Biometrics 70, 802–811
December 2014

DOI: 10.1111/biom.12211

Hierarchical Functional Data with Mixed Continuous and Binary Measurements

Haocheng Li,[1,*] John Staudenmayer,[2] and Raymond J. Carroll[1]

[1]Department of Statistics, Texas A&M University, 3143 TAMU, College Station, Texas, U.S.A.
[2]Department of Mathematics and Statistics, University of Massachusetts, Amherst, Massachusetts, U.S.A.
[*]*email:* haochengli@stat.tamu.edu

Summary. Motivated by objective measurements of physical activity, we take a functional data approach to longitudinal data with simultaneous measurement of a continuous and a binary outcomes. The regression structures are specified as smooth curves measured at various time-points with random effects that have a hierarchical correlation structure. The random effect curves for each variable are summarized using a few important principal components, and the association of the two longitudinal variables is modeled through the association of the principal component scores. We use penalized splines to model the mean curves and the principal component curves, and cast the proposed model into a mixed effects model framework for model fitting, prediction and inference. Via a quasilikelihood type approximation for the binary component, we develop an algorithm to fit the model. Data-based transformation of the continuous variable and selection of the number of principal components are incorporated into the algorithm. The method is applied to the motivating physical activity data and is evaluated empirically by a simulation study. Extensions for different types of outcomes are also discussed.

Key words: Accelerometry; Binary longitudinal data; Longitudinal data; Mixed-effects model; Penalized splines; Physical activity; Principal components; Sedentary behavior.

1. Introduction

We propose a new methodology for modeling paired functional data that consist of simultaneous measurements of a continuous and a binary variable. Our methodology is motivated by data recording the physical activity of individuals over time. The paired factors, energy expenditure and interruptions to sedentary behavior (sitting or lying down), were measured simultaneously every 5 minutes for 3 hours in sixty individuals. The main purposes of this study are to model the functional pattern of the two measurements, and to explore their correlation structures. The unit of energy expenditure recorded by the device is the metabolic equivalent (MET), a continuous measurement. In addition, the interruption of sedentary behavior measurement (INT) is binary, recording whether sedentary behavior was interrupted at least once in the corresponding time interval. Figure 1a displays a sample data from one subject. Figure 1b shows the averaged METs level across subjects with or without interruption of sedentary behavior in each time point, respectively. The two curves have large differences, which implies a possible correlation between METs and interruption of sedentary behavior. Figures 1c–f are histograms and Q–Q plots for the METs level at two time points. The plots suggest the continuous measurement are skewed.

Current studies have mainly focused on either continuous or binary variables treated separately. For example, Yao, Müller, and Wang (2005b) discuss functional linear regression, and Hall, Müller, and Yao (2008) develop a modeling strategy for non-Gaussian longitudinal observations. However, there is very limited methodology for joint analysis of continuous and binary variables. In addition, as in our application to measurements of physical activity, data are sometimes very skewed, and transformation is required. In this article we develop methodology to handle both binary and skewed continuous variables. This problem has not been addressed previously.

For the analysis of paired data with different types, a major issue is to model the correlation structure between the two measurements. A naive way is to ignore the correlation and fit the two variables separately, but this potentially gives less insight into the data than a paired analysis. In particular, our application requires us to understand the association patterns between two measurements over time, and, as we will show, ignoring the correlation between the two responses can lead to biased predictions.

To solve this problem, Gueorguieva and Agresti (2001) suggest using random effects to link the variables. In the case of functional data though, a large number of random effects are required to model the smooth curves, and this can cause problems in computation and model interpretation. To circumvent this issue, Zhou, Huang, and Carroll (2008) develop a method to reduce the dimension of the problem by using a few important principal components to summarize random curves. The correlation between the paired variables is then modeled by the correlation across the principal component scores. We address our problem with this general approach, but the existing methodology is limited to paired *continuous* data.

To facilitate the study of the correlation structure, we propose a paired "pseudo" normal distribution strategy. We use a penalized quasilikelihood method to approximate the binary variable by a "pseudo" normal variable. The skewed-normal variable with Box–Cox transformation can also be handled

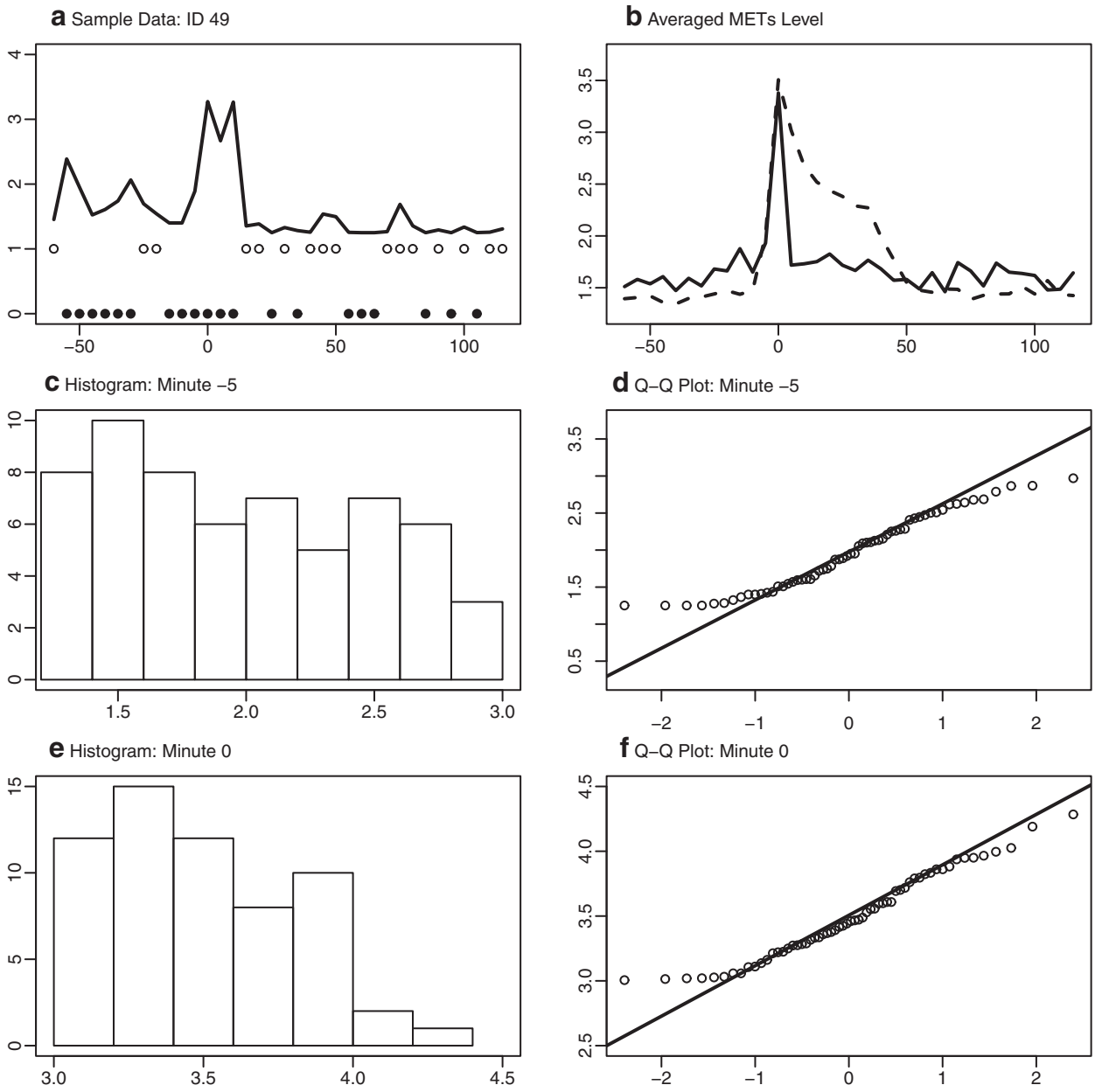

Figure 1. (a) Sample data from subject with ID 49. The solid line displays the METs level. Unfilled and filled dots display whether or not interruption of sedentary behavior occurs from minute −60 to 115, respectively. See Section 5 for more explanation of the data. (b) Averaged METs level across subjects. Solid and dashed lines represent the averaged METs level across subjects with and without interruption of sedentary behavior from minute −60 to 115, respectively. (c) and (d) histogram and Q–Q plot for the METs level on minute −5. (e) and (f) histogram and Q–Q plot for the METs level on minute 0.

as normal. Thus two variables can be treated as a paired "pseudo" normal distribution, which facilitates estimation of the correlation structure. Our methodological strategy can be extended to other types of variables with exponential distributions, as we will show, see Section 6 and Online Supplemental Materials.

Current computational techniques for the mixed continuous and binary problems mainly use Monte Carlo or Gauss–Hermite quadrature approximations (Gueorguieva and Agresti, 2001), which can be computationally expensive in many scenarios. Based on our paired "pseudo" normal distribution, a new efficient algorithm is proposed. The algorithm has to estimate the principal component vectors, and it includes features of both the model-based approach proposed by Zhou et al. (2008) and the eigen-decomposition method discussed in Yao, Müller, and Wang (2005a). Since our problem involves skewed continuous and binary variables, the likelihood approach is attractive, but the eigen-decomposition

approach is much more computationally efficient. As a result, the new algorithm includes both features of likelihood and eigen-decomposition.

The article is organized as follows. Section 2 describes the model, while Section 3 describes our algorithm for model fitting. Section 4 gives results from a simulation study. Section 5 analyzes the motivating physical activity data set. Concluding remarks and the model extensions are in Section 6.

2. Model

2.1. *The Mixed Effects Model for Continuous and Binary Data*

Let $\{Y_i(t), W_i(t)\}$ be the paired continuous and binary outcome observations at time t for subject $i = 1, \ldots, n$, and let t_{ij} for $j = 1, \ldots, m_i$ be the observation times. Let $g_{\text{tr}}(\cdot;\lambda)$ be the Box–Cox transformation with transformation parameter λ, and let logit$(\cdot)$ denote the logit transformation. Our general model is

$$g_{\text{tr}}\{Y_i(t);\lambda\} = \mu(t) + \mathbf{U}_i(t) + \epsilon_{yi}(t), \tag{1}$$

$$\text{logit}[\text{pr}\{W_i(t) = 1\}] = \nu(t) + \mathbf{V}_i(t), \tag{2}$$

where $\mu(t)$ and $\nu(t)$ are fixed curves, $\mathbf{U}_i(t)$ and $\mathbf{V}_i(t)$ are correlated random effects curves, and $\epsilon_{yi}(t)$ denotes independent random noise with mean zero and variance σ^2. We assume that given $\mathbf{U}_i(t)$ and $\mathbf{V}_i(t)$, the paired observations are independent. Therefore, the correlation structure between the two variables comes from the random effects curves $\mathbf{U}_i(t)$ and $\mathbf{V}_i(t)$.

We further model $\mathbf{U}_i(t)$ and $\mathbf{V}_i(t)$ by k_y and k_w principal components, so that

$$\mathbf{U}_i(t) = \sum_{\ell=1}^{k_y} f_{y,\ell}(t)\alpha_{yi,\ell};\ \mathbf{V}_i(t) = \sum_{\ell=1}^{k_w} f_{w,\ell}(t)\alpha_{wi,\ell}. \tag{3}$$

Here $\{f_{y,1}(t), \ldots, f_{y,k_y}(t)\}$ and $\{f_{w,1}(t), \ldots, f_{w,k_w}(t)\}$ are orthogonal principal component functions, which have $\int f_{y,\ell}(t) f_{y,\ell^*}(t)\,\mathrm{d}t = \int f_{w,\ell}(t) f_{w,\ell^*}(t)\,\mathrm{d}t = I(\ell = \ell^*)$, where $I(\cdot)$ is an indicator function. $\{\alpha_{yi,1}, \ldots, \alpha_{yi,k_y}\}$ and $\{\alpha_{wi,1}, \ldots, \alpha_{wi,k_w}\}$ are independent respective principal component scores.

In practice, the fixed curves $\{\mu(t), \nu(t)\}$, number of principal components (k_y, k_w) as well as principal component functions $\{f_{y,\ell}(t), f_{w,\ell^*}(t); \ell = 1, \ldots, k_y, \ell^* = 1, \ldots, k_w\}$ are unknown, and need to be estimated. In addition, $(\alpha_{yi,\ell}, \alpha_{wi,\ell^*}; \ell = 1, \ldots, k_y, \ell^* = 1, \ldots, k_w)$ are also unknown and we treat them as random effects following normal distributions with zero means and a covariance matrix from which we obtain the correlation structure between paired observations. The detailed model specification will be given in Section 2.2.

2.2. *Modeling with B-Splines*

We employ a set of smooth basis functions to represent the functions $\{\mu(t), \nu(t), \mathbf{U}_i(t), \mathbf{V}_i(t)\}$. Other basis expansion approximation of functions suggested in Ruppert, Wand, and Carroll (2003) could of course be used in our methodology. Let $b(t) = \{b_1(t), \ldots, b_q(t)\}^{\mathrm{T}}$ be the vector of orthogonal B-spline basis functions evaluated at t, which can be computed using an exact approach found in the R package "orthogonalsplinebasis." This means that $\int b(t)b^{\mathrm{T}}(t)\,\mathrm{d}t = \mathbf{I}_q$, where $\mathbf{I}_q$ is the $q \times q$ identity matrix. In this approach, and similar to Zhou et al. (2008) we smooth the fixed curves and principal component functions by writing

$$\mu(t) = b^{\mathrm{T}}(t)\beta_y;\ \nu(t) = b^{\mathrm{T}}(t)\beta_w;$$
$$f_{y,\ell}(t) = b^{\mathrm{T}}(t)\theta_{y,\ell};\ f_{w,\ell}(t) = b^{\mathrm{T}}(t)\theta_{w,\ell}, \tag{4}$$

where β_y and β_w are $q \times 1$ spline coefficients vectors for fixed effects, and $\theta_{y,\ell}$ and $\theta_{w,\ell}$ are $q \times 1$ orthogonal spline coefficients vectors for principal component functions which have $\theta_{y,\ell}^{\mathrm{T}}\theta_{y,\ell^*} = \theta_{w,\ell}^{\mathrm{T}}\theta_{w,\ell^*} = I(\ell = \ell^*)$.

Combining (3) and (4), we write model (1) and (2) as

$$g_{\text{tr}}\{Y_i(t);\lambda\} = b^{\mathrm{T}}(t)\beta_y + b^{\mathrm{T}}(t)\Theta_y\alpha_{yi} + \epsilon_{yi}(t); \tag{5}$$

$$\text{logit}[\text{pr}\{W_i(t) = 1\}] = b^{\mathrm{T}}(t)\beta_w + b^{\mathrm{T}}(t)\Theta_w\alpha_{wi}, \tag{6}$$

where $\Theta_y = (\theta_{y,1}, \ldots, \theta_{y,k_y})$ and $\Theta_w = (\theta_{w,1}, \ldots, \theta_{w,k_w})$, $\alpha_{yi} = (\alpha_{yi,1}, \ldots, \alpha_{yi,k_y})^{\mathrm{T}}$ and $\alpha_{wi} = (\alpha_{wi,1}, \ldots, \alpha_{wi,k_w})^{\mathrm{T}}$. Based on the orthogonality in (4), $\Theta_y^{\mathrm{T}}\Theta_y = \mathbf{I}_{k_y}$ and $\Theta_w^{\mathrm{T}}\Theta_w = \mathbf{I}_{k_w}$.

We assume the random effect principal component scores α_{yi} and α_{wi} follow normal distributions with mean zero and covariance matrices $\Delta_{yy} = \text{diag}(\Delta_{yy,1}, \ldots, \Delta_{yy,k_y})$ and $\Delta_{ww} = \text{diag}(\Delta_{ww,1}, \ldots, \Delta_{ww,k_w})$, where $\Delta_{yy,\ell} = \text{var}(\alpha_{yi,\ell})$ and $\Delta_{ww,\ell} = \text{var}(\alpha_{wi,\ell})$. For identifiability (Zhou et al., 2008), we require $\Delta_{yy,1} > \cdots > \Delta_{yy,k_y}$ and $\Delta_{ww,1} > \cdots > \Delta_{ww,k_w}$. In our model settings, $\Theta_y\alpha_{yi}$ and $\Theta_w\alpha_{wi}$ are identifiable, but the signs of Θ_y and Θ_w in each of their columns are not identifiable. Identifiability can be achieved using the sign constraint criteria discussed in Zhou et al. (2008, 2010).

To study the correlated paired observations, we have to study the association structure between α_{yi} and α_{wi}. Define the random effects $\alpha_i = (\alpha_{yi}^{\mathrm{T}}, \alpha_{wi}^{\mathrm{T}})^{\mathrm{T}} = \text{Normal}(0, \Delta)$ with the $(k_y + k_w) \times (k_y + k_w)$ matrix Δ including diagonal elements Δ_{yy} and Δ_{ww} and off-diagonal element Δ_{yw}. In particular, Δ_{yw} determines the covariance of $\mathbf{U}_i(t)$ and $\mathbf{V}_i(t)$ by

$$\text{cov}\{\mathbf{U}_i(t), \mathbf{V}_i(t)\} = b^{\mathrm{T}}(t)\Theta_y\Delta_{yw}\Theta_w^{\mathrm{T}}b(t).$$

Therefore, the modeling with B-splines involves six sets of parameters to be estimated: (a) the Box–Cox transformation parameter: λ; (b) the B-spline coefficients for the fixed effects: β_y and β_w; (c) the random noise variance: σ^2; (d) the number of principal components: k_y and k_w; (e) the B-spline coefficients for principal component functions: Θ_y and Θ_w; and (f) the principal component scores covariance matrix: Δ_{yy}, Δ_{ww} and Δ_{yw}.

2.3. *Paired "Pseudo" Normal Distribution*

Section 2.2 specifies the model and the parameters that are to be estimated, but estimation is complicated by the binary variable. To solve this problem, we approximate the binary variable $W_i(t)$ using a penalized quasilikelihood strategy that includes a second order approximation term. This method was introduced in Goldstein and Rasbash (1996) and greatly improves over the methods that were discussed in Breslow and Clayton (1993). The method is as follows. Let $H(\cdot)$ be the inverse logit function, and let the first and second derivatives of $H(\cdot)$ be $H'(\cdot)$ and $H''(\cdot)$, respectively. Based on the

binary data model (6), given known values of $(\widehat{\beta}_w, \widehat{\Theta}_w, \widehat{\alpha}_{wi})$ and $\widehat{\eta}_i(t) = b^{\mathrm{T}}(t)\widehat{\beta}_w + b^{\mathrm{T}}(t)\widehat{\Theta}_w\widehat{\alpha}_{wi}$, we have the approximate model

$$\begin{aligned} W_i^*(t) &= [1/H'\{\widehat{\eta}_i(t)\}][W_i(t) - H\{\widehat{\eta}_i(t)\}] + \widehat{\eta}_i(t) \\ &\quad -[1/2H'\{\widehat{\eta}_i(t)\}]H''\{\widehat{\eta}_i(t)\}\{b^{\mathrm{T}}(t)\widehat{\Theta}_w\widehat{\mathrm{var}}(\alpha_{wi} - \widehat{\alpha}_{wi})\widehat{\Theta}_w^{\mathrm{T}}b(t)\} \\ &= b^{\mathrm{T}}(t)\beta_w + b^{\mathrm{T}}(t)\Theta_w\alpha_{wi} + \epsilon_{wi}(t), \end{aligned} \tag{7}$$

where $\epsilon_{wi}(t) = \mathrm{Normal}[0, 1/H'\{\widehat{\eta}_i(t)\}]$.

Therefore, models (5) and (6) become

$$Y_i^*(t) = b^{\mathrm{T}}(t)\beta_y + b^{\mathrm{T}}(t)\Theta_y\alpha_{yi} + \epsilon_{yi}(t); \tag{8}$$

$$W_i^*(t) = b^{\mathrm{T}}(t)\beta_w + b^{\mathrm{T}}(t)\Theta_w\alpha_{wi} + \epsilon_{wi}(t). \tag{9}$$

where $Y_i^*(t) = g_{\mathrm{tr}}\{Y_i(t); \lambda\}$. Equations (8) and (9) play important roles in inference because they indicate that, the Box–Cox transformed continuous variable is normal, the transformed binary variable is approximately normal, and the random effect terms α_{yi} and α_{wi} are also normal. Therefore, we obtain a paired "pseudo" normal distribution. The bivariate formulation also leads to convenient estimation of model parameters; see Section 3.

3. Model Fitting Procedure

3.1. *Link to the Mixed Effects Model*

To facilitate model fitting, we first rewrite the principal component part of the models (8) and (9) by $\gamma_{yi} = \Theta_y\alpha_{yi}$, $\gamma_{wi} = \Theta_w\alpha_{wi}$, and further denote $\gamma_i = (\gamma_{yi}^{\mathrm{T}}, \gamma_{wi}^{\mathrm{T}})^{\mathrm{T}}$ and covariance matrix $\Psi = \mathrm{cov}(\gamma_i) = \mathrm{cov}(\gamma_{yi}, \gamma_{wi}) = \mathrm{cov}(\Theta_y\alpha_{yi}, \Theta_w\alpha_{wi})$ with block diagonal elements $\Psi_{yy} = \Theta_y\Delta_{yy}\Theta_y^{\mathrm{T}}$ and $\Psi_{ww} = \Theta_w\Delta_{ww}\Theta_w^{\mathrm{T}}$ and off-diagonal elements $\Psi_{yw} = \Theta_y\Delta_{yw}\Theta_w^{\mathrm{T}}$ and its transpose.

Then (8) and (9) become

$$Y_i^*(t) = b^{\mathrm{T}}(t)\beta_y + b^{\mathrm{T}}(t)\gamma_{yi} + \epsilon_{yi}(t); \tag{10}$$

$$W_i^*(t) = b^{\mathrm{T}}(t)\beta_w + b^{\mathrm{T}}(t)\gamma_{wi} + \epsilon_{wi}(t). \tag{11}$$

The newly introduced random effect γ_i and its covariance matrix Ψ are not included in the list of model parameters in Section 2.2, but they are important ancillary components in our model fitting. In particular, the special case of $k_y = k_w = q$ makes Ψ full rank. That in turn makes (10) and (11) equivalent to the commonly used mixed effect model or multivariate mixed effects model. Our estimation algorithm takes advantage of that equivalence. We describe the method in more detail in the next section.

3.2. *Joint Estimation Algorithm*

We estimate the parameters by extending the idea of an ECME algorithm (Schafer, 1998). The ECME algorithm updates the fixed effects parameters by the Newton–Raphson approach, and updates the random effects parameters by the EM method. We provide a brief sketch of the model fitting here.

We set the initial numbers of principal components to be $k_y = k_w = q$ and thus the initial value of Ψ is a full rank covariance matrix. We also give initial values for parameters λ, β_y, β_w and σ^2. Then the iteration procedure is

1. update λ, β_y, β_w, and σ^2 by a Newton–Raphson approach,
2. update Ψ by the EM method, and
3. update k_y, k_w, Θ_y, Θ_w, Δ_{yy}, Δ_{ww}, and Δ_{yw} with an eigendecomposition of Ψ.

The entire procedure is iterated until convergence. The details of Steps 1 and 2 are given in the Online Supplementary Material. Section 3.3 displays the details for Step 3.

3.3. *Eigenvalue Decomposition Procedure*

For the Ψ obtained in the EM step, we use eigenvalue decomposition for Ψ_{yy} and Ψ_{ww} as

$$\Psi_{yy} = \widetilde{\Theta}_y\widetilde{\Delta}_{yy}\widetilde{\Theta}_y^{\mathrm{T}}; \ \Psi_{ww} = \widetilde{\Theta}_w\widetilde{\Delta}_{ww}\widetilde{\Theta}_w^{\mathrm{T}},$$

where $\widetilde{\Delta}_{yy} = \mathrm{diag}\{\widetilde{\Delta}_{yy,1}, \ldots, \widetilde{\Delta}_{yy,q}\}$ and $\widetilde{\Delta}_{ww} = \mathrm{diag}\{\widetilde{\Delta}_{ww,1}, \ldots, \widetilde{\Delta}_{ww,q}\}$. Components in $\widetilde{\Delta}_{yy}$ and $\widetilde{\Delta}_{ww}$ are sorted in decreasing order, respectively. $\widetilde{\Theta}_y$ and $\widetilde{\Theta}_w$ are orthogonal matrices, respectively. We also obtain $\widetilde{\Delta}_{yw} = \widetilde{\Theta}_y^{\mathrm{T}}\Psi_{yw}\widetilde{\Theta}_w$ and its transpose.

To select the number of principal components (k_y, k_w), we set a threshold P and choose k_y and k_w as

$$k_y = \min\left\{k : \frac{\widetilde{\Delta}_{yy,1} + \cdots + \widetilde{\Delta}_{yy,k}}{\widetilde{\Delta}_{yy,1} + \cdots + \widetilde{\Delta}_{yy,q}} \geq P\right\};$$

$$k_w = \min\left\{k : \frac{\widetilde{\Delta}_{ww,1} + \cdots + \widetilde{\Delta}_{ww,k}}{\widetilde{\Delta}_{ww,1} + \cdots + \widetilde{\Delta}_{ww,q}} \geq P\right\}.$$

Then we maintain the selected components and remove unselected components from the covariance structure by taking Θ_y and Θ_w to be the first k_y and k_w columns of $\widetilde{\Theta}_y$ and $\widetilde{\Theta}_w$, respectively; Δ_{yy} and Δ_{ww} to be the first k_y and k_w rows and columns of $\widetilde{\Delta}_{yy}$ and $\widetilde{\Delta}_{ww}$, respectively; and Δ_{yw} to be the first k_y rows and k_w columns of $\widetilde{\Delta}_{yw}$. Thus, we have updated parameters k_y, k_w, Θ_y, Θ_w, Δ_{yy}, Δ_{ww} and Δ_{yw}.

Finally, the reduced-rank Ψ^{RR} can be obtained by

$$\Psi^{\mathrm{RR}} = \boldsymbol{\Theta}\Delta\boldsymbol{\Theta}^{\mathrm{T}},$$

where $\boldsymbol{\Theta}$ is the block diagonal matrix of updated Θ_y and Θ_w, and Δ has diagonal elements with updated Δ_{yy} and Δ_{ww}, and off-diagonal element with updated Δ_{yw}. The updated reduced-rank Ψ^{RR} is used in the next iteration.

A subjective choice of P is often satisfactory. We use $P = 0.85$ in all our numerical studies to follow, simulations and data analysis, with good results. Of course, other approaches such as AIC and BIC could be used instead.

3.4. *Maximum Penalized Likelihood*

The previous discussion focuses on the modeling of the response variables using basis functions. It is helpful however to introduce roughness penalties to regularize the fits of the functions (Eilers and Marx, 1996).

We penalize the loglikelihood and update the parameters in each iteration to maximize

$$\mathcal{L}_{pen} = \mathcal{L} - \tau_\mu \beta_y^{\mathrm{T}} D\beta_y - \tau_v \beta_w^{\mathrm{T}} D\beta_w - \tau_{\mathbf{u}} \textstyle\sum_{\ell=1}^{k_y} \theta_{y,\ell}^{\mathrm{T}} D\theta_{y,\ell} - \tau_{\mathbf{v}} \textstyle\sum_{\ell=1}^{k_w} \theta_{w,\ell}^{\mathrm{T}} D\theta_{w,\ell}, \quad (12)$$

where the loglikelihood $\mathcal{L}$ is defined by the model (8) and (9), with formula presented in the Online Supplementary Material, τ_μ, τ_v, $\tau_{\mathbf{u}}$ and $\tau_{\mathbf{v}}$ are penalty parameters, and the penalty matrix is $D = \int b''(t) b''(t)^{\mathrm{T}} \, \mathrm{d}t$. Using maximum penalized likelihood has only a minor effect on the *form* of the algorithm in Section 3.2, although of course it has a major effect on the estimation results. In the Online Supplementary Material, we describe the updated algorithm for maximum penalized likelihood.

3.5. *Tuning Parameter Selection*

We use fivefold crossvalidation to choose penalty parameters τ_μ, τ_v, $\tau_{\mathbf{u}}$ and $\tau_{\mathbf{v}}$ in (12). Since the computational cost of searching over a four dimensional grid for τ_μ, τ_v, $\tau_{\mathbf{u}}$ and $\tau_{\mathbf{v}}$ is non-trivial though, we separate the search into two parts, as follows. First, the tuning parameters for the continuous variable, τ_μ and $\tau_{\mathbf{u}}$, are obtained by maximizing the crossvalidated likelihood based on the marginal likelihood for Y_i from model (8).

Second, we find τ_v and $\tau_{\mathbf{v}}$ by maximizing the crossvalidated likelihood for the binary variable. However, Molenberghs and Verbeke (2005) note that the likelihood value in (9) for $W_i^*(t)$ with "pseudo" normal distribution is different from model (6) for the actual binary data $W_i(t)$. As a result, for crossvalidation we use the crossvalidated likelihood from actual binary data $W_i(t)$ model

$$\int \left(\textstyle\prod_{j=1}^{m_i} H\{b^{\mathrm{T}}(t)\beta_w + b^{\mathrm{T}}(t)\Theta_w \alpha_{wi}\}^{W_i(t_{ij})} [1 - H\{b^{\mathrm{T}}(t)\beta_w + b^{\mathrm{T}}(t)\Theta_w \alpha_{wi}\}]^{1-W_i(t_{ij})} \right) f(\alpha_{wi}; \Delta_{ww}) \, \mathrm{d}\alpha_{wi}, \quad (13)$$

where $f(\alpha_{wi}; \Delta_{ww})$ is the normal density function for α_{wi} defined in Section 2.2. We compute (13) with a Gauss–Hermite quadrature method with 21 quadrature points.

An anonymous referee suggested that a REML approach might be able to be developed to estimate these turning parameters as variance components. Our crossvalidation method is well established in the non-/semi-parametric literature, and our results indicate that it works well in our particular situation, as well as being straightforward. Future work could develop a REML approach in our context.

4. Simulation Studies

In this section, we use simulation studies to illustrate the performance of our method (labeled as JOINT). As a comparison, a naive approach is also explored, which separately fits the two variables by assuming them to be independent (labeled as INDEPENDENT).

In the simulation study of 500 runs, we have $n = 60$ subjects and each subject contributes data at 36 time points. The time points for each subject are equally spaced, $t_{ij} = j$, $j = 1, \ldots, 36$. At time t, subject i has two observations $\{Y_i(t), W_i(t)\}$, where

$$g_{\mathrm{tr}}\{Y_i(t); \lambda\} = \mu(t) + f_{y,1}(t)\alpha_{yi,1} + f_{y,2}(t)\alpha_{yi,2} + \epsilon_{yi}(t);$$
$$\mathrm{logit}\{W_i(t) = 1\} = \nu(t) + f_{w,1}(t)\alpha_{wi,1},$$

with $\mathrm{var}\{\epsilon_{yi}(t)\} = \sigma^2 = 0.25$ and $\lambda = 0.5$. The fixed effects curves have the form $\mu(t) = 3 + t/25 + \exp\{-(t-18)^2/25\}$ and $\nu(t) = \exp(t/4 - 5.5)/\{1 + \exp(t/4 - 5.5)\} - 1.5$. The principal component functions are $f_{y,1}(t) = \sin\{2\pi(t-1)/35\}/\sqrt{17.5}$, $f_{y,2}(t) = \cos\{2\pi(t-1)/35\}/\sqrt{17.5}$ and $f_{w,1}(t) = 1/\sqrt{35}$, so that $\int f_{y,1}^2(t)\,\mathrm{d}t = \int f_{y,2}^2(t)\,\mathrm{d}t = \int f_{w,1}^2(t)\,\mathrm{d}t = 1$ and $\int f_{y,1}(t) f_{y,2}(t)\,\mathrm{d}t = 0$. We set $\Delta_{yy,1} = 12$, $\Delta_{yy,2} = 6$, $\Delta_{ww,1} = 36$. In addition, $\Delta_{yw,1} = 10$, and $\Delta_{yw,2} = 7$. In this setting, approximately 30% of the binary responses equal 1.0.

We use a cubic B-spline basis function with 10 equispaced knots to fit the data in both the JOINT and INDEPENDENT approaches. We also use both model fitting methods to estimate bivariate the conditional mean curve $E\{Y_i(t)|W_i(t-1) = 1, W_i(t-2) = 1\}$. Figure 2 summarizes the results graphically. Figure 2a–e shows the true fixed and random curves, and the averaged estimates of the two methods. The JOINT and INDEPENDENT methods have similar estimates which capture the true curve patterns. Figure 2f shows the estimate of $E\{Y_i(t)|W_i(t-1) = 1, W_i(t-2) = 1\}$. Our JOINT approach is obviously the much less biased of the two, indicating the need for joint modeling. Table 1 presents the means and the mean squared errors (MSE) of the parameter estimates.

Other methods based on pre-smoothing the individual data are discussed in the Online Supplementary Material.

5. Application to Physical Activity Data

In this section we apply our methods to data from Kozey-Keadle et al. (in press) who measured both energy expenditure and interruptions to sedentary behavior (sitting or lying down). Those data are part of a larger project that investigated the metabolic effects of several interventions to increase exercise and reduce sedentary time in moderately overweight but healthy office workers. The device they used was an ActivPAL™ (www.paltech.plus.com), a wearable monitor that uses accelerometers to measure movement and the angle of the wearer's leg simultaneously over time and on a dense time scale. The movement measurements are used to estimate energy expenditure, and the leg angle detects when the wearer stands up, which indicates an interruption of sedentary behavior.

The unit of energy expenditure recorded by the device is the metabolic equivalent (MET), which is a relative measure that is defined as the ratio of a person's energy cost during an activity to that person's resting energy expenditure. For instance, sitting still while awake is 1 MET, and walking is approximately 2–4 METs, depending on the speed. An activity that requires at least 3 METs is termed moderate to vigorous physical activity (MVPA). METs is a continuous measurement, but it has an approximate floor (1 MET), so

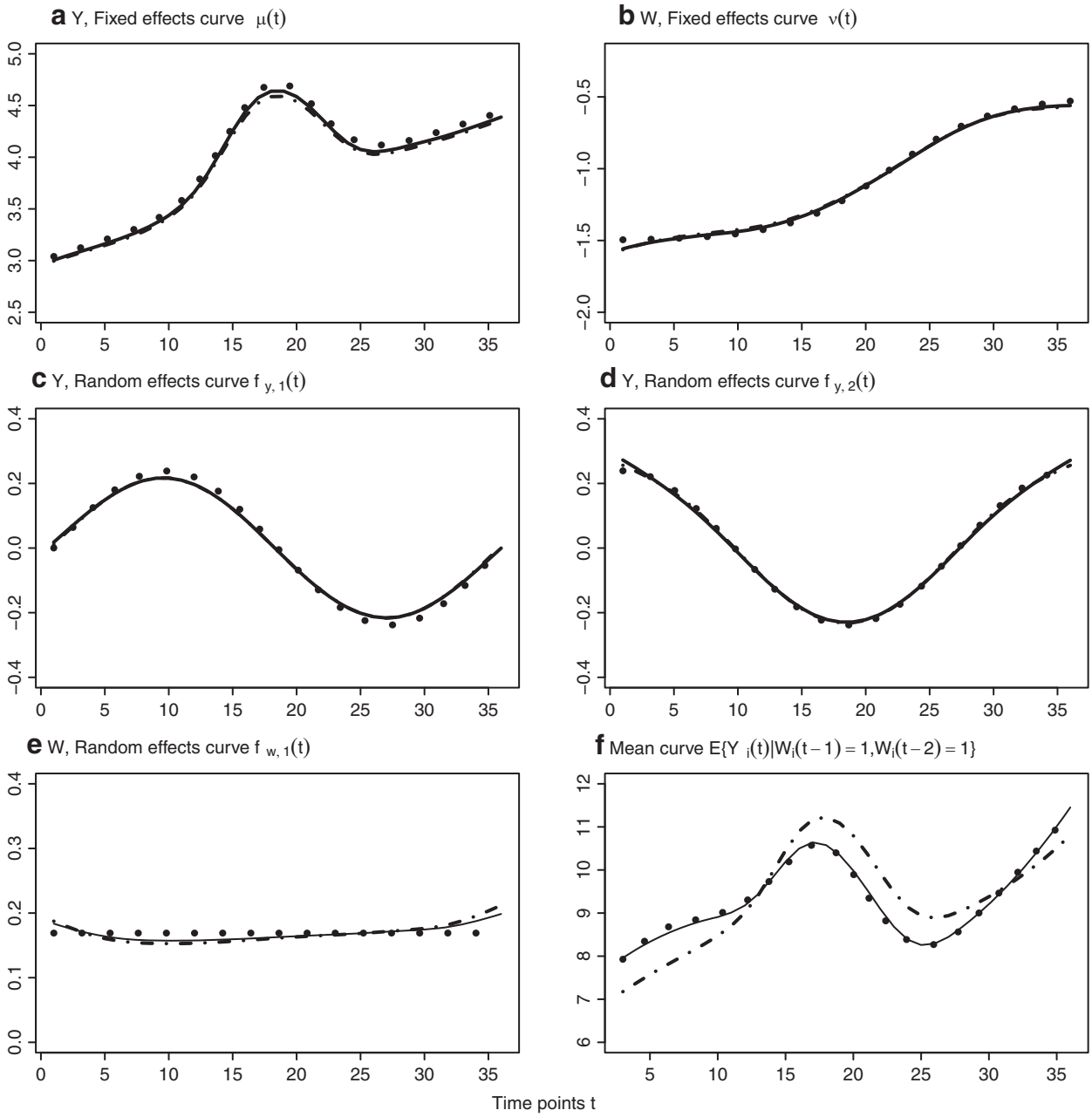

Figure 2. Fitted fixed effects curves and principal component functions for 500 simulated data sets: (a) fixed effects curve $\mu(t)$ of Y, (b) fixed effects curve $\nu(t)$ of W, (c) principal component function $f_{y,1}(t)$ for Y, (d) principal component function $f_{y,2}(t)$ for Y, (e) principal component function $f_{w,1}(t)$ for W, (f) mean curve for $E\{Y_i(t)|W_i(t-1)=1, W_i(t-2)=1\}$. Dotted lines denote the true curves. Solid and dot-dashed lines represent the averaged values of the fitted curves for the methods JOINT and INDEPENDENT, respectively.

measurements of METs tend to be skewed to the right. Referring back to the notation of Section 2, $Y_i(t)$ is mean METs minus 1.24 for subject i in the tth time interval. The interruption of sedentary behavior measurement (INT) is binary, and $W_i(t) = 1$ if sedentary behavior was interrupted at least once in the tth time interval and is zero otherwise.

We have summarized the data in five minute intervals. To illustrate our methods, we study the pattern of METs and interruptions to sedentary behavior in the time around a bout of MVPA. To do this, we selected one day from each individual and found the first MVPA bout. Then we extracted data for 1 hour before and 2 hours after that bout (36 five-minute intervals). In our figures, minute −60 denotes 1 hour before the first MVPA bout, minute 0 represents the first MVPA bout and minute 115 ends 2 hours after the first MVPA bout. Since physical activity and sedentary behavior are related, we expect correlation between METs and interruption of sedentary behavior across time.

Table 1
Results for simulation. Displayed are the average estimates and mean squared errors (MSE) of the parameters for the JOINT and INDEPENDENT methods. The numbers marked with an asterisk means that the actual number is multiplied by 1000.

Parameter	σ^2	λ	$\Delta_{yy,1}$	$\Delta_{yy,2}$
True	0.25	0.50	12.00	6.00
JOINT mean	0.24	0.49	11.54	5.32
INDEPENDENT mean	0.23	0.48	11.28	5.25
JOINT MSE	1.08*	0.92*	5.61	1.99
INDEPENDENT MSE	0.79*	0.73*	4.90	1.81
Parameter	$\Delta_{ww,1}$	$\Delta_{yw,1}$	$\Delta_{yw,2}$	
True	36.00	10.00	7.00	
JOINT mean	36.17	9.62	6.35	
INDEPENDENT mean	36.03	NA	NA	
JOINT MSE	70.44	15.20	5.80	
INDEPENDENT MSE	69.67	NA	NA	

The model used cubic B-splines with ten equally spaced interior knots. Fivefold crossvalidation was used to select the penalty parameters. Figure 3a and b presents fixed effects for METs, $\mu(t)$, and interruption of sedentary behavior, and $\nu(t)$. The plots include the estimated curves and 90% bootstrap confidence intervals. The plots illustrate that mean energy expenditure (METs) increases dramatically at about 15 minutes before the first MVPA bout, and decreases back to just above the starting level by an hour after the bout. Similarly, the probability of interrupting sedentary behavior increases before the bout of MVPA, and decreases after the bout has started. The transformation parameter is $\lambda = 0.073$ which indicates severe skewness for METs.

The approach described in Section 3 results in $k_y = k_w = 2$ principal components. Figure 3c–f displays the principal component curves for METs $\{f_{y,1}(t),\ f_{y,2}(t)\}$ and interruption of sedentary behavior, $\{f_{w,1}(t),\ f_{w,2}(t)\}$, along with corresponding 90% bootstrap confidence intervals. All of the principal component curves suggest that the subject-to-subject variability of METs and interruption of sedentary behavior can be divided approximately into three intervals: 0–60 minutes before the first MVPA bout, 0–60 minutes after the bout started, and 60–120 minutes after the bout started. The first component for METs, $f_{y,1}(t)$, rises to a peak about 40 minutes after the start of the MVPA bout, decreases after that, but it stays positive. This suggests that main mode of subject to subject variability in METs is how much energy was expended and how long the bout of activity lasted. On the other hand, $f_{y,2}(t)$ has one positive and two negative peaks, which indicates variability only in the METs cost of the bout, not the duration. For interruptions to sedentary behavior, $f_{w,1}(t)$ does not cross zero, and this shows that the variation is largely characterized by more or fewer interruptions at all time points. Moreover, the variations are positively correlated, which implies that each subject was more (or less) likely to interrupt sedentary behavior in each interval. The second component for interruption of sedentary behavior, $f_{w,2}(t)$, has one positive and two negative peaks which suggests that subjects who were more likely to have breaks from sedentary shortly after the MVPA bout were less likely to interrupt sedentary behavior before or after the bout, or vice versa.

We obtained the estimates for $\Delta_{yy,1} = 21.41$, $\Delta_{yy,2} = 5.15$, $\Delta_{ww,1} = 9.26$, $\Delta_{ww,2} = 3.85$, $\Delta_{yw,11} = 3.54$, $\Delta_{yw,12} = -4.03$, $\Delta_{yw,21} = -2.31$, and $\Delta_{yw,22} = 1.11$. They show that the primary principal component scores in the two variables are positively correlated with correlation coefficient 0.25, which indicates that the energy expenditure and sedentary behavior interruptions can be positively associated at some time points.

To illustrate explicitly the covariance structure of the principal components curves, we take $\text{cov}\{\mathbf{U}_i(t), \mathbf{V}_i(s)\}$ as 3D plot in Figure 4. The covariance surface appears to have four features. First, METs and interruption of sedentary behavior are negatively correlated when the MVPA bout occurred (t and s both zero). This makes sense since sedentary to non-sedentary transitions are less likely to occur during a bout of moderate to vigorous activity. Next, when $t > 0$ and $s > 0$, METs and interruption of sedentary behavior are positively correlated. Again, this might be expected since an interruption of sedentary behavior is likely to be associated with an increase in energy expenditure. Third, the correlation is also positive when $t > 0$ and $s < 0$.

This is probably the most interesting feature of the plot, and it suggests that people who are more likely to interrupt sedentary behavior (i.e., stand up more often) at a time when they are generally inactive are also more likely to be more active after a bout of MVPA. Finally, before the bout of MVPA ($t < 0$), there is little correlation between METs and interruption of sedentary behavior. This is likely because, by construction, METs were relatively low and constant before the MVPA bout.

Suppose someone has made consecutive sedentary behavior interruptions in the previous 10 minutes. It is of interest to estimate the expectation of energy expenditure in the next time period. More generally, there is interest in estimating the current mean of METs given the sedentary behavior interruption history in the past m time periods, which is $E\{Y_i(t)|W_i(t-1), \ldots, W_i(t-m)\}$. Our fitted model can provide such mean curves. Figure 5 shows the mean curve with/without sedentary behavior interruptions in the previous 10 minutes. It can be seen that without previous sedentary behavior interruptions, the energy expenditure is higher around the MVPA bout.

6. Extensions and Discussion

We have proposed a joint modeling and estimation strategy for functional data with both continuous and binary variables. Because the physical activity data in Section 5 is highly skewed, our algorithm estimates a Box–Cox transformation, while at the same time employing a data-based method to select the number of principal components for both variables. The simulation results are encouraging and show that our method has little bias and outperforms separate marginal analyses of the two responses. The analysis of the physical activity data using the our method demonstrates its utility in applications.

In Section 2.2, our model constructs the correlation structure between two variables by using correlated principal

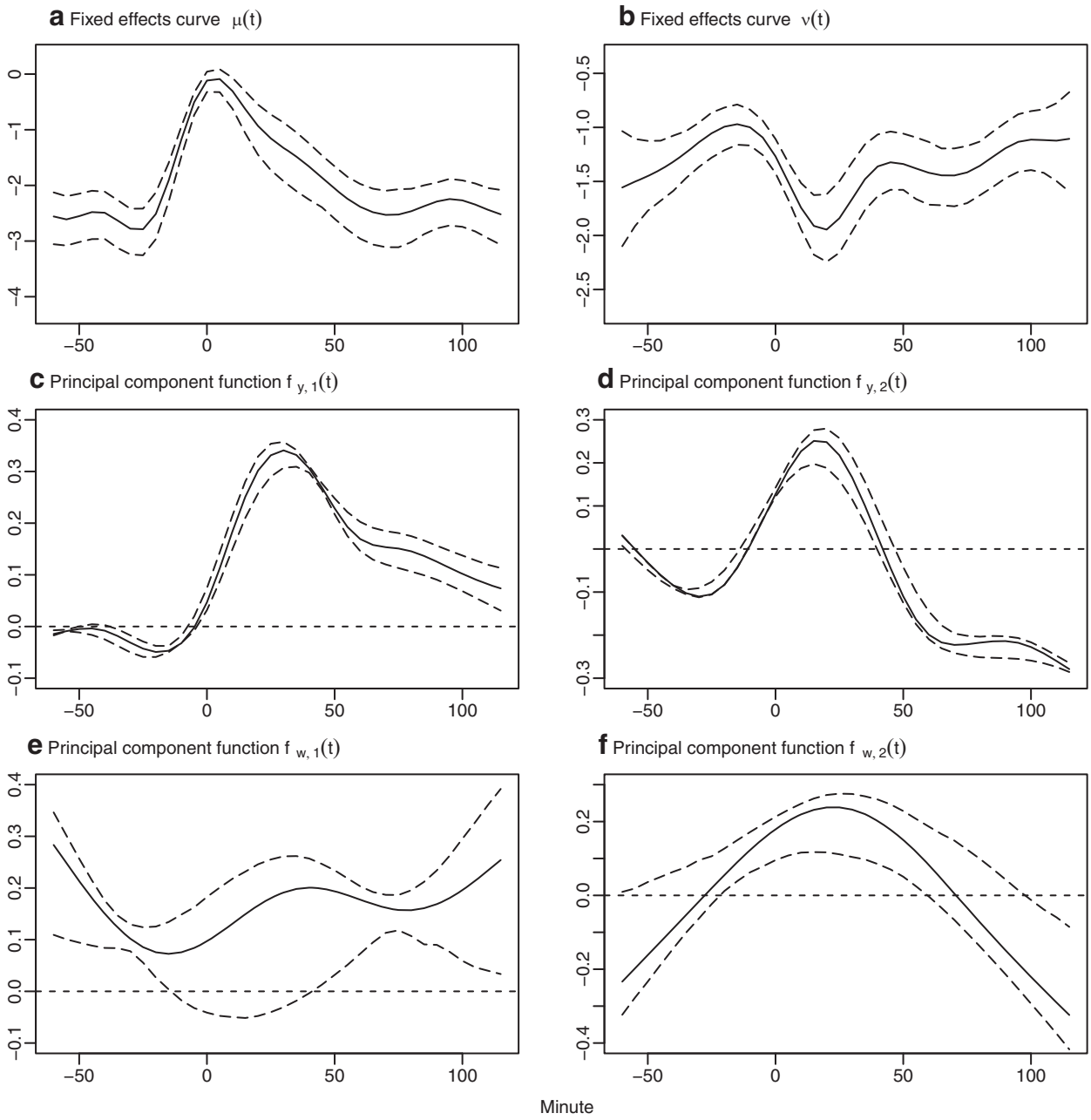

Figure 3. Fixed effects and mean structure for the physical activity data in Section 5: (a) fixed effects curve of METs, (b) fixed effects curve of interruption of sedentary behavior, (c) and (d) principal component function for METs, (e) and (f) principal component function for interruption of sedentary behavior. Solid lines represent the averaged values of the fitted curves. The upper and lower dashed lines are the 5% and 95% quantiles of the fitted values across 500 bootstrap estimates.

component scores. There are alternative models to postulate the association, such as conditional formulations. For example, we can extend our model to have

$$g_{\mathrm{tr}}\{Y_i(t)|W_i(t);\lambda\} = I\{W_i(t)=0\}\mu_0(t) + I\{W_i(t)=1\}\mu_1(t) + \mathbf{U}_i(t) + \epsilon_{yi}(t),$$
$$\mathrm{logit}[\mathrm{pr}\{W_i(t)=1\}] = \nu(t) + \mathbf{V}_i(t),$$

where $g_{\mathrm{tr}}\{Y_i(t)|W_i(t);\lambda\}$ denotes that the Box–Cox transformed continuous variable $Y_i(t)$ depends on the observation of $W_i(t)$, and $\mu_0(t)$ and $\mu_1(t)$ are the fixed effect curves under $W_i(t)$=0 and $W_i(t)$=1, respectively. Other settings follow the model in Section 2. Therefore, given different observations of $W_i(t)$, the Box–Cox transformed $Y_i(t)$ may involve different fixed effect curves. Similarly, another conditional formulation is

$$\mathrm{logit}[\mathrm{pr}\{W_i(t)=1|Y_i(t)\}] = \nu_0(t) + \nu_1\{t, Y_i(t)\} + \mathbf{V}_i(t),$$
$$g_{\mathrm{tr}}\{Y_i(t);\lambda\} = \mu(t) + \mathbf{U}_i(t) + \epsilon_{yi}(t),$$

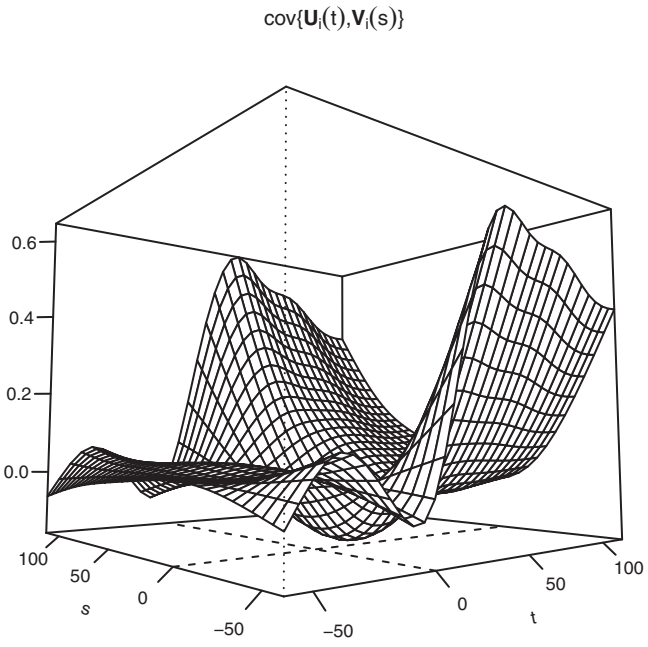

Figure 4. The estimates of covariance surfaces for $\text{cov}\{\mathbf{U}_i(t), \mathbf{V}_i(s)\}$.

where $\text{pr}\{W_i(t) = 1|Y_i(t)\}$ is the probability of $W_i(t) = 1$ given the observation of $Y_i(t)$, and $\nu_0(t)$ and $\nu_1\{t, Y_i(t)\}$ are the fixed effect curves depending on time t and the observation of $\{t, Y_i(t)\}$, respectively. Both conditional models can be modeled by similar B-spline and the paired "pseudo" normal

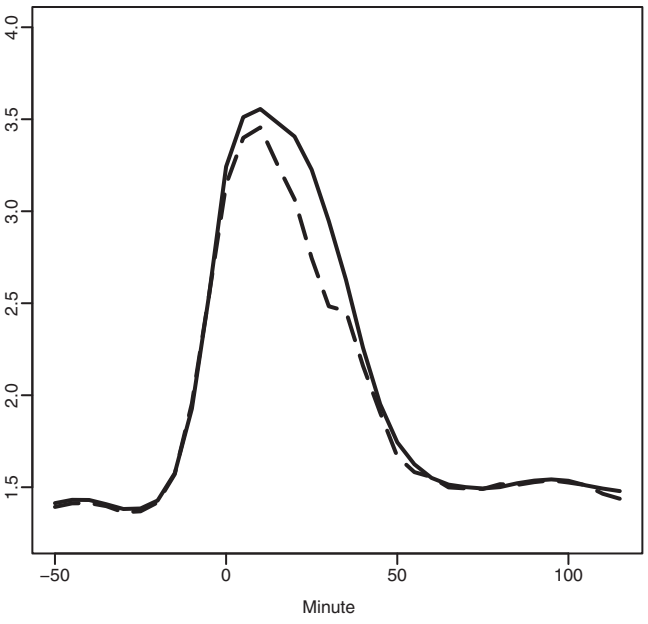

Figure 5. The estimates of mean curves for METs level. Solid line represents the expected curve without sedentary behavior interruptions in the past 10 minutes, while dashed line displays the curve with sedentary behavior interruptions in the past 10 minutes.

distribution framework discussed in Section 2, and our algorithm can fit both models with minor modifications. We study the two models in detail and use them to fit our physical activity data. The results are included in the Online Supplementary Materials.

Although our method is developed to handle physical activity data with mixed continuous and binary variables, it can be extended to handle data of other types. In particular, when the variable belongs to the exponential family with canonical link, the normalized approximation in (7) can be conducted by taking $H(\cdot)$ to be the corresponding inverse canonical link function, and $H'(\cdot)$ and $H''(\cdot)$ as the first and second derivatives of $H(\cdot)$, respectively. For example, consider count-binary mixed data where $Y_i(t)$ is replaced with a count observation with log link for subject i at time point t, and $W_i(t)$ follows the definition of binary variable as earlier. This model can be written as

$$\log\{Y_i(t)\} = \mu(t) + \mathbf{U}_i(t), \;\; \text{logit}[\text{pr}\{W_i(t) = 1\}] = \nu(t) + \mathbf{V}_i(t).$$

Again, our modeling strategy and computation algorithm can easily handle these data with minor modification. We study this model via a simulation study illustrated in the Online Supplementary Materials. The simulation results have little bias, which suggests our methods are flexible enough to handle data of different types.

Finally, we have used one parameter λ for data transformation. It would be interesting to consider letting the data transformation parameter vary with time, use $\lambda(t)$. This has the same flavor of spatially adaptive penalized regression splines, for example, Crainiceanu et al. (2007), where the penalty parameter is allowed to vary with time. Because different Box–Cox transformations are on very different scales, some normalization to alleviate this phenomena will be useful, for example, see Hinkley and Runger (1984).

7. Supplementary Materials

R programs, technical details, tables, and figures referenced in Sections 2, 3, 4, and 6 are available with this paper at the *Biometrics* website on Wiley Online Library.

Acknowledgements

We thank the Associate Editor and the referees and Dr. Sarah Kozey-Keadle for their helpful comments. Our work was supported by grants from the National Cancer Institute (R37-CA057030, R01-CA121005) and partially by the Spanish Ministry of Science and Innovation (project MTM 2011-22664).

References

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9–25.

Crainiceanu, C. M., Ruppert, D., Carroll, R. J., Joshi, A., and Goodner, B. (2007). Spatially adaptive Bayesian penalized splines with heteroscedastic errors. *Journal of Computational and Graphical Statistics* **16**, 265–288.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* **11**, 89–121.

Goldstein, H. and Rasbash, J. (1996). Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society, Series A* **159**, 505–513.

Gueorguieva, R. V. and Agresti, A. (2001). A correlated probit model for joint modeling of clustered binary and continuous responses. *Journal of the American Statistical Association* **96**, 1102–1112.

Hall, P., Müller, H.-G., and Yao, F. (2008). Modelling sparse generalized longitudinal observations with latent Gaussian processes. *Journal of the Royal Statistical Society, Series B* **70**, 703–723.

Hinkley, D. V. and Runger, G. (1984). The analysis of transformed data. *Journal of the American Statistical Association* **79**, 302–309.

Kozey-Keadle, S., Staudenmayer, J., Libertine, A., Mavilia, M., Lyden, K., Braun, B., and Freedson, P. (2013). Changes in Sedentary Time and Physical Activity in Response to an Exercise Training and/or Lifestyle Intervention. *Journal of Physical Activity and Health*. [Epub ahead of print].

Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York, USA: Springer.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. New York: Cambridge University Press.

Schafer, J. L. (1998). Some improved procedures for linear mixed models. Technical Report, The Methodological Center, The Pennsylvania State University.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005a). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**, 577–590.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005b). Functional linear regression analysis for longitudinal data. *The Annals of Statistics* **33**, 2873–2903.

Zhou, L., Huang, J. Z., and Carroll, R. J. (2008). Joint modelling of paired sparse functional data using principal components. *Biometrika* **95**, 601–619.

Zhou, L., Huang, J. Z., Martinez, J. G., Maity, A., Baladandayuthapani, V., and Carroll, R. J. (2010). Reduced rank mixed effects models for spatially correlated hierarchical functional data. *Journal of the American Statistical Association* **105**, 390–400.

Received June 2013. *Revised May* 2014. *Accepted June* 2014.