ (n + qm) \log(2\pi) + \log |W| + \log |\mathfrak{D}| \right. \\ &\quad \left. + \left[ (\tilde{y} - \tilde{X}\beta - \tilde{Z}u)^T W^{-1} (\tilde{y} - \tilde{X}\beta - \tilde{Z}u) + u^T \mathfrak{D}^{-1} u \right] \right\}. \end{aligned} \tag{7}$$

As the results of removing the fixed term, taking the log function into account and computing the partial derivatives of equation (7) in respond to the  $\beta$  and  $u$  to zero and using  $\hat{\beta}_{Ke}$  and  $\hat{u}_{Kp}$  to demonstrate the Kernel estimator ( $Ke$ ) and the Kernel predictor ( $Kp$ ), the solutions are given as

$$\tilde{X}^T W^{-1} (\tilde{y} - \tilde{X} \hat{\beta}_{Ke}) - \tilde{X}^T W^{-1} \tilde{Z} \hat{u}_{Kp} = 0 \tag{8}$$

$$\tilde{Z}^T W^{-1} (\tilde{y} - \tilde{X} \hat{\beta}_{Ke}) - \tilde{Z}^T W^{-1} \tilde{Z} + \mathfrak{D}^{-1} \hat{u}_{Kp} = 0. \tag{9}$$

Equations (8) and (9) are parallel to Henderson's MMEs obtained by [17] and [16], with a distinction that equations (8) and (9) are practiced to  $\hat{\beta}_{Ke}$  and  $\hat{u}_{Kp}$  where Henderson's MMEs are practiced to the best linear unbiased estimator (BLUE) and the best linear unbiased predictor (BLUP).

Equations (8) and (9) can compactly be rewritten in matrix as

$$\begin{pmatrix} \tilde{X}^T W^{-1} \tilde{X} & \tilde{X}^T W^{-1} \tilde{Z} \\ \tilde{Z}^T W^{-1} \tilde{X} & \tilde{Z}^T W^{-1} \tilde{Z} + \mathfrak{D}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta}_{Ke} \\ \hat{u}_{Kp} \end{pmatrix} = \begin{pmatrix} \tilde{X}^T W^{-1} \tilde{y} \\ \tilde{Z}^T W^{-1} \tilde{y} \end{pmatrix}. \tag{10}$$

Using [18]'s approach, equation (10) can be written as

$$C\hat{\varphi} = \gamma^T W^{-1} \tilde{y} \tag{11}$$

where  $\hat{\varphi} = (\hat{\beta}_{Ke}^T, \hat{u}_{Kp}^T)^T$ ,  $\gamma = (\tilde{X}, \tilde{Z})$  and  $C = \gamma^T W^{-1} \gamma + \mathfrak{D}^{**}$  with  $\mathfrak{D}^* = \begin{bmatrix} I_p & 0 \\ 0 & \mathfrak{D} \end{bmatrix}$  and  $G^{**} = \begin{bmatrix} I_p & 0 \\ 0 & \mathfrak{D}^{-1} \end{bmatrix}$  where the ‘+’ indicates the Moore–Penrose inverse.

By resolving equation (11), the following equation is found as

$$\hat{\varphi} = C^{-1} \gamma^T W^{-1} \tilde{y} \tag{12}$$

where  $C^{-1}$  is attained with the help of the inverse of the partitioned matrix (see [19]) as

$$C^{-1} = \begin{pmatrix} (\tilde{X}^T V^{-1} \tilde{X})^{-1} & -(\tilde{X}^T V^{-1} \tilde{X})^{-1} \tilde{X}^T V^{-1} \tilde{Z} \mathfrak{D} \\ -\mathfrak{D} \tilde{Z}^T V^{-1} \tilde{X} (\tilde{X}^T V^{-1} \tilde{X})^{-1} & (\tilde{Z}^T W^{-1} \tilde{Z} + \mathfrak{D}^{-1})^{-1} + \mathfrak{D} \tilde{Z}^T V^{-1} \tilde{X} (\tilde{X}^T V^{-1} \tilde{X})^{-1} \tilde{X}^T V^{-1} \tilde{Z} \mathfrak{D} \end{pmatrix}.$$

After algebraic simplifications and  $C^{-1}$  is replaced in equation (12), we suggest the Kernel estimator and the Kernel predictor, respectively, as

$$\hat{\beta}_{Ke} = (\tilde{X}^T V^{-1} \tilde{X})^{-1} \tilde{X}^T V^{-1} \tilde{y} \tag{13}$$

$$\hat{u}_{Kp} = \mathfrak{D} \tilde{Z}^T V^{-1} (\tilde{y} - \tilde{X} \hat{\beta}_{Ke}). \tag{14}$$

If  $\beta$  were known, the estimator of  $g(t) = E(Y - X\beta|T = t)$  can be defined as

$$\hat{g}(t) = \sum_{i=1}^n \omega_{ij}(t) (y_i - X_i \hat{\beta}_{Ke}). \tag{15}$$

### 2.1. Asymptotic property of kernel estimator

In this subsection, we will examine the asymptotic property of Kernel estimator.

**Theorem 2.1** Under the assumptions that the  $(y_i, X_i, T_i)$  are independent and identically distributed (i.i.d) triplets,  $g^{(r)}(\cdot)$  is the  $r$ th derivative of any function  $g(\cdot)$ ,  $v^{kl}$  is the  $(k, l)$ th element of  $V^{-1}$ ,  $f_k(t)$  is density of  $T_k$ , the Kernel density function  $K(\cdot)$  is assumed to has mean 0, unit variance,  $h \propto n^{-\alpha}$ ,  $\frac{1}{5} \leq \alpha \leq \frac{1}{3}$  and  $n \rightarrow \infty$  are held,  $\hat{\beta}$  converges in distribution

$$\sqrt{n} \left\{ \hat{\beta} - \beta + \frac{h^2 b_1(\beta, g)}{2} \right\} \xrightarrow{D} N(0, V_k)$$

where the bias term  $b_1(\beta, g) = A^{-1} E\{\tilde{X}^T V^{-1} g^{(2)}(t)\}$  and  $V_k = A^{-1} E[(J_1 - J_2)^T V_0 (J_1 - J_2)] A^{-1}$  for  $A = \lim_{n \rightarrow \infty} A_n = E(\tilde{X}^T V^{-1} \tilde{X})$ ,  $V_0 = Var(y|x, t, z)$ ,  $J_1 = V^{-1} \tilde{X}$ ,  $J_2 = [J_{21}, \dots, J_{2n}]$ , where  $J_{2i} = [J_{2i1}, \dots, J_{2in}]^T$  is  $J_{2ij} = \frac{[\sum_{k=1}^n \sum_{l=1}^n E(\tilde{X}_k V^{kl} | t_l = t_{ij}) f_j(t_{ij})]}{\sum_{l=1}^n f_l(t_{ij})}$  for  $\mu_0 = E(y|x, t, z)$ .

**Proof.**  $\hat{\beta}_{Ke}$  is found as  $\hat{\beta}_{Ke} = (\tilde{X}^T V^{-1} \tilde{X})^{-1} \tilde{X}^T V^{-1} \tilde{y}$  given by Eq. (13) where  $\tilde{X} = X - \omega_x(T)$ . Then,

$$\sqrt{n} \{ \hat{\beta} - \beta \} = \sqrt{n} \left[ \left( \sum_{i=1}^n \tilde{X}_i^T V_i^{-1} \tilde{X}_i \right)^{-1} \left( \sum_{i=1}^n \tilde{X}_i^T V_i^{-1} \tilde{y}_i \right) - \beta \right]$$

$$= \left[ n^{-1} \tilde{X}_i^T V_i^{-1} \tilde{X}_i \right]^{-1} \left[ \sqrt{nn}^{-1} \sum_{i=1}^n \tilde{X}_i^T V_i^{-1} (\tilde{y}_i - (\tilde{X}_i \beta + \tilde{g}(t_i) + \tilde{Z}_i u_i)) \right] = A_n^{-1} (\sqrt{n} \mathfrak{C}_n)$$

where  $A = \lim_{n \rightarrow \infty} A_n = E(\tilde{X}^T V^{-1} \tilde{X})$ ,

$$\begin{aligned} \mathfrak{C}_n &= n^{-1} \sum_{i=1}^n \tilde{X}_i^T V_i^{-1} (y_i - (X_i \beta + \hat{g}(t_i) + g(t_i) - g(t_i) + Z_i u_i)) \\ &= n^{-1} \sum_{i=1}^n \tilde{X}_i^T V_i^{-1} (y_i - (X_i \beta + g(t_i) + Z_i u_i) - (\hat{g}(t_i) - g(t_i))) \\ &= n^{-1} \sum_{i=1}^n \tilde{X}_i^T V_i^{-1} (y_i - \mu_i) - n^{-1} \sum_{i=1}^n \tilde{X}_i^T V_i^{-1} (\hat{g}(t_i, \beta) - g(t_i)) + O_p(1) = \mathfrak{C}_{1n} - \mathfrak{C}_{2n} + O_p(1) \end{aligned}$$

where  $\mu_i = X_i \beta + g(t_i) + Z_i u_i$  and  $\mathfrak{C}_{1n} = n^{-1} \sum_{i=1}^n \tilde{X}_i^T V_i^{-1} (y_i - \mu_i) = n^{-1} \sum_{i=1}^n J_{1i}^T (y_i - \mu_i)$  for  $J_{1i} = V_i^{-1} \tilde{X}_i$  and  $\mathfrak{C}_{2n} = n^{-1} \sum_{i=1}^n \tilde{X}_i^T V_i^{-1} (\hat{g}(t_i, \beta) - g(t_i))$ .

Since the derivation of the asymptotic distribution of  $\sqrt{n} \mathfrak{C}_{1n}$  is easy, we now think the asymptotic distribution of  $\sqrt{n} \mathfrak{C}_{2n}$ . By following [4],  $\mathfrak{C}_{2n}$  is found as  $\mathfrak{C}_{2n} = n^{-1} \sum_{i=1}^n J_{21i}^T (y_i - \mu_i) + \frac{h^2}{2} E \left( \tilde{X}_i^T V_i^{-1} g^{(2)}(t) \right) + O_p(1)$  where  $J_{2i} = [J_{2i1}, \dots, J_{2in}]^T$  for  $J_{2ij} = \frac{[\sum_{k=1}^n \sum_{l=1}^n E(\tilde{X}_k^T V_k^{-1} |t_l = t_{ij}) f_j(t_{ij})]}{\sum_{l=1}^n f_l(t_{ij})}$  and  $v^{kl}$  is the  $(k, l)$ th element of  $V^{-1}$ ,  $f_k(t)$  is density of  $T_k$ . Then,

$$\sqrt{n} \{\hat{\beta} - \beta\} = A^{-1} n^{-1/2} \left\{ \sum_{i=1}^n (J_{1i} - J_{2i}) (y_i - \mu_i) + \frac{(nh^4)^{1/2}}{2} b_1(\beta, g) \right\} + O_p(1)$$

where the bias term  $b_1(\beta, g) = A^{-1} E\{\tilde{X}^T V^{-1} g^{(2)}(t)\}$ . Equivalently,  $\sqrt{n} \{\hat{\beta} - \beta + \frac{h^2 b_1(\beta, g)}{2}\} \overset{D}{\rightsquigarrow} N(0, V_k)$  where  $V_k = A^{-1} E[(J_1 - J_2)^T V_0 (J_1 - J_2)] A^{-1}$  for  $V_0 = Var(y|x, t, z)$  and  $\mu_0 = E(y|x, t, z)$ . Thus, the proof of the Theorem 2.1 is completed.

### 3. A Monte Carlo Simulation Study

In this article, we will investigate a Monte Carlo simulation study to confront the performance of  $\hat{\beta}_{Ke}$  and  $\hat{u}_{Kp}$  in respect of the estimated mean square error (EMSE) and the predicted mean square error (PMSE), respectively. Then mean square error (MSE) of successful models which have minimum EMSE and PMSE values are calculated to demonstrate the best model.

We get  $m = 10, 30, 60$  subjects and  $n_i = 10$  observations for every subject. By following [20], we choose  $\beta = (\beta_1, \dots, \beta_p)^T$  as the normalized eigenvector corresponding to the largest eigenvalue of  $\tilde{X}^T V^{-1} \tilde{X}$  so that  $\beta^T \beta = 1$ . The  $x_{ij}$  covariates are generated from the standard normal distribution and  $t_{ij}$  is generated from uniform distribution ( $U(0,1)$ ). Then, the model is written for  $p = 2$  fixed effects and  $q = 2$  random effects as

$$y_{ij} = \beta_1 x_{ij1} + \beta_2 x_{ij2} + g(t_{ij}) + u_1 + u_2 time_{ij} + \varepsilon_i, \quad u_i \overset{iid}{\sim} N_q(0, \sigma^2 D), \quad \varepsilon_i \overset{iid}{\sim} N_{n_i}(0, \sigma^2 I_{n_i}), \quad i = 1, \dots, m$$

where  $D = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$  is the AR(1) process with  $\rho = 0.30, 0.60, 0.90$  and  $time_{ij}$  indicates the time which was given as the same set of occasions,  $\{time_{ij} = j \text{ for } i = 1, \dots, m, j = 1, \dots, n_i\}$ . To simulate our results we thought both supersmooth and ordinary smooth functions. Hence we create 2 different functions  $g_1$  and  $g_2$  respectively given

as  $S(t) = \begin{cases} -1, & t < -1 \\ t, & t \in [-1,1] \\ 1, & t > 1 \end{cases}$  and  $erf(t) = \frac{2}{\pi} \int_0^t e^{-s^2} ds$  while error function shows supersmooth function,  $S(t)$

function shows ordinary smooth function. For kernel smoothing we use quartic kernel function  $K(t) = \frac{15}{16} (1 - t^2)^2 I(|t| \leq 1)$  and  $h_n^{-1} = 1.2(\ln n)^{0.25}$ .

**Table 1.** The estimated  $\hat{\beta}, \hat{u}$  and the real  $\beta, u$  values

Function	Parameters	n = 100			n = 300			n = 600		
		$\rho = 0.30$	$\rho = 0.60$	$\rho = 0.90$	$\rho = 0.30$	$\rho = 0.60$	$\rho = 0.90$	$\rho = 0.30$	$\rho = 0.60$	$\rho = 0.90$
<b>Partially Linear Mixed Model</b>										
$g_1(t)$	$\beta_1$	-0.1383	-0.1997	-0.1779	0.5754	0.5554	0.5067	0.8250	0.8497	0.7731
		-0.2254*	-0.2258*	-0.2283*	0.6291*	0.6229*	0.6036*	0.7822*	0.7954*	0.7671*
	$\beta_2$	-1.0411	-1.0610	-1.0884	-0.8416	-0.7985	-0.7444	-0.6570	-0.6296	-0.6635
		-0.9743*	-0.9742*	-0.9736*	-0.7774*	-0.7823*	-0.7973*	-0.6231*	-0.6060*	-0.6415*
$u_1$	-56.4091	-87.8467	-54.2041	-0.0772	-0.1754	-0.0542	-0.1229	-0.1970	-0.3093	
	0.7973*	0.7973*	0.7973*	0.9931*	0.9931*	0.9931*	-2.3814*	-2.3814*	-2.3814*	
$u_2$	-58.8106	-88.1406	-52.1021	0.00496	0.1420	0.2580	-0.3973	-0.6017	-0.7800	
	-0.3654*	-0.0287*	0.4413*	0.2255*	0.5351*	0.8607*	-1.1995*	-1.8357*	-2.3649*	
$g_2(t)$	$\beta_1$	-0.1444	-0.1693	-0.1764	0.5657	0.5373	0.5031	0.7778	0.7937	0.7724
		-0.2254*	-0.2258*	-0.2283*	0.6195*	0.5968*	0.6289*	0.7715*	0.7806*	0.7735*
	$\beta_2$	-1.0363	-1.0880	-1.0700	-0.8335	-0.7994	-0.7533	-0.6752	-0.6439	-0.6585
		-0.9743*	-0.9742*	-0.9736*	-0.7850*	-0.8024*	-0.7775*	-0.6362*	-0.6251*	-0.6338*
$u_1$	-63.0259	-79.1291	-15.3698	-0.0345	0.0026	-0.1398	-0.1075	0.0158	-0.1465	
	0.7973*	0.7973*	0.7973*	0.9931*	0.9931*	0.9931*	-2.3814*	-2.3814*	-2.3814*	
$u_2$	-65.3904	-79.4588	-13.3486	0.0556	0.1601	0.2464	-0.4011	-0.5902	-0.7792	
	-0.3654*	-0.0287*	0.4413*	0.2255*	0.5351*	0.8607*	-1.1995*	-1.8357*	-2.3649*	
<b>Partially Linear Model</b>										
$g_1(t)$	$\beta_1$	-0.2087	-0.2341	-0.1724	0.5766	0.5346	0.5063	0.8058	0.7752	0.7208
		-0.2254*	-0.2258*	-0.2283*	0.6114*	0.6298*	0.6060*	0.7784*	0.7566*	0.7456*
	$\beta_2$	-1.0755	-1.0553	-1.0845	-0.8228	-0.8107	-0.7385	-0.6601	-0.7056	-0.7395
		-0.9743*	-0.9742*	-0.9736*	-0.7913*	-0.7767*	-0.7954*	-0.6278*	-0.6539*	-0.6664*
$u_1$	-90.2572	-95.4439	-77.2981	-0.1160	-0.0520	-0.1182	-0.0695	-0.1697	-0.2989	
	0.7973*	0.7973*	0.7973*	0.9931*	0.9931*	0.9931*	-2.3814*	-2.3814*	-2.3814*	
$u_2$	-92.6514	-95.7318	-75.1554	0.0450	0.1496	0.2557	-0.3843	-0.6070	-0.8016	
	-0.3654*	-0.0287*	0.4413*	0.2255*	0.5351*	0.8607*	-1.1995*	-1.8357*	-2.3649*	
$g_2(t)$	$\beta_1$	-0.1541	-0.1692	-0.2404	0.5682	0.5542	0.5147	0.7918	0.8142	0.6971
		-0.2254*	-0.2258*	-0.2283*	0.6071*	0.6203*	0.6363*	0.7658*	0.7953*	0.7353*
	$\beta_2$	-1.0571	-1.0746	-1.0626	-0.8326	-0.8039	-0.7587	-0.6698	-0.6155	-0.7268
		-0.9743*	-0.9742*	-0.9736*	-0.7947*	-0.7843*	-0.7714*	-0.6431*	-0.6062*	-0.6777*
$u_1$	-38.7551	-62.1012	-52.8449	-0.0588	-0.1286	-0.1139	-0.0553	-0.0715	-0.1227	
	0.7973*	0.7973*	0.7973*	0.9931*	0.9931*	0.9931*	-2.3814*	-2.3814*	-2.3814*	
$u_2$	-41.0706	-62.3697	-50.7557	0.0548	0.1475	0.2480	-0.3856	-0.5977	-0.7806	
	-0.3654*	-0.0287*	0.4413*	0.2255*	0.5351*	0.8607*	-1.1995*	-1.8357*	-2.3649*	
<b>Linear Mixed Model</b>										
$g_1(t)$	$\beta_1$	-0.1946	-0.1908	-0.2080	0.5691	0.5240	0.5144	0.7811	0.7967	0.7921
		-0.2254*	-0.2258*	-0.2283*	0.6167*	0.6100*	0.6275*	0.7675*	0.7764*	0.7786*
	$\beta_2$	-1.1265	-1.0509	-1.0505	-0.8322	-0.7817	-0.7523	-0.6636	-0.6746	-0.6294
		-0.9743*	-0.9742*	-0.9736*	-0.7872*	-0.7924*	-0.7581*	-0.6410*	-0.6303*	-0.6276*
$u_1$	-71.1029	-78.1316	-33.6736	-0.1175	-0.1044	-0.1270	-0.0545	-0.1857	0.0943	
	0.7973*	0.7973*	0.7973*	0.9931*	0.9931*	0.9931*	-2.3814*	-2.3814*	-2.3814*	
$u_2$	-73.5222	-78.4585	-31.6024	0.0490	0.1550	0.2521	-0.3860	-0.6089	-0.7374	
	-0.3654*	-0.0287*	0.4413*	0.2255*	0.5351*	0.8607*	-1.1995*	-1.8357*	-2.3649*	
$g_2(t)$	$\beta_1$	-0.1385	-0.1923	-0.2115	0.5774	0.5551	0.5047	0.7847	0.7872	0.7638
		-0.2254*	-0.2258*	-0.2283*	0.6302*	0.6138*	0.6357*	0.7520*	0.7719*	0.7609*
	$\beta_2$	-1.0250	-1.0495	-1.0657	-0.8485	-0.8072	-0.7500	-0.7276	-0.6632	-0.7011
		-0.9743*	-0.9742*	-0.9736*	-0.7765*	-0.7995*	-0.7719*	-0.6592*	-0.6357*	-0.6489*
$u_1$	-27.8043	-52.6802	-9.3439	-0.1053	-0.0816	-0.0999	-0.1330	-0.1856	-0.2027	
	0.7973*	0.7973*	0.7973*	0.9931*	0.9931*	0.9931*	-2.3814*	-2.3814*	-2.3814*	
$u_2$	-30.1215	-52.9815	-7.2430	0.0509	0.1576	0.2511	-0.3952	-0.6185	-0.7855	
	-0.3654*	-0.0287*	0.4413*	0.2255*	0.5351*	0.8607*	-1.1995*	-1.8357*	-2.3649*	

\* demonstrates the real values.

The experiment is replicated 500 times by producing response variable. We compare our model with the partially linear models and the linear mixed models under the AR(1) process. The estimated  $\hat{\beta}, \hat{u}$  and the real  $\beta, u$  values are compared in Table 1.

The EMSE for any estimator  $\hat{\beta}$  of  $\beta$  and the PMSE for any predictor  $\hat{u}$  of  $u$  are computed for each  $m, n_i, \rho$ , and 500 replicated experiments, respectively, as

$$EMSE(\hat{\beta}) = \frac{1}{500} \sum_{r=1}^{500} (\hat{\beta}_{(r)} - \beta)^T (\hat{\beta}_{(r)} - \beta) \text{ and } PMSE(\hat{u}) = \frac{1}{500} \sum_{r=1}^{500} (\hat{u}_{(r)} - u)^T (\hat{u}_{(r)} - u)$$

where the subscript ( $r$ ) demonstrates to the  $r$ th replication and the performances are given in Table 2.

**Table 2.** The EMSE and PMSE values

Function	$n = 300$			$n = 600$			
	$\rho = 0.30$	$\rho = 0.60$	$\rho = 0.90$	$\rho = 0.30$	$\rho = 0.60$	$\rho = 0.90$	
<b>Partially Linear Mixed Model</b>							
$g_1(t)$	EMSE	<b>0.0081</b>	<b>0.0132</b>	0.0298	<b>0.0037</b>	0.0045	<b>0.0059</b>
	PMSE	<b>1.1813</b>	1.4730	<b>1.3950</b>	<b>5.8783</b>	6.5728	<b>7.0004</b>
$g_2(t)$	EMSE	0.0093	0.0173	0.0293	0.0050	0.0054	0.0060
	PMSE	<b>1.1511</b>	<b>1.2358</b>	1.7122	5.9866	7.3714	7.6027
<b>Partially Linear Model</b>							
$g_1(t)$	EMSE	0.0096	0.0141	0.0319	0.0039	0.0041	0.0093
	PMSE	1.2036	<b>1.2510</b>	1.5556	6.0365	6.6212	7.0593
$g_2(t)$	EMSE	0.0097	<b>0.0127</b>	<b>0.0259</b>	0.0043	0.0055	0.0105
	PMSE	1.1737	1.3550	1.6076	6.2039	6.8727	7.8799
<b>Linear Mixed Model</b>							
$g_1(t)$	EMSE	0.0092	0.0198	<b>0.0252</b>	0.0052	<b>0.0041</b>	0.0066
	PMSE	1.2895	1.3530	1.4792	6.2763	<b>6.4763</b>	8.9287
$g_2(t)$	EMSE	<b>0.0078</b>	0.0132	0.0289	<b>0.0039</b>	<b>0.0049</b>	<b>0.0054</b>
	PMSE	1.2951	1.3307	<b>1.5378</b>	<b>5.8891</b>	<b>6.4500</b>	<b>7.4195</b>

Then to compare successful models we compute MSE values which are given in Table 3 using the following equation

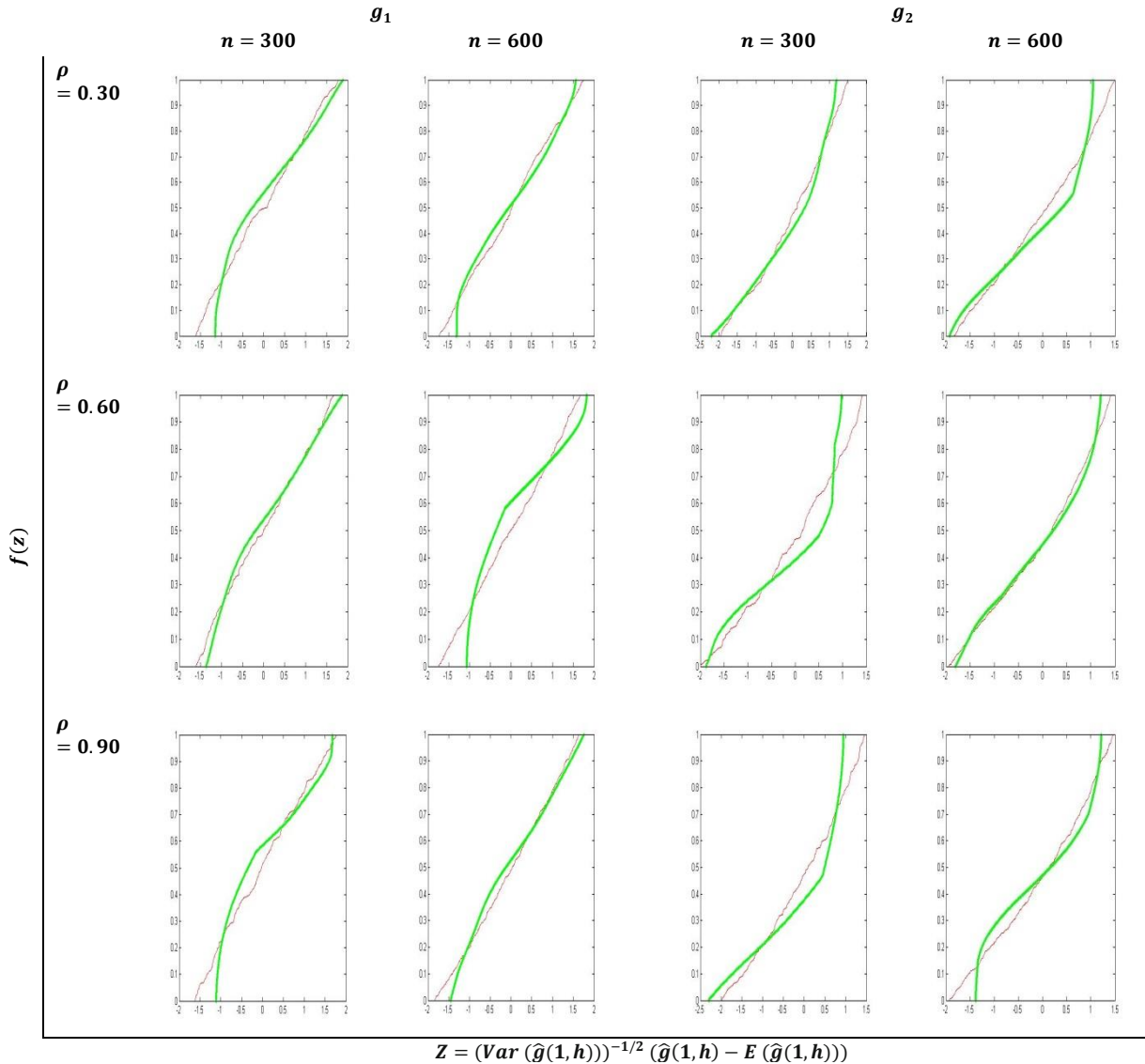
$$MSE = \frac{1}{500} \sum_{r=1}^{500} (\hat{y}_{(r)} - y)^T (\hat{y}_{(r)} - y).$$

**Table 3.** MSE values of models

Function	$n = 300$			$n = 600$		
	$\rho = 0.30$	$\rho = 0.60$	$\rho = 0.90$	$\rho = 0.30$	$\rho = 0.60$	$\rho = 0.90$
<b>Partially Linear Mixed Model</b>						
$g_1(t)$	0.8180	2.2366	6.6358	33.5532	76.8170	132.1835
$g_2(t)$	0.8920	3.0549	6.0778	33.1504	66.2158	120.4612
<b>Linear Mixed Model</b>						
$g_1(t)$	4.7880	9.5019	17.1337	55.4555	102.0681	170.6203
$g_2(t)$	4.5935	9.2114	16.8390	53.7372	101.2066	155.6432

Table 2 and 3 are generated under  $n = 300, 600$  and  $\rho = 0.30, 0.60, 0.90$  conditions. Since the estimated  $\hat{u}$  values are influenced from small samples sizes ( $n = 100$ ), the difference is arised between the estimated and the real values of  $u$ . And then, we have derived the EMSE, PMSE and MSE values for  $n = 300, 600$ .

We also investigate comparison between the finite sample and the asymptotic distributions of our estimator. In Figure 1 the ordinate is probability and the abscissa is  $Z = (Var(\hat{g}(1, h)))^{-1/2} (\hat{g}(1, h) - E(\hat{g}(1, h)))$ . The empirical c.d.f. of the estimator shown as a dashed line agrees very well with the normal c.d.f. shown as a solid line.



$$Z = (\text{Var}(\hat{g}(\mathbf{1}, h)))^{-1/2} (\hat{g}(\mathbf{1}, h) - E(\hat{g}(\mathbf{1}, h)))$$

Figure 1. Comparison of the finite sample and asymptotic distributions of the estimators of functions.

#### 4. Conclusion

This article presents a new approach which is called as Henderson's method approach to obtain the kernel estimator and predictor at the same time in PLMM. After the kernel estimator and the kernel predictor are suggested, asymptotic normality of the proposed estimator is also derived. Then, a Monte Carlo simulation study is done to support the theoretical results in the article.

The simulation study shows that the PLMMs have generally the best and the LMMs have the second best performances in resulting of having smaller EMSE and PMSE values. To compare the performances of the PLMMs and LMMs, we find their MSE values of the response variables. It is easily seen that PLMM has better MSE values which means that results show the superiority of the PLMMs when we think both estimators  $\hat{\beta}$  and  $\hat{u}$  at the same time. We also investigate comparison between the finite sample and the asymptotic distributions of estimator  $\hat{g}$  of PLMM. This demonstrated that empirical c.d.f. of the estimator agrees very well with the normal c.d.f.



## Acknowledgment

The authors would like to thank the editor and anonymous referees for their constructive comments.

## References

- [1] Laird, N. M. and Ware, J. H., Random-effects models for longitudinal data, *Biometrics*, 38 (1982) 963-974.
- [2] Engle, R. F., Granger, C. W. J., Rice, J. and Weiss, A., Semiparametric estimates of the relation between weather and electricity sales, *J. Am. Stat. Assoc.*, 81 (1986) 310-320.
- [3] Severini, T. A. and Staniswalis, J. G., Quasi-likelihood estimation in semi-parametric models, *J. Am. Stat. Assoc.*, 89 (1994) 510-512.
- [4] Lin, X. and Carroll, R. J., Semiparametric regression for clustered data using generalised estimating equations, *J. Am. Stat. Assoc.*, 96 (2001) 1045–1056.
- [5] Wang, N., Marginal nonparametric kernel regression accounting for within-subject correlation, *Biometrika*, 90 (2003) 43–52.
- [6] Rice, J. A., Convergence rates for partially splined models, *Stat. Probabil. Lett.*, 4 (1986) 204–208.
- [7] Speckman, P. E., Regression analysis for partially linear models, *J. R. Stat. Soc. B.*, 50 (1988) 413–436.
- [8] Opsomer, J. D. and Ruppert, D., A root-n consistent backfitting estimator for semiparametric additive modeling, *J. Comput. Graph. Stat.*, 8 (1999) 715–732.
- [9] Zeger, S. L. and Diggle, P. J., Semi-parametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters, *Biometrics*, 50 (1994) 689–699.
- [10] Qin, G. Y. and Zhu, Z. Y., Robust estimation in generalized semiparametric mixed models for longitudinal data, *J. Multivar. Anal.*, 98 (2007) 1658–1683.
- [11] Qin, G. Y. and Zhu, Z. Y., Robustified maximum likelihood estimation in generalized partial linear mixed model for longitudinal data, *Biometrics*, 65 (2009) 52–59.
- [12] Li, Z. and Zhu, L., On variance components in semiparametric mixed models for longitudinal data, *Scand. J. Stat.*, 37 (2010) 442–457.
- [13] Zhang, D., Generalized linear mixedmodels with varying coefficients for longitudinal data, *Biometrics*, 60 (2004) 8–15.
- [14] Liang, H., Generalized partially linear mixed-effects models incorporating mismeasured covariates, *Ann. I. Stat. Math.*, 61 (2009) 27–46.
- [15] Taavoni, M. and Arashi, M., Kernel estimation in semiparametric mixed effect longitudinal modeling, *Stat. Papers*, (2019) <https://doi.org/10.1007/s00362-019-01125-8>.
- [16] Henderson, C. R., Estimation of genetic parameters. *Ann. I. Stat. Math.*, 21 (1950) 309–310 .
- [17] Henderson, C. R., Kempthorne, O., Searle, S. R., and Von Krosig, C. N., Estimation of environmental and genetic trends from records subject to culling, *Biometrics*, 15 (1959) 192–218.
- [18] Gilmour, A. R., Thompson, R., and Cullis, B. R., Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models, *Biometrics*, 51 (1995) 1440–1450.
- [19] Searle, S. R., Matrix Algebra Useful for Statistics, New York: John Wiley and Sons, 1982.
- [20] Newhouse, J. P. and S. D. Oman., An evaluation of ridge estimators, *Rand Corporation R-716-PR*, (1971) 1–28.