A Marginal Approach to Longitudinal Functional Data for Analyzing Daily Physical Activity Patterns

Oliver Y. Chén[1], Luo Xiao[2], Brian S. Caffo[1], Martin A. Lindquist[1], Jennifer A. Schrack[3], Luigi Ferrucci[4], Ciprian M. Crainiceanu[1]

[1]Department of Biostatistics
Johns Hopkins University

[2]Department of Statistics
North Carolina State University

[3]Department of Epidemiology
Johns Hopkins University

[4] National Institute on Aging

January 31, 2016

Abstract

Objective measurement of physical activity using wearable devices such as accelerometers provides detailed information on patterns and intensities of daily physical activity, which can be potential biomarkers of human aging. Accelerometers record quasi-continuous activity information for many days and for hundreds of individuals. For example, in the Baltimore Longitudinal Study on Aging (BLSA), daily physical activity was recorded for about 300 adults during each visit for several days and each subject has two to four visits. An interesting problem that naturally arises is how to quantify daily physical activity patterns change with age, gender, body mass index, among other covariates. In this paper, we propose a longitudinal functional data model where the parameters of interest are bivariate functions of time and age. To deal with the complex correlation structure in the data, we use a GEE-type approach for model estimation. For efficient parameters and covariance estimation, we introduce a two-step procedure. Our results reveal several interesting and previously unknown daily activity patterns associated with human aging.

Keywords accelerometry, bivariate smoothing, covariance function, FACE, *P*-splines.

1 Introduction

Functional data analysis has been pervasive in the life and social sciences, and has found widespread applications in fields where functional and longitudinal data are present, including gerontology (e.g. study of physical activity in relation to aging), meteorology (e.g. hourly weather forcasting), and polysomnography (e.g. study of sleep quality in relation to health).

Depending on the inferential methods used, the analysis of functional data can be divided into two areas: functional principal component analysis (functional PCA) and functional linear models (including functional analysis of variance, functional ANOVA). Functional PCA models aim at capturing the principal direction of variation while conducting direction reduction, with eigenfunctions describing the variation of the data and principal component scores summarizing subject-specific features (Kirkpatrick and Heckman (1989); Ramsay and Dalzell (1991); Silverman et al. (1996); Ramsay and Silverman (2005); James et al. (2000); Yao et al. (2003); Hall and Hosseini-Nasab (2006); Di et al. (2009)) Functional ANOVA models make inference on level-specific functional means, given a hierachy of units (e.g. smoothing spline models for nested and crossed curves (Brumback and Rice, 1998); functional mixed effect models (Guo, 2002); Bayesian wavelet models (Morris and Carroll (2006); Morris et al. (2003); Morris et al. (2008)); Bayesian adaptive regression splines (Bigelow and Dunson, 2007); Byesian models for spatially correlated functional data (Baladandayuthapani et al., 2008); multilevel fPCA (Di et al., 2009)). For further reading, (Müller, 2005) overviews the application of functional data analysis to logitudinal data, the book (Ramsay (2006)) proivdes a broad overview of functional data analysis methods with applications to curve and image analysis; and the book (Ruppert et al. (2003)) overviews functional data analysis in the semiparametric framwork in detail.

Depending on whether the research interest is in estimating the population effect or subject-specific effect, the analysis of functional data can be divided into two areas: marginal ap-

proaches and mixed effect approaches. Most of the aforementioned models are considered as mixed effect approaches through a functional mixed effect model with an interest in estimating subject-specific effects. During the past decade, marginal approaches to multilevel functional data has begun to receive a great deal of attention in the statistical literature (e.g. Lin and Carroll (2000); Welsh et al. (2002); Lin et al. (2004); Chen et al. (2013)). In particular, (Chen et al., 2013) proposed a marginal approach to estimate the population average effect by modeling multivariate functional data through penalized spline GEE. (Diggle et al., 2002) discusses over the pros and and cons between the marginal models and mixed effect models.

One interesting research area in functional data analysis that has drawn significant attention from the scientific field is the study of physical activity across the life course. Higher physical activity is associated with fewer chronic diseases, better physical performance, and longer active life expectancy (Yorston et al. (2012); Pate et al. (1995); Ferrucci and Alley (2007); Leveille et al. (1999)). Due to the endogenously complex structure of the physical data and exogenous measurement errors, little is known about how intensity and daily patterns of free-living activity vary with age and other variables such as gender and body mass index. Therefore, on one hand, a more insightful understanding of these critical aspects of activity in older populations may help revealing the complex relationship between the intensity and duration of activity, and aging; and on the other hand, a better analyzing, modeling, and understanding of physical activity data is essential for conducting sustainable future studies, providing appropriate treatment, and promoting adequate public health policies, with a goal to improve health and prolong life expectancy in an aging population.

Currently, one of the most popular approaches to capture physical activity data is via using accelerometers. Compared to traditional approaches, using accelerometers allows researchers to obtain relatively more objective and detailed measurement of physical activity. A number of previous studies have used accelerometers in observational studies and clinical trials to inves-

tigate the effect of treatment on activity, and the relationship between activity and behavioral outcome, such as aging (Troiano et al. (2008); Bussmann et al. (2001); Culhane et al. (2005); Bai et al. (2013); He et al. (2014); Xiao et al. (2015)). Due to the complexity of the data collected in physical activity studies, there are three fundemental challenges: (i) the data collected from accelerometers are enormously large; (ii) the data are multilevel and functional; and hence are inevitably present with random effects that have a complex correlation structure; and (iii) the activity count depends not only on the time, but also on variables such as age, gender, and body mass index. To extract information from large multilevel functional data, (Di et al., 2009) proposed a multivariate functional principal component model; to seperate the bivariate systematic and random circadian patterns of physical activity, (Xiao et al., 2015) proposed a mixed effect bivariate structured functional model; and to study the population effect of multilevel functional data, (Chen et al., 2013) proposed a penalized spline generalized estimation equation model. Little is known, however, about how to estimate the marginal fixed effects when (i) the multilevel functional data is large, (ii) the functions are bivariate with additional presence of multiple covariates, and (iii) when the random effects have a complex correlation structure. To address these questions, we propose a longitudinal functional data model where the parameters of interest are bivariate functions of time and age.

The rest of the paper is organized as follows. In Section 2, we give a detailed description of the Baltimore Longitudinal Study on Aging (BLSA) and its data organization. In Section 3, we introduce a marginal model with specified working covariance for the activity data and estimation procedures. In Section 4, we evaluate our methods via extensive simulations. In Section 5, we apply the proposed methods to the BLSA data. Section 6 concludes the paper with a discussion.

2 Data Description

The data used in this paper were collected from the Baltimore Longitudinal study of Aging (BLSA). The BLSA is a study of normative human aging. It is conducted by the National Institute of Aging Intrammural Research Program; and it is America's longest-running scientific study of human aging. The participants of the BLSA are free of major choronic conditions and cognitive and functional impairment upon enrollment. The sample data we focus on in this paper consists of men and women aging between 60 and 90. For every clinical visit, the participants underwent a comprehensive physical examination and health history assessment, and were fitted with the Actiheart activity monitor positioned horizontally on the chest at thre third intercostal space using two standard electrocardiogram electordes. Accelerometry counts were measured in 1-minute epochs for the following 7 days in a free-living environment. Participants were asked to wear the monitor at all times, except when bathing or swimming. Actiheart data was downloaded using commmercial software (Actilife, version 4.0.32) to derive activity counts per minute. Days with more than 5% of data missing were excluded from the analysis. For the remaning days, missing values were imputed as the average activity counts per minute over all available days for each participant. To explore activity patterns associated with age-related difference, the sample was divided into four age strata: 60-70 year-old (group 1), 70-80 year-old (group 2), 80-90 year-old (group 3), >90 year-old (group 4). Other covariates collected include gender, body mass index (BMI), functional status (gait speed), employment status, race, education, and comorbidities.

Trained and certified technicians administered all assessments following standardized protocols. The Internal Review Board of the Medstar Research Institute approved the study protocol and all participants provided written informed consent. A general description of the sample and enrollemnt procedures and criteria of BLSA can be found in (Stone and Norris, 1966). Addi-

tional demographic information and study procedure can be found in (Schrack et al., 2013).

The scientific goal of this paper is to reveal daily activity patterns associated with human aging. The statistical goal of this paper is to, when the multilevel functional data is present with complex correlation structure, (1) estimate the bivariate mean and slope functions of age and time; and (2) investigate the effect of covariates (e.g. BMI index) on activity.

3 Methods

Suppose the physical data take the form $\{y_{ij}(t_k), x_i, 1 \leq i \leq I, 1 \leq j \leq n_i, 1 \leq k \leq m\}$, where n denotes the number of subjects, n_i is the number of visits for subject i, $t_k = k$ (hour unit), and $m = 24$. Let $\mathbf{y}_{ij} = \{y_{ij}(t_1), \ldots, y_{ij}(t_m)\}^T$, $\mathbf{Y}_i = [\mathbf{y}_{i1}, \ldots, \mathbf{y}_{in_i}]$, and $\mathbf{Y} = [\mathbf{Y}_1, \ldots, \mathbf{Y}_I]$. Then $\mathbf{Y}$ is a data matrix of dimension $m \times n$ with $n = \sum_{i=1}^{I} n_i$. Moreover, we let $\mathbf{y}_i = \text{vec}(\mathbf{Y}_i)$ and $\mathbf{y} = \text{vec}(\mathbf{Y})$, where vec is an operator that vertically stacks the columns of a matrix into a column vector. Throughout the paper, we use bold lower case symbols for columns vectors, and bold upper case symbols for matrices.

3.1 Model formulation

We consider a marginal model for the activity data. The mean model for the activity data is

$$\mathbb{E}\{y_{ij}(t)|(x_{ij}, z_{ij})\} = \beta_0(t, x_{ij}) + z_{ij}\beta_1(t, x_{ij}), \tag{1}$$

where $t \in [0, 24]$ denotes time of day in the unit of hours, i is subject index, j is the visit index, $y_{ij}(t)$ is the observed log counts at time t, averaged over multiple days, x_{ij} is age, and z_{ij} is a covariate such as body mass index. For simplicity, we only consider one covariate z_{ij}, but the model can be easily extended for multiple covariates. Here $\beta_0(t, x)$ and $\beta_1(t, x)$ are two bivariate smooth functions of time of day and age, representing time-and-age adjusted baseline effect and effect of covariate z, respectively.

To specify a reasonable working covariance for the activity data, we use the following working model for $e_{ijk} = y_{ij}(t_k) - E\{y_{ij}(t_k)|(x_{ij}, z_{ij})\}$,

$$e_{ij}(t_k) = b_i(t_k) + w_{ij}(t_k) + \epsilon_{ijk}, \tag{2}$$

where $b_i(t)$ models the subject-specific mean function of daily activity, $w_{ij}(t)$ models the deviation of the jth visit from the subject-specific mean function, and $\epsilon_{ijk} \sim N(0, \sigma_\epsilon^2)$ are random measurement errors. We model $b_i(t)$ and $w_{ij}(t)$ by two zero mean Gaussian processes with covariance functions $\mathcal{C}^b(s,t)$ and $\mathcal{C}^w(s,t)$, respectively. Finally, we assume $b_i(t), w_{ij}(t)$ and ϵ_{ijk} are mutually independent.

Model (2) can be regarded as a functional ANOVA model for multilevel functional data. The implied covariance structure is that data are independent across subjects but correlated across visits within the same subject. It can be derived that

$$\text{Cov}\{y_{ij}(s), y_{ij'}(t)|(\mathbf{x}_i, \mathbf{z}_i)\} = \mathcal{C}^b(s,t) + 1_{\{j=j'\}}\mathcal{C}^w(s,t) + 1_{\{j=j'\, s=t\}}\sigma_\epsilon^2, \tag{3}$$

where $\mathbf{x}_i = (x_{i1}, \ldots, x_{in_i})'$ and $\mathbf{z}_i = (z_{i1}, \ldots, z_{in_i})'$. In (3), $\mathcal{C}^b(s,t)$ and $\mathcal{C}^w(s,t)$ are two covariance operators smooth in s and t that model the between-subject covariance and the within-subject covariance of the functional observations, respectively; and σ_ϵ^2 is the variance of idiosyncratic noises.

3.2 Model for bivariate smooth functions

We use the bivariate *P*-splines (Eilers and Marx, 2003) to model the two smooth functions $\beta_0(t,x)$ and $\beta_1(t,x)$. Let $\{B_1(t), \ldots, B_{c_t}(t)\}$ be the collection of univariate B-spline basis functions along t, where c_t is the number of interior knots plus the order (degree plus 1) of the B-splines. Similarly, let $\{\breve{B}_1(x), \ldots, \breve{B}_{c_x}(x)\}$ be the collection of B-spline basis functions along x. We model $\beta_1(t,x)$ by $\sum_{1\le\kappa\le c_t, 1\le\ell\le c_x} B_\kappa(t)\breve{B}_\ell(x)\gamma_{\kappa\ell}$, where $\mathbf{\Gamma} = (\gamma_{\kappa\ell})_{1\le\kappa\le c_t, 1\le\ell\le c_x}$

is a $c_t \times c_x$ coefficient matrix. Let $\mathbf{B}(t) = \{B_1(t), \ldots, B_{c_t}(t)\}'$ be a vector of length c_t and $\breve{\mathbf{B}}(x) = \{\breve{B}_1(x), \ldots, \breve{B}_{c_x}(x)\}'$ be a vector of length c_x. Then we have

$$\beta_1(t,x) = \sum_{1\le\kappa\le c_t, 1\le\ell\le c_x} B_\kappa(t)\breve{B}_\ell(x)\gamma_{\kappa\ell} = \{\breve{\mathbf{B}}(x)\otimes\mathbf{B}(t)\}'\boldsymbol{\gamma},$$

where $\otimes$ is the tensor product of matrices and $\boldsymbol{\gamma} = \text{vec}(\boldsymbol{\Gamma})$, a vector of length $c_t c_x$.

Similarly we model $\beta_0(t,x)$ by

$$\beta_0(t,x) = \{\breve{\mathbf{B}}(x)\otimes\mathbf{B}(t)\}'\boldsymbol{\alpha},$$

where $\boldsymbol{\alpha}$ is also a coefficient vector of length $c_t c_x$.

For subject i at visit j, model (1) can be written as

$$\mathbb{E}(\mathbf{y}_{ij}) = \mathbf{X}_{ij}\boldsymbol{\beta},$$

where $\mathbf{X}_{ij} = (1, z_{ij})\otimes\mathbf{B}(;x_{ij})$, $\mathbf{B}(;x_{ij}) = \left[\breve{\mathbf{B}}(x_{ij})\otimes\mathbf{B}(t_1), \ldots, \breve{\mathbf{B}}(x_{ij})\otimes\mathbf{B}(t_m)\right]'$, and $\boldsymbol{\beta} = (\boldsymbol{\alpha}', \boldsymbol{\gamma}')'$. The dimensions of $\mathbf{y}_{ij}$, $\mathbf{B}(;x_{ij})$, and $\boldsymbol{\beta}$ are $m\times 1$, $m\times(c_t c_x)$, and $2c_t c_x \times 1$, respectively. Then for all visits from subject i, we have

$$\mathbb{E}(\mathbf{y}_i) = \mathbf{X}_i\boldsymbol{\beta}$$

where $\mathbf{y}_i = \text{vec}(\mathbf{Y}_i)$ is of dimension $(mn_i)\times 1$ and $\mathbf{X}_i = (\mathbf{X}'_{i1}, \ldots, \mathbf{X}'_{in_i})'$ is of dimension $(mn_i)\times(2c_t c_x)$.

3.3 GEE and penalized GEE

Let $\mathbf{V}_i = \text{Cov}\{\mathbf{y}_i|(\mathbf{x}_i, \mathbf{z}_i)\}$, then $\mathbf{V}_i$ is of dimension $(mn_i)\times(mn_i)$. It can be shown that

$$\mathbf{V}_i = \mathbf{J}_{n_i}\otimes\boldsymbol{\Sigma}^b + \mathbf{I}_{n_i}\otimes\boldsymbol{\Sigma}^w + \sigma^2_\epsilon\mathbf{I}_{mn_i} \tag{4}$$

where $\boldsymbol{\Sigma}^b = \begin{bmatrix} \mathcal{C}^b(t_1,t_1), & \ldots & ,\mathcal{C}^b(t_1,t_m) \\ \vdots & \ddots & \vdots \\ \mathcal{C}^b(t_m,t_1) & \ldots & ,\mathcal{C}^b(t_m,t_m) \end{bmatrix}$, $\boldsymbol{\Sigma}^w = \begin{bmatrix} \mathcal{C}^w(t_1,t_1), & \ldots & ,\mathcal{C}^w(t_1,t_m) \\ \vdots & \ddots & \vdots \\ \mathcal{C}^w(t_m,t_1) & \ldots & ,\mathcal{C}^w(t_m,t_m) \end{bmatrix}$ and $\mathbf{J}_{n_i} = \mathbf{1}_{n_i}\mathbf{1}'_{n_i}$.

Therefore, the generalized estimating equation (GEE) is

$$I^{-1}\sum_{i=1}^{I}\mathbf{X}_i'\mathbf{V}_i^{-1}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})=0.$$

To obtain smooth estimate of the bivariate functions, similar to Eilers and Marx (2003), we impose penalty on the coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$:

$$\boldsymbol{\alpha}'\mathbf{P}_\alpha\boldsymbol{\alpha}+\boldsymbol{\gamma}'\mathbf{P}_\gamma\boldsymbol{\gamma},$$

where $\mathbf{P}_\alpha=\lambda_1(\mathbf{I}_{c_x}\otimes\mathbf{D}_1'\mathbf{D}_1)+\lambda_2(\mathbf{D}_2'\mathbf{D}_2\otimes\mathbf{I}_{c_t})$ and $\mathbf{P}_\gamma=\lambda_3(\mathbf{I}_{c_x}\otimes\mathbf{D}_1'\mathbf{D}_1)+\lambda_4(\mathbf{D}_2'\mathbf{D}_2\otimes\mathbf{I}_{c_t})$. The penalty $\mathbf{P}_\gamma$ penalizes both the rows and columns of $\boldsymbol{\Gamma}$. Then we obtain the penalized generalized estimating equation

$$I^{-1}\sum_{i=1}^{I}\mathbf{X}_i'\mathbf{V}_i^{-1}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})+\mathbf{P}\boldsymbol{\beta}=0, \tag{5}$$

where $\mathbf{P}=\text{blockdiag}(\mathbf{P}_\alpha,\mathbf{P}_\gamma)$.

Note that solving (5) is equivalent to minimizing the penalized generalized least squares

$$I^{-1}\sum_{i=1}^{I}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})'\mathbf{V}_i^{-1}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})+\boldsymbol{\beta}'\mathbf{P}\boldsymbol{\beta} \tag{6}$$

and we obtain the estimate

$$\hat{\boldsymbol{\beta}}=\left(I^{-1}\sum_{i=1}^{I}\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i+\mathbf{P}\right)^{-1}\left(I^{-1}\sum_{i=1}^{I}\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{y}_i\right). \tag{7}$$

Let $\mathbf{H}_0=\text{blockdiag}(\mathbf{I}_{c_tc_x},\mathbf{0}_{c_tc_x})$ and $\mathbf{H}_1=\text{blockdiag}(\mathbf{0}_{c_tc_x},\mathbf{I}_{c_tc_x})$. Then

$$\hat{\beta}_i(t,x)=\{\breve{\mathbf{B}}(x)\otimes\mathbf{B}(t)\}'\mathbf{H}_i\hat{\boldsymbol{\beta}},\quad i=0,1. \tag{8}$$

Given $\{(\mathbf{x}_i,\mathbf{z}_i),i=1,\ldots,I\}$, we obtain by (7) the sandwich covariance formula,

$$\text{Cov}(\hat{\boldsymbol{\beta}})=\left(I^{-1}\sum_{i=1}^{I}\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i+\mathbf{P}\right)^{-1}\left(I^{-2}\sum_{i=1}^{I}\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i\right)\left(I^{-1}\sum_{i=1}^{I}\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i+\mathbf{P}\right)^{-1}.$$

It follows that

$$\text{Var}\{\hat{\beta}_i(t,x)\} = \{\breve{\mathbf{B}}(x) \otimes \mathbf{B}(t)\}' \mathbf{H}_i \text{Cov}(\hat{\boldsymbol{\beta}}) \mathbf{H}_i \{\breve{\mathbf{B}}(x) \otimes \mathbf{B}(t)\}.$$

Hence a 95% point-wise confidence interval for $\hat{\beta}_i(t,x)$ can be approximated by

$$\left(\hat{\beta}_i(t,x) - 1.96\sqrt{\text{Var}\{\hat{\beta}_i(t,x)\}}, \quad \hat{\beta}_i(t,x) + 1.96\sqrt{\text{Var}\{\hat{\beta}_i(t,x)\}} \right).$$

3.4 Estimation

We use a two-stage procedure for model estimation. In the first stage, we let $\mathbf{V}_i$ be identity matrices and then obtain an estimate of $\boldsymbol{\beta}$ by (7). Then given the estimate of $\boldsymbol{\beta}$, we estimate the two covariance functions $\mathcal{C}^b$, $\mathcal{C}^w$ and σ^2. In the second stage, we update $\mathbf{V}_i$ and refit (5) to obtain improved estimate of $\boldsymbol{\beta}$ in (7).

For stage 1, we adopt the following procedure.

1. By plugging $\mathbf{V}_i = \mathbf{I}_{mn_i}$ into (7), we obtain the estimate

$$\hat{\boldsymbol{\beta}} = \left(I^{-1} \sum_{i=1}^{I} \mathbf{X}_i' \mathbf{X}_i + \mathbf{P} \right)^{-1} \left(I^{-1} \sum_{i=1}^{I} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{y}_i \right);$$

2. Compute residuals $\hat{\mathbf{e}}_{ij} = \{\hat{e}_{ij}(t_1), \ldots, \hat{e}_{ij}(t_m)\}' = \mathbf{y}_{ij} - \mathbf{X}_{ij}\hat{\boldsymbol{\beta}}$, $1 \le j \le n_i, 1 \le i \le I$;

3. Based on $\hat{\mathbf{e}}_{ij}$'s, we estimate $\mathcal{C}^b(s,t)$, $\mathcal{C}^w(s,t)$ and σ_ϵ^2; see Section 3.5 for details. Denote the resulting estimates by $\tilde{\mathcal{C}}^b(s,t)$, $\tilde{\mathcal{C}}^w(s,t)$ and $\tilde{\sigma}_\epsilon^2$, respectively.

For stage 2, we have two steps:

1. Given the estimates $\tilde{\mathcal{C}}^b(s,t)$, $\tilde{\mathcal{C}}^w(s,t)$ and $\tilde{\sigma}^2$, we calculate via (4)

$$\tilde{\mathbf{V}}_i = \mathbf{J}_{n_i} \otimes \tilde{\boldsymbol{\Sigma}}^b + \mathbf{I}_{n_i} \otimes \tilde{\boldsymbol{\Sigma}}^w + \tilde{\sigma}_\epsilon^2 \mathbf{I}_{mn_i}, 1 \le i \le I;$$

2. Plug $\tilde{\mathbf{V}}_i$ into (7) and obtain the updated estimate $\tilde{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$.

3.5 Covariance smoothing

3.5.1 Empirical estimates of covariance functions

We first construct empirical estimates of the two covariance functions $\mathcal{C}^b(s,t)$ and $\mathcal{C}^w(s,t)$ via method-of-moment approaches. Consider the model specification of covariance in (3). When $j \neq j'$,

$$\text{Cov}\left\{e_{ij}(s), e_{ij'}(t)\right\} = \mathcal{C}^b(s,t) = \mathbb{E}\left\{I^{-1}\sum_{i=1}^{I}\frac{1}{n_i(n_i-1)}\sum_{j\neq j'}e_{ij}(s)e_{ij'}(t)\right\}.$$

Hence an unbiased estimator of $\mathcal{C}^b(s,t)$ is

$$\hat{\mathcal{C}}^b(s,t) = I^{-1}\sum_{i=1}^{I}\frac{1}{n_i(n_i-1)}\sum_{j\neq j'}e_{ij}(s)e_{ij'}(t).$$

Next note that $\text{Cov}\left\{e_{ij}(s), e_{ij}(t)\right\} = \mathcal{C}^b(s,t) + \mathcal{C}^w(s,t) + 1_{\{s=t\}}\sigma^2$, where $1_{\{\cdot\}}$ equals if the statement in the bracket is true and 0 otherwise. Hence, if $s \neq t$, $\text{Cov}\left\{e_{ij}(s), e_{ij}(t)\right\} = \mathcal{C}^b(s,t) + \mathcal{C}^w(s,t)$, an unbiased empirical estimate of $\mathcal{C}^b(s,t)$ is

$$\hat{\mathcal{C}}^w(s,t) = I^{-1}\sum_{i=1}^{I}\frac{1}{n_i}\sum_{j=1}^{n_i}e_{ij}(s)e_{ij}(t) - \hat{\mathcal{C}}^b(s,t).$$

For notational convenience, we also use the above definition of $\hat{\mathcal{C}}^2(s,t)$ for $s = t$.

3.5.2 Smoothing of empirical covariances

The constructed empirical estimates of covariance functions, $\hat{\mathcal{C}}^b(s,t)$ and $\hat{\mathcal{C}}^w(s,t)$, are usually noisy and need to be smoothed. We follow the idea in Xiao et al. (2014) for fast covariance smoothing.

First let $\hat{\mathbf{\Sigma}}^b = \left[\hat{\mathcal{C}}^b(t_\kappa, t_\ell)\right]_{1\leq\kappa,\ell\leq m}$ and $\hat{\mathbf{\Sigma}}^w = \left[\hat{\mathcal{C}}^w(t_\kappa, t_\ell)\right]_{1\leq\kappa,\ell\leq m}$. Then both $\hat{\mathbf{\Sigma}}^b$ and $\hat{\mathbf{\Sigma}}^w$ are of dimensions $m \times m$. Note that the nugget effect due to σ^2_ϵ presents in $\hat{\mathbf{\Sigma}}^w$. It is easy to show that

$$\hat{\mathbf{\Sigma}}^b = \sum_{i=1}^{I}\mathbf{E}_i\mathbf{H}_i\mathbf{E}_i',$$

where $\mathbf{E}_i = [\mathbf{e}_{i1}, \mathbf{e}_{i2}, \ldots, \mathbf{e}_{in_i}]$ is of dimension $m \times n_i$, $\mathbf{e}_{ij} = \{e_{ij}(t_1), e_{ij}(t_2), \ldots, e_{ij}(t_m)\}'$ is a vector of length m, and $\mathbf{H}_i = w_i(\mathbf{J}_{n_i} - \mathbf{I}_{n_i})$ is a matrix of dimension $n_i \times n_i$, $w_i = \{In_i(n_i - 1)\}^{-1}$, and $\mathbf{J}$ indicates a square matrix of all 1's. Similarly we can show that

$$\hat{\mathbf{\Sigma}}^w = \sum_{i=1}^{I}(\tilde{w}_i \mathbf{E}_i \mathbf{E}_i' - \mathbf{E}_i \mathbf{H}_i \mathbf{E}_i') = \sum_{i=1}^{I} \mathbf{E}_i \tilde{\mathbf{H}}_i \mathbf{E}_i',$$

where $\tilde{\mathbf{H}}_i = \tilde{w}_i \mathbf{I}_{n_i} - w_i(\mathbf{J}_{n_i} - \mathbf{I}_{n_i})$ is of dimension $n_i \times n_i$ with $\tilde{w}_i = (In_i)^{-1}$.

The fast covariance estimation (FACE) algorithm in Xiao et al. (2014) can efficiently smooth any matrix that can be written in a sandwich form as $\mathbf{YHY}'$, where $\mathbf{Y}$ is a data matrix and $\mathbf{H}$ is a positive semi-definite symmetric matrix.

Theorem 1. *FACE is invariant to orthogonal transformation of data.*

See Appendix for a proof.

Therefore, $\hat{\mathbf{\Sigma}}^b$ can be smoothed conveniently using the FACE algorithm and we denote the resulting smooth estimate by $\tilde{\mathcal{C}}^b(s,t)$. To smooth $\hat{\mathbf{\Sigma}}^w$, however, we need to deal with the nugget effect in $\hat{\mathbf{\Sigma}}^w$. The idea is as follows. Suppose we fill in the diagonal entries of $\hat{\mathbf{\Sigma}}^w$ by some estimates, for example, average of neighboring non-diagonal elements in $\hat{\mathbf{\Sigma}}^w$. With a little abuse of notation, we still denote the estimate by $\hat{\mathbf{\Sigma}}^w$. Then we conduct a spectral decomposition of $\hat{\mathbf{\Sigma}}^w = \sum_{\kappa=1}^{m} d_\kappa \mathbf{u}_\kappa \mathbf{u}_\kappa'$, where $d_1 \geq d_2 \geq ...$ are ordered eigenvalues and $\mathbf{u}_\kappa$'s are the corresponding eigenvectors. Note that since $\hat{\mathbf{\Sigma}}^w$ is in general not positive-definite, there exists some integer $K_0 < m$ such that $d_\kappa > 0$ if $\kappa \leq K_0$ and $d_\kappa \leq 0$ if $\kappa > K_0$. Then we let $\hat{\mathbf{\Sigma}}^{w,\text{new}} = \sum_{\kappa=1}^{K_0} d_\kappa \mathbf{u}_\kappa \mathbf{u}_\kappa' = \mathbf{U}_{K_0} \text{diag}(d_1, \ldots, d_{K_0}) \mathbf{U}_{K_0}'$, where $\mathbf{U}_{K_0} = [\mathbf{u}_1, \ldots, \mathbf{u}_{K_0}]$. Now we could apply the FACE algorithm to $\hat{\mathbf{\Sigma}}^{w,\text{new}}$ and the estimate is denoted by $\tilde{\mathcal{C}}^w(s,t)$. Note that the spectral approximation of $\hat{\mathbf{\Sigma}}^w$ by $\hat{\mathbf{\Sigma}}^{w,\text{new}}$ is always beneficial for the covariance estimation since covariance functions are positive semi-definite.

3.5.3 Estimation of σ_ϵ^2

By the equality $\text{Cov}\{e_{ij}(t), e_{ij}(t)\} = \mathcal{C}^b(t,t) + \mathcal{C}^w(t,t) + \sigma_\epsilon^2$, σ_ϵ^2 can be estimated by

$$\tilde{\sigma}_\epsilon^2 = \frac{1}{I}\sum_{i=1}^{I}\frac{1}{mn_i}\sum_{j,t}\left\{e_{ij}(t)^2 - \tilde{\mathcal{C}}^b(t,t) - \tilde{\mathcal{C}}^w(t,t)\right\},$$

where $\tilde{\mathcal{C}}^b$ and $\tilde{\mathcal{C}}^w$ are the smooth estimates obtained from Section 3.5.2.

3.6 Selection of smoothing parameters

We use generalized cross validation for selecting the smoothing parameters, following the approach in Wood (2000). Let $\tilde{\mathbf{y}}_i = \mathbf{V}_i^{-1/2}\mathbf{y}_i$, $\tilde{\mathbf{X}}_i = \mathbf{V}_i^{-1/2}\mathbf{X}_i$, then equation (6) becomes

$$I^{-1}\sum_{i=1}^{I}(\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta})'(\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta}) + \boldsymbol{\beta}'\mathbf{P}\boldsymbol{\beta},$$

which is equivalent to equation (4) in Wood (2000).

4 Simulations

Here we conduct a simulation study to investigate the performance of the proposed methodology.

4.1 Bivariate intercept and slope model

We generate data from the following model

$$y_{ij}(t_k) = \beta_0(t_k, x_{ij}) + z_{ij}\beta_1(t_k, x_{ij}) + \sum_{s=1}^{N_1}\xi_{is}\phi_s^{(1)}(t_k) + \sum_{l=1}^{N_2}\zeta_{ijl}\phi_l^{(2)}(t_k) + \epsilon_{ij}(t_k), \qquad (9)$$

where $\{\phi_1^{(1)}, \phi_2^{(1)}, \ldots, \phi_{N_1}^{(1)}\}$ is a set of orthonormal eigenfunctions, $\{\phi_1^{(2)}, \phi_2^{(2)}, \ldots, \phi_{N_2}^{(2)}\}$ is another set of orthonormal eigenfunctions, $\xi_{ik} \sim N\{0, \lambda_k^{(1)}\}$, $\zeta_{ijl} \sim N\{0, \lambda_l^{(2)}\}$, $\epsilon_{ij}(t_k) \sim N(0, \sigma_\epsilon^2)$, $t_k \in \{1/m, 2/m, \ldots, m/m\}$, $1 \le j \le n_i$ and $1 \le i \le I$. Here $\lambda_1^{(1)} \ge \lambda_2^{(1)} \ge \cdots \ge \lambda_{N_1}^{(1)} > 0$

and also $\lambda_1^{(2)} \geq \lambda_2^{(2)} \geq \cdots \geq \lambda_{N_2}^{(2)} > 0$. Then the implied true between and within covariances are $\mathcal{C}^b(s,t) = \sum_{k=1}^{N_1} \lambda_k^{(1)} \phi_k^{(1)}(s)\phi_k^{(1)}(t)$ and $\mathcal{C}^w(s,t) = \sum_{k=1}^{N_2} \lambda_k^{(2)} \phi_k^{(2)}(s)\phi_k^{(2)}(t)$.

We let $m = 100$ and the number of visits, n_i, varies between 2 and 4. We generate covariates x_{ij} and z_{ij} from the uniform distribution in the unit interval. We let $\beta_0(t,x) = 2 - x + \cos(2\pi t)$ and $\beta_1(t,x) = 2 - x + \sin(2\pi t)$, such that they are periodic in t and are nonnegative. We let $N_1 = N_2 = 4$ with $\lambda_k^{(1)} = \lambda_k^{(2)} = 0.5^{k-1}$ for $k \in \{1,2,3,4\}$.

We study a few combinations of levels of idiosyncratic noises and eigenfunctions. We let σ_ϵ^2 be either 1 (moderate noise) or 4 (large noise). As for eigenfunctions, we consider the following two cases:

Mutually orthogonal bases:

$\{\phi_k^{(1)}(t)\}_{k=1}^4 = \left\{\sqrt{2}\sin(2\pi t), \sqrt{2}\cos(2\pi t), \sqrt{2}\sin(4\pi t), \sqrt{2}\cos(4\pi t)\right\}$ and

$\{\phi_k^{(2)}(t)\}_{k=1}^4 = \left\{\sqrt{2}\sin(6\pi t), \sqrt{2}\cos(3\pi t), \sqrt{2}\sin(4\pi t), \sqrt{2}\cos(4\pi t)\right\}$.

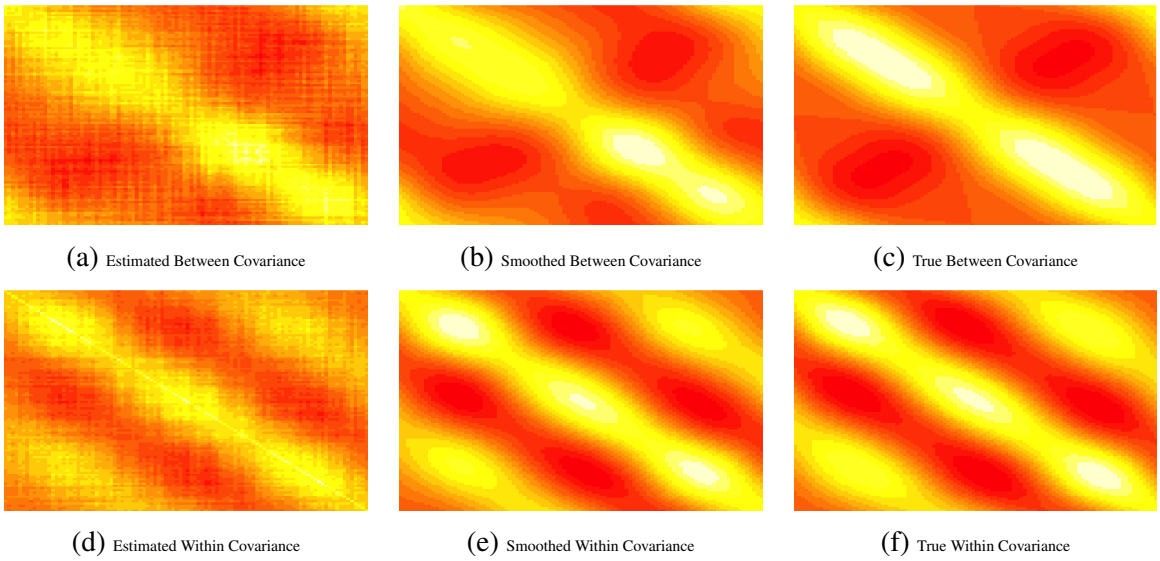
Mutually nonorthogonal bases:

$\{\phi_k^{(1)}(t)\}_{k=1}^4 = \left\{\sqrt{2}\sin(2\pi t), \sqrt{2}\cos(2\pi t), \sqrt{2}\sin(4\pi t), \sqrt{2}\cos(4\pi t)\right\}$ and $\{\phi_k^{(2)}(t)\}_{k=1}^4 = \big\{1, \sqrt{3}(2t-1), \sqrt{5}(6t^2-6t+1), \sqrt{7}(20t^3-30t^2+12t-1)\big\}$.

To assess the performance of our method under different sample sizes, we let I to be either 50 or 100. There are 8 model conditions to examine and for each condition, we conduct 500 simulations.

4.2 Simulation results

To visually demonstrate the performance of our method in estimating and smoothing the bivariate covariaces, as well as it in estimating of the bivariate functions of the fixed effect, we choose one set of simulation as an example. In this example, we use one simulation with mutually orthogonal bases, idosyncratic noise $\sigma^2 = 1$, and let the rest input follow from Section 4.1.

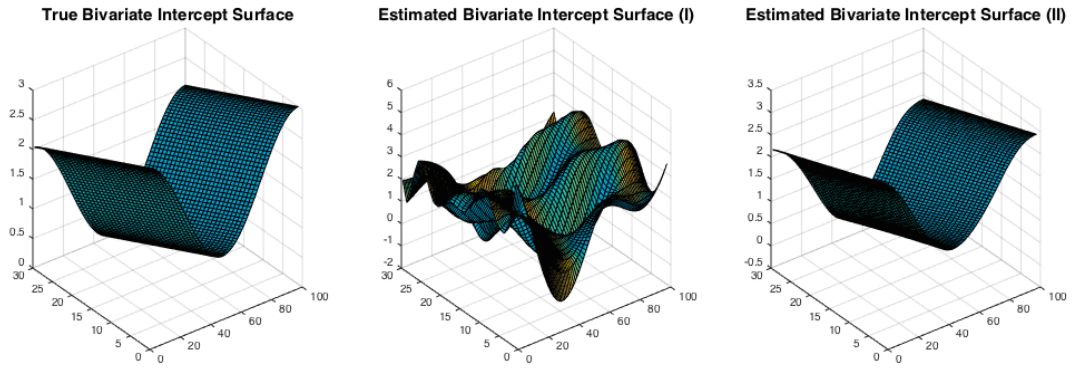


(a) Estimated Between Covariance (b) Smoothed Between Covariance (c) True Between Covariance

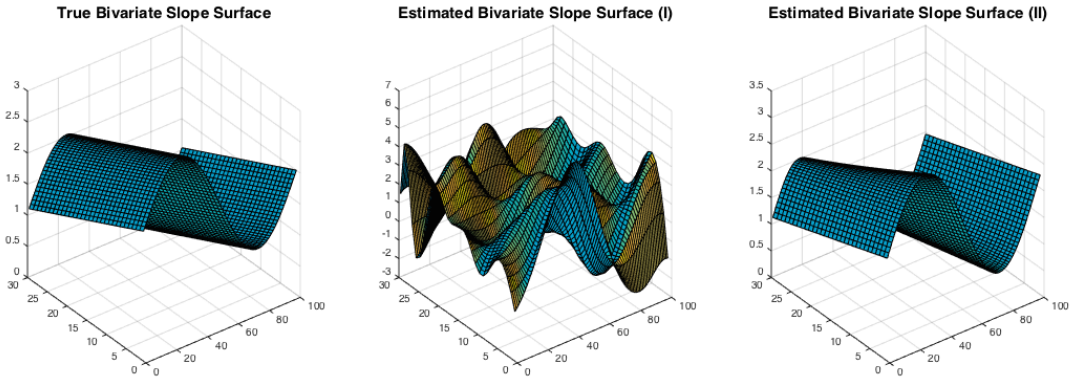
(d) Estimated Within Covariance (e) Smoothed Within Covariance (f) True Within Covariance

Figure 1: Bivariate covariates estimation and smoothing

Consider the same simulation scheme as in Figure 2. Consider 50 subjects each has $2-4$ visits. The true between and within covariances are $\mathcal{C}^b(s,t) = \sum_{k=1}^4 \lambda_k^{(1)} \phi_k^{(1)}(s)\phi_k^{(1)}(t)$ and $\mathcal{C}^w(s,t) = \sum_{k=1}^4 \lambda_k^{(2)} \phi_k^{(2)}(s)\phi_k^{(2)}(t)$, where $t \in \{0.01, 0.02, \cdots, 1\}$, $\lambda_k^{(1)} = 0.5^{k-1}$, for $k \in \{1,2,3,4\}$, and $\lambda_k^{(1)} = 0.5^{k-1}$, for $k \in \{1,2,3,4\}$. Consider mutually orthogonal bases $\{\phi_k^{(1)}(t)\}_{k=1}^4 = \left\{\sqrt{2}\sin(2\pi t), \sqrt{2}\cos(2\pi t), \sqrt{2}\sin(4\pi t), \sqrt{2}\cos(4\pi t)\right\}$ and $\{\phi_k^{(2)}(t)\}_{k=1}^4 = \left\{\sqrt{2}\sin(3\pi t), \sqrt{2}\cos(3\pi t), \sqrt{2}\sin(4\pi t), \sqrt{2}\cos(4\pi t)\right\}$. Let the idosyncratic noises be $\sigma^2 = 1$. (a)-(c) show the heatmaps for the estimated, smoothed, and true bivariate between covariances; and (d)-(f) show the heatmaps for the estimated, smoothed, and true bivariate within covariance.



(a) True and estimated bivariate intercept surfaces



(b) True and estimated bivariate slope surface

Figure 2: Bivariate Surface Estimates

Consider 200 subjects each has 2-4 varing trials whose physical activities are simulated. During each trial, physical activity data are obtained at 100 distinct time points. The activity data at time t_k are generated from, for $i = 1$, $y_{ij}(t_k) = \beta_0(t_k, x_{ij}) + z_{ij}\beta_1(t_k, x_{ij}) + \left(\sum_{s=1}^{N_1} \xi_{is}\phi_s^{(1)}(t_k) + \sum_{l=1}^{N_2} \zeta_{ijl}\phi_l^{(2)}(t_k) \right) + \epsilon_{ij}(t_k)$, where $x_i \sim U[0,1]$, $z_i \sim U[0,1]$, $t \in \{0.01, 0.02, \cdots, 1\}$, the bivariate intercept and slope functions are $b_0(x,t) = 2 - x + \cos(2\pi t)$ and $b_1(x,t) = 2 - x + \sin(2\pi t)$, and $\epsilon_{ij} \sim \mathcal{N}(0,1)$. The eigenfunctions at both subject and trial level are chosed to be $\lambda_k^{(1)} = 0.5^{k-1}$, for $k \in \{1,2,3,4\}$, and $\lambda_l^{(1)} = 0.5^{k-1}$, for $l \in \{1,2,3,4\}$; and the bases functions are chosen to be orthogonal: $\{\phi_k^{(1)}(t)\}_{k=1}^4 = \left\{\sqrt{2}\sin(2\pi t), \sqrt{2}\cos(2\pi t), \sqrt{2}\sin(4\pi t), \sqrt{2}\cos(4\pi t)\right\}$ and $\{\phi_k^{(2)}(t)\}_{k=1}^4 = \left\{\sqrt{2}\sin(3\pi t), \sqrt{2}\cos(3\pi t), \sqrt{2}\sin(4\pi t), \sqrt{2}\cos(4\pi t)\right\}$. (a) from left to right are the true, estimated (first step), and estimated (second step) bivariate intercept surface $\beta_0(x,t)$; and (b) from left to right are the true, estimated (first step), and estimated (second step) bivariate slope surface $\beta_1(x,t)$, evaluated at 88 selected equally spaced age points $x = \{x_0, \cdots, x_{88}\}$ and 100 equally spaced time points $t = \{0.01, 0.02, \cdots, 1\}$, where $x_0 = \min(x_{ij})$ and $x_{88} = \max(x_{ij})$.

Figure 1 provides plots of the estimated, smoothed, and true bivariate covariances. The top row displays plots of between covariaces; and the bottom row displays plots of within covariaces. Both plots of estimated between and withinin covariances ((a) and (d)) represent the general pattern of the plots of their corresponding true covariates ((c) and (f)) resonably well; but the edges are noticeably coarse. In particular, there is an obvious diagonal line in the estimated within covariance plot, represting the nugget effect. The plots of smoothed between and within covariances ((b) and (e)) show significant improvement of the coarse edges seen in ((a) and (d)); and their patterns are graphically much closer to those of their corresponding true covariates. In addition, the nugget effect in (d) is smoothed out in the smoothed within covariance plot.

Figure 2 shows bivariate surface plots of the truth, stage I estimation, and stage II estimation. The top row displays surface plots of bivariate intercept; and the bottom row displays surface plots of bivariate slope. Both intercept and slope surface plots using the first stage estimates represent the corresponding true surfaces poorly, because the working covariance used is assumed to be only diagonal. The updated working covariance in the second stage greatly improves the performance of bivariate functions estimation, and the surface plots are exceedingly close to their corresponding true plots.

To numerically evaluate the performance of our estimates under different combination of iodsyncratic noises, eigenfunctions, and sample sizes, and at different estimating stages (stage I and stage II), we consider the integrated square erros (ISEs) of β_0 and β_1, as follows. For every simulation $1 \leq j \leq 500$, we have:

$$ISE(\beta_i^{(j)}) = \frac{1}{c_t c_x} \sum_{\substack{1 \leq t \leq c_t \\ 1 \leq x \leq c_x}} [\hat{\beta}_i^{(j)}(t,x) - \beta_i(t,x)]^2, \quad i = 0, 1.$$

The means and standard deviations of the ISEs for β_0 and β_1 under different simulation settings are provided in Table 1. Table 1 also summarizers the means and standard deviations

of the estimated square error of σ^2, for every setting.

The simulation results show that our method performs well for both orthogonal and unorthogonal bases. In addition, the means and standard deviation of ISEs of the estimates in stage II are noticeably smaller than those in stage I, indicating that the correction of the working covariance $\mathbf{V}_i$ has significantly improved our estimation of the bivariate functions. As the idiosyncratic noises increase, the means and standard deviations of the ISEs of the stage II estimates of the bivariate functions increase, but only moderately; whereas the increase in $\hat{\sigma}^2$ is quite pronunced. This indicates that the estimation of the bivariate function is not sensitive to underlying idiosyncratic noises. Finally, for every simulation setting, when sample size increases, the means and standard deviations of ISEs of the bivariate functions at two stages all decrease, the standard deviations of all squared errors of σ^2 all decrease, but the means of the squared errors for σ^2 increase for $n = 50$, revealing possible instability of estimation of σ^2 when the sample size is small.

			Stage I		Stage II		
			$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\sigma}^2$
Orthogonal Bases	$n = 50$	$\sigma = 1$	3.19 (1.03)	5.53 (1.88)	0.63 (0.20)	0.75(0.25)	3.00 (3.44)
		$\sigma = 2$	3.12 (0.95)	5.40 (1.81)	0.69 (0.20)	0.91 (0.27)	13.71 (6.22)
	$n = 100$	$\sigma = 1$	1.38 (0.26)	2.36 (0.44)	0.33 (0.10)	0.39 (0.13)	3.77 (1.96)
		$\sigma = 2$	1.38 (0.26)	2.36 (0.46)	0.36 (0.10)	0.47 (0.13)	13.46 (4.33)
Unorthogonal Bases	$n = 50$	$\sigma = 1$	3.12 (0.93)	5.44 (1.66)	0.95 (0.27)	1.52 (0.46)	0.16 (5.55)
		$\sigma = 2$	3.10 (1.03)	5.38 (1.83)	1.01 (0.32)	1.62 (0.57)	12.17 (7.25)
	$n = 100$	$\sigma = 1$	1.36 (0.24)	2.35 (0.42)	0.48 (0.13)	0.76 (0.22)	0.33 (4.58)
		$\sigma = 2$	1.36 (0.25)	2.33 (0.43)	0.50 (0.13)	0.80 (0.23)	10.81 (5.76)

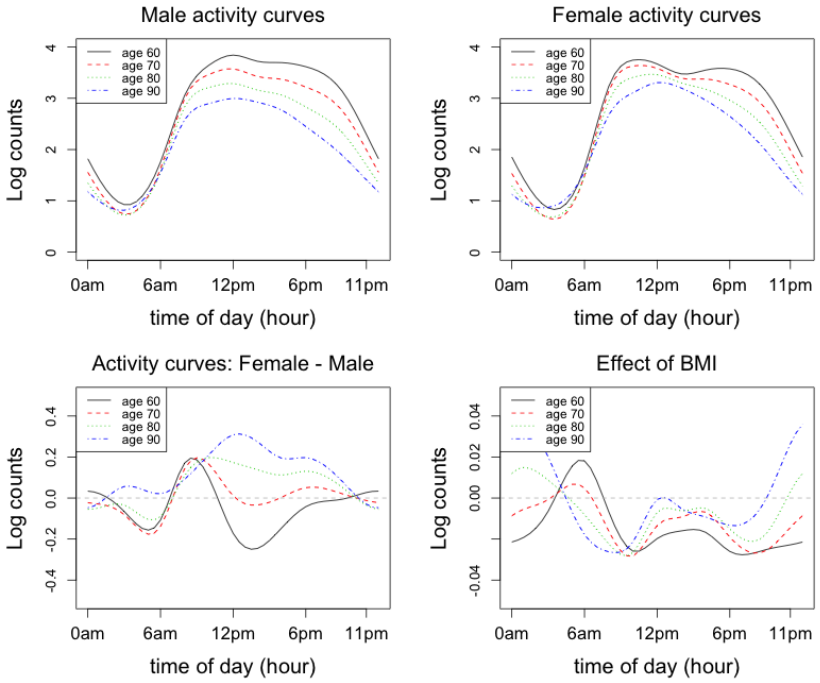
Table 1: Mean (standard deviation) $\times 100$ of the integrated square errors (ISEs) of the estimates

Simulations are conducted under a variety of combinations of idiosyncratic noises and eigenfunctions: we consider the idiosyncratic noises σ^2 to be 1^2 (moderate-noise) and 2^2 (large noise), and eigenfunctions that are mutually orthogonal and nonorthogonal. For each combination, we consider 50 and 100 subjects. Means (standard deviations) $\times 100$ of the ISEs for bivariate intercept (β_0), and bivariate slope (β_1) are reported for both stage I and stage II. Also included are Means (standard deviations) $\times 100$ of the ISE of the estimated idiosyncratic noises. 500 simulations are run for each combination.

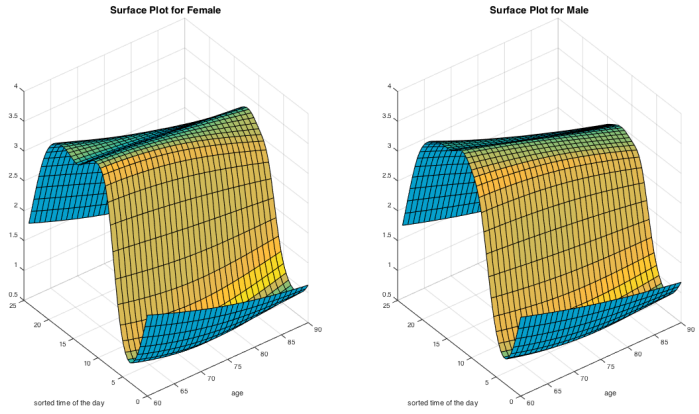
5 Application to Physical Activity Data from BLSA

Due to the complex between and wintin covariance structure present in the BLSA data, we apply our proposed method in Section 3 to address the following research interests, across time of a day: (1) the general activity patterns for different age groups; (2) the gender- and age- specific activity patterns; (3) the gender activity difference for each age group; and (4) the covariate (BMI) effect on activity for diferrent age groups.

The top two plots in Figure 3 (a) shows the smoothed activity curves for four age groups of women and men: $<$ 60 years old (solid black lines), 60-70 years old (dashed red lines), 70-80 years old (dashed green lines), and 90 years old (dashed blue lines). We observe that, first, for both gneders and all age groups, the activity curves of a day follow the similar pattern. One is less active during night than during the day. In particular, one's activity starts to fall from around 6 pm and reaches the bottom at around 4 am; from 4 am to 9 am, there is a steep increase in physical activity; and from 9 am to 6 pm, the activity is relatively high and stable,



(a) Estimated activity curves across time



(b) Fixed effect surface plots of female and male

Figure 3: Top plots in (a): Smoothed mean profiles for four age groups of women and men: < 60 years old (solid black lines), 60-70 years old (dashed red lines), 70-80 years old (dashed green lines), and 90 years old (dashed blue lines); bottom left plot in (a): Smoothed mean profiles of gender difference for four age groups; bottom right plot in (a): BMI effect to physical activity for four age groups.

with the peak achieves around noon. It is interesting to see that for younger age groups of both genders, there is a slight drop in activity count after noon and before 4 pm with the lowert point occurs around 2 pm. This is particularly obvious for females in the age 60 group. Within each gender group, the younger groups have more intense physical activity. But the activity intensity difference is more visible during the day than at night. For men, the activity intensities between different age groups are relatively noticeable from 7 am to 11 pm, with 11 am and around 6-7 pm the most apparent; and for women, the activity intensities between different age groups are relatively noticeable from 7 am to noon, and from noon to 11 pm, with 11 am and around 6-7 pm the most apparent. In addition, we found that the age 90 group for both male and female are more active than their corresponding age 70 and 80 groups, and in particular the age 90 group in female is more active than their corresponding age 60 group.

The lower left plot in Figure 3 (a) shows the activity count difference between female and male for different age groups, across time of a day. From the plot, we see that, first, the females in age group 90 are more active than their male counterparts throughout the entire day, especially from 6 am to 6 pm, with an exception from 10 pm to 1 am. Second, the gender activity difference for age groups 60-80 are very similar from 0 am to 11 am: males are more active than females from 0 am to 7 am, but less active from 7 am to 11 am. The gender activity difference for age groups 60-80 from 11 am to 0 am is still relatively similar in pattern, but somewhat different in magnitude compared to the 0 reference line: the younger males and older females are more active compared to their gender counterparts from 11 am to 0 am. Therefore, males younger than 90 are in general more active during the night; and olderly females are more active than their male counterpart: the older the age group, the more obvious the gender effect appear.

The lower right plot in Figure 3 (a) is a plot of BMI effect on activity for different age groups aross time of a day. For each age group, we see that in general an increase in BMI leads to a decrease in activity, which accords with our common sense. It is, however, interesting that

for elderly people, an increase in BMI improves thier activity during the night. One possible explanation for this is that in the BLSA study there is a minor representation of normally- and overly- weighted subjects above 80 year-old; hence an increase of BMI in subjects that mostly underweighted will increase their activity. This is particularly pronounced at night. For people more than 80 year-old, an increase of BMI does not increase their activity during the day; from the plot, however, an increase of BMI for people more than 80 year-old does not reduce their activity counts as much as people younder than 80 do.

The two plots in Figure 3 (b) extend the activity curves between male and female to average activity surfaces between male and females across sorted time of the day and age.

6 Discussion

This paper addresses the problem in longitudinal functional data analysis where the functional curves have a complex correlation structure. As a reasonable generalization of this problem, a bivariate functional model that allows for a complex correlation structure and multiple covariates is introduced; and a two-step procedure and a GEE-like approach are proposed for covariance estimation, covariate smoothing, and fixed effect estimation.

References and Notes

Bai, J., B. He, H. Shou, V. Zipunnikov, T. A. Glass, and C. M. Crainiceanu (2013). Normalization and extraction of interpretable metrics from raw accelerometry data. *Biostatistics*, kxt029.

Baladandayuthapani, V., B. K. Mallick, M. Young Hong, J. R. Lupton, N. D. Turner, and R. J. Carroll (2008). Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinogenesis. *Biometrics 64*(1), 64–73.

Bigelow, J. L. and D. B. Dunson (2007). Bayesian adaptive regression splines for hierarchical data. *Biometrics 63*(3), 724–732.

Brumback, B. A. and J. A. Rice (1998). Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of the American Statistical Association 93*(443), 961–976.

Bussmann, J., W. Martens, J. Tulen, F. Schasfoort, H. Van Den Berg-Emons, and H. Stam (2001). Measuring daily behavior using ambulatory accelerometry: the activity monitor. *Behavior Research Methods, Instruments, & Computers 33*(3), 349–356.

Chen, H., Y. Wang, M. C. Paik, and H. A. Choi (2013). A marginal approach to reduced-rank penalized spline smoothing with application to multilevel functional data. *Journal of the American Statistical Association 108*(504), 1216–1229.

Culhane, K., M. OâĂŹConnor, D. Lyons, and G. Lyons (2005). Accelerometers in rehabilitation medicine for older adults. *Age and ageing 34*(6), 556–560.

Di, C., C. M. Crainiceanu, B. S. Caffo, and N. Punjabi (2009). Multilevel functional principal component analysis. *Ann. Appl. Statist. 3*, 458–488.

Diggle, P., P. Heagerty, K.-Y. Liang, and S. Zeger (2002). *Analysis of longitudinal data*. Oxford University Press.

Eilers, P. and B. Marx (2003). Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. *Chemometrics and Intelligent Laboratory Systems 66*, 159–174.

Ferrucci, L. and D. Alley (2007). Obesity, disability, and mortality: a puzzling link. *Archives of internal medicine 167*(8), 750–751.

Guo, W. (2002). Functional mixed effects models. *Biometrics 58*(1), 121–128.

Hall, P. and M. Hosseini-Nasab (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68*(1), 109–126.

He, B., J. Bai, V. V. Zipunnikov, A. Koster, P. Caserotti, B. Lange-Maia, N. W. Glynn, T. B. Harris, and C. M. Crainiceanu (2014). Predicting human movement with multiple accelerometers using movelets. *Medicine and science in sports and exercise 46*(9), 1859–1866.

James, G. M., T. J. Hastie, and C. A. Sugar (2000). Principal component models for sparse functional data. *Biometrika 87*(3), 587–602.

Karhunen, K. (1947). *Über lineare Methoden in der Wahrscheinlichkeitsrechnung*, Volume 37. Universitat Helsinki.

Kirkpatrick, M. and N. Heckman (1989). A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters. *Journal of mathematical biology 27*(4), 429–450.

Leveille, S. G., J. M. Guralnik, L. Ferrucci, and J. A. Langlois (1999). Aging successfully until death in old age: opportunities for increasing active life expectancy. *American Journal of Epidemiology 149*(7), 654–664.

Lin, X. and R. J. Carroll (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American statistical Association 95*(450), 520–534.

Lin, X., N. Wang, A. H. Welsh, and R. J. Carroll (2004). Equivalent kernels of smoothing splines in nonparametric regression for clustered/longitudinal data. *Biometrika 91*(1), 177–193.

Morris, J. S., P. J. Brown, R. C. Herrick, K. A. Baggerly, and K. R. Coombes (2008). Bayesian analysis of mass spectrometry proteomic data using wavelet-based functional mixed models. *Biometrics 64*(2), 479–489.

Morris, J. S. and R. J. Carroll (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68*(2), 179–199.

Morris, J. S., M. Vannucci, P. J. Brown, and R. J. Carroll (2003). Wavelet-based nonparametric modeling of hierarchical functions in colon carcinogenesis. *Journal of the American Statistical Association 98*(463), 573–583.

Müller, H.-G. (2005). Functional modelling and classification of longitudinal data*. *Scandinavian Journal of Statistics 32*(2), 223–240.

Pate, R. R., M. Pratt, S. N. Blair, W. L. Haskell, C. A. Macera, C. Bouchard, D. Buchner, W. Ettinger, G. W. Heath, A. C. King, et al. (1995). Physical activity and public health: a recommendation from the centers for disease control and prevention and the american college of sports medicine. *Jama 273*(5), 402–407.

Ramsay, J. and C. J. Dalzell (1991). Some tools for functional data analysis (with Discussion). *J. R. Statist. Soc. B 53*, 539–572.

Ramsay, J. and B. Silverman (2005). *Functional data analysis*. New York: Springer.

Ramsay, J. O. (2006). *Functional data analysis*. Wiley Online Library.

Ruppert, D., M. P. Wand, and R. J. Carroll (2003). *Semiparametric regression*. Number 12. Cambridge university press.

Schrack, J. A., V. Zipunnikov, J. Goldsmith, J. Bai, E. M. Simonsick, C. Crainiceanu, and L. Ferrucci (2013). Assessing the âĂIJphysical cliffâĂİ: detailed quantification of age-related

differences in daily patterns of physical activity. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, glt199.

Silverman, B. W. et al. (1996). Smoothed functional principal components analysis by choice of norm. *The Annals of Statistics 24*(1), 1–24.

Stone, J. L. and A. H. Norris (1966). Activities and attitudes of participants in the baltimore longitudinal study. *Journal of gerontology 21*(4), 575–580.

Troiano, R. P., D. Berrigan, K. W. Dodd, L. C. Masse, T. Tilert, M. McDowell, et al. (2008). Physical activity in the united states measured by accelerometer. *Medicine and science in sports and exercise 40*(1), 181.

Welsh, A. H., X. Lin, and R. J. Carroll (2002). Marginal longitudinal nonparametric regression: locality and efficiency of spline and kernel methods. *Journal of the American Statistical Association 97*(458), 482–493.

Wood, S. N. (2000). Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 413–428.

Xiao, L., L. Huang, J. A. Schrack, L. Ferrucci, V. Zipunnikov, and C. M. Crainiceanu (2015). Quantifying the lifetime circadian rhythm of physical activity: a covariate-dependent functional approach. *Biostatistics 16*(2), 352–367.

Xiao, L., D. Ruppert, V. Zipunnikov, and C. Crainiceanu (2014). Fast covariance function estimation for high-dimensional functional data. *Stat. Comput.*. To appear.

Yao, F., H.-G. Müller, A. J. Clifford, S. R. Dueker, J. Follett, Y. Lin, B. A. Buchholz, and

J. S. Vogel (2003). Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate. *Biometrics 59*(3), 676–685.

Yorston, L. C., G. S. Kolt, and R. R. Rosenkranz (2012). Physical activity and physical function in older adults: the 45 and up study. *Journal of the American Geriatrics Society 60*(4), 719–725.

7 Appendix

Proof of Theorem 1.

Proof. Denote the empirical covariance as $\hat{\Sigma}$. It is positive semi-definite. If we can find a matrix Y such that $\hat{\Sigma} = YY'$; then $(YO)(YO)' = \hat{\Sigma}$ when $O'O = OO' = I$.

It suffices to check the following two conditions.

- The FACE estimator of the covriance matrix has the form $\tilde{\Sigma} = S\hat{\Sigma}S = SYY'S = SYO(YO)'S$, where S is a symmetric smoother matrix. This implies if the algorithm for selecting the smoothing parameter λ is the same, then we have the same smoothing covariance regardless of such an orthogonal transformation;

- The smoothing parameter λ chosen by minimizing the pooled generalized cross validation (PGCV), a functional extension of the GCV: $\| Y - SY \|^2 / \{1 - \frac{tr(S)}{m}\}^2$ is invariant of orthogonal transformation, where $\| \cdot \|$ is the Euclidean norm of a vector. This is true because (a) the denominator does not depend on O; and (b) $\| Y - SY \|^2 = \| YO - SYO \|^2$ since O is a orthogonal matrix.

□