Article

Application of an Intensive Longitudinal Functional Model with Multiple Time Scales in Objectively Measured Children's Physical Activity

Mostafa Zahed [1,*]**, Trent Lalonde** [2] **and Maryam Skafyan** [3]

1   Department of Mathematics and Statistics, East Tennessee State University, Johnson City, TN 37614, USA
2   Colorado Department of Human Services, Denver, CO 80203, USA; trent.lalonde@state.co.us
3   Department of Applied Statistics and Research Methods, University of Northern Colroado, Greeley, CO 80639, USA; skaf9868@bears.unco.edu
*   Correspondence: zahedm@etsu.edu

Abstract: This study proposes an intensive longitudinal functional model with multiple time-varying scales and subject-specific random intercepts through mixed model equivalence that includes multiple functional predictors, one or more scalar covariates, and one or more scalar covariates. An estimation framework is proposed for estimating a time-varying coefficient function that is modeled as a linear combination of time-invariant functions with time-varying coefficients. The model takes advantage of the information structure of the penalty, while the estimation procedure utilizes the equivalence between penalized least squares estimation and linear mixed models. A number of simulations are conducted in order to empirically evaluate the process. In the simulation, it was observed that mean square errors for functional coefficients decreased with increasing sample size and level of association. Additionally, sample size had a greater impact on a smaller level of association, and level of association also had a greater impact on a smaller sample size. These results provide empirical evidence that ILFMM estimates of functional coefficients are close to the true functional estimate (basically unchanged). In addition, the results indicated that the AIC could be used to guide the choice of ridge weights. Moreover, when ridge weight ratios were sufficiently large, there was minimal impact on estimation performance. Studying two time scales is important in a wide range of fields, including physics, chemistry, biology, engineering, economics, and more. It allows researchers to gain a better understanding of complex systems and processes that operate over different time frames. Consequently, studying physical activities with two time scales is critical for advancing our understanding of human performance and health and for developing effective strategies to optimize physical activity and exercise programs. Therefore, the proposed model was applied to analyze the physical activity data from the Active Schools Institute of the University of Northern Colorado to determine what kind of time-structure patterns of activities could adequately describe the relationship between daily total magnitude and kids' daily and weekly physical activity.

Keywords: functional data analysis; intensive longitudinal data; functional penalized regression; decomposition based penalty; generalized ridge regression; activity pattern; accelerometery

MSC: 62P99

1. Introduction

Functional Data Analysis (FDA) deals with the analysis and theory of data that are in the form of functions, images and shapes, or more general objects [1,2]. The atom of functional data is a function, where for each subject in a random sample one or several functions are recorded. While the term "functional data analysis" was coined by [1,3], the history of this area is much older and dates back to Greven et al. (2020) [4] and Rao et al. (1950) [5]. There are several ways that the data can called functional. For

instance, in archaeological studies the form of a three-dimensional image of each bone can be presented as functional data [6].

FDA has many applications in areas such as biostatistics, environmental monitoring, finance, and image analysis. For example, in biostatistics, FDA can be used to study the patterns of gene expression over time or to model the progression of a disease. In finance, FDA can be used to analyze the dynamics of financial markets over time, and to develop models for predicting future stock prices or other financial variables [7].

The functional data comes in many forms; however, their definition is a based on functions which are often smooth curves [8]. The smoothness is a characteristic that filter the noises inside the raw data as efficiently as possible [9]. They are several approaches for representing functional data in smoothing form: basis expansion, least-squares, and roughness penalty. One of basis expansions is based on B-splines. In this approach, both functional coefficient and functional predictor expand in the form of B-splines and thereafter the regularized estimate of functional coefficient will be applied [10]. Müller (2005) [11] expressed that the regression coefficient can be expressed in orthonormal basis determined by the eigenfunctions of the covariance of functional regressor. Goldmsith [12] in 2011, combined both approaches introduced by Ramsay [10] and Müller [11]. This method has been considered in this study for smoothness and it is called roughness penalty because it penalizes fits that are too rough [13].

In many cases, data are observed longitudinally and collected over time. Longitudinal data is epansion of a time series, which is a sequence of data points that are ordered based on time, such that each data point corresponds to a specific moment in time [14]. Longitudinal data analysis can provide more accurate estimates of the effects of interventions or treatments over time [15] .The mixed-effects regression models have been used to study the longitudinal data [16]. These models allow us to make subject-specific conclusions and present statements about individuals changes, trends, or effects over time.

Sometimes in practice, the data has both characteristics of functional and longitudinal. One of the case can be when the outcome is not functional; however, covariates are collected functionally and longitudinally over one-time scale. People often use two approaches. One is based on function models where functional covariates have been smoothed according to roughness penalty and it is called Longitudinal Penalized Functional Regression (LPFR) [17]. The other one is built on according to principal component analysis for longitudinal functional data [18]. One of the main disadvantage of both approaches is the functional covariates are remained constant over time. But sometime in practice, we might have situation where functional covariates and scalar outcome have been changed over time which is considered in this study. Longitudinal functional models have not been studied to incorporate longitudinal functional for multiple time scales; however; Kundu et al. (2016) [19] published their work about the same data situation for one time-scale. Therefore, this study is an extension of Kund's work to incorporate multilevel longitudinal functional covariates.

Studying two time scales is important in a wide range of fields, including physics, chemistry, biology, engineering, economics, and more. It allows researchers to gain a better understanding of complex systems and processes that operate over different time frames [20].

One of the key benefits of studying two time scales is that it can help researchers identify the underlying mechanisms that contribute to the behavior of a system. For example, in physics, studying the behavior of particles over different time scales can help researchers understand how those particles interact with each other and how their interactions lead to the formation of larger structures.

In chemistry, studying chemical reactions over different time scales can help researchers identify the intermediates and transition states that are involved in the reaction, and help them design more efficient and effective reactions [21].

Studying two time scales can also help researchers design and optimize systems for specific applications. For example, in engineering, studying the behavior of materials

over different time scales can help researchers design materials with specific mechanical, electrical, or chemical properties.

Overall, studying two time scales is critical for advancing our understanding of complex systems and processes, and for developing new technologies and applications that can improve our lives.

Studying physical activities with two time scales is important because it allows us to understand how different processes operate at different timescales and how they interact with each other. Physical activities involve a wide range of processes that occur at different timescales, from short-term processes such as muscle activation and joint movements, to longer-term processes such as training adaptations and recovery [22].

By studying physical activities with two time scales, researchers can better understand the underlying mechanisms that contribute to performance and health outcomes [23]. For example, studying the short-term and long-term effects of exercise on muscle function and metabolism can help us understand how exercise training improves overall health and athletic performance [24].

Furthermore, studying physical activities with two time scales can help inform the development of effective training programs and interventions. By understanding how different processes interact over time, we can design training programs that are optimized for different goals, such as improving strength or endurance, and minimizing the risk of injury or overtraining [25].

Overall, studying physical activities with two time scales is critical for advancing our understanding of human performance and health and for developing effective strategies to optimize physical activity and exercise programs [26].

The proposed model was appropriate to apply on the data collected by the Active Schools Institute of the University of Northern Colorado. They were interested in the associations between daily and weekly activity profile, as measured by accelerometers, and academic and behavioral outcomes. To address these interests, students from one primary school in a suburban area in the western United States wore accelerometers during the school day for 5 consecutive days over 4 different weeks of the year. Corresponding academic and behavioral data were also obtained.

The data collected from the Active Schools Institute assumed daily student activity crossed within weeks. Due to nature of this data, establishing a multilevel mixed-effects regression model, including demographics or teacher effects, and accelerometer wear-time was essential. Hence, the proposed model was suited to apply on the physical activity data set.

As a consequence, this article addresses the following questions:

According to Section 2, the following question has been addressed: How can we represent a longitudinal functional regression model with time-varying regression functions for multiple time scales when time components are crossed?

Under Section 2.3, the following question is addressed: How can the mean parameters of the propsed model be estimated?

A response to this question can be found in Section 5: How can the model and estimation be implemented in software?

A detailed answer to this question can be found in Section 3: How do the proposed model compare to similar existing models with single time scales in terms of model MSE, across sample sizes and levels of association?

2. Intensive Longitudinal Functional Model with Multiple Time Scales

The intensive longitudinal method represents sequences of repeated measurements frequently recorded to characterize a separate change process for each subject [27].

Intensive longitudinal data refers to data that is collected repeatedly and frequently over a relatively short period of time, often at a high frequency (e.g., several times a day). This type of data is particularly useful for studying processes that unfold over time, such as mood fluctuations, stress reactivity, or medication effects [27,28].

Intensive longitudinal data can be collected using a variety of methods, including ecological momentary assessment (EMA), experience sampling methods (ESM), ambulatory monitoring, and diary or journal methods [29]. These methods typically involve asking participants to report on their experiences, thoughts, and behaviors at multiple points throughout the day, either through self-reports on electronic devices (such as smartphones or smartwatches), or by filling out paper or online forms [30].

The analysis of intensive longitudinal data requires specialized statistical methods that can capture the dynamic nature of the data, such as time-series analysis, multilevel modeling, and dynamic systems modeling. These methods allow researchers to examine how variables change over time, how they are related to each other, and how they are influenced by external factors [31].

Intensive longitudinal data has been used in a variety of fields, including psychology, medicine, and public health, to study phenomena such as mood disorders, substance use, physical activity, and sleep patterns.

This study considers a longitudinal model where the time scales of functional predictors cross. Additionally, the proposed model is suitable when data is collected frequently over a short period of time, often at a high frequency (e.g., several times a day). As a consequence, the proposed model may be referred to as an Intensive Longitudinal Functional Model with Multiple Time Scales (ILFMM) [32].

A specific estimation process is also included in the ILFMM. Estimating parameters is performed using a generalized ridge estimate. As a result of this estimation process, a linear mixed model representation can be derived. By using a mixed model approach, tuning parameters can be automatically selected. It is also possible to estimate parameters by fitting a linear mixed model based on this linear mixed model representation. It is worthwhile to note that the model presented has some advantages. There is no restriction on the time course of regression functions. In addition, it is possible to incorporate the structure of the regression function directly into the estimation process [19]. The approach described by Kundu et al. (2016)[19] was extended here to incorporate longitudinal functional predictors spanning multiple time scales.

2.1. Statistical Model of ILFMM

To represent a longitudinal functional regression model with time-varying regression functions for multiple time scales when time components are crossed, assume $\mathbf{X}$ presents univariate predictors, $\mathbf{W}_{it_w}$ denotes a predictor function for the ith subject at timepoint t_w, and $\mathbf{D}_{it_d}$ denotes a predictor function for the ith subject at timepoint t_d, where $i = \{1, 2, \ldots, N\}$, $t = \{t_1, t_2, \ldots, t_{n_i}\}$, and longitudinal-time points t can be decomposed in terms of t_w and t_d, where each t_k ($k = \{1, 2, \ldots, n_i\}$) corresponds to one value each from t_{w_k} and t_{d_k}. It is assumed that each observed predictor is sampled at the same p locations, $s_1, \ldots, s_p \in \Omega$, and there are an equal number of observations per subject (i.e., n_i are equal). Let $\mathbf{W}_{it_w} := \left[\mathbf{W}_{it_w}(s_1), \cdots, \mathbf{W}_{it_w}(s_p) \right]^\top$ and $\mathbf{D}_{it_d} := \left[\mathbf{D}_{it_d}(s_1), \cdots, \mathbf{D}_{it_d}(s_p) \right]^\top$ be the $p \times 1$ vector of values sampled from the realized functions $\mathbf{W}_{it_w}$ and $\mathbf{D}_{it_d}$, respectively. Then, the observed data have a form of

$$\{y_{it}; \mathbf{X}_{it}; \mathbf{W}_{it_w}; \mathbf{D}_{it_d}\},$$

where y_{it} is a scalar outcome, $\mathbf{X}_{it}$ is a $K \times 1$ column vector of measurements on K scalar predictors, $\mathbf{W}_{it_w}$ is the sampled predictor from the ith subject corresponding to time t_w, and $\mathbf{D}_{it_d}$ is the sampled predictor from the ith subject corresponding to time t_d. The longitudinal functional linear model with the scalar outcome and functional predictor along with multiple time scales could be written as

$$y_{it} = \mathbf{X}_{it}^\top \boldsymbol{\beta} + \int_\Omega \boldsymbol{W}_{it_w}(s)\boldsymbol{\gamma}(t_w, s)ds + \int_\Omega \boldsymbol{D}_{it_d}(s)\boldsymbol{\eta}(t_d, s)ds + \mathbf{z}_{it}^\top \mathbf{b}_i + \varepsilon_{it_w t_d}, \tag{1}$$

where $\gamma \equiv \gamma(t_w, s)$ are functional coefficients at time t_w, $\eta \equiv \eta(t_d, s)$ are functional coefficients at time t_d, $\mathbf{b}_i$ is the vector of r random effects pertaining to subject i, and $\varepsilon_{it_w t_d}$ denote the subject-specific random effect and random error term, respectively.

Similar to a linear mixed model with time-related slope for longitudinal data, it is assumed that γ and η can be decomposed into several time-invariant component functions $\gamma_0, \cdots, \gamma_W$ and $\gamma_0, \cdots, \gamma_D$ as follows:

$$\gamma(t_w, s) = \gamma_0(s) + f_1(t_1)\gamma_1(s) + \ldots + f_W(t_W)\gamma_W(s), \tag{2}$$

$$\eta(t_d, s) = \eta_0(s) + g_1(t_1)\eta_1(s) + \ldots + g_D(t_D)\eta_D(s), \tag{3}$$

where $f_1, \ldots, f_W$ and $g_1, \ldots, g_D$ are functions of t_w and t_d, respectively. Additionally, γ and η are functions of s. In general, $\gamma(t_w, s)$ and $\eta(t_d, s)$ have three components: t_w, t_d, and s. The time component t_w enters into γ through $f_1(t_1), \ldots, f_W(t_W)$ and the functional component enters into $\gamma(t_w, s)$ through $\gamma_0(s), \cdots, \gamma_W(s)$. Additionally, the time component t_d enters into $\eta(t_d, s)$ through $g_1(t_1), \ldots, g_D(t_D)$ and the functional component enters into $\eta(t_d, s)$ through $\eta_0(s), \cdots, \eta_D(s)$. The model is flexible with any function of t_w and t_d with $f(0) = 0$ (e.g., $f(t_w) = t_w$ or $t_w e_w^t$ and $f(t_d) = t_d$ or $t_d e_d^t$). Additionally, $\varepsilon_{it_w t_d} \sim N(0, \sigma_\varepsilon^2)$ and $\mathbf{b}_i$ is distributed as $N(0, \Sigma_{\mathbf{b}_i})$. It is assumed that $\varepsilon_{it_w t_d}$ and $\mathbf{b}_i$ are independent, $\varepsilon_{it_w t_d}$ and $\varepsilon_{i't_w't_d'}$ are independent whenever $i \neq i'$ or $t_w \neq t_w'$ or $t_d \neq t_d'$ or all, and $\mathbf{b}_i$ and $\mathbf{b}_{i'}$ are independent if $i \neq i'$.

In Model (1) $\mathbf{X}_{it}^\top \boldsymbol{\beta}$ is the standard fixed effect. In addition, $\mathbf{Z}_{it}\mathbf{b}_i$ is the standard multilevel random effect and $\int_\Omega \mathbf{W}_{it_w}(s)\gamma(t_w, s)ds$ is the subject and time specific functional effect corresponding to level w. Here $\int_\Omega \mathbf{D}_{it_d}(s)\eta(t_d, s)ds$ is the subject and time specific functional effect corresponding to level d. And both the variances and covariances across time are assumed to be constant as $\text{Var}(y_{it_w t_d}) = \sigma_b^2 + \sigma_\varepsilon^2$ and $\text{Cov}(y_{it_w t_d}, y_{it_w' t_d'}) = \sigma_b^2$, where $t_w \neq t_w'$ and $t_d \neq t_d'$. This implies a compound symmetry assumption for the variances and covariances.

Model (1) can be written based on Equations (2) and (3) as

$$
\begin{aligned}
y_{it} &= \mathbf{X}_{it}^\top \boldsymbol{\beta} + \int_\Omega \mathbf{W}_{it_w}(s)(\gamma_0(s) + f_1(t_1)\gamma_1(s) + \ldots + f_W(t_W)\gamma_W(s))ds \\
&+ \int_\Omega \mathbf{D}_{it_d}(s)(\eta_0(s) + g_1(t_1)\eta_1(s) + \ldots + g_D(t_D)\eta_D(s))ds \\
&+ \mathbf{z}_{it}^\top \mathbf{b}_i + \varepsilon_{it_w t_d}.
\end{aligned}
\tag{4}
$$

A primary goal in developing Model (4) is the estimation of $\boldsymbol{\beta}$, $\gamma_w = [\gamma_w(s_1), \ldots, \gamma_w(s_p)]^\top$, $0 \leq w \leq W$, and $\eta_d = [\eta_d(s_1), \ldots, \eta_d(s_p)]^\top$, $0 \leq d \leq D$.

2.2. Matrix Form of ILFMM

In the following an explanation will be given demonstrating how the proposed model can be written in matrix form. For this purpose, the observed values and the design matrix of fixed effect of Model (1) can be considered as:

$$
\mathbf{y} = \begin{pmatrix} y_{1t_1} \\ y_{1t_2} \\ \vdots \\ y_{1t_{n_1}} \\ y_{2t_1} \\ y_{2t_2} \\ \vdots \\ y_{2t_{n_2}} \\ \vdots \\ y_{Nt_1} \\ y_{Nt_2} \\ \vdots \\ y_{Nt_{n_N}} \end{pmatrix} \in \mathbb{R}^{\sum_{i=1}^N n_i \times 1}, \tag{5}
$$

$$
\mathbf{X} = \begin{pmatrix} X_{1t_1} & X_{1t_2} & \dots & X_{1t_{n_1}} \\ X_{2t_1} & X_{2t_2} & \dots & X_{2t_{n_2}} \\ \vdots & \vdots & \dots & \vdots \\ X_{Nt_1} & X_{Nt_2} & \dots & X_{Nt_{n_N}} \end{pmatrix} \in \mathbb{M}_{\sum_{i=1}^N n_i \times k}(\mathbb{R}). \tag{6}
$$

It is important to point out that it is assumed there are an equal number of observations per subject (i.e., n_i are equal). In addition, the matrices of functional coefficients, $\mathbf{W}$ and $\mathbf{D}$, have the form

$$
\mathbf{W} = \begin{pmatrix} \mathbf{W}_1 \\ \vdots \\ \mathbf{W}_N \end{pmatrix} \in \mathbb{M}_{\sum_{i=1}^N n_i \times (W+1)p}(\mathbb{R}) \quad \text{and} \quad \mathbf{D} = \begin{pmatrix} \mathbf{D}_1 \\ \vdots \\ \mathbf{D}_N \end{pmatrix} \in \mathbb{M}_{\sum_{i=1}^N n_i \times (D+1)p}(\mathbb{R}), \tag{7}
$$

where

$$
\mathbf{W}_i = \begin{pmatrix} \mathbf{W}_{it_1}^\top & f_1(t_1)\mathbf{W}_{it_1}^\top & \cdots & f_W(t_1)\mathbf{W}_{it_1}^\top \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{W}_{it_{n_i}}^\top & f_1(t_{n_i})\mathbf{W}_{it_{n_i}}^\top & \cdots & f_W(t_{n_i})\mathbf{W}_{it_i}^\top \end{pmatrix} \tag{8}
$$

and

$$
\mathbf{D}_i = \begin{pmatrix} \mathbf{D}_{it_1}^\top & g_1(t_1)\mathbf{D}_{it_1}^\top & \cdots & g_D(t_1)\mathbf{D}_{it_1}^\top \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{D}_{it_{n_i}}^\top & g_1(t_{n_i})\mathbf{D}_{it_{n_i}}^\top & \cdots & g_D(t_{n_i})\mathbf{D}_{it_i}^\top \end{pmatrix}. \tag{9}
$$

So Model (4) would be

$$
\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\gamma} + \mathbf{D}\boldsymbol{\eta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}, \tag{10}
$$

where $\boldsymbol{\beta}$ is the associated coefficient vector, $\boldsymbol{\gamma} = \left[\gamma_0^\top, \gamma_1^\top, \cdots, \gamma_W^\top\right]^\top \in \mathbb{R}^{(W+1)p \times 1}$ and $\boldsymbol{\eta} = \left[\eta_0^\top, \eta_1^\top, \cdots, \eta_D^\top\right]^\top \in \mathbb{R}^{(D+1)p \times 1}$ are coefficient vectors of function predictors, $\mathbf{b} \in \mathbb{R}^{rN \times 1}$ is a vector of random effects and $\mathbf{Z} \in \mathbb{M}_{\sum_{i=1}^N n_i \times rN}(\mathbb{R})$ is the corresponding design matrix.

In this framework, the estimation approach is an extension of generalized ridge regression that takes the form of a penalty to the longitudinal setting in a manner that allows the estimated regression functions $\gamma(t_w, s)$ and $\eta(t_d, s)$ to vary with time.

2.3. Estimation Based on the Mixed Model Representation

The goal of this section is to estimate parameters based on the matrix form of Model (4). This model was described in Section 2.2 and it can be written as:

$$Y = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\gamma} + \mathbf{D}\boldsymbol{\eta} + \mathbf{Z}b + \boldsymbol{\epsilon}, \tag{11}$$

where $\mathbf{X}$, $\mathbf{W}$, $\mathbf{D}$, and $\mathbf{Z}$ are known matrices, $\boldsymbol{\beta}$ is an unobservable vector of fixed effects, $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ are unobservable functional vectors of fixed effects of two time scales, and b is an unobservable vector of random effects with $E(b) = \mathbf{0}, Cov(b) = \mathbf{B}$, and $Cov(b, \boldsymbol{\epsilon}) = \mathbf{0}$. Let $Cov(\boldsymbol{\epsilon}) = \mathbf{R}$ and $E(\boldsymbol{\epsilon}) = \mathbf{0}$, then $\mathbf{Z}b + \varepsilon$ is distributed as $N(\mathbf{0}, \mathbf{V})$ because $E(\mathbf{Z}b + \varepsilon) = E(\varepsilon) = \mathbf{0}$ and $Cov(\mathbf{Z}b + \varepsilon) = \mathbf{ZBZ'} + \mathbf{R} \equiv \mathbf{V}$. In addition, it is assumed that $\mathbf{B}$ and $\mathbf{R}$ are known.

The well-known mixed model equation, developed by Henderson (1950) [33], can be applied at this time. Henderson's mixed model equation is similar in spirit to normal equations, but the mixed model equation simultaneously provides Best Linear Unbiased Estimators (BLUEs) and Best Linear Unbiased Predictors (BLUPs) by estimating the variance component $\mathbf{V}$.

Theorem 1. *If $\left(\hat{\boldsymbol{\alpha}}' \quad \hat{\mathbf{b}}'\right)$ is a solution to the mixed model*

$$\mathbf{Y} = \mathbf{C}\boldsymbol{\alpha} + \mathbf{Z}b + \boldsymbol{\epsilon}, \tag{12}$$

where $\mathrm{C} = [\mathbf{X} \quad \mathbf{W} \quad \mathbf{D}]$, $\boldsymbol{\alpha} = (\boldsymbol{\beta}' \quad \boldsymbol{\gamma}' \quad \boldsymbol{\eta}')$, $Cov(\mathbf{b}) = \mathbf{B}$, and $Cov(\boldsymbol{\epsilon}) = \mathbf{R}$, then $\mathbf{C}\hat{\boldsymbol{\alpha}}$ is a BLUE of $\mathbf{C}\boldsymbol{\alpha}$ and simultaneously $\hat{\mathbf{b}}$ is a BLUP of $\mathbf{b}$ if $\mathbf{B}$ and $\mathbf{R}$ are known and also by estimating the variance component $\mathbf{V}$.

Theorem 1 demonstrates how the mixed Model (11) can provide BLUEs and BLUPs simultaneously. Appendix A.1 includes the proof of the Theorem 1.

2.4. Roughness Penalties Estimate

In this section a roughness penalties estimate based on the matrix form of Model (4) is discussed, which was described in Section 2.2. As mentioned, the mixed model can be written as:

$$Y = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\gamma} + \mathbf{D}\boldsymbol{\eta} + \mathbf{Z}b + \boldsymbol{\epsilon}, \tag{13}$$

where $\mathbf{X}$, $\mathbf{W}$, $\mathbf{D}$, and $\mathbf{Z}$ are known matrices, $\boldsymbol{\beta}$ is an unobservable vector of fixed effects, $\boldsymbol{\gamma}$ and $\boldsymbol{\eta}$ are unobservable functional vectors of fixed effects for two time scales, and b is an unobservable vector of random effects with $E(b) = \mathbf{0}, Cov(b) = \mathbf{B}$, and $Cov(b, \boldsymbol{\epsilon}) = \mathbf{0}$. Let $Cov(\boldsymbol{\epsilon}) = \mathbf{R}$ and $E(\boldsymbol{\epsilon}) = \mathbf{0}$, then $\mathbf{Z}b + \varepsilon$ is distributed as $N(\mathbf{0}, \mathbf{V})$ (see Section 2.3).

It is also assumed that $\mathbf{V}$ is known.

Theorem 2. *For each $w = 0, \ldots, W$ and $d = 0, \ldots, D$, let $\mathcal{L}_d$ and $\mathcal{L}_w$ be penalty operators for $\boldsymbol{\eta}_w$ and $\boldsymbol{\gamma}_d$ with τ_w^2 and λ_d^2 as the associated tuning parameters. A generalized ridge estimate associated with Model (13) for minimizing $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\eta}$*

$$\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{W}\boldsymbol{\gamma} - \mathbf{D}\boldsymbol{\eta}\|_{\mathbf{v}^{-1}}^2 \quad + \quad \tau_0^2\|\boldsymbol{\eta_0}\|_{\mathcal{L}_0^\top \mathcal{L}_0}^2 + \cdots + \tau_W^2\|\boldsymbol{\eta}_W\|_{\mathcal{L}_W^\top \mathcal{L}_W}^2$$
$$+ \quad \lambda_0^2\|\boldsymbol{\gamma_0}\|_{\mathcal{L}_0^\top \mathcal{L}_0}^2 + \cdots + \lambda_D^2\|\boldsymbol{\gamma}_D\|_{\mathcal{L}_D^\top \mathcal{L}_D}^2, \tag{14}$$

is obtained as

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\eta}} \\ \hat{\boldsymbol{\gamma}} \end{pmatrix} = \left(\mathbf{C}^\top \mathbf{V}^{-1}\mathbf{C} - \mathbf{M}\right)^{-1} \mathbf{C}^\top \mathbf{V}^{-1}\mathbf{y}, \tag{15}$$

where

$$\mathrm{C} = [\mathbf{X} \quad \mathbf{W} \quad \mathbf{D}], \mathbf{M} = blockdiag\left\{0, \mathcal{L}_\mathbf{\Gamma}^\top \mathcal{L}_\mathbf{\Gamma}, \mathcal{L}_\mathbf{H}^\top \mathcal{L}_\mathbf{H}\right\},$$

$$\mathcal{L}_{\boldsymbol{\Gamma}} = \textit{blockdiag}\ \{\tau_0 \mathcal{L}_0, \cdots, \tau_W \mathcal{L}_W\},$$

and

$$\mathcal{L}_{\mathrm{H}} = \textit{blockdiag}\ \{\gamma_0 \mathcal{L}_0, \cdots, \gamma_D \mathcal{L}_D\}.$$

In Theorem 2, an expression for the generalized ridge estimate of Model (13) will be derived. The results of both theorems illustrated that the ridge estimate is BLUP from an equivalent mixed model; hence, the estimation procedure takes advantage of the equivalence between penalized least-squares estimation and linear mixed model representation. An Appendix A.2 contains the proof of the Theorem 2.

2.5. Selection of Time-Structure in $\gamma(t_w, s)$ and $\eta(t_d, s)$

In the proposed approach there is freedom to choose a time-structure for both $\gamma(t_w, s)$ and $\eta(t_d, s)$. The matter of selecting appropriate time-structure in $\gamma(t_w, s)$ and $\eta(t_d, s)$ are analogous, in essence, to that of selecting time structure in a linear mixed-effects model.

For example, selection of time-structure $\eta_1 + t_d \eta_2$ or $\eta_1 + t_d \eta_1 + t_d^2 \eta_2$ or $\eta_1 + t_d \eta_1 + t_d^2 \eta_2 + t_d^3 \eta_3$ is similar to selection of variables in mutiple linear regression models based on linear mixed-effects model.

When comparing different time structures, it is common to apply Akaike Information Criterion (AIC). The formula $-2l + 2p$, where l refers to a Restricted, or a Residual Maximum Likelihood (REML) estimation of mixed models and p refers to the number of parameters of the model, for information criterion (AIC) was used to determine time-structure in $\gamma(t_w, s)$ and $\eta(t_d, s)$.

2.6. Selection of a, b, c, and d for Decomposition Penalties

The set of values $\{1, 2, 3\}$ for each of a, b, c, and d are considered in this study. Those are treated as weights of a trade-off between preferred and non-preferred subspaces with $ab =$ constant and $cd =$ constant.

Additionally, REML is used to estimate tuning parameters when b and d are considered as fixed values and a grid search determines a and c values. These values are used to jointly select the tuning parameters.

Then, the selected values of a and c can be applied to the preferred subspace. The formula for information criterion (AIC) $-2l + 2p$ creates values for a and c, where l refers to a Restricted, or a Residual Maximum Likelihood (REML) estimation of mixed models and p refers to the number of parameters of the model.

3. Simulation Steps

In this section, the process of simulating the proposed model is explored. Without loss of generality, it was assumed that $b = 1$ and $d = 1$; however, a and c could be varied on the set values of $\{1, 2, 3\}$. The largest values of a and c indicate there was much prior information in the estimation process. The following was used to generate the response data for 100 subjects by simulating functional covariates W in 4 time points ($t_w = 1, 2, 3, 4$) and functional covariates D in 5 time points ($t_d = 1, 2, 3, 4, 5$):

$$
\begin{aligned}
y_{it} = \ & \beta_0 + \beta_1 t + \int_0^1 W_{it_w}(s)\gamma(s, t_w)ds \\
& + \int_0^1 D_{it_d}(s)\eta(s, t_d)ds + b_i + \varepsilon_{it_w t_d}, \quad i = 1, \cdots, 100,
\end{aligned}
\tag{16}
$$

where $\gamma(s, t_w) = \gamma_1(s) + t_w \gamma_2(s)$, $\eta(s, t_d) = \eta_1(s) + t_d \eta_2(s)$. Functional covariates W and D have the following form: $W = (W_1 \quad W_2)$ and $D = (D_1 \quad D_2)$. Consequently, a rewritten of Model (16) would be

$$
\begin{aligned}
y_{it} &= \beta_0 + \beta_1 t + \int_0^1 W_{1_{it_w}}(s)\gamma_1(s)\,ds + \int_0^1 t_w W_{2_{it_w}}(s)\gamma_2(s)\,ds \\
&+ \int_0^1 D_{1_{it_d}}(s)\gamma_1(s)\,ds + \int_0^1 t_d D_{2_{it_d}}(s)\gamma_2(s)\,ds \\
&+ b_i + \varepsilon_{it_w t_d}, \quad i = 1, \cdots, 100.
\end{aligned}
\tag{17}
$$

Functional covariates W were generated by considering three sets of sampling points: wide, moderate, and narrow corresponding to one, five, and two bumps, respectively [19].

As an example, the first set of 100 columns in W was represented as follows:

a. We have recorded all of the observed values for the first subject at the first time point in the first 100×1 row of the W matrix.

b. As a result, all observation values for W at all time points and all subjects in each following row have been arranged. This is so that each row represents an observation value for W from a specific time point and subject. In accordance with y, the rows were ordered by time and subject.

Having constructed the first set of 100 columns, we next proceeded to construct the second set of 100 columns by multiplying the first 100 columns by t_w.

Since it was impossible to generate the curves with functional form only by considering discrete observations at different points of the domain interval, it was necessary to consider some set of sampling points on the interval $[0,1]$ [19]. To simulate the true functional forms of the curves, prior knowledge about the functional variable was applied. Due to this, the observed value of W was calculated considering the degrees of curvature at the wide, moderate, and narrow bumps [19]. White noise was added to the predictor functions to account for the instrumental measurement noise [19].

Functional covariates D were generated by considering three sets of sampling points: wide, moderate, and narrow corresponding to one, five, and two bumps respectively as $H_{\text{wide}} = \{0.40\}, H_{\text{moderate}} = \{0.10, 0.25, 0.65, 0.75, 0.80\}$, and $H_{\text{narrow}} = \{0.50, 0.85\}$. The first set of 100 columns in D were represented similar to process of presenting W as discussed ealier. Consequently, the observed value of D was generated in the same way as W.

The functional coefficients γ_1, γ_2 were generated according to [19] considering different degrees of curvature at the wide, moderate, and narrow bumps. But η_1, and η_2 were generated as follows with bumps centered at $H_{\eta_1} = \{0.10, 0.45, 0.75\}$, and $H_{\eta_2} = \{0.25, 0.65\}$,

$$
\begin{aligned}
\eta_1(s) &= \sum_{h \in H_{\eta_1}} a_w(h) \exp\left[-c_w(h) * \left(\frac{s-h}{100}\right)^2\right], \\
\eta_2(s) &= \sum_{h \in H_{\eta_n}} a_n(h) \exp\left[-c_n(h) * \left(\frac{s-h}{100}\right)^2\right], \forall s \in [0,1],
\end{aligned}
\tag{18}
$$

where $a_w(h)$, $c_w(h)$, $a_n(h)$, and $c_n(h)$ correspond to amplitude and degree of curvature respectively. In fact, "bump" is the same as the "turning" of the polynomial graph. While the proper term is local maximum, the term "bump" referred to the "turning" of functional coefficients. The amplitude is the height from the centerline to the peak (dip) of the polynomial, and it measures the height from highest to lowest points divided by two. The amplitude values as $0.1, 0.4, 0.5$, and 0.6 were considered to simulate functional coefficients [19]. Another term that played an important role in simulating was degree of curvature. It was referred to the degree of a polynomial [34]. The amplitude and curvature used to generate the functional coefficients η_1, and η_2 are specified in Table 1.

Table 1. The amplitude and curvature parameters for η_1 and η_2.

h	Amplitude $a_w(h)$	Curvature $c_w(h)$	Amplitude $a_m(h)$	Curvature $c_m(h)$
5	0.15			
15			0.15	500
30	−0.10	250		
50				
70	0.10	250		
80			−0.15	1000
90				
95				

The second and third columns are parameters for η_1, and the last two columns are for η_2.

The model was applied to analyze the physical activities of kids with logitudonal functional data with two-time scale. The results suggested that $\beta_0 = 0.06$ and $\beta_1 = 0.4$ were reasonable choices for fixed parameters in the simulation study. Having made these choices the fixed effect, the random effect, and errors were generated and are shown in Table 2.

Table 2. Choices for fixed effects, random effect, and errors in the simulation study.

Fixed Effect	β_0	0.06
Fixed Effect	β_1	0.04
Random Effect	b_i	$\sim \mathcal{N}(0, 0.05^2)$
Errors	$\epsilon_{it_w t_d}$	$\sim \mathcal{N}(0, 0.02^2)$

Throughout this simulation the following decomposition penalties, an extension of Kundu et al. (2016) [19] for multiple time scales, were considered: $b = 1$ and $d = 1$ while a and c were varied over $1, 2$, and 3, also structured penalties were $\mathcal{L}_{\mathbf{Q}_w} = 10^a (\mathbf{I} - P_{\mathbf{Q}_w}) + bP_{\mathbf{Q}_w}$ and $\mathcal{L}_{\mathbf{Q}_d} = 10^c (\mathbf{I} - P_{\mathbf{Q}_d}) + dP_{\mathbf{Q}_d}$. R 3.5.2 GUI 1.70 El Capitan build (7612) and RStudio Version 1.1.383 were used to analyze the simulated data. The method discussed in this section was implemented in the R package `nlme` via the `lme` function. The R code has been uploaded to GitHub and it is viewable at https://github.com/Mostafa2020-bit/PackageILFMM since November 2020.

About 100 data were generated for two different sample sizes(100 and 200) along with two different R^2 (0.6 and 0.9).

The primary interests were in the estimation of the functional coefficients and the measurement of the squared distance between the fitted and the true value of **y**. Estimation error was summarized in terms of the Mean Squared Errors (MSE) of the functional coefficients. Also, the MSE of $\tilde{\mathbf{y}}$ was obtained.

With respect to the MSE of both the functional covariates and the response, we investigated the effect of increasing the sample size and R^2 on these values of *MSE*. If these were increased, would $MSE(\gamma_1)$, $MSE(\gamma_2)$, $MSE(\eta_1)$, $MSE(\eta_2)$, and $MSE(\tilde{\mathbf{y}})$ decrease or not? In this scenario, the following criteria were computed: $MSE(\gamma_1) = \|\gamma_1 - \tilde{\gamma_1}\|^2 = \sum_{s=1}^{p} \int_\Omega (\gamma_{1s} - \tilde{\gamma}_{1s})^2 ds$, where $\tilde{\gamma_1}$ is the estimate of γ_1 and each observed predictor γ_1 is sampled at the same p locations, $s_1, \ldots, s_p \in \Omega$. Similarly, $MSE(\gamma_2)$, $MSE(\eta_1)$, $MSE(\eta_2)$ were computed; however, the calculation of $MSE(\tilde{\mathbf{y}})$ was different because of its discrete characteristics and it was computed as $MSE(\tilde{\mathbf{y}}) = \frac{\|\mathbf{y} - \tilde{\mathbf{y}}\|^2}{N} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{n_i} (y_{it_j} - \tilde{y}_{it_j})^2}{N}$, where $\tilde{\mathbf{y}}$ denotes the model fitted values of the true $\mathbf{y}$.

The $MSE(\gamma_1)$, $MSE(\gamma_2)$, $MSE(\eta_1)$, and $MSE(\eta_2)$ could provide empirical evidence on whether the estimates are close to the corresponding true function or not.

To determine the effect of the decomposed penalty weights a, b, c, and d without loss of generality, b and d were considered a fixed value of 1. Then a and c were increased up to 3 on an exponential scale (i.e., $\{1, 2, 3\}$). The reason for this was to determine which

combination of *a*, *b*, *c*, and *d* values would improve the estimation of γ_1, γ_2, η_1, and η_2. In this manner, an evaluation of estimation performance would remain almost unchanged.

Decomposition penalty values were selected according to the $-2l + 2p$ formula for AIC. It was expected that minimized AIC for selected decomposition penalty values leads to a minimized MSE for different sample sizes and R^2s. AIC values were computed for all combinations of *a* and *c* values. Consequently, the simulation parameters were *a* and *c* values, the four sample sizes, and the four R^2s.

4. Simulation Conclusions

The simulation study illustrated the potential advantage of an Intensive Longitudinal Functional Model with Multiple time scales (ILFMM) estimate in exploiting an informed, structured penalty. The simulation suggested that as the sample size and R^2 increased, the MSE for both functional coefficients and the fitted values decreased, Figure 1 and Table 3.
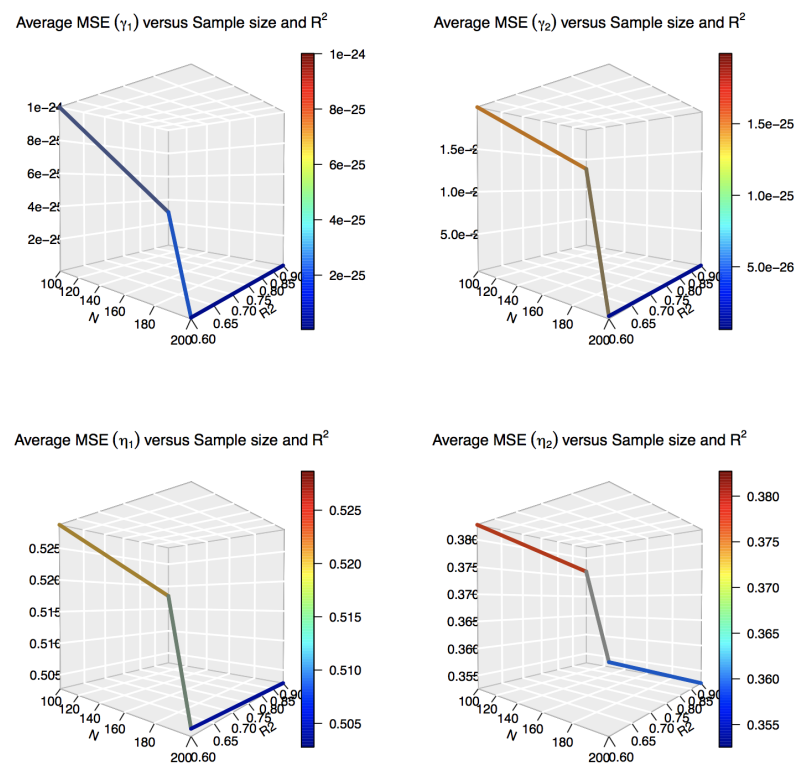

Figure 1. Average MSE of functional coefficients versus simulation cases (3-D plot).

Table 3. Averages $MSE(\gamma_1)$, $MSE(\gamma_2)$, $MSE(\eta_1)$, $MSE(\eta_2)$, and $MSE(\tilde{y})$ for four cases of simulation.

Case	N	R^2	$MSE(\gamma_1)$	$MSE(\gamma_2)$	$MSE(\eta_1)$	$MSE(\eta_2)$	$MSE(\tilde{y})$
Case 1	100	0.6	1.000522×10^{-24}	1.991303×10^{-25}	0.5035614	0.3827327	0.01625859
Case 2	100	0.9	2.912084×10^{-25}	1.343186×10^{-24}	0.5027712	0.3765218	0.01378987
Case 3	200	0.6	8.249432×10^{-27}	8.167412×10^{-27}	0.5286615	0.3614322	0.01240092
Case 4	200	0.9	3.903604×10^{-27}	6.183226×10^{-27}	0.5178613	0.3525342	0.01182925

Furthermore, sample size had a larger impact for smaller R^2, and also R^2 had a greater impact for smaller sample size, Figure 2.
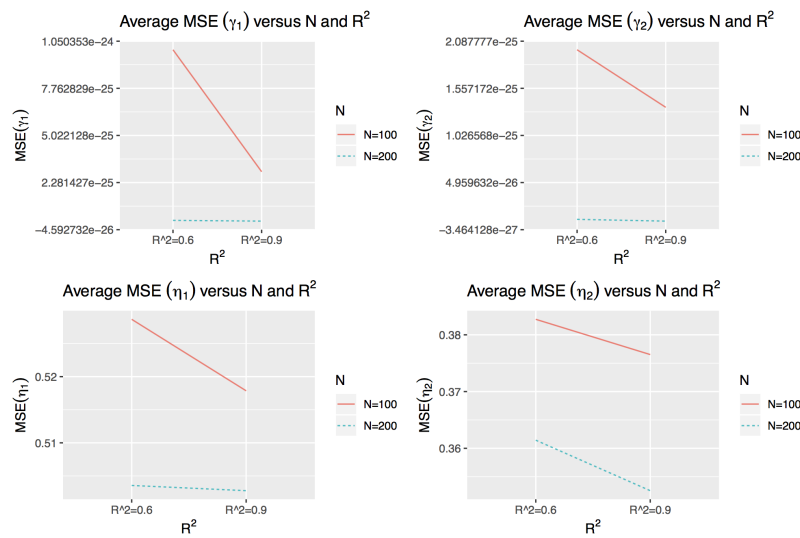
Figure 2. Average MSE of functional coefficients versus simulation cases (2-D plot).

These results provided empirical evidence that the ILFMM estimates of functional coefficients were close to the true functional estimate (basically unchanged). In other words, the estimation of the functional coefficients rose to the level of the estimation of true functions. These results suggested that AIC could guide the choice of ridge weights, Figures 3 and 4.

The results for AIC are displayed graphically in Figure 3. It has been demonstrated, when c = 3, AIC is minimized among all cases. Therefore, the choice of decomposition penalties is a = 3, b = 1, c = 3, and d = 1. Also with sufficiently large values there was minimal impact on the estimation performance.

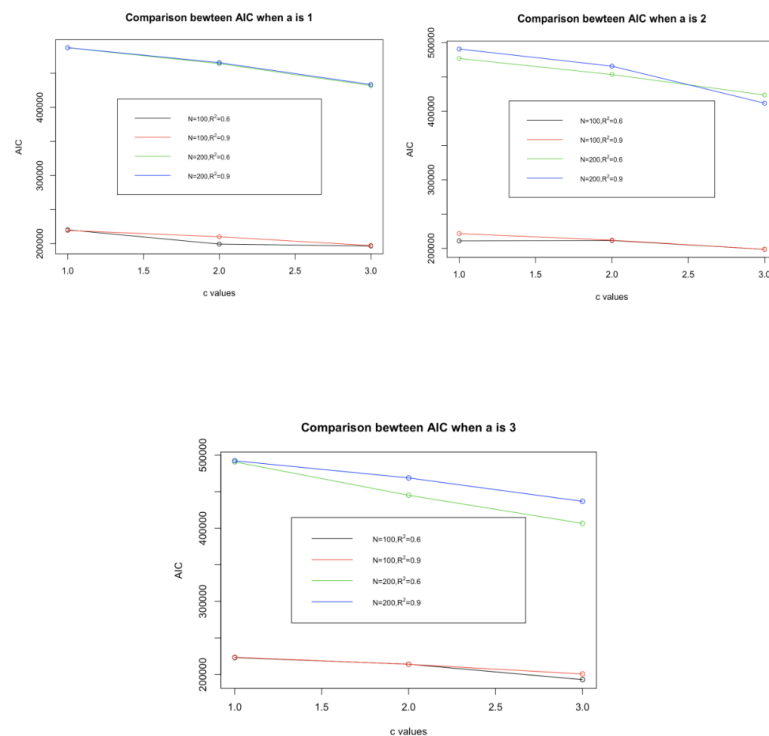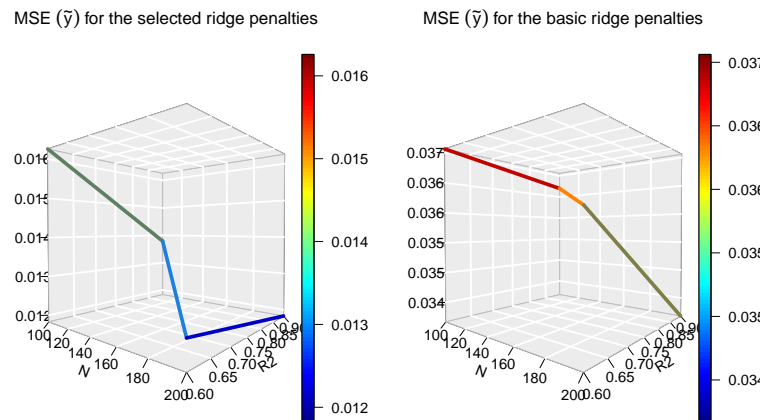**Figure 3.** Comparison of AIC for selection of c ($a = 3, b = 1, d = 1$).

Figure 4. Average MSEs of $MSE(\bar{y})$ for the defined selection of ridge penalties.

In addition, the model with the selected ridge penalties had the lowest MSE values of $\bar{y}$ as compared to the model with the defined selection of ridge penalties, Figure 4 and Table 3. These results implied that the model with the selected ridge penalties would perform better than the model with the basic selection of ridge penalties.

Furthermore, for smaller sample sizes and R^2s, the ILFMM estimate may oversmooth the estimated regression function. However, by increasing the sample size to 200 or R^2 to 0.9, it was observed that the average ILFMM estimate of the functional coefficients approached the true functions.

All in all, the model with ridge penalties $a = 3, b = 1, c = 3$, and $d = 1$ appears to perform better than the remaining models with other selections of the ridge penalties.

In the absence of a model with two time scales for comparison, it was decided to evaluate the effect of sample size and R^2 on MSE of functional coefficients, the fitted and AIC values, which would indicate whether the patterns are consistent with those seen in previous models.

Kundu's approach was based on a single time scale. The results of the proposed approach was lined up with consideration to a single time scale comparing AIC values [19] (i.e., single versus multiple time scales).

Also, Kudu concluded the selected ridge penalties had the lowest MSE values as when compared to the model with the basic selection. This demonstrated that the selected ridge penalties were reasonable. In addition, when the sample size and R^2 were increased, MSE values were decreased. Furthermore, sample size had a larger impact for smaller R^2, and also R^2 had a greater impact for smaller sample size [19].

5. Physical Activity Study Application

The proposed method could be implemented in a `ILFMM()` (i.e., **I**ntensive **L**ongitudinal **F**unctional **M**odel with **M**ultiple time scales) package in R. The author has written the `ILFMM()` package for a longitudinal functional model with multiple time scales along with scalar outcome, multiple functional predictors, one or more scalar covariates, and subject-specific random intercepts through mixed model equivalence. This package also works for a longitudinal functional model with a single time scale. The `ILFMM()` package has been uploaded to GitHub and it is viewable at https://github.com/Mostafa2020-bit/PackageILFMM sicne Novebmer 2020.

It is more common in the Physical Activity (PA) field to calculate the amount of PA a child gets in a single day, rather than a week. So this might be why it has been difficult to find information on weekly amounts of PA. Consequently, it is standard in PA projects to get week data by taking the average of all weekdays [35,36].

Up until about five years ago most physical activity researchers only measured and reported five days of activity at a time. Most reviews of handling accelerometer data focus on the epochs, the device or the type of outcome reported (Moderate-to-Vigorous Physical Activity (MVPA) or light PA) [37].

In PA field, the daily accelerometer-assessed time in MVPA over a week is usually calculated as the mean of data over 7 days. For participants with less than 7 valid days of data, the following formula is common to use for standardizing measurements to one week for all participants [37]:

$[(5\times$ mean daily weekday MVPA time $+2\times$ mean daily weekend MVPA time $)]/7$.

The dataset was stored in a secure password-protected Super R machine server utilized by IT at University of Northern Colorado for the Active Schools Institute. The author had permission from the Active Schools Institute to access the dataset. The data used included the five-second records from all students in school 1 for all weeks 1, 2, 3, and 4 over five days of school time. The standard approach was used to get week data.

For all weeks the "Axis-1" measurements ranged approximately from 0 to 2647.0, with a mean of 20.75, where "Axis-1" referred to the physical activities over x-axis coordinate.

For week 1, many observations were irrelevant (i.e., zeros recorded late in the evening), so observations were truncated to consider only the time between 8:00 AM and 3:30 PM. For example, by considering the first day of data collection the number of observations were 1,059,840. Further limitation on the number of observations came from the restriction of the hours from 8:00 AM to 3:30 PM on the first day. This resulted in a final data set of 496,800 observations. This is around 5900 observations per student. This number is close but a little high for 5-second observations. It would be anticipated that there would be 5400 observations per student (12 per minute, 60 min. per hour, 7.5 h). However, 8 students show twice as many observations, suggesting that they have two days labeled as 9 November 2016. For all other weeks there was no such a problem.

For all weeks, the Stata default for time-stamps was New Year's 1960. All official time stamps showed dates around this time, but for analysis, dates were inherited from the "Date" variable that the research team included.

Time stamps were translated in an unexpected way. For Day 1, all activity appeared to begin at 1:00 AM for each student. We assumed this should have been 8:00 AM and translated the time-stamps accordingly. For all other days, recording appeared to begin at 5:00 PM, but all activity appeared to begin at midnight. We again assumed these times should have been 8:00 AM and translated time-stamps accordingly. We then removed all observations outside the hours of 8:00 AM through 3:30 PM. Plots of the five-second data are usually "smoothed" to show the true patterns in the data. Smoothing is a method of taking the data in "windows of time" to estimate means or regression trends. The size of the window has a strong impact on the graphic produced but is subjective, and it is usually determined by a span value.

In designing the matrix for weekly activities, 162,000 data points per subject were considered for each week. The researcher chose 162,000 data points because it was large enough to capture physical activity patterns from 8:00 AM to 3:30 PM. This implies 684,000 observations per subject for four weeks (i.e., 162,000×4 = 684,000). And since 80 subjects were considered, the total number of observations were 51,840,000. These design matrices are graphically shown for 11 subjects in Figures 5 and 6. It provides us with information on the overall patterns of physical activity on a daily and weekly basis.
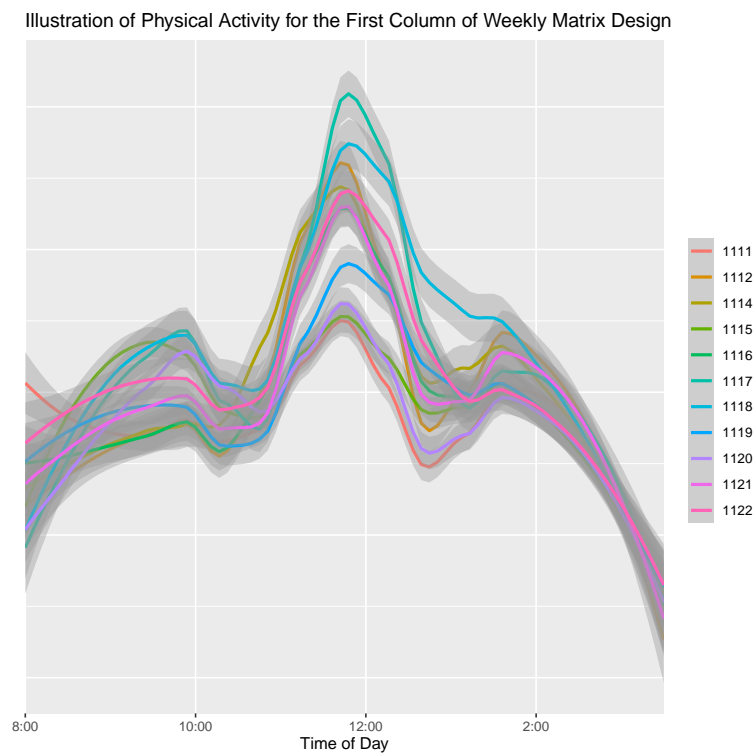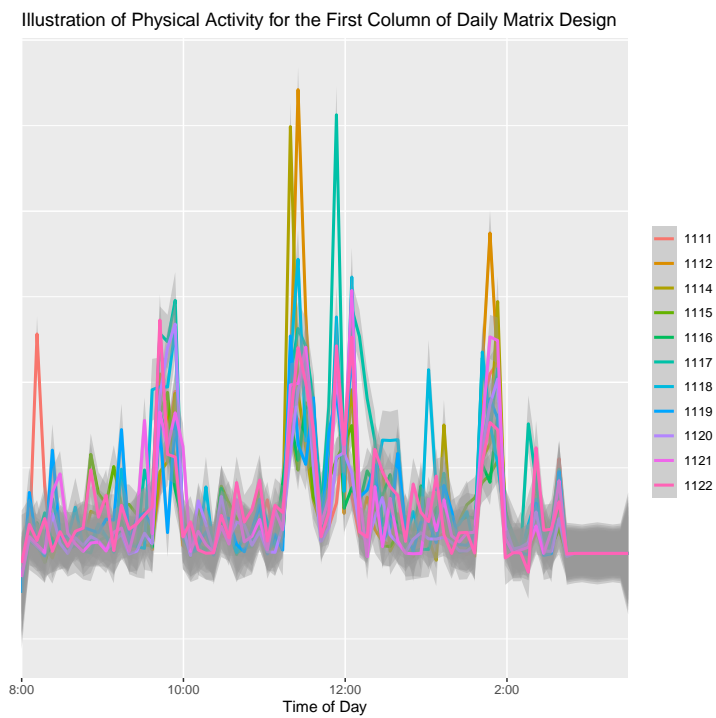
Figure 5. First column of weekly activity for 11 subjects over five days.
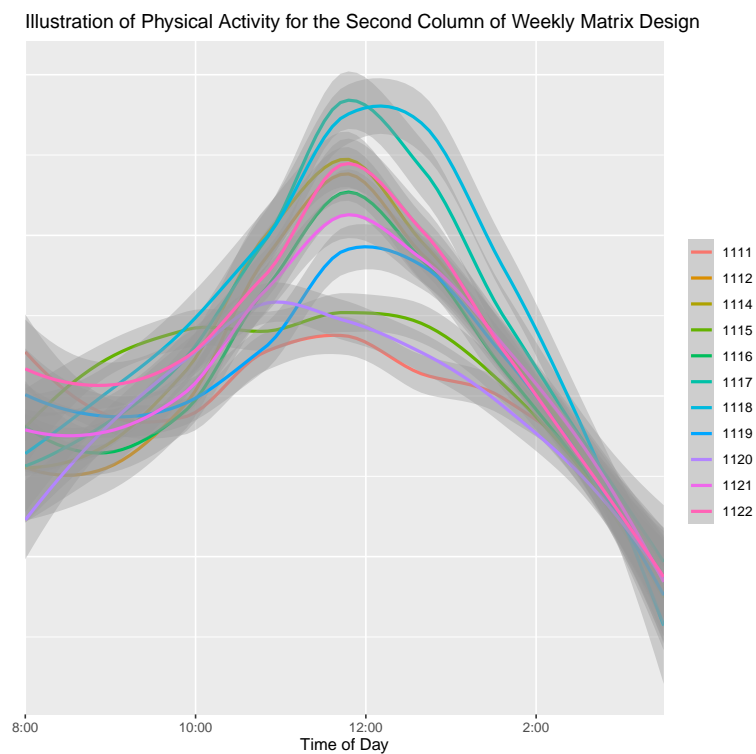
Figure 6. Second column of weekly activity for 11 subjects over five days.

For daily physical activity analysis, we chose 162,000 because it was large enough to capture physical activity patterns every five seconds from 8:00 AM to 3:30 PM. This implies 3,240,000 observations per subject for five days across four weeks (i.e., $162{,}000 \times 5 \times 4 = 3{,}240{,}000$). And since 80 subjects were considered the total number of observations were 259,200,000. These design matrices are graphically shown in Figures 7 and 8.

Illustration of Physical Activity for the First Column of Daily Matrix Design

Figure 7. First column of daily activity for 11 subjects across four weeks.

Illustration of Physical Activity for the Second Column of Daily Matrix Design
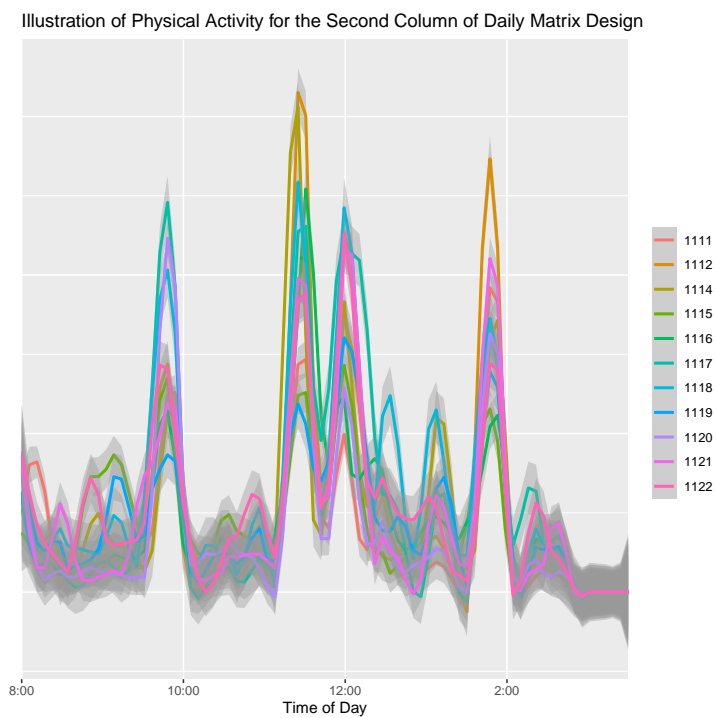
Figure 8. Second column of daily activity for 11 subjects across four weeks.

The proposed model with two-time scales along with $a = 3$, $b = 1$, $c = 3$, and $d = 1$ as ridge penalties was applied to real data as

$$y_{it_w} = \beta_0 + \beta_p Teacher_p + \int_0^1 \boldsymbol{W}_{it_w}(s)\gamma(s,t_w)ds + \int_0^1 \boldsymbol{D}_{it_d}(s)\boldsymbol{\eta}(s,t_d)ds + b_i + \varepsilon_{it_w t_d}, \quad (19)$$

where $\gamma(s,t_w) = \gamma_1(s) + t_w\gamma_2(s)$, $\eta(t_d,s) = \eta_1(s) + t_d\eta_2(s) + t_d^2\eta_2(s) + t_d^3\eta_4(s)$, $i = 1,\cdots,100$, and $p = 1,2,3,4,5,6$. The expanded form of Model (19) would be

$$
\begin{aligned}
y_{it_w t_d} =\ & \beta_0 + \beta_2 Teacher_2 + \beta_3 Teacher_3 + \beta_4 Teacher_4 \\
+\ & \beta_5 Teacher_5 + \beta_6 Teacher_6 \\
+\ & \int_0^1 \boldsymbol{W_{1}}_{it_w}(s)\boldsymbol{\gamma_1}(s)ds + \int_0^1 t_w \boldsymbol{W_{2}}_{it_w}(s)\boldsymbol{\gamma_2}(s)ds \\
+\ & \int_0^1 \boldsymbol{D_{1}}_{it_d}(s)\boldsymbol{\eta_1}(s)ds + \int_0^1 t_d \boldsymbol{D_{2}}_{it_d}(s)\boldsymbol{\eta_2}(s)ds + \int_0^1 t_d^2 \boldsymbol{D_{3}}_{it_d}(s)\boldsymbol{\eta_3}(s)ds \\
+\ & \int_0^1 t_d^3 \boldsymbol{D_{4}}_{it_d}(s)\boldsymbol{\eta_4}(s) \\
+\ & b_i + \varepsilon_{it_w}.
\end{aligned}
\quad (20)
$$

The ILFMM estimates were obtained as BLUP assuming $\varepsilon_{it_w t_d} \sim N(0,\sigma_\varepsilon^2)$ and the subject-specific random intercepts $b_i \sim N(0,\sigma_b^2)$. The model was fitted into `ILFMM()` R package and the estimates of σ_ε^2 and σ_b^2 were 0.000320 and 0.7281, respectively.

We compared models based on AIC values with a different selection of weight penalties and time structures (linear, quadratic, and cubic). The results are summarized in Table 4.

Table 4. Comparison of AIC for selection of ridge weights and time structures.

Model	Time Structure for Days	Time Structure for Weeks	a	c	AIC
Model 1	$\eta_1(s)$	$\gamma_1(s)$	1	1	2,096,978
Model 2	$\eta_1(s)$	$\gamma_1(s) + t\gamma_2(s)$	1	1	1,978,254
Model 3	$\eta_1(s) + t\eta_2(s)$	$\gamma_1(s)$	1	1	1,947,720
Model 4	$\eta_1(s) + t\eta_2(s)$	$\gamma_1(s) + t\gamma_2(s)$	1	1	1,947,720
Model 5	$\eta_1(s) + t\eta_2(s) + t^2\eta_3(s)$	$\gamma_1(s)$	1	1	1,947,720
Model 6	$\eta_1(s) + t\eta_2(s) + t^2\eta_3(s)$	$\gamma_1(s) + t\gamma_2(s)$	1	1	1,947,720
Model 7	$\eta_1(s) + t\eta_2(s) + t^2\eta_3(s) + t^3\eta_4(s)$	$\gamma_1(s)$	1	1	1,947,720
Model 8	$\eta_1(s) + t\eta_2(s) + t^2\eta_3(s) + t^3\eta_4(s)$	$\gamma_1(s) + t\gamma_2(s)$	1	1	1,947,720
Model 9	$\eta_1(s)$	$\gamma_1(s)$	3	3	1,789,975
Model 10	$\eta_1(s)$	$\gamma_1(s) + t\gamma_2(s)$	3	3	16,898,224
Model 11	$\eta_1(s) + t\eta_2(s)$	$\gamma_1(s)$	3	3	1,574,037
Model 12	$\eta_1(s) + t\eta_2(s)$	$\gamma_1(s) + t\gamma_2(s)$	3	3	1,442,020
Model 13	$\eta_1(s) + t\eta_2(s) + t^2\eta_3(s)$	$\gamma_1(s)$	3	3	1,347,340
Model 14	$\eta_1(s) + t\eta_2(s) + t^2\eta_3(s)$	$\gamma_1(s) + t\gamma_2(s)$	3	3	1,304,740
Model 15	$\eta_1(s) + t\eta_2(s) + t^2\eta_3(s) + t^3\eta_4(s)$	$\gamma_1(s)$	3	3	1,300,120
Model 16	$\eta_1(s) + t\eta_2(s) + t^2\eta_3(s) + t^3\eta_4(s)$	$\gamma_1(s) + t\gamma_2(s)$	3	3	1,235,622

Based on the AIC, Model 16, with a cubic time structure for days and linear time structure for weeks and weight penalties $a = 3$, $b = 1$, $c = 3$, and $d = 1$ appears to be better than the other models.

6. Discussion

The proposed intensive longitudinal functional model with multiple time-varying scales with scalar outcome, multiple functional predictors, one or more scalar covariates, and subject-specific random intercepts through mixed model equivalence has been defined. There were three primary advantages of this framework. First, this model estimated a time-dependent regression function. Second, it was also able to incorporate structural information into the estimation process. And third, it was easily implemented through linear mixed model equivalence.

One of the significant limitations of this study was the hardware needs for the organization and analysis of the data. Storage space to house the data, networking bandwidth to transfer it to and from analytics systems, and computing resources to perform those analytics were very time-consuming.

The physical activity data study results indicate that the proposed method is an appropriate one for the tested application. By using the proposed model, we could investigate what kind of time-structure for activity patterns would adequately describe the relationship between the subjects' daily total magnitude and their weekly physical activities. Likewise, it described this relationship for daily physical activities.

So, the interest of the Active Schools Institute has been addressed through the proposed model. The proposed model concluded that the time structure of weekly activity was linear, but for daily activity was cubic. This conclusion was reasonable because the general patterns of movement throughout the day had more fluctuations than weekly movement. The weekly movements showed more smoothed patterns. Also, several dips in data points around the sedentary category for daily movement were observed. However, for weekly movement, it appeared this sedentary behavior was changed (i.e., it is needed to consider two separate definitions for sedentary behavior for weekly and daily activities). This analysis also gives us a slightly better sense of how movement intensity changes over time for both time scales.

The following are recommendations for future studies. One area of further investigation is the effect of different time structures. In this study, a linear form for simulation was considered; however, other forms of $f(t)$, such as $e^{(t)} - 1$ or $log(t + 1)$ could be explored. A relative improvement in AIC due to the use of these complex structures would be anticipated. For instance, it would be expected to have a minimal AIC associated with $\gamma(t, s) = \gamma_0(s) + [e^{(t)} - 1]\gamma_1(s)$ in comparison to AIC for $\gamma(t, s) = \gamma_0(s) + t\gamma_1(s)$.

In the proposed model, the ridge weights were set as $b = d = 1$ while a and c varied on an exponential scale. Larger values of a and c indicated greater emphasis of prior information on the estimation process. Future study could investigate other scales of ridge weights to determine more accurately and quickly the estimation process.

A possible extension of this work could be to incorporate multiple functional predictors. For example, in PA projects, the physical activity and patterns might be studied over an entire year. This implies there would be three-time scales: days, weeks, and years. Consequently, an extension of the proposed approach would be needed.

Additionally, an interesting and powerful use of the framework underlying the proposed model is when the response variable is binary. In this case when the response variable is binary, different problems arise. Multicollinearity and high dimensionality prejudice the estimation of the model and the interpretation of its parameters. This prejudice can be overcome by using an intensive longitudinal multilevel functional logistic regression model with principal component analysis. This study could be extended to any differentiable Hilbert space for analyzing of image, spatial, and spatial-temporal data.

Abbreviations

The following abbreviations are used in this manuscript:

LPFR	Longitudinal Penalized Functional Regression
ILFMM	Intensive Longitudinal Functional Model with Multiple time scales
BLUEs	Best Linear Unbiasted Estimators
BLUPs	Best Linear Unbiased Predictors
AIC	Akaike Information Criterion
REML	Residual Maximum Likelihood
MSE	Mean Squared Errors
PA	Physical Activity
MVPA	Moderate-to-Vigorous Physical Activity

Appendix A

Appendix A.1. Proof of Theorem 1

Proof. If $\hat{\boldsymbol{\alpha}}$ is a solution to $\mathbf{C}'\mathbf{V}^{-1}\mathbf{C}\boldsymbol{\alpha} = \mathbf{C}'\mathbf{V}^{-1}\mathbf{Y}$, then $\mathbf{C}\hat{\boldsymbol{\alpha}}$ will be a BLUE of $\mathbf{C}\boldsymbol{\alpha}$. To implement this equation it is necessary to know a form for $\mathbf{V}^{-1}$. According to Proposition B.56 page 356 of [38], the inverse of $\mathbf{V}$ in terms of $\mathbf{Z}$, $\mathbf{B}$, and $\mathbf{R}$ is as follows:

$$\mathbf{V}^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}\left[\mathbf{B}^{-1} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\right]^{-1}\mathbf{Z}'R^{-1}.$$

If $\hat{\boldsymbol{\alpha}}$ and $\hat{\mathbf{b}}$ are solutions, then the second row of the normal equation for the mixed model Equation (12) gives

$$\mathbf{Z}'\mathbf{R}^{-1}\mathbf{C}\hat{\boldsymbol{\alpha}} + \left[\mathbf{B}^{-1} + \mathbf{C}'\mathbf{R}^{-1}\mathbf{C}\right]\hat{\mathbf{b}} = \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y} \tag{A1}$$

or

$$\hat{\mathbf{b}} = \left[\mathbf{B}^{-1} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\right]^{-1}\mathbf{Z}'\mathbf{R}^{-1}(\mathbf{Y} - \mathbf{C}\hat{\boldsymbol{\alpha}}). \tag{A2}$$

The first row the normal equation of Equation (12) is

$$\mathbf{C}'\mathbf{R}^{-1}\mathbf{C}\hat{\boldsymbol{\alpha}} + \mathbf{C}'\mathbf{R}^{-1}\mathbf{C}\hat{\mathbf{b}} = \mathbf{C}'\mathbf{R}^{-1}\mathbf{Y}. \tag{A3}$$

Substituting for $\hat{\mathbf{b}}$ gives

$$\mathbf{C}'\mathbf{R}^{-1}\mathbf{C}\hat{\boldsymbol{\alpha}} + \mathbf{C}'\mathbf{R}^{-1}\mathbf{Z}\left[\mathbf{B}^{-1} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\right]^{-1}\mathbf{Z}'\mathbf{R}^{-1}(\mathbf{Y} - \mathbf{C}\hat{\boldsymbol{\alpha}}) = \mathbf{C}'\mathbf{R}^{-1}Y \tag{A4}$$

or

$$\begin{aligned}\mathbf{C}'\mathbf{R}^{-1}\mathbf{C}\hat{\boldsymbol{\alpha}} - \mathbf{C}'\mathbf{R}^{-1}\mathbf{Z}\left[\mathbf{B}^{-1} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\right]^{-1}&\mathbf{Z}'\mathbf{R}^{-1}\mathbf{C}\hat{\boldsymbol{\alpha}} \\ &= \mathbf{C}'\mathbf{R}^{-1}\mathbf{Y} - \mathbf{C}'\mathbf{R}^{-1}\mathbf{Z}\left[\mathbf{B}^{-1} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\right]^{-1}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y},\end{aligned} \tag{A5}$$

which is $\mathbf{C}'\mathbf{V}^{-1}\mathbf{C}\hat{\boldsymbol{\alpha}} = \mathbf{C}'\mathbf{V}^{-1}\mathbf{Y}$. Therefore, $\hat{\boldsymbol{\alpha}}$ is a generalized least squares solution and $\mathbf{C}\hat{\boldsymbol{\alpha}}$ is a BLUE.

Equation ([A2]) can be rewritten as follows:

$$
\begin{aligned}
\hat{\mathbf{b}} &= \left(\mathbf{B}\left[\mathbf{B}^{-1} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} \right] - \mathbf{B}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} \right)\left[\mathbf{B}^{-1} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} \right]^{-1} \mathbf{Z}'\mathbf{R}^{-1}(\mathbf{Y} - \mathbf{C}\hat{\boldsymbol{\alpha}}) \\
&= \left(\mathbf{B}\mathbf{Z}'\mathbf{R}^{-1} - \mathbf{B}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\left[\mathbf{B}^{-1} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} \right]^{-1}\mathbf{Z}'\mathbf{R}^{-1} \right)(\mathbf{Y} - \mathbf{C}\hat{\boldsymbol{\alpha}}) \\
&= \mathbf{B}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{C}\hat{\boldsymbol{\alpha}}),
\end{aligned}
\tag{A6}
$$

which is the BLUP of $\hat{\mathbf{b}}$ of $\mathbf{b}$. $\square$

Appendix A.2. Proof of Theorem [2]

Proof. Let $\mathbf{U} = \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \\ \boldsymbol{\eta} \end{pmatrix}$. By knowing that $\|\mathbf{a}\|_{\mathbf{B}}^2 = \mathbf{a}^\top \mathbf{B}\mathbf{a}$ Equation ([14]) can be written as

$$
(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{W}\boldsymbol{\gamma} - \mathbf{D}\boldsymbol{\eta})^\top \mathbf{v}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{W}\boldsymbol{\gamma} - \mathbf{D}\boldsymbol{\eta}) + \mathbf{U}^\top \mathbf{M}\mathbf{U}
$$

or

$$
(\mathbf{y} - \mathbf{C}\mathbf{U}^\top)^\top \mathbf{v}^{-1}(\mathbf{y} - \mathbf{C}\mathbf{U}^\top) + \mathbf{U}^\top \mathbf{M}\mathbf{U}.
$$

Now, consider $l = (\mathbf{y} - \mathbf{C}\mathbf{U}^\top)^\top \mathbf{v}^{-1}(\mathbf{y} - \mathbf{C}\mathbf{U}^\top) + \mathbf{U}^\top \mathbf{M}\mathbf{U}$. By taking the partial derivative with respect to $\mathbf{U}^\top$ the following would be true:

$$
\frac{\partial l}{\partial \mathbf{U}^\top} = 2\mathbf{C}^\top \mathbf{V}^{-1}(\mathbf{y} - \mathbf{C}\mathbf{U}^\top) + 2\mathbf{M}\mathbf{U}^\top.
$$

Set

$$
\frac{\partial l}{\partial \mathbf{U}^\top} = \mathbf{0}.
$$

Then, the following verdict becomes apparent:

$$
2\mathbf{C}^\top \mathbf{V}^{-1}(\mathbf{y} - \mathbf{C}\mathbf{U}^\top) = -2\mathbf{M}\mathbf{U}^\top.
$$

It implies that

$$
\begin{aligned}
\mathbf{C}^\top \mathbf{V}^{-1}\mathbf{y} - \mathbf{C}^\top \mathbf{V}^{-1}\mathbf{C}\mathbf{U}^\top &= -\mathbf{M}\mathbf{U}^\top \\
\mathbf{C}^\top \mathbf{V}^{-1}\mathbf{y} &= \mathbf{C}^\top \mathbf{V}^{-1}\mathbf{C}\mathbf{U}^\top - \mathbf{M}\mathbf{U}^\top \\
\mathbf{C}^\top \mathbf{V}^{-1}\mathbf{y} &= (\mathbf{C}^\top \mathbf{V}^{-1}\mathbf{C} - \mathbf{M})\mathbf{U}^\top.
\end{aligned}
$$

So,

$$
\hat{\mathbf{U}}^\top = (\mathbf{C}^\top \mathbf{V}^{-1}\mathbf{C} - \mathbf{M})^{-1}\mathbf{C}^\top \mathbf{V}^{-1}\mathbf{y}.
$$

Consequently, a generalized ridge estimate of β, γ, and η can be obtained

$$
\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\eta}} \\ \hat{\boldsymbol{\gamma}} \end{pmatrix} = \left(\mathbf{C}^\top \mathbf{V}^{-1}\mathbf{C} - \mathbf{M} \right)^{-1} \mathbf{C}^\top \mathbf{V}^{-1}\mathbf{y}.
\tag{A7}
$$

$\square$

References

1. Ramsay, J.O. When the data are functions. *Psychometrika* **1982**, *47*, 379–396. [CrossRef]
2. Ramsay, J.O.; Hooker, G.; Graves, S. *Functional Data Analysis with R and MATLAB*; Springer: Berlin/Heidelberg, Germany, 2021.

3. Ramsay, J.O.; Dalzell, C. Some tools for functional data analysis. *J. R. Stat. Soc. Ser. B (Methodol.)* **1991**, *53*, 539–572. [CrossRef]

4. Greven, S.; Crainiceanu, C.; Caffo, B.; Reich, D. Longitudinal functional principal component analysis. *Electron. J. Stat.* **2010**, *4*, 1022–1054. [CrossRef] [PubMed]

5. Rao, C.R. Some statistical methods for comparison of growth curves. *Biometrics* **1950**, *14*, 1–17. [CrossRef]

6. Wang, J.L.; Chiou, J.M.; Müller, H.G. Functional data analysis. *Annu. Rev. Stat. Its Appl.* **2016**, *3*, 257–295. [CrossRef]

7. Di, C.; Crainiceanu, C.M.; Mueller, H.G. Functional Data Analysis. In *Handbook of Big Data Analytics*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 409–435.

8. Cardot, H.; Ferraty, F.; Sarda, P. Functional linear model. *Stat. Probab. Lett.* **1999**, *45*, 11–22. [CrossRef]

9. Cuevas, A.; Febrero, M.; Fraiman, R. Linear functional regression: The case of fixed design and functional response. *Can. J. Stat.* **2002**, *39*, 285–300. [CrossRef]

10. Ramsay, J.O.; Heckman, N.; Silverman, B.W. Spline smoothing with model-based penalties. *Behav. Res. Methods Instruments Comput.* **1997**, *29*, 99–106. [CrossRef]

11. Müller, H.G. Functional modelling and classification of longitudinal data. *Scand. J. Stat.* **2005**, *32*, 223–240. [CrossRef]

12. Goldsmith, J.; Bobb, J.; Crainiceanu, C.; Caffo, B.; Reich, D. Penalized functional regression. *J. Comput. Graph. Stat.* **2011**, *4*, 453–469. [CrossRef]

13. Ruppert, D.; Wand, M.; Carroll, R. *Semiparametric Regression*; Cambridge Series in Statistical and Probabilistic Mathematics; Cambridge University Press: Cambridge, UK, 2003.

14. Zahed, M. Forecasting tourist's arrivals to the USA with SARIMA models, Paper 61-2018, SAS 9.4-2018. In Proceedings of the Western Users of SAS Software (WUSS) 2018, Sacramento, CA, USA, 5–7 September 2018.

15. Hailemariam, M.; O'Neill, S.; O'Hare, G.M. Longitudinal data analysis of mental health outcomes among postpartum women using a Bayesian multilevel model. *J. Affect. Disord.* **2021**, *292*, 699–707.

16. Hedeker, D.; Gibbons, R.D. *Longitudinal Data Analysis*; Wiley-Interscience: Hoboken, NJ, USA, 2006.

17. Goldsmith, J.; Crainiceanu, C.M.; Caffo, B.; Reich, D. Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **2012**, *3*, 453–469. [CrossRef]

18. Gertheiss, J.; Goldsmith, J.; Crainiceanu, C.; Greven, S. Longitudinal scalar-on-functions regression with application to tractography data. *Biostatistics* **2013**, *3*, 447–461. [CrossRef]

19. Kundu, M.G.; Harezlak, J.; Randolph, T.W. Longitudinal functional models with structured penalties. *Biostatistics* **2016**, *2*, 114–139. [CrossRef]

20. Fukagata, K.; Iwamoto, K.; Kasagi, N. Contribution of Reynolds stress distribution to the skin friction in wall-bounded flows. *Phys. Fluids* **2002**, *14*, L73–L76. [CrossRef]

21. Sokolov, I.M. Models of anomalous diffusion in crowded environments. *Soft Matter* **2012**, *8*, 9043–9052. [CrossRef]

22. Costell, M.H.; Ancellin, N.; Bernard, R.E.; Zhao, S.; Upson, J.J.; Morgan, L.A.; Maniscalco, K.; Olzinski, A.R.; Ballard, V.L.; Herry, K.; et al. Comparison of soluble guanylate cyclase stimulators and activators in models of cardiovascular disease associated with oxidative stress. *Front. Pharmacol.* **2012**, *3*, 128. [CrossRef]

23. Zahed, M.; Skafyan, M. Application of Feature Selection and Dimension Reduction Techniques on Large-Scale CT Dataset for Lung Cancer Diagnosis Based on Radiomics, Paper 222-2023, 2018. In Proceedings of the Southeast SAS Users Group (SESUG), Mobile, AL, USA, 14–17 October 2018.

24. Burt, D.; Lamb, K.; Nicholas, C.; Twist, C. Lower-volume muscle-damaging exercise protects against high-volume muscle-damaging exercise and the detrimental effects on endurance performance. *Eur. J. Appl. Physiol.* **2015**, *115*, 1523–1532. [CrossRef]

25. Meeusen, R.; Duclos, M.; Foster, C.; Fry, A.; Gleeson, M.; Nieman, D.; Raglin, J.; Rietjens, G.; Steinacker, J.; Urhausen, A. Prevention, diagnosis and treatment of the overtraining syndrome: Joint consensus statement of the European College of Sport Science (ECSS) and the American College of Sports Medicine (ACSM). *Eur. J. Sport Sci.* **2013**, *13*, 1–24. [CrossRef]

26. Seiler, S.; Jøranson, K.; Olesen, B.; Hetlelid, K. Adaptations to aerobic interval training: Interactive effects of exercise intensity and total work duration. *Scand. J. Med. Sci. Sport.* **2013**, *23*, 74–83. [CrossRef]

27. Bolger, N.; Laurenceau, J.P. *Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research*; Guilford Press: New York, NY, USA, 2013.

28. Stone, A.; Shiffman, S.; Atienza, A.; Nebeling, L. *The Science of Real-Time Data Capture: Self-Reports in Health Research*; Oxford University Press: Oxford, UK, 2007.

29. Trull, T.; Ebner-Priemer, U. Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: Introduction to the special section. *Psychol. Assess.* **2009**, *21*, 457–462. [CrossRef] [PubMed]

30. Fahrenberg, J.; Myrtek, M.; Pawlik, K.; Perrez, M. Ambulatory assessment-monitoring behavior in daily life settings. *Eur. J. Psychol. Assess.* **2007**, *23*, 206–213. [CrossRef]

31. Hektner, J.M.; Schmidt, J.A.; Csikszentmihalyi, M. *Experience Sampling Method: Measuring the Quality of Everyday Life*; Sage: Thousand Oaks, CA, USA, 2007.

32. Zahed, M. *An Intensive Longitudinal Functional Linear Model with Multiple Time Scales*; University of Northern Colorado, ProQuest LLC: Greeley, CO, USA, 2020.

33. Henderson, C. Estimation of genetic parameters (abstract). *Ann. Math. Stat.* **1950**, *1*, 817–827.

34. Carden, R.L.; Jahromi, M.Z. The inverse q-numerical range problem and connections to the Davis–Wielandt shell and the pseudospectra of a matrix. *Linear Algebra Appl.* **2017**, *531*, 479–497. [CrossRef]

35. Laird, D.A. Relative performance of college students as conditioned by time of day and day of week. *J. Exp. Psychol.* **1925**, *8*, 50. [CrossRef]

36. McLellan, G.; Arthur, R.; Donnelly, S.; Buchan, D.S. Segmented sedentary time and physical activity patterns throughout the week from wrist-worn ActiGraph GT3X+ accelerometers among children 7–12 years old. *J. Sport Health Sci.* **2020**, *9*, 179–188. [CrossRef]

37. Menai, M.; Van Hees, V.T.; Elbaz, A.; Kivimaki, M.; Singh-Manoux, A.; Sabia, S. Accelerometer assessed moderate-to-vigorous physical activity and successful ageing: Results from the Whitehall II study. *Sci. Rep.* **2017**, *7*, 45772. [CrossRef]

38. Christensen, R. *Plane Answers to Complex Questions*; Springer: New York, NY, USA, 2002; Volume 35.