A semiparametric stochastic mixed model for bivariate longitudinal data

Kexin Ji and Joel A. Dubin

Department of Statistics and Actuarial Science, University of Waterloo, 200 University Ave. W. Waterloo, Ontario N2L 3G1, Canada
kji@uwaterloo.ca, jdubin@uwaterloo.ca

Abstract

We propose and consider inference for a semiparametric stochastic mixed model for bivariate longitudinal data. The approach models the mean of responses by parametric fixed effects and a smooth nonparametric function for the underlying time effects, and the relationship across the bivariate responses by a bivariate Gaussian random field and a joint distribution of random effects. The proposed model not only can model complicated individual hormone profiles, but also allows for more flexible within-subject and between-response correlations. The fixed effects regression coefficients and the nonparametric time functions are estimated using maximum penalized likelihood, where the resulting estimator for the nonparametric time function is a cubic smoothing spline. The smoothing parameters and all variance components are estimated simultaneously using restricted maximum likelihood. Simulation results show that the parameter estimates are close to the true values with small standard errors. The fit of the proposed model on a real bivariate longitudinal dataset of pre-menopausal women performs well.

Keywords: Bivariate longitudinal data, Gaussian field, Penalized likelihood, Smoothing spline

1. Introduction

The distinctive feature of longitudinal data is that measurements of each subject are collected repeatedly over time, which induces a correlation structure among observations for the same subject. For multivariate longitudinal data, repeated measurements are observed jointly for two or more responses. To better understand the relationships among the responses at the same time or different times, the correlation structure among the responses needs to be studied.

The model we proposed is motivated by a longitudinal hormone study on estrogen and progesterone in pre-menopausal women, where assays of daily urine samples for metabolites of estrogen and progesterone are collected over one menstrual cycle. We are interested in modelling time courses for the estrogen and progesterone metabolite profiles for a single cycle, the effects of covariates on the hormone excretion, and the potential correlation between the two hormones. Joint modeling of the hormone profiles of estrogen and progesterone is challenging. First, the time courses of the univariate hormones profiles is complex such that to model using a simple parametric function, such as linear mixed effects model, is insufficient; see Figure 1. Second, multiple layers of correlation structures, say within-subject correlation between the bivariate hormones at different time points, also present a challenge.

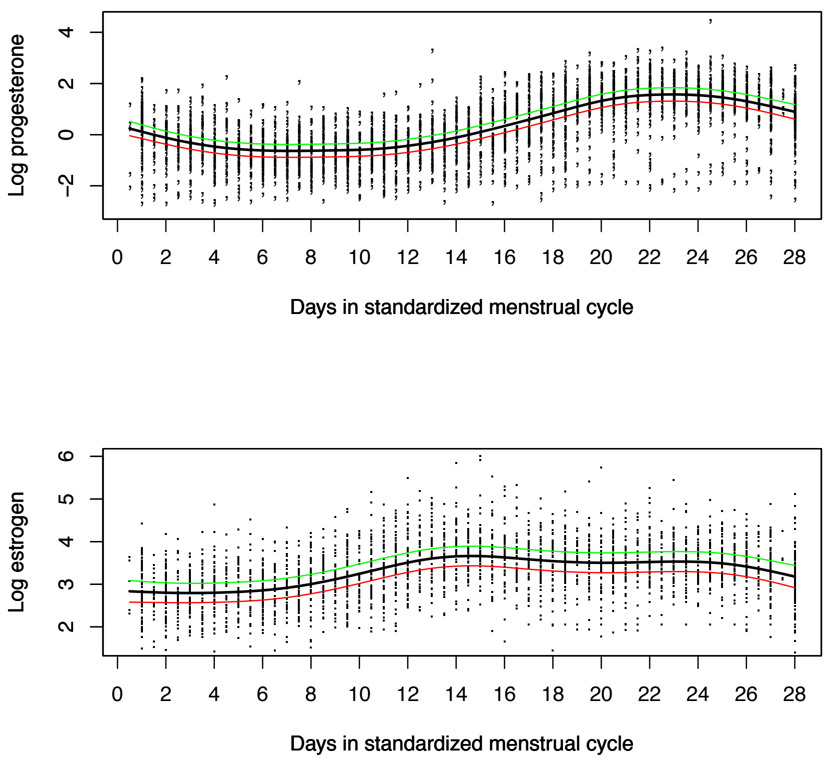

Figure 1: Plots of log progesterone and log estrogen levels against days in a standardized menstrual cycle, superimposed by estimated population mean curve $\hat{f}_1$ and $\hat{f}_2$.

There have been many extensions to allow for further flexibility in specifying the mean structure for the modelling of longitudinal data. For example, Brumback & Rice [1] used natural cubic splines to model the mean structure of the linear mixed model and specified the design matrices associated with fixed effects and random effects by bases of functions. Verbyla et al. [11] used data-based determination of the smoothing parameters and applied to the analysis of designed experiments. Some efforts are also geared towards enriching the specifications of the correlation structure for the modelling of longitudinal data. Taylor et al. [10] used an integrated Ornstein-Uhlenbeck process to model the covariance structure. [15] and Zhang et al. [16] proposed a semiparametric stochastic mixed model for regular and periodic longitudinal data, respectively, where covariate effects are modelled parametrically and the underlying complex (periodic) time courses are modelled nonparametrically; and used a stochastic process and random effects to model the within-subject and between-subject correlation respectively. Instead of smoothing splines, [13] modelled cyclic longitudinal data using mixed model L-splines. Wood [14] used a penalized cubic regression spline to model a cyclic smooth function. These models, however at this stage to our knowledge, are only applicable to univariate longitudinal analysis.

Some of the univariate techniques have been extended to the multivariate case. For example, Sy et al. [9] employed multivariate stochastic processes to jointly model bivariate longitudinal data. More recently, [8] modelled bivariate longitudinal responses, from outcomes of different data types, via a mixed effects hidden Markov modeling approach. And, most relevant in terms of the type of data focused upon in this paper, [7] extended the univariate state space model in time series analysis, and proposed a bivariate hierarchical state space model to bivariate circadian rhythmic longitudinal responses; each response is modelled by a hierarchical state space model, with both population-average and subject-specific components; and the bivariate model is constructed by linking the univariate models based on the hypothesized relationship between the responses.

In the current paper, we extend Zhang et al. [15] and propose a bivariate semiparametric stochastic mixed model for bivariate repeated measures data. The bivariate model uses parametric fixed effects and smooth nonparametric functions for each of the two underlying time effects. The between-subject correlations are modelled using separate but correlated random effects and the within-subject correlations by a bivariate Gaussian random field. The model allows us to investigate the relationship of the two responses through the correlation of the random effects and the bivariate Gaussian fields, which can not only describe the concurrent relationship of the two responses but also allows for characterizations of the relationship across time points. We derive maximum penalized likelihood estimators for both the fixed effects regression coefficients and the nonparametric time functions. The smoothing parameters and all variance components are estimated simultaneously using restricted maximum likelihood.

The paper is organized as followed. Section 2 specifies the proposed model with assumptions. Section 3 provides estimation and inference procedures. Specifically, Sections 3.1 gives estimation procedures for the model parameters, the nonparametric components, random effects and the Gaussian fields. Section 3.2 specifies the biases and covariances for all the estimators given in Section 3.1; and Section 3.3 concludes this section by providing the estimation procedures of the smoothing parameters and variance components. Section 4 investigates the proposed methodology through a simulation study. Section 5 illustrates the model by analyzing bivariate longitudinal female hormone data collected daily over a single menstrual cycle, and finally, Section 6 provides a summary of our proposed model, and discusses challenges and future work.

2. The Bivariate Semiparametric Stochastic Mixed Effects Model

2.1. The Proposed Model Specifications and Assumptions

Denote $\boldsymbol{Y}_{ij} = (Y_{1ij}, Y_{2ij})^T$ to be the bivariate metabolite levels of estrogen and progesterone for the ith subject at time point t_{ij}, $i = 1, \ldots, m$ and $j = 1, \ldots, n_i$. The bivariate model is

$$\boldsymbol{Y}_{ij} = \boldsymbol{X}_{ij}^T \boldsymbol{\beta} + \boldsymbol{f}(t_{ij}) + \boldsymbol{Z}_{ij}^T \boldsymbol{b}_i + \boldsymbol{U}_i(t_{ij}) + \boldsymbol{\epsilon}_{ij}, \tag{1}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)^T$ are $(p_1 + p_2) \times 1$ vectors of fixed effects regression coefficients associated with known covariates $\boldsymbol{X}_{ij} = \mathrm{diag}(\boldsymbol{X}_{1ij}^T, \boldsymbol{X}_{2ij}^T)$; $\boldsymbol{b} = (\boldsymbol{b}_{1i}, \boldsymbol{b}_{2i})^T$ are $2q \times 1$ vectors of random effects, respectively, associated with known covariates $\boldsymbol{Z}_{ij} = \mathrm{diag}(\boldsymbol{Z}_{1ij}^T, \boldsymbol{Z}_{2ij}^T)$; $\boldsymbol{f}(t) = (f_1(t), f_2(t))^T$ are twice-differentiable smooth functions of time; $\boldsymbol{U}_i(t) = (U_{1i}(t), U_{2i}(t))^T$ is a mean zero bivariate Gaussian field with covariance matrix

$$\boldsymbol{C}_i(s,t) \quad = \quad \begin{pmatrix} \sqrt{\xi_1(s)\xi_1(t)}\eta_1(\rho_1; s, t) & \sqrt{\xi_1(s)\xi_2(t)}\eta_3(\rho_3; s, t) \\ \sqrt{\xi_2(s)\xi_1(t)}\eta_3(\rho_3; t, s) & \sqrt{\xi_2(s)\xi_2(t)}\eta_2(\rho_2; s, t) \end{pmatrix} \tag{2}$$

where $\xi_1(t)$ and $\xi_2(t)$ are variance functions; $\mathrm{corr}(U_{1i}(t), U_{1i}(s)) = \eta_1(\rho_1; s, t)$, $\mathrm{corr}(U_{2i}(t), U_{2i}(s)) = \eta_2(\rho_2; s, t)$, and $\mathrm{corr}(U_{1i}(t), U_{2i}(s)) = \eta_3(\rho_3; s, t)$ are correlation functions, where ρ_1, ρ_2 and ρ_3 are correlation coefficients; and the measurement errors $\epsilon_{ij} = (\epsilon_{1ij}, \epsilon_{2ij})^T$ are bivariate normal with mean $\mathbf{0}$ and variance $\mathrm{diag}(\sigma_1^2, \sigma_2^2)$. We assume $\boldsymbol{b}_i$ to be $2q$-dimensional normal with mean zero and unstructured covariance matrix $\boldsymbol{G}(\boldsymbol{\phi})$, and that the random effects, the stochastic process and the measurement error to be mutually independent.

This model (1) is an extension to the model proposed in Zhang et al. [15], where a univariate semiparametric stochastic mixed model for longitudinal data was proposed. The challenge here is that we are modeling a bivariate longitudinal response model, which is achieved by modeling a joint distribution of the random effects and the bivariate Gaussian field. The distributions of the two random effects can be potentially distinct, with different distributions or the same distribution with different parameters; but the two random effects are assumed separate but correlated.

2.2. Matrix Notation

To make inferences from the model (1), we will write the model in matrix form - first, in subject level; then, over all subjects.

Denote $\boldsymbol{Y}_i = (\boldsymbol{Y}_{i1}, \ldots, \boldsymbol{Y}_{in_i})^T$ and similarly for $\boldsymbol{X}_i$, $\boldsymbol{Z}_i$, $\boldsymbol{U}_i$ and $\boldsymbol{\epsilon}_i$. Let $\boldsymbol{t'} = (t'_1, \ldots, t'_r)$ be a vector of ordered distinct values of $t_{ij}, i = 1, \ldots, m$ and $j = 1 \ldots n_i$ and define $\tilde{\boldsymbol{N}}_i$ to be the $n_i \times r$ incidence matrix for the i^{th} subject connecting $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{in_i})^T$ and $\boldsymbol{t'}$ such that

$$\tilde{\boldsymbol{N}}_i[j, \ell] = \begin{cases} 1 & \text{if } t'_{ij} = t'_\ell \\ 0 & \text{otherwise,} \end{cases}$$

where $\tilde{\boldsymbol{N}}_i[j,\ell]$ denotes the $(j,\ell)^{\text{th}}$ entry of matrix $\tilde{\boldsymbol{N}}_i$ for $j = 1,\dots,n_i$ and $\ell = 1,\dots,r_1$. Let $\boldsymbol{N}_{1i} = \boldsymbol{A}_{1i}\tilde{\boldsymbol{N}}_i$ and $\boldsymbol{N}_{2i} = \boldsymbol{A}_{2i}\tilde{\boldsymbol{N}}_i$, be the incidence matrices for the first and second response, respectively, where

$$\boldsymbol{A}_{1i} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & & \ddots & & \\ 0 & \dots & 0 & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix} \in \mathbb{R}^{2n_i \times n_i}, \quad \boldsymbol{A}_{2i} = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & & \ddots & & \\ 0 & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & 1 \end{pmatrix} \in \mathbb{R}^{2n_i \times n_i}.$$

Then, the proposed bivariate semiparametric stochastic mixed model (1) can be written as

$$\boldsymbol{Y}_i = \boldsymbol{X_i\beta} + \boldsymbol{N}_{1i}\boldsymbol{f}_1 + \boldsymbol{N}_{2i}\boldsymbol{f}_2 + \boldsymbol{Z_i b_i} + \boldsymbol{U}_i + \boldsymbol{\epsilon}_i$$

for subject i, where $\boldsymbol{f}_1 = (f_1(t'_1),\dots,f_1(t'_r))^T$ and $\boldsymbol{f}_2 = (f_2(t'_1),\dots,f_2(t'_r))^T$. Note that here we implicitly assume that each subject has distinct and potentially unequally spaced time points and that the bivariate responses are observed at the same time point for the same subject which is often realized in actual data applications, though this assumption can be easily modified if needed.

Further denoting $\boldsymbol{Y} = (\boldsymbol{Y}_1^T,\dots,\boldsymbol{Y}_m^T)^T$ and $\boldsymbol{X}$, $\boldsymbol{N}_1$, $\boldsymbol{N}_2, \boldsymbol{b}, \boldsymbol{U}, \boldsymbol{\epsilon}$ similarly and letting $n = \sum_{i=1}^m n_i$, then the bivariate semiparametric stochastic mixed effects model over all subjects is

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{N}_1\boldsymbol{f}_1 + \boldsymbol{N}_2\boldsymbol{f}_2 + \boldsymbol{Zb} + \boldsymbol{U} + \boldsymbol{\epsilon}, \tag{3}$$

with assumptions

$$\begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{U} \\ \boldsymbol{\epsilon} \end{pmatrix} \sim \boldsymbol{N}\left(\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{D}(\boldsymbol{\phi}) & \boldsymbol{0} & 0 \\ \boldsymbol{0} & \boldsymbol{\Gamma}(\boldsymbol{\xi},\rho) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Sigma}(\boldsymbol{\sigma}^2) \end{pmatrix} \right)$$

where $\boldsymbol{D}(\boldsymbol{\phi}) = \text{diag}(\boldsymbol{G},\dots,\boldsymbol{G})$; $\boldsymbol{\Gamma}(\boldsymbol{\xi},\rho) = \text{diag}(\boldsymbol{\Gamma}_1(\boldsymbol{t}_1,\boldsymbol{t}_1),\dots,\boldsymbol{\Gamma}_m(\boldsymbol{t}_m,\boldsymbol{t}_m))$ and the $(k,k')^{\text{th}}$ entry of $\boldsymbol{\Gamma}_i(\boldsymbol{t}_i,\boldsymbol{t}_i)$ is $\boldsymbol{C}_i(k,k')$; and $\boldsymbol{\Sigma}(\boldsymbol{\sigma}^2)$ is the diagonal matrix with alternating entries σ_1^2 and σ_2^2.

2.3. Covariance Structures

(a) Mean and Covariance of the Proposed Model

The marginal or population-averaged mean of $\boldsymbol{Y}$ is

$$E(\boldsymbol{Y}) = \boldsymbol{X\beta} + \boldsymbol{N}_1\boldsymbol{f}_1 + \boldsymbol{N}_2\boldsymbol{f}_2,$$

and the marginal covariance of $\boldsymbol{Y}$, averaged over the distribution of subject-specific effects $\boldsymbol{b}$ is

$$\text{cov}(\boldsymbol{Y}) = \boldsymbol{ZDZ}^T + \boldsymbol{\Gamma} + \boldsymbol{\Sigma}.$$

The mean response for a specific subject is

$$E(\boldsymbol{Y}|\boldsymbol{b}) = \boldsymbol{X\beta} + \boldsymbol{N}_1\boldsymbol{f}_1 + \boldsymbol{N}_2\boldsymbol{f}_2 + \boldsymbol{Zb},$$

and the covariance among the longitudinal observations for a specific subject is

$$\mathrm{cov}(\boldsymbol{Y}|\boldsymbol{b}) = \mathrm{cov}(\boldsymbol{U}) + \mathrm{cov}(\boldsymbol{\epsilon}) = \boldsymbol{\Gamma} + \boldsymbol{\Sigma},$$

which describes the covariance of the subject deviations from the subject-specific mean response $E(\boldsymbol{Y}|\boldsymbol{b})$. The within-subject correlation $(\boldsymbol{\Gamma} + \boldsymbol{\Sigma})$ structure is enhanced by the addition of bivariate Gaussian field into the model.

(b) Modelling the Relationship of the Bivariate Responses

Moreover, the proposed model also allows for explicit analysis of the relationship of the bivariate responses, which is explained by the covariance of the random effects and and the covariance of the bivariate Gaussian field. Specifically, the covariance between the bivariate responses at time points t_{ij} and t_{ik} for individual i is given by

$$\begin{aligned}
& \mathrm{cov}(Y_{1ij}, Y_{2ik}) \\
= \;& \mathrm{cov}(\boldsymbol{X_{1ij}^T\beta_1} + f_1(t_{ij}) + \boldsymbol{Z_{1ij}^T b_{1i}} + U_{1i}(t_{ij}) + \epsilon_{1ij}, \boldsymbol{X_{2ik}^T \beta_2} + f_2(t_{ik}) + \boldsymbol{Z_{2ik}^T b_{2i}} + U_{2i}(t_{ik}) + \epsilon_{2ik}) \\
= \;& \mathrm{cov}(\boldsymbol{Z_{1ij}^T b_{1i}} + U_{1i}(t_{ij}) + \epsilon_{1ij}, \boldsymbol{Z_{2ik}^T b_{2i}} + U_{2i}(t_{ik}) + \epsilon_{2ik}) \\
= \;& \boldsymbol{Z_{1ij}^T}\mathrm{cov}(\boldsymbol{b_{1i}}, \boldsymbol{b_{2i}})\boldsymbol{Z_{2ik}} + \mathrm{cov}(U_{1i}(t_{ij}), U_{2i}(t_{ik})) \\
= \;& \boldsymbol{Z_{1ij}^T G' Z_{2ik}} + \sqrt{\xi_1(t_{ij})\xi_2(t_{ik})}\eta_3(\rho_3; t_{ij}, t_{ik}),
\end{aligned}$$

where $\boldsymbol{G}'$ is the q by q upper off-diagonal block matrix of covariance matrix $\boldsymbol{G}$ for the random effects $\boldsymbol{b}_i$. The result shows that the inclusion of bivariate Gaussian field allows for modelling the covariance of the bivariate responses at different time points.

2.4. The Gaussian Field Specification

To accommodate for more complicated within-subject correlation and potential correlation between the bivariate responses, we propose to include various stationary and nonstationary bivariate Gaussian fields to model serial correlation. This allows for the within-subject covariance and the correlation between bivariate responses to be a function of time.

There are potentially many choices available: Wiener process or Brownian motion [10]; an integrated Wiener process and so on. One particular Gaussian process/field worthy of mentioning is the Ornstein-Uhlenbeck (OU) process [6] which has a correlation function that decays exponentially over time $\mathrm{corr}(U_i(t), U_i(s))$ $= \exp\{-\alpha|s-t|\}$. The variance function for OU process $\xi(t) = \sigma^2/2a$ is a constant, thus the process is strictly stationary. When $\xi(t)$ varies over time, then the process becomes nonhomogenesous (NOU) and, for example, we can assume $\xi(t) = \exp(a_0 + a_1 t + a_1 t^2)$.

3. Estimation and Inference

3.1. Estimation of Model Coefficients, Nonparametric Function, Random Effects and Gaussian Fields

The proposed model (3) implies the *marginal model*

$$Y = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{N}_1\boldsymbol{f}_1 + \boldsymbol{N}_2\boldsymbol{f}_2 + \boldsymbol{\epsilon}^*, \boldsymbol{\epsilon}^* \sim N_{2n}(\boldsymbol{0}, \boldsymbol{V})$$

where $\boldsymbol{V} = \boldsymbol{Z}\boldsymbol{D}\boldsymbol{Z}^T + \boldsymbol{\Gamma} + \boldsymbol{\Sigma}$. Thus, the *log-likelihood* function for $(\boldsymbol{\beta}, \boldsymbol{f}_1, \boldsymbol{f}_2)$ is :

$$\ell(\boldsymbol{\beta}, \boldsymbol{f}_1, \boldsymbol{f}_2; \boldsymbol{Y}) \propto -\frac{1}{2}\log|\boldsymbol{V}| - \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{\beta} - \boldsymbol{N}_1\boldsymbol{f}_1 - \boldsymbol{N}_2\boldsymbol{f}_2)^T\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{\beta} - \boldsymbol{N}_1\boldsymbol{f}_1 - \boldsymbol{N}_2\boldsymbol{f}_2)$$

for given fixed variance parameters. We estimate the parameters $\boldsymbol{\beta}$, $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ by maximizing the penalized likelihood [12]:

$$\ell(\boldsymbol{\beta}, \boldsymbol{f}_1, \boldsymbol{f}_2; \boldsymbol{Y}) - \lambda_1\int_a^b [f_1''(t)]^2 dt - \lambda_2\int_a^b [f_2''(t)]^2 dt = \ell(\boldsymbol{\beta}, \boldsymbol{f}_1, \boldsymbol{f}_2; \boldsymbol{Y}) - \lambda_1\boldsymbol{f}_1^T\boldsymbol{K}\boldsymbol{f}_1 - \lambda_2\boldsymbol{f}_2^T\boldsymbol{K}\boldsymbol{f}_2 \quad (4)$$

where λ_1 and λ_2 are smoothing parameters; a and b is the range of time t; and $\boldsymbol{K}$ is the nonnegative definite smoothing matrix, defined in Equation (2.3) in Green & Silverman [5]. Since observation time points t_{ij} are assumed to be the same for both responses, the smoothing matrix $\boldsymbol{K}$, which is determined by time increments, is also the same. The resulting estimators for the nonparametric functions are the natural cubic spline estimators of $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$.

Differentiation of (4) with respect to $\boldsymbol{\beta}$, $\boldsymbol{f}_1$, $\boldsymbol{f}_2$ gives the estimators $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{f}}_1, \hat{\boldsymbol{f}}_2)$ that solves

$$\begin{pmatrix} \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}_1 & \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}_2 \\ \boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{X} & \boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K} & \boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_2 \\ \boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{X} & \boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_1 & \boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{f}_1 \\ \boldsymbol{f}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y} \\ \boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{Y} \\ \boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{Y} \end{pmatrix}, \quad (5)$$

where $\boldsymbol{W} = \boldsymbol{V}^{-1}$. To study the theoretical properties of the estimates, such as bias and covariance, we derive the closed-form solutions for $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{f}}_1$ and $\hat{\boldsymbol{f}}_2$

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{Y} \quad (6)$$

$$\hat{\boldsymbol{f}}_1 = (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{Y} \quad (7)$$

$$\hat{\boldsymbol{f}}_2 = (\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{Y}, \quad (8)$$

where $\boldsymbol{W}_x = \boldsymbol{W}_1 - \boldsymbol{W}_1\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_1$, $\boldsymbol{W}_{f_1} = \boldsymbol{W}_2 - \boldsymbol{W}_2\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_2$, and $\boldsymbol{W}_{f_2} = \boldsymbol{W}_1 - \boldsymbol{W}_1\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1$ are weight matrices with $\boldsymbol{W}_1 = \boldsymbol{W} - \boldsymbol{W}\boldsymbol{N}_1(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}$ and $\boldsymbol{W}_2 = \boldsymbol{W} - \boldsymbol{W}\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}$.

Estimation of the subject-specific random effects $\boldsymbol{b}_i$ and the subject-specific Gaussian field $\boldsymbol{U}_i(\boldsymbol{s}_i)$ is obtained by calculating their conditional expectations given the data $\boldsymbol{Y}$. Therefore,

$$\hat{\boldsymbol{b}}_i = E(\boldsymbol{b}|\boldsymbol{Y}) = \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{f}}_{1i} - \hat{\boldsymbol{f}}_{2i}) \quad (9)$$

and similarly,

$$\hat{\boldsymbol{U}}_i(\boldsymbol{s}_i) = \boldsymbol{\Gamma}(\boldsymbol{s}_i, \boldsymbol{t}_i)\boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{f}}_{1i} - \hat{\boldsymbol{f}}_{2i}) \quad (10)$$

where $\hat{\boldsymbol{f}}_{1i} = \boldsymbol{N}_{1i}\hat{\boldsymbol{f}}_1$ and $\hat{\boldsymbol{f}}_{2i} = \boldsymbol{N}_{2i}\hat{\boldsymbol{f}}_2$. Technical details for this subsection is included in Appendix A.

3.2. Biases and Covariances of Model Coefficients, Nonparametric Function, Random Effects and Gaussian Fields

From closed-form solutions of estimators from equation (6) (7) and (8) in Section 3.1, the biases of the estimators $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{f}}_1$ and $\hat{\boldsymbol{f}}_2$ can be easily calculated (Appendix B), and we have

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta} &= (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x(\boldsymbol{N}_1\boldsymbol{f}_1 + \boldsymbol{N}_2\boldsymbol{f}_2) \qquad (11)\\ E(\hat{\boldsymbol{f}}_1) - \boldsymbol{f}_1 &= (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_2\boldsymbol{f}_2 - \lambda_1\boldsymbol{K}\boldsymbol{f}_1) \qquad (12)\\ E(\hat{\boldsymbol{f}}_2) - \boldsymbol{f}_2 &= (\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_1\boldsymbol{f}_1 - \lambda_2\boldsymbol{K}\boldsymbol{f}_2). \qquad (13)\end{aligned}$$

Similarly, the expected values of the estimators in (9) and (10) for the subject-specific random effects $\boldsymbol{b}_i$ and for the subject-specific Gaussian field $\boldsymbol{U}_i(\boldsymbol{s}_i)$ are

$$\begin{aligned} E(\hat{\boldsymbol{b}}_i) &= \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i[\lambda_1\boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{K} - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{N}_1\\ &\qquad -\boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_1]\boldsymbol{f}_1\\ &\quad +\boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i[\lambda_2\boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{K} - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{N}_2\\ &\qquad -\boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_2]\boldsymbol{f}_2\end{aligned}$$

and

$$\begin{aligned} E\left[\hat{\boldsymbol{U}}_i(\boldsymbol{s}_i)\right] &= \boldsymbol{\Gamma}_i(\boldsymbol{s}_i,\boldsymbol{t}_i)[\lambda_1\boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{K} - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{N}_1\\ &\qquad -\boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_1]\boldsymbol{f}_1\\ &\quad +\boldsymbol{\Gamma}_i(\boldsymbol{s}_i,\boldsymbol{t}_i)[\lambda_2\boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{K} - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{N}_2\\ &\qquad -\boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_2]\boldsymbol{f}_2.\end{aligned}$$

It can be shown that the biases of $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{f}}_1$, $\hat{\boldsymbol{f}}_2$, $\hat{\boldsymbol{b}}_i$ and $\hat{\boldsymbol{U}}_i$ all go to $\boldsymbol{0}$ as both smoothing parameters $\lambda_1 \to 0$ and $\lambda_2 \to 0$, see Lemma 1 in Appendix B.

For covariances, simple calculation using (6) (7) and (8) gives the covariance of $\hat{\boldsymbol{\beta}}$

$$\mathrm{cov}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{V}\boldsymbol{W}_x\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}$$

and the respective covariances of $\hat{\boldsymbol{f}}_1$ and $\hat{\boldsymbol{f}}_2$

$$\mathrm{cov}(\hat{\boldsymbol{f}}_1) = (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{V}\boldsymbol{W}_{f_1}\boldsymbol{N}_1(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}$$

$$\mathrm{cov}(\hat{\boldsymbol{f}}_2) = (\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{V}\boldsymbol{W}_{f_2}\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}.$$

The covariances of the estimators in (9) and (10) for the subject-specific random effects $\boldsymbol{b}_i$ and for the subject-specific Gaussian field $\boldsymbol{U}_i(\boldsymbol{s}_i)$ are

$$\mathrm{cov}(\hat{\boldsymbol{b}}_i - \boldsymbol{b}_i) = \boldsymbol{D} - \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D} + \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i\boldsymbol{\chi}_i\boldsymbol{C}^{-1}\boldsymbol{\chi}^T\boldsymbol{W}\boldsymbol{\chi}\boldsymbol{C}^{-1}\boldsymbol{\chi}_i^T\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D} \qquad (14)$$

and

$$\mathrm{cov}(\hat{\boldsymbol{U}}_i(\boldsymbol{s}_i) - \boldsymbol{U}_i(\boldsymbol{s}_i)) = \boldsymbol{\Gamma}(\boldsymbol{s}_i, \boldsymbol{s}_i) - \boldsymbol{\Gamma}(\boldsymbol{s}_i, \boldsymbol{t}_i)\boldsymbol{W}_i\boldsymbol{\Gamma}(\boldsymbol{s}_i, \boldsymbol{t}_i)^T + \boldsymbol{\Gamma}(\boldsymbol{s}_i, \boldsymbol{t}_i)\boldsymbol{W}_i\boldsymbol{\chi}_i\boldsymbol{C}^{-1}\boldsymbol{\chi}^T\boldsymbol{W}\boldsymbol{\chi}\boldsymbol{C}^{-1}\boldsymbol{\chi}_i^T\boldsymbol{W}_i\boldsymbol{\Gamma}(\boldsymbol{s}_i, \boldsymbol{t}_i)^T,$$

where $\boldsymbol{\chi}_i = \begin{pmatrix} \boldsymbol{X}_i & \boldsymbol{N}_{1i} & \boldsymbol{N}_{2i} \end{pmatrix}$ and $\boldsymbol{\chi} = \begin{pmatrix} \boldsymbol{X} & \boldsymbol{N}_1 & \boldsymbol{N}_2 \end{pmatrix}$.

3.3. Estimation of the Smoothing Parameters and Variance Parameters

To estimate the smoothing parameters and variance components jointly using the restricted maximum likelihood (REML), we rewrite the proposed semiparametric model as a modified linear mixed model. Specifically, by [4], the nonparametric functions $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ under a one-to-one linear transformation are

$$\begin{aligned} \boldsymbol{f}_1 &= \boldsymbol{T}\boldsymbol{\delta}_1 + \boldsymbol{B}\boldsymbol{a}_1 \\ \boldsymbol{f}_2 &= \boldsymbol{T}\boldsymbol{\delta}_2 + \boldsymbol{B}\boldsymbol{a}_2 \end{aligned}$$

where δ_1 and δ_2 are vectors of dimensions 2; $a_1$and a_2 are of dimensions $r-2$; $\boldsymbol{B} = \boldsymbol{L}(\boldsymbol{L}^T\boldsymbol{L})^{-1}$ and $\boldsymbol{L}$ is $r \times (r-2)$ full-rank matrix satisfying $\boldsymbol{K} = \boldsymbol{L}\boldsymbol{L}^T$ and $\boldsymbol{L}^T\boldsymbol{T} = 0$. Thus the proposed semiparametric mixed model (3) can be rewritten as a modified linear mixed model [15],

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{N}_1\boldsymbol{T}\boldsymbol{\delta}_1 + \boldsymbol{N}_1\boldsymbol{B}\boldsymbol{a}_1 + \boldsymbol{N}_2\boldsymbol{T}\boldsymbol{\delta}_2 + \boldsymbol{N}_2\boldsymbol{B}\boldsymbol{a}_2 + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{U} + \boldsymbol{\epsilon}, \tag{15}$$

where $\boldsymbol{\beta}_* = (\boldsymbol{\beta}^T, \boldsymbol{\delta}_1^T, \boldsymbol{\delta}_2^T)^T$ are the regression coefficients and $\boldsymbol{b}_* = (\boldsymbol{a}_1^T, \boldsymbol{a}_2^T, \boldsymbol{b}^T, \boldsymbol{U}^T)^T$ are mutually independent random effects with $\boldsymbol{a}_1$ distributed as normal $(0, \tau_1\boldsymbol{I})$, $\boldsymbol{a}_2$ distributed as normal$(0, \tau_2\boldsymbol{I})$ where $\tau_1 = 1/\lambda_1$ and $\tau_2 = 1/\lambda_2$, and $(\boldsymbol{b}, \boldsymbol{U})$ having the same distribution as specified before. The marginal variance of $\boldsymbol{Y}$ under the modified mixed model representation becomes $\boldsymbol{V}_* = \tau_1\boldsymbol{B}_{1*}\boldsymbol{B}_{1*}^T + \tau_2\boldsymbol{B}_{2*}\boldsymbol{B}_{2*}^T + \boldsymbol{V}$, where $\boldsymbol{B}_{1*} = \boldsymbol{N}_1\boldsymbol{B}$ and $\boldsymbol{B}_{2*} = \boldsymbol{N}_2\boldsymbol{B}$.

Under the modified linear mixed model (15), the REML log-likelihood of $(\tau_1, \tau_2, \boldsymbol{\theta})$ is

$$\ell_R(\tau_1, \tau_2, \boldsymbol{\theta}; \boldsymbol{Y}) = -\frac{1}{2}\left[\log|\boldsymbol{V}_*| + \log|\boldsymbol{X}_*^T\boldsymbol{V}_*^{-1}\boldsymbol{X}_*| + (\boldsymbol{Y} - \boldsymbol{X}_*\hat{\boldsymbol{\beta}}_*)^T\boldsymbol{V}_*^{-1}(\boldsymbol{Y} - \boldsymbol{X}_*\hat{\boldsymbol{\beta}}_*)\right],$$

where $\boldsymbol{X}_* = [\boldsymbol{X}, \boldsymbol{N}_1\boldsymbol{T}, \boldsymbol{N}_2\boldsymbol{T}]$. Taking the derivative of ℓ_R with respect to τ_1, τ_1, and $\boldsymbol{\theta}$ and using the identity $\boldsymbol{V}_*^{-1}(\boldsymbol{Y} - \boldsymbol{X}_*\hat{\boldsymbol{\beta}}_*) = \boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1\hat{\boldsymbol{f}}_1 - \boldsymbol{N}_2\hat{\boldsymbol{f}}_2)$, the estimating equations for smoothing parameters τ_1, τ_2 and variance components $\boldsymbol{\theta}$ can be obtained:

$$\frac{\partial\ell_R}{\partial\tau_1} = -\frac{1}{2}\mathrm{tr}(\boldsymbol{P}_*\boldsymbol{B}_{1*}\boldsymbol{B}_{1*}^T) + \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1\hat{\boldsymbol{f}}_1 - \boldsymbol{N}_2\hat{\boldsymbol{f}}_2)^T\boldsymbol{V}^{-1}\boldsymbol{B}_{1*}\boldsymbol{B}_{1*}^T\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1\hat{\boldsymbol{f}}_1 - \boldsymbol{N}_2\hat{\boldsymbol{f}}_2), \tag{16}$$

$$\frac{\partial\ell_R}{\partial\tau_2} = -\frac{1}{2}\mathrm{tr}(\boldsymbol{P}_*\boldsymbol{B}_{2*}\boldsymbol{B}_{2*}^T) + \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1\hat{\boldsymbol{f}}_1 - \boldsymbol{N}_2\hat{\boldsymbol{f}}_2)^T\boldsymbol{V}^{-1}\boldsymbol{B}_{2*}\boldsymbol{B}_{2*}^T\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1\hat{\boldsymbol{f}}_1 - \boldsymbol{N}_2\hat{\boldsymbol{f}}_2), \tag{17}$$

and

$$\frac{\partial\ell_R}{\partial\theta_j} = -\frac{1}{2}\mathrm{tr}(\boldsymbol{P}_*\frac{\partial\boldsymbol{V}}{\partial\theta_j}) + \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1\hat{\boldsymbol{f}}_1 - \boldsymbol{N}_2\hat{\boldsymbol{f}}_2)^T\boldsymbol{V}^{-1}\frac{\partial\boldsymbol{V}}{\partial\theta_j}\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1\hat{\boldsymbol{f}}_1 - \boldsymbol{N}_2\hat{\boldsymbol{f}}_2), \tag{18}$$

where $\boldsymbol{P}_* = \boldsymbol{V}_*^{-1} - \boldsymbol{V}_*^{-1}\boldsymbol{X}_*(\boldsymbol{X}_*^T\boldsymbol{V}_*^{-1}\boldsymbol{X}_*)^{-1}\boldsymbol{X}_*^T\boldsymbol{V}_*^{-1}$ is the projection matrix.

The covariance of the smoothing parameters τ_1, τ_2 and variance components $\boldsymbol{\theta}$ can be estimated using a Fisher-scoring algorithm, where the Fisher information matrix is obtained using (16), (17) and (18),

$$\boldsymbol{I} = \begin{pmatrix} \boldsymbol{I}_{\tau_1\tau_1} & \boldsymbol{I}_{\tau_1\tau_2} & \boldsymbol{I}_{\tau_1\boldsymbol{\theta}} \\ \boldsymbol{I}_{\tau_2\tau_1} & \boldsymbol{I}_{\tau_2\tau_2} & \boldsymbol{I}_{\tau_2\boldsymbol{\theta}} \\ \boldsymbol{I}_{\boldsymbol{\theta}\tau_1} & \boldsymbol{I}_{\boldsymbol{\theta}\tau_2} & \boldsymbol{I}_{\boldsymbol{\theta}\boldsymbol{\theta}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_{\tau_1\tau_1} & \boldsymbol{I}_{\tau_1\tau_2} & \boldsymbol{I}_{\tau_1\boldsymbol{\theta}} \\ \boldsymbol{I}_{\tau_1\tau_2}^T & \boldsymbol{I}_{\tau_2\tau_2} & \boldsymbol{I}_{\tau_2\boldsymbol{\theta}} \\ \boldsymbol{I}_{\tau_1\boldsymbol{\theta}}^T & \boldsymbol{I}_{\tau_2\boldsymbol{\theta}}^T & \boldsymbol{I}_{\boldsymbol{\theta}\boldsymbol{\theta}} \end{pmatrix},$$

where

$$\boldsymbol{I}_{\tau_1\tau_1} = \frac{1}{2}\mathrm{tr}(\boldsymbol{P}_*\boldsymbol{B}_{1*}\boldsymbol{B}_{1*}^T\boldsymbol{P}_*\boldsymbol{B}_{1*}\boldsymbol{B}_{1*}^T), \qquad \boldsymbol{I}_{\tau_2\tau_2} = \frac{1}{2}\mathrm{tr}(\boldsymbol{P}_*\boldsymbol{B}_{2*}\boldsymbol{B}_{2*}^T\boldsymbol{P}_*\boldsymbol{B}_{2*}\boldsymbol{B}_{2*}^T),$$

$$\boldsymbol{I}_{\tau_1\theta_j} = \frac{1}{2}\mathrm{tr}\left(\boldsymbol{P}_*\boldsymbol{B}_{1*}\boldsymbol{B}_{1*}^T\boldsymbol{P}_*\frac{\partial \boldsymbol{V}}{\partial\theta_j}\right), \qquad \boldsymbol{I}_{\tau_2\theta_j} = \frac{1}{2}\mathrm{tr}\left(\boldsymbol{P}_*\boldsymbol{B}_{2*}\boldsymbol{B}_{2*}^T\boldsymbol{P}_*\frac{\partial \boldsymbol{V}}{\partial\theta_j}\right),$$

and

$$\boldsymbol{I}_{\tau_1\tau_2} = \frac{1}{2}\mathrm{tr}(\boldsymbol{P}_*\boldsymbol{B}_{1*}\boldsymbol{B}_{1*}^T\boldsymbol{P}_*\boldsymbol{B}_{2*}\boldsymbol{B}_{2*}^T), \qquad \boldsymbol{I}_{\theta_j\theta_k} = \frac{1}{2}\mathrm{tr}\left(\boldsymbol{P}_*\frac{\partial \boldsymbol{V}}{\partial\theta_j}\boldsymbol{P}_*\frac{\partial \boldsymbol{V}}{\partial\theta_k}\right).$$

4. Simulations

4.1. A Simulation Study using NOU

We conduct a simulation study to evaluate the performance of the estimates of the model regression parameters and nonparametric function using the REML estimates of the smoothing parameters and the variance parameters. Bivariate longitudinal data are generated according to the following model:

$$\begin{aligned} Y_{1ij} &= \mathrm{age}_i^T\beta_1 + f_1(t_{ij}) + b_{1i} + U_{1i}(t_{ij}) + \epsilon_{1ij} \\ Y_{2ij} &= \mathrm{age}_i^T\beta_2 + f_2(t_{ij}) + b_{2i} + U_{2i}(t_{ij}) + \epsilon_{2ij} \\ & \qquad i = 1,\ldots,30;\ j = 1,\ldots,28;\ t_{ij} \in \{1,\ldots,28\} \end{aligned}$$

where b_{1i} and b_{2i} are independent but correlated random intercepts following a bivariate normal distribution with mean $\mathbf{0}$ and unstructured covariance matrix $\boldsymbol{D}(\phi_1, \phi_{1,2}, \phi_2)$; U_{1i} and U_{2i} are simulated from mean $\mathbf{0}$ bivariate NOU fields modeling serial correlation, with variance function $\mathrm{var}(U_{1i}(t)) = \exp\{a_{10}+a_{11}t+a_{12}t^2\}$, var $(U_{2i}(t)) = \exp\{a_{20} + a_{21}t + a_{22}t^2\}$ and $\mathrm{corr}(U_{1i}(t), U_{1i}(s)) = \rho_1^{|s-t|}$ $\mathrm{corr}(U_{2i}(t), U_{2i}(s)) = \rho_2^{|s-t|}$, i.e. the covariance function for the bivariate NOU field is

$$\boldsymbol{C}_i(s,t) = \begin{pmatrix} \rho_1^{|s-t|}\exp\{a_{10} + \frac{1}{2}[a_{11}(s+t) + a_{12}(s^2+t^2)]\} & 0 \\ 0 & \rho_2^{|s-t|}\exp\{a_{20} + \frac{1}{2}[a_{21}(s+t) + a_{22}(s^2+t^2)]\} \end{pmatrix};$$

ϵ_{1ij} and ϵ_{2ij} are simulated from a mean $\mathbf{0}$ bivariate normal distribution with covariance $\mathrm{diag}(\sigma_1^2, \sigma_2^2)$; and the nonparametric smooth functions are generated from $f_1(t) = 5\sin(2\pi/28)\,t$ and $f_2(t) = 3\cos(2\pi/28)\,t$.

Table 1 records the simulation results for estimates of model parameters based on 500 simulation replicates and 30 subjects. The Bias is defined as the bias of the parameter estimated divided by its true value, i.e., relative bias. The parameter estimates of the regression coefficients β_1 and β_2, and the variance estimates of

Table 1: Estimates of regression coefficients, variance parameter and smoothing parameter for the progesterone and estrogen data.

Model parameters	True Value	Parameter estimate	Bias	SE	Model SE
β_1	1.00	0.9987	0.0013	0.0271	0.0267
β_2	0.75	0.7496	0.0005	0.0282	0.0267
τ_1	1.00	0.7478	0.2522	0.1460	
τ_2	1.00	0.7388	0.2612	0.1535	
ϕ_1	1.00	0.9946	0.0054	0.0895	
$\phi_{1,2}$	-0.50	-0.5019	-0.0038	0.0730	
ϕ_2	1.00	0.9971	0.0029	0.0868	
σ_1^2	1.00	0.9989	0.0011	0.0173	
σ_2^2	1.00	0.9994	0.0006	0.0185	
ρ_1	0.20	0.1620	0.1900	0.1034	
a_{10}	-0.44	-0.4936	-0.1218	0.7143	
a_{11}	0.30	0.3530	0.1767	0.7607	
a_{12}	-0.20	-0.2151	-0.0755	0.1823	
ρ_2	0.15	0.1483	0.0113	0.1531	
a_{20}	-1.60	-1.8383	-0.1489	0.9225	
a_{21}	0.30	0.4771	0.5903	0.6754	
a_{22}	-0.10	-0.1298	-0.2980	0.1187	

the random intercepts and measurement errors are nearly unbiased, whereas the estimates of the smoothing parameters and the NOU variance parameters are slightly biased.

The biases for the nonparametric functions $\hat{f}_1$ and $\hat{f}_2$ are both minimal and center around 0, see Figure 2. Figure 3 shows that model standard errors of estimates of $\hat{f}_1$ and $\hat{f}_2$ agree quite well with the empirical standard errors.

Figure 4 shows the estimated pointwise 95% coverage probabilities of the true nonparametric functions f_1 and f_2. The means for the estimated coverage probabilities are 95% and 93% for $\hat{f}_1$ and $\hat{f}_2$. Overall, our simulation study results are good.

4.2. Misspecification of Gaussian Fields

We further conduct simulation studies when the Gaussian fields are incorrectly specified and study the effect of this misspecification on fixed effects, variance, and smoothing parameter estimations. Specifically, we use OU and Wiener bivariate Gaussian fields, respectively, to analyze datasets generated by NOU bivariate Gaussian field with the same specification as above.

Based on 400 simulations results for each choice of Gaussian field, the estimates of regression coefficients and random intercepts are fairly robust with bias close to zero even when the bivariate Gaussian field is

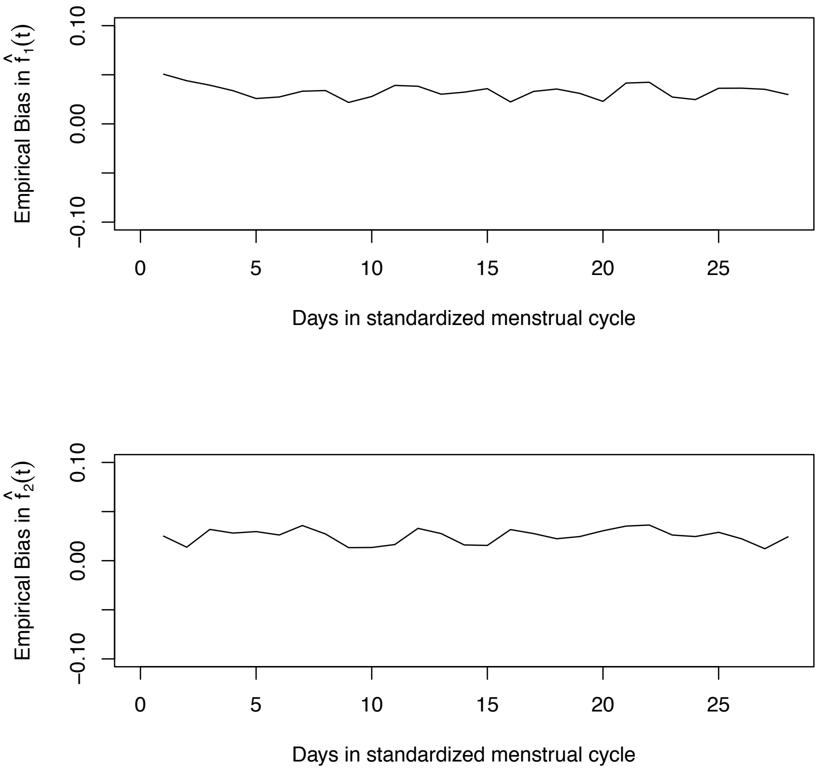

Figure 2: Empirical Bias in estimated nonparametric functions $\hat{f}_1$ and $\hat{f}_2$ based on 500 simulation replications.

misspecified as bivariate OU or Wiener field. The estimates for the smoothing parameters is much more biased for both bivariate OU or Wiener field, though misspecification in OU field would lead to less bias than that in Wiener. The estimates for variance of the measurement error are almost unbiased with the misspecification of bivariate Wiener field; whereas it is 20% more biased in the case of misspecification of bivariate OU field. In conclusion, misspecification of Gaussian field does not have a major influence if more emphasis is placed on the estimates of regression coefficients, yet the estimates of smoothing parameters and some variance components can vary significantly from the true values in the presence of misspecification of Gaussian field.

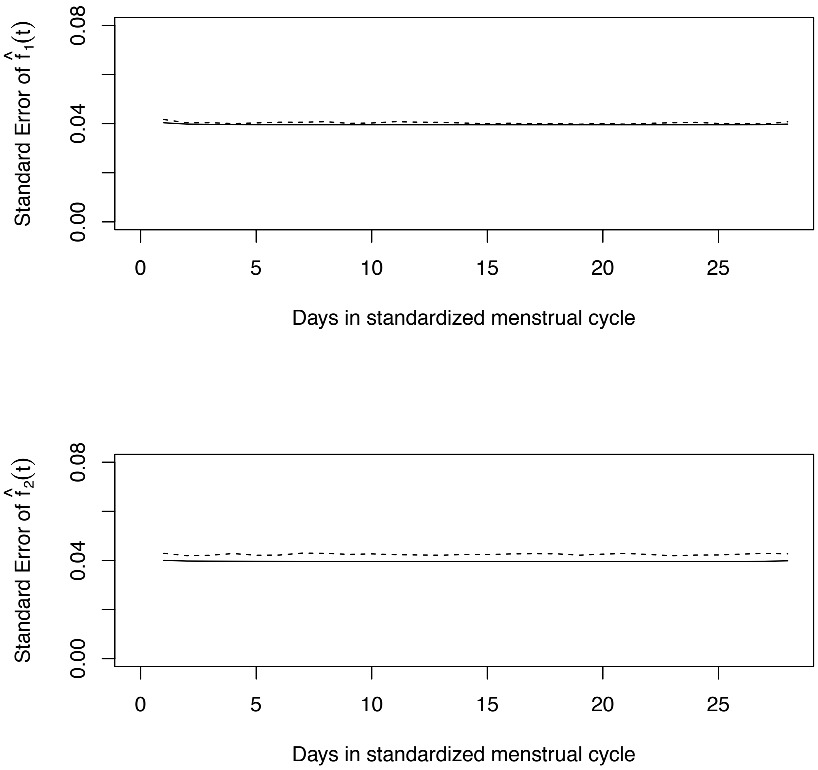

Figure 3: Pointwise empirical (dashes) and frequentist (solid) standard errors of the estimated nonparametric functions $\hat{f}_1$ and $\hat{f}_2$ based on 500 simulation replications.

5. Bivariate Longitudinal Hormone Data Analysis

The model we proposed was motivated by a bivariate longitudinal hormone dataset on progesterone and estrogen. Daily urine samples were collected from 403 employed women aged 20 to 44 years who completed a median of five consecutive menstrual cycles of collection each [3]. Of these, 338 women collected daily urine samples for at least one complete menstrual cycle, had fewer than three days of missing data in any five-day rolling window, did not have a conception in the analyzed cycles, and had complete covariate information. One menstrual cycle was randomly selected from each of the 338 women. Risk factor data were obtained by in-person interview at baseline. The details of the study design and assay methods are described in detail previously [3] and [2].

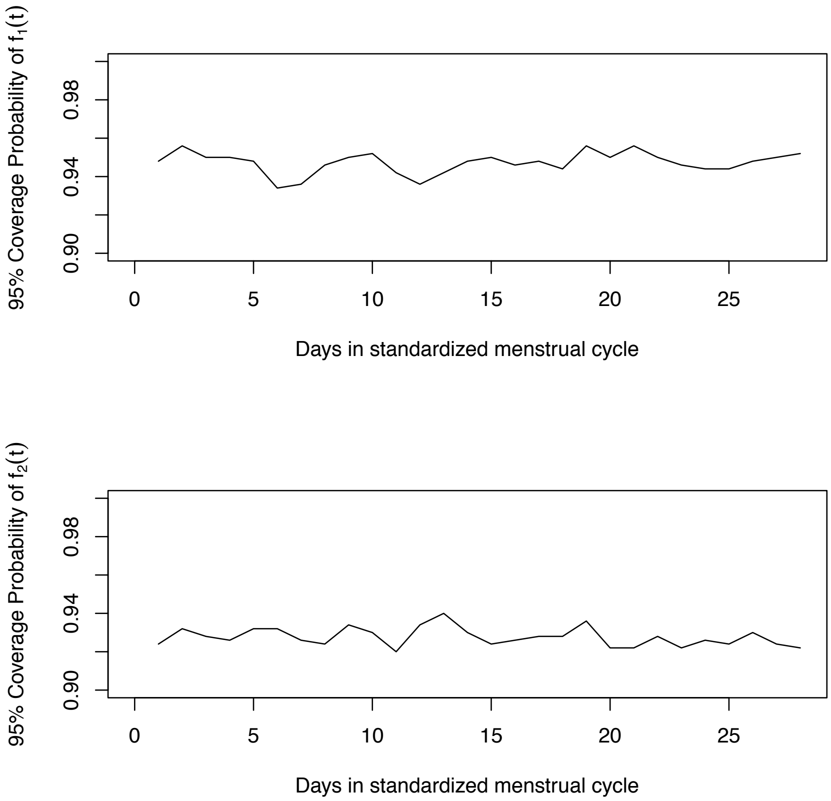

Figure 4: A graph showing the estimated 95% coverage probabilities of the true nonparametric functions f_1 and f_2 based on 500 simulation replications.

For demonstration purposes, we randomly select 100 study participants from the study, with a total of 5498 observations for both responses. Each woman contributes from 16 to 43 observations over a menstrual cycle, resulting an average of 28 observations per woman. In order for the results to be biologically meaningful, the menstrual cycle length for each women has been standardized to a reference of 28 days, based on the assumption that the change of hormone level for each woman depends on the time of the menstrual cycle relative to the cycle length. The standardization generates 56 distinct time points with increments between time points of 1/2 day each. To make the normality assumption more appropriate, a log transformation was used for each of the two hormones.

Figure 1 plots the log-transformed progesterone and estrogen levels during a standardized menstrual cycle. Figure 5 plots their empirical sample variances calculated at each distinct time point.

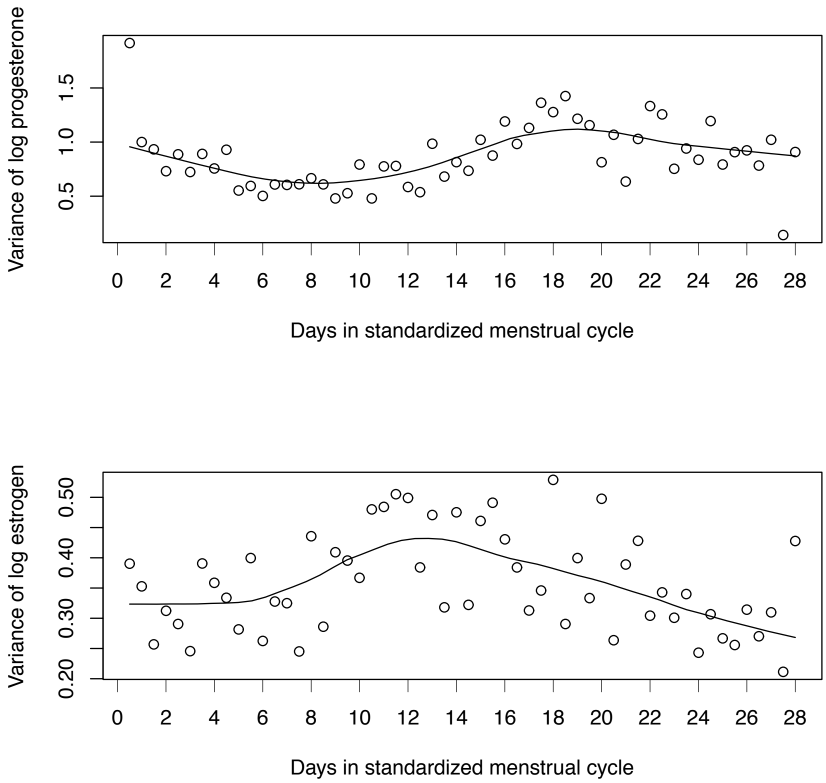

Figure 5: Plots of empirical sample variance of log progesterone and log estrogen levels at each distinct time points in a standardized menstrual cycle.

Denoting $\{(Y_{1ij}, Y_{2ij})\}$ the j^{th} log-transformed progesterone and estrogen values measured at standardized day t_{ij} since menstruation for the i^{th} woman, we consider the following bivariate semiparametric stochastic mixed model:

$$
\begin{aligned}
Y_{1ij} &= \text{age}_i^T\beta_{11} + \text{underWeight}_i^T\beta_{12} + \text{overWeight}_i^T\beta_{13} + f_1(t_{ij}) + b_{1i} + U_{1i}(t_{ij}) + \epsilon_{1ij} \\
Y_{2ij} &= \text{age}_i^T\beta_{21} + \text{underWeight}_i^T\beta_{22} + \text{overWeight}_i^T\beta_{23} + f_2(t_{ij}) + b_{2i} + U_{2i}(t_{ij}) + \epsilon_{2ij} \\
& \quad i = 1, \ldots, 100; j = 1, \ldots, n_i; t_{ij} \in \{0.5, 1.0, \ldots, 28\}
\end{aligned}
$$

where the specifications of random intercepts b_{1i} and b_{2i}, the bivariate Gaussian field U_{1i} and U_{2i}, the measurement errors ϵ_{1ij} and ϵ_{2ij}, and the nonparametric smooth functions are the same as those in the simulation study. Covariates underWeight and overWeight are indicator variables, which is characterized by Body Mass Index (BMI) where if BMI is less than 19.0 then the person is categorized as underWeight whereas if BMI is greater than 25.7, then overWeight. For computational stability, standardized days were centered at the median day 14 and divided by 10; covariate age is also centered at median 33 years old and divided by 100. Thus, $f_1(t)$ and $f_2(t)$ represents the progesterone and estrogen curves, respectively, for women of 33 years old with normal weight.

Table 2: Estimates of regression coefficients, variance parameter and smoothing parameter for the progesterone and estrogen data.

Model parameters	Parameter estimate	Standard Error
β_{11}	-1.2651	1.8674
β_{12}	-0.1687	0.2995
β_{13}	-0.1837	0.2009
β_{21}	-0.1455	1.7131
β_{22}	0.0068	0.2747
β_{23}	0.0765	0.1843
τ_1	4.6081	
τ_2	1.8475	
ϕ_1	0.6455	
$\phi_{1,2}$	-0.2755	
ϕ_2	0.6208	
σ_1^2	0.6499	
σ_2^2	0.6019	
ρ_1	0.2368	
a_{10}	-0.7699	
a_{11}	0.2894	
a_{12}	-0.1673	
ρ_2	0.0917	
a_{20}	-1.7431	
a_{21}	0.5172	
a_{22}	-0.0800	

Table 2 records the results of estimates of regression coefficients, smoothing parameters and variance components. The standard errors of our fixed effects parameter estimates are sufficiently large such that none of the fixed effects parameter estimates are statistically significant. That said, in terms of point estimates, we find a negative association on both responses with age, and both overweight and underweight (compared to regular BMI) are also negatively associated with progesterone only. The estimated correlation of the bivariate responses can be calculated from variance estimates from Table 2:

$$
\begin{aligned}
\operatorname{corr}(Y_{1ij}, Y_{2ik}) &= \frac{\operatorname{cov}(Y_{1ij}, Y_{2ik})}{\sqrt{\operatorname{var}(Y_{1ij})}\sqrt{\operatorname{var}(Y_{2ik})}} \\
&= \frac{\phi_{1,2}}{\sqrt{\sigma_1^2 + \phi_1 + \rho_1^0 \exp\{a_{10} + a_{11}t_{ij} + a_{12}t_{ij}^2\}}\sqrt{\sigma_2^2 + \phi_2 + \rho_2^0 \exp\{a_{20} + a_{21}t_{ik} + a_{22}t_{ik}^2\}}} \\
&= \frac{-0.2755}{\sqrt{0.6499 + 0.6455 + \exp\{-0.7699 + 0.2894t_{ij} - 0.1673t_{ij}^2\}}\sqrt{0.6019 + 0.6208 + \exp\{-1.7431 + 0.5172t_{ik} - 0.0800t_{ik}^2\}}},
\end{aligned}
$$

since $\operatorname{cov}(Y_{1ij}, Y_{2ij}) = \operatorname{cov}(b_1, b_2)$ in this case. For example, if $t_{ij} = 5$ and $t_{ik} = 6$, then $\operatorname{corr}(Y_{1ij}, Y_{2ik}) = -0.2343$, which indicates that that the two hormones are negatively correlated when progesterone is at $t_{ij} = 5$ and estrogen is at $t_{ik} = 6$. The estimates of nonparametric function $\hat{f}_1$ and $\hat{f}_2$ and their 95% confidence interval are superimposed over the log responses in Figure 1, respectively, which accurately captures the underlying trends of the bivariate longitudinal responses.

6. Discussion

We propose and build a model for analysis of bivariate cyclic longitudinal data and provide inference procedures. The model is proposed in the likelihood framework and the regression parameters and nonparametric functions are estimated by maximizing a penalized likelihood function. The smoothing parameter and variance components are numerically estimated using the Fisher-scoring algorithm based on restricted maximum likelihood. Modelling the time effect nonparametrically gives more flexibility in specifying the response mean structure, and the Gaussian field allows for additional flexibility in specifying the within-subject correlation structure, including possibly non-stationarity. The correlation of the two responses were explained only through the correlation of the random effects in the data analysis in Section 5, though the proposed model 1 can accommodate more complicated correlation structure of the responses through the covariance matrix (2) of the bivariate Gaussian field, as illustrated in subsection 2.3. Simulation results show that inference procedure performs well in all estimation results.

The bivariate semiparametric stochastic longitudinal model we proposed can be readily extended to multivariate longitudinal data. Dimensionality can pose as a challenge during the extension however. In the bivariate studies, we employed both C++ and parallel computing in the simulation study. Despite the effort, there is still computational burden to this methodology's estimation. Also, special attention was given to parameter initialization as some initialization of parameters may lead to infinity in some entries of variance-covariance matrix, thus causing the matrix degenerate. This said, in the analysis of the real dataset, we tried three very different initializations of the parameters from one another, and all estimates of the regression parameters, the variance components, and smoothing parameters are qualitatively the same, which is reassuring.

We would like to further explore sensitivity/robustness to the model assumptions. We have investigated the impact of Gaussian field misspecification in the simulation studies, which show that the choice of Gaussian field has little impact on the fixed effect parameters of interest. However, if we were interested in the underlying biological process, a deeper understanding of the choice of the Gaussian field is needed. Also, non-Guassian models can be generalized under the proposed framework if needed. In spite of further work to consider, this is a flexible and informative method for modeling bivariate longitudinal response data and we look forward to further extensions of this work in the above and possibly other directions.

Appendix A.

Proof of (5). Taking derivative of the log-likelihood function (4) with respect to $\boldsymbol{\beta}$, $\boldsymbol{f}_1$, and $\boldsymbol{f}_1$, we have

$$\begin{aligned}
\ell_{\boldsymbol{\beta}} &= \boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{N}_1\boldsymbol{f}_1-\boldsymbol{N}_2\boldsymbol{f}_2)\\
\ell_{\boldsymbol{f}_1} &= \boldsymbol{N}_1^T\boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{N}_1\boldsymbol{f}_1-\boldsymbol{N}_2\boldsymbol{f}_2)-\lambda_1\boldsymbol{K}\boldsymbol{f}_1\\
\ell_{\boldsymbol{f}_2} &= \boldsymbol{N}_2^T\boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{N}_1\boldsymbol{f}_1-\boldsymbol{N}_2\boldsymbol{f}_2)-\lambda_2\boldsymbol{K}\boldsymbol{f}_2.
\end{aligned}$$

Set $\ell_{\boldsymbol{\beta}}, \ell_{\boldsymbol{f}_1}$ and $\ell_{\boldsymbol{f}_2}$ to be zero, we have

$$\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{N}_1\hat{\boldsymbol{f}}_1+\boldsymbol{N}_2\hat{\boldsymbol{f}}_2) = \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y} \tag{A.1}$$

$$\boldsymbol{N}_1^T\boldsymbol{W}(\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{N}_1\hat{\boldsymbol{f}}_1+\boldsymbol{N}_2\hat{\boldsymbol{f}}_2)+\lambda_1\boldsymbol{K}\hat{\boldsymbol{f}}_{\boldsymbol{1}} = \boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{Y} \tag{A.2}$$

$$\boldsymbol{N}_2^T\boldsymbol{W}(\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{N}_1\hat{\boldsymbol{f}}_1+\boldsymbol{N}_2\hat{\boldsymbol{f}}_2)+\lambda_2\boldsymbol{K}\hat{\boldsymbol{f}}_2 = \boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{Y}, \tag{A.3}$$

which can be rewritten as (5). □

Proof of (6), (7) and (8). From equations (A.1), (A.2) and (A.3), we can reexpress the parameter estimators

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{N}_1\hat{\boldsymbol{f}}_1-\boldsymbol{N}_2\hat{\boldsymbol{f}}_2)$$

$$\hat{\boldsymbol{f}}_1 = (\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1+\lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{N}_2\hat{\boldsymbol{f}}_2) \tag{A.4}$$

$$\hat{\boldsymbol{f}}_2 = (\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2+\lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{N}_1\hat{\boldsymbol{f}}_1). \tag{A.5}$$

To solve explicitly for $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{f}}_{\boldsymbol{1}}$ and $\hat{\boldsymbol{f}}_{\boldsymbol{2}}$, we first plug $\hat{\boldsymbol{f}}_{\boldsymbol{1}}$ (A.4) into equations (A.1) and (A.3), rearrange and obtain

$$\begin{aligned}
&\boldsymbol{X}^T\boldsymbol{W}\big[\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{N}_1(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1+\lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}(\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{N}_2\hat{\boldsymbol{f}}_2)+\boldsymbol{N}_2\hat{\boldsymbol{f}}_2\big]\\
=\ &\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y}-\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}_1(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1+\lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{Y};
\end{aligned}$$

and

$$\begin{aligned}
&\boldsymbol{N}_2^T\boldsymbol{W}\big[\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{N}_1(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1+\lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}(\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{N}_2\hat{\boldsymbol{f}}_2)+\boldsymbol{N}_2\hat{\boldsymbol{f}}_2\big]\\
=\ &\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{Y}-\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_1(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1+\lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{Y}
\end{aligned}$$

respectively; which can be rewritten as

$$\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{N}_2\hat{\boldsymbol{f}}_2=\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{Y} \tag{A.6}$$

and

$$\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{X}\hat{\boldsymbol{\beta}}+(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2+\boldsymbol{\lambda}_2\boldsymbol{K})\hat{\boldsymbol{f}}_2=\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{Y} \tag{A.7}$$

respectively, where $\boldsymbol{W}_1=\boldsymbol{W}-\boldsymbol{W}\boldsymbol{N}_1(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1+\lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}$. Or equivalently as

$$\hat{\boldsymbol{\beta}}=(\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1(\boldsymbol{Y}-\boldsymbol{N}_2\hat{\boldsymbol{f}}_2), \tag{A.8}$$

and

$$\hat{\boldsymbol{f}}_2=(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2+\lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_1(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}) \tag{A.9}$$

respectively. Then plugging (A.8) into (A.7) and (A.9) into (A.6) and rearrange, we have

$$\begin{aligned}&(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2+\boldsymbol{\lambda}_2\boldsymbol{K})\hat{\boldsymbol{f}}_2-\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{N}_2\hat{\boldsymbol{f}}_2\\ =\;&\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{Y}-\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{Y},\end{aligned}$$

and

$$\begin{aligned}&\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2+\lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{X}\hat{\boldsymbol{\beta}}\\ =\;&\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{Y}-\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2+\lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{Y},\end{aligned}$$

respectively. Therefore, after rearranging and regrouping terms, the closed-form solutions for $\hat{\boldsymbol{f}}_2$ and $\hat{\boldsymbol{\beta}}$ are

$$\hat{\boldsymbol{f}}_2=(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2+\lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{Y},$$

and

$$\hat{\boldsymbol{\beta}}=(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{Y},$$

where $\boldsymbol{W}_{f_2}=\boldsymbol{W}_1-\boldsymbol{W}_1\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1$, and $\boldsymbol{W}_x=\boldsymbol{W}_1-\boldsymbol{W}_1\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2+\lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_1$. Similarly, to obtain the closed-form solution for $\hat{\boldsymbol{f}}_{\boldsymbol{1}}$, we plug (A.5) into equation (A.2) and obtain

$$\boldsymbol{N}_1^T\boldsymbol{W}_2\boldsymbol{X}\hat{\boldsymbol{\beta}}+(\boldsymbol{N}_1^T\boldsymbol{W}_2\boldsymbol{N}_1+\boldsymbol{\lambda}_1\boldsymbol{K})\hat{\boldsymbol{f}}_1=\boldsymbol{N}_1^T\boldsymbol{W}_2\boldsymbol{Y}, \tag{A.10}$$

where $\boldsymbol{W}_2=\boldsymbol{W}-\boldsymbol{W}\boldsymbol{N}_2(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1+\lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}$. Plugging $\hat{\boldsymbol{\beta}}=(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_2(\boldsymbol{Y}-\boldsymbol{N}_1\hat{\boldsymbol{f}}_1)$ into (A.10), the closed-form solution for $\hat{\boldsymbol{f}}_1$ is

$$\hat{\boldsymbol{f}}_{\boldsymbol{1}}=(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1+\lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{Y},$$

where $\boldsymbol{W}_{f_1}=\boldsymbol{W}_2-\boldsymbol{W}_2\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_2$. □

Proof of (9) and (10). From model assumptions, it follows that

$$\boldsymbol{Y}\sim N_{2n}\left(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{N}_1\boldsymbol{f}_1+\boldsymbol{N}_2\boldsymbol{f}_2,\boldsymbol{V}\right),\;\boldsymbol{b}\sim N_{2m}(\boldsymbol{0},\boldsymbol{D}).$$

Since the covariance of $\boldsymbol{Y}$ and $\boldsymbol{b}$ is

$$\begin{aligned}
\mathrm{cov}(\boldsymbol{Y},\boldsymbol{b}) &= \mathrm{cov}(\boldsymbol{X\beta}+\boldsymbol{N}_1\boldsymbol{f}_1+\boldsymbol{N}_2\boldsymbol{f}_2+\boldsymbol{Zb}+\boldsymbol{U}+\boldsymbol{\epsilon},\boldsymbol{b}) \\
&= \mathrm{cov}(\boldsymbol{X\beta},\boldsymbol{b})+\mathrm{cov}(\boldsymbol{N}_1\boldsymbol{f}_1,\boldsymbol{b})+\mathrm{cov}(\boldsymbol{N}_2\boldsymbol{f}_2,\boldsymbol{b})+\boldsymbol{Z}\mathrm{cov}(\boldsymbol{b},\boldsymbol{b})+\mathrm{cov}(\boldsymbol{U},\boldsymbol{b})+\mathrm{cov}(\boldsymbol{\epsilon},\boldsymbol{b}) \\
&= ZD,
\end{aligned}$$

the joint distribution of $\boldsymbol{Y}$ and $\boldsymbol{b}$ is

$$\begin{pmatrix}\boldsymbol{Y}\\ \boldsymbol{b}\end{pmatrix} \sim \boldsymbol{N}_{2n+2m}\left(\begin{pmatrix}\boldsymbol{X\beta}+\boldsymbol{N}_1\boldsymbol{f}_1+\boldsymbol{N}_2\boldsymbol{f}_2\\ \boldsymbol{0}\end{pmatrix}, \begin{pmatrix}\boldsymbol{V} & \boldsymbol{ZD}\\ \boldsymbol{DZ}^T & D\end{pmatrix}\right).$$

Therefore, by the property of normality, the conditional expectation results follows. □

Appendix B.

Proof of (11) and (12). For regression coefficient estimator $\hat{\boldsymbol{\beta}}$,

$$\begin{aligned}
\mathrm{E}(\hat{\boldsymbol{\beta}}) &= (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x E[\boldsymbol{Y}] \\
&= (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x(\boldsymbol{X\beta}+\boldsymbol{N}_1\boldsymbol{f}_1+\boldsymbol{N}_2\boldsymbol{f}_2) \\
&= \boldsymbol{\beta}+(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x(\boldsymbol{N}_1\boldsymbol{f}_1+\boldsymbol{N}_2\boldsymbol{f}_2)
\end{aligned}$$

For nonparametric function estimator $\hat{\boldsymbol{f}}_1=(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1+\lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{Y}$,

$$\begin{aligned}
\mathrm{E}(\hat{\boldsymbol{f}}_1) &= (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1+\lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}(\boldsymbol{X\beta}+\boldsymbol{N}_1\boldsymbol{f}_1+\boldsymbol{N}_2\boldsymbol{f}_2) \\
&= \boldsymbol{0}+(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1+\lambda_1\boldsymbol{K})^{-1}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1+\lambda_1\boldsymbol{K}-\lambda_1\boldsymbol{K})\boldsymbol{f}_1 \\
&\qquad +(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1+\lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_2\boldsymbol{f}_2 \\
&= \boldsymbol{f}_1-\lambda_1(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1+\lambda_1\boldsymbol{K})^{-1}\boldsymbol{K}\boldsymbol{f}_1+(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1+\lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_2\boldsymbol{f}_2
\end{aligned}$$

where $(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1+\lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{X\beta}=\boldsymbol{0}$ since

$$\begin{aligned}
& (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1+\lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{X\beta} \\
= & (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1+\lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\left[\boldsymbol{W}_2-\boldsymbol{W}_2\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_2\right]\boldsymbol{X\beta} \\
= & (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1+\lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_2\boldsymbol{X\beta}-(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1+\lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_2\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X\beta} \\
= & \boldsymbol{0}.
\end{aligned}$$

□

Remark 1. The bias of nonparametric function estimator $\hat{\boldsymbol{f}}_2$ in (13) can be derived similarly as that of $\hat{\boldsymbol{f}}_1$.

Lemma 1. *The biases of $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{f}}_1$, $\hat{\boldsymbol{f}}_2$, $\hat{\boldsymbol{b}}_i$ and $\hat{\boldsymbol{U}}_i$ all go to $\boldsymbol{0}$ as both smoothing parameters $\lambda_1 \to 0$ and $\lambda_2 \to 0$.*

Proof. As $\lambda_1 \to 0$ and $\lambda_2 \to 0$ simultaneously, then

$$\boldsymbol{W}_x \to \boldsymbol{W}_1 - \boldsymbol{W}_1\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_1, \tag{B.1}$$

where

$$\boldsymbol{W}_1 \to \boldsymbol{W} - \boldsymbol{W}\boldsymbol{N}_1(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\boldsymbol{W} \tag{B.2}$$

Plugging $\boldsymbol{W}_x$ in (B.1) into bias of $\hat{\boldsymbol{\beta}}$ in (11), we have

$$\begin{aligned}
& \mathrm{E}(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta} \\
= \; & (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\left[\boldsymbol{W}_1 - \boldsymbol{W}_1\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_1\right](\boldsymbol{N}_1\boldsymbol{f}_1 + \boldsymbol{N}_2\boldsymbol{f}_2) \\
= \; & (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1(\boldsymbol{N}_1\boldsymbol{f}_1 + \boldsymbol{N}_2\boldsymbol{f}_2) \\
& \qquad -(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_1(\boldsymbol{N}_1\boldsymbol{f}_1 + \boldsymbol{N}_2\boldsymbol{f}_2) \\
= \; & (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{N}_1\boldsymbol{f}_1 \\
& +(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{N}_2\left[\boldsymbol{f}_2 - (\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2\boldsymbol{f}_2\right] \\
& -(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_1\boldsymbol{f}_1 \\
= \; & (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{N}_1\boldsymbol{f}_1 - (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_1\boldsymbol{f}_1
\end{aligned} \tag{B.3}$$

Further, plugging $\boldsymbol{W}_1$ in (B.2) into(B.3), we have

$$\begin{aligned}
& \mathrm{E}(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta} \\
= \; & (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\left[\boldsymbol{W} - \boldsymbol{W}\boldsymbol{N}_1(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\boldsymbol{W}\right]\boldsymbol{N}_1\boldsymbol{f}_1 \\
& -(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\left[\boldsymbol{W} - \boldsymbol{W}\boldsymbol{N}_1(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\boldsymbol{W}\right]\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T \\
& \qquad \left[\boldsymbol{W} - \boldsymbol{W}\boldsymbol{N}_1(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\boldsymbol{W}\right]\boldsymbol{N}_1\boldsymbol{f}_1 \\
= \; & (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}_1\boldsymbol{f}_1 - (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}_1(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1\boldsymbol{f}_1 \\
& -(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_1\boldsymbol{f}_1 \\
& +(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_1(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1\boldsymbol{f}_1 \\
& +(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}_1(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_1\boldsymbol{f}_1 \\
& -(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}_1(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T \\
& \qquad \boldsymbol{W}\boldsymbol{N}_1(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1\boldsymbol{f}_1 \\
= \; & \boldsymbol{0}
\end{aligned}$$

Therefore, the bias of $\hat{\boldsymbol{\beta}}$ goes to $\boldsymbol{0}$ as $\lambda_1 \to 0$ and $\lambda_2 \to 0$.

As $\lambda_1 \to 0$ and $\lambda_2 \to 0$ simultaneously, the bias of $\hat{\boldsymbol{f}}_{\boldsymbol{1}}$

$$\mathrm{E}(\hat{\boldsymbol{f}}_{\boldsymbol{1}}) - \boldsymbol{f}_{\boldsymbol{1}} \to (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_2\boldsymbol{f}_2 \tag{B.4}$$

where

$$\boldsymbol{W}_{f_1} = \boldsymbol{W}_2 - \boldsymbol{W}_2\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_2 \tag{B.5}$$

and

$$\boldsymbol{W}_2 \to \boldsymbol{W} - \boldsymbol{W}\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W} \tag{B.6}$$

Plugging $\boldsymbol{W}_{f_1}$ in (B.5) into bias of $\hat{\boldsymbol{f}}_{\boldsymbol{1}}$ in (B.4) and plugging $\boldsymbol{W}_2$ in (B.6) into $\boldsymbol{W}_{f_1}$ in (B.5), we have

$$\begin{aligned}
& \mathrm{E}(\hat{\boldsymbol{f}}_{\boldsymbol{1}}) - \boldsymbol{f}_{\boldsymbol{1}} \\
= \; & (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_2\boldsymbol{N}_2\boldsymbol{f}_2 - (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_2\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{N}_2\boldsymbol{f}_2 \\
= \; & (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_2\boldsymbol{f}_2 - (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2\boldsymbol{f}_2 \\
& -(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\left[\boldsymbol{W} - \boldsymbol{W}\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}\right]\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}^T \\
& \qquad\qquad\qquad\qquad\qquad\qquad \left[\boldsymbol{W} - \boldsymbol{W}\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}\right]\boldsymbol{N}_2\boldsymbol{f}_2 \\
= \; & \boldsymbol{0} - (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}_2\boldsymbol{f}_2 \\
& \quad +(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2\boldsymbol{f}_2 \\
& +(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}_2\boldsymbol{f}_2 \\
& -(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}^T \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \boldsymbol{W}\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2\boldsymbol{f}_2 \\
= \; & \boldsymbol{0}
\end{aligned}$$

Therefore, the bias of $\hat{\boldsymbol{f}}_1$ goes to $\boldsymbol{0}$ as $\lambda_1 \to 0$ and $\lambda_2 \to 0$. Similar results can be shown for the bias of $\hat{\boldsymbol{f}}_2$. Since

$$\boldsymbol{X}^T\boldsymbol{W}_x(\boldsymbol{N}_1\boldsymbol{f}_1 + \boldsymbol{N}_2\boldsymbol{f}_2) \to \boldsymbol{0},$$

and

$$\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_1\boldsymbol{f}_1 \to \boldsymbol{0}, \qquad \boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_2\boldsymbol{f}_2 \to \boldsymbol{0}$$

as $\lambda_1 \to 0$ and $\lambda_2 \to 0$ as shown before, both the biases in the estimators of the random effects $\hat{\boldsymbol{b}}_i$ and the stochastic process $\hat{\boldsymbol{U}}_i$ go to zero. □

Remark 2. The covariance result for the random effect in (14) is non-trivial and long, and is available upon request.

References

[1] Brumback, B. A. & Rice, J. A. (1998). Semiparametric stochastic mixed models for longitudinal data. *J. Am. Statist. Assoc.* **93**, 961–976.

[2] Gold, E. B., Eskenazi, B., Hammond, S. K., Lasley, B. L.,Samuels, S. J., O'Neill Rasor, M., Hines, C. J., Overstreet, J. W. & Schenker, M. B. (1995a). Prospectively assessed menstrual cycle characteristics in female wafer-fabrication and nonfabrication semiconductor employees. *Am. J. Industrial Medicine* **28**, 799–815.

[3] Gold, E. B., Eskenazi, B., Lasley, B. L., Samuels, S. J.,O'Neill Rasor, M., Overstreet, J. W. & Schenker, M. B. (1995b). Epidemiologic methods for prospective assessment of menstrual cycle and reproductive characteristics in female semiconductor workers. *Am. J. Industrial Medicine* **28**, 783–797.

[4] Green, P. J. (1987). Penalized likelihood for general semi-parametric regression models. *International Statistical Review* **55**, 245–260.

[5] Green, P. J. & Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: A roughness penalty approach.* London, UK: Chapman and Hall/CRC.

[6] Koralov, L. B. & Sinai, Y. G. (2007). *Theory of Probability and Random Processes.* New York, NY: Springer, 2 ed.

[7] Liu, Z., Cappola, A. R., Crofford, L. J. & Guo, W. (1999). Modeling bivariate longitudinal hormone profiles by hierarchical state space models. *J. Am. Statist. Assoc.* **109**, 108–118.

[8] Raffa, J. D., & Dubin, J. A. (2015). Multivariate longitudinal data analysis with mixed effects hidden Markov models. *Biometrics* **71**, 821–831.

[9] Sy, J. P., Taylor, J. M. G. & Cumberland, W. G. (1997). A stochastic model for the analysis of bivariate longitudinal AIDS data. *Biometrics* **53**, 542–555.

[10] Taylor, J. M. G., Cumberland, W. G. & Sy, J. P. (2000). A stochastic model for analysis of longitudinal AIDS data. *J. Am. Statist. Assoc.* **89**, 727–736.

[11] Verbyla, A. P., Cullis, B. R., Kenward, M. G. & Welham, S. J. (1999). The analysis of designed experiments and longitudinal data by using smoothing splines. *J. R. Statist. Soc.* C **48**, 269–311.

[12] Wang, Y., Guo, W. & Brown, M. B. (2000). Spline smoothing for bivariate data with applications to association between hormones. *Statistica Sinica.* 377–497.

[13] Welham, S. J., Cullis, B. R., Kenward, M. G. & Thompson, R. (2006). The analysis of longitudinal data using mixed model L-splines. *Biometrics.* **62**, 392–401.

[14] Wood, S. (2006). *Generalized Additive Models: An Introduction with R.* Boca Raton, Florida: Chapman and Hall/CRC.

[15] Zhang, D., Lin, X., Raz, J. & Sowers, M. (1998). Semiparametric stochastic mixed models for longitudinal data. *J. Am. Statist. Assoc.* **93**, 710–719.

[16] Zhang, D., Lin, X. & Sowers, M. (2000). Semiparametric regression for periodic longitudinal hormone data from multiple menstrual cycles. *Biometrics* **56**, 31–39.