Biometrika (2017), 103, 1, pp. 1–18
Printed in Great Britain

Advance Access publication on 31 July 2016

A bivariate semiparametric stochastic mixed model

By 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

Summary

We propose and consider inference for a semiparametric stochastic mixed effects model for bivariate periodic repeated measures data. The bivariate model uses parametric fixed effects for modeling covariate effects and periodic smooth nonparametric functions for each of the two underlying time effects. The between-subject and within-subject correlations are modelled using separate but correlated random effects and a bivariate Gaussian random field, respectively. We derive estimators for both the fixed effects regression coefficients and the nonparametric time functions using maximum penalized likelihood, where the resulting estimator for the nonparametric time function is a periodic 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 longitudinal dataset of pre-menopausal women agree with the real dataset.

Some key words: Bivariate longitudinal data; Cyclic; 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 for at least one menstrual cycle. We are interested in modelling time courses for the estrogen and progesterone metabolite profiles, the effects of demographic and lifestyle factors on the hormone excretion, and the potential correlation between the two hormones. One difficulty in modeling is that the time courses of the two hormones profiles vary complicated to model using a simple parametric function and the data is periodic, see Figure 4. Multiple layers of correlation structures, say within-subject correlation for the same hormone and between hormones at the same time point or different time points, also present a challenge.

For univariate longitudinal analysis, there have been many extensions to allow for further flexibility in the mean. For example, Brumback & Rice (1998) 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. (1999) used data-based determination of the smoothing parameters and applied to the analysis of designed experiments. Some efforts are also geared towards enriching the modeling of the correlation structure. Taylor et al. (1994) used an integrated Ornstein-Uhlenbeck process to model the covariance structure in a longitudinal data. In particular, Zhang et al. (1998) and Zhang et al. (2000) proposed a semiparametric stochastic mixed model for periodic longitudinal data, where covariate effects are modeled parametrically and the underlying complex periodic time course are modeled nonparametrically; and used a stochastic process and random effects to model the within-subject correlation. Instead of smoothing spline, Welham et al. (2006) modelled cyclic longitudinal data using mixed model L-splines. Wood (2006) modified penalized cubic regression spline to model a cyclic smooth function.

Some of the univariate techniques have been extended to multivariate case. For example, Sy et al. (1997) employed multivariate stochastic processes to jointly model bivariate longitudinal data. More recently, Raffa & Dubin (2015) modeled 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 in this paper, Liu et al. (2014) 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.

In this paper, we extend Zhang et al. (1998) and Zhang et al. (2000) and propose a bivariate semiparametric stochastic mixed model for bivariate periodic repeated measures data. The bivariate model uses parametric fixed effects for modeling covariate effects and periodic smooth nonparametric functions for each of the two underlying time effects. The between-subject and within-subject correlations are modeled using separate but correlated random effects and a bivariate Gaussian random field, respectively. Correlation between the two responses are modeled through the specification of the correlation structure for the random effects, though can also be modeleed through the bivariate Gasussian field, we will not consider this case here. 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.

2. The Bivariate Semiparametric Stochastic Mixed Effects Model

2·1. General Bivariate Longitudinal Model with Joint Distribution of Random Effects

Denote $\{Y_{1ij}, Y_{2ij}\}$ to be the bivariate metabolit 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

$$\begin{aligned} Y_{1ij} &= \boldsymbol{X}_{1ij}^T \boldsymbol{\beta}_1 + f_1(t_{ij}) + \boldsymbol{Z}_{1ij}^T \boldsymbol{b}_{1i} + U_{1i}(t_{ij}) + \epsilon_{1ij} \\ Y_{2ij} &= \boldsymbol{X}_{2ij}^T \boldsymbol{\beta}_2 + f_2(t_{ij}) + \boldsymbol{Z}_{2ij}^T \boldsymbol{b}_{2i} + U_{2i}(t_{ij}) + \epsilon_{2ij}, \end{aligned} \tag{1}$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are $p_1 \times 1$ and $p_2 \times 1$ vectors of fixed effects regression coefficients, respectively, associated with known covariates $(\boldsymbol{X}_{1ij}, \boldsymbol{X}_{2ij})$; $\boldsymbol{b}_{1i}$ and $\boldsymbol{b}_{2i}$ are $q_1 \times 1$ and $q_2 \times 1$ vectors of random effects, respectively, associated with known covariates $(\boldsymbol{Z}_{1ij}, \boldsymbol{Z}_{2ij})$; $f_1(t)$ and $f_2(t)$ are twice-differentiable periodic smooth functions of time with periods T_1 and T_2, respectively; $(U_{1i}(t_{ij}), U_{2i}(t_{ij}))$ is a mean zero bivariate Gaussian field with covariance matrix

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

where $\xi_1(t)$ and $\xi_2(t)$ are periodic 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 $(\epsilon_{1ij}, \epsilon_{2ij})^T$ are bivariate normal with mean $\mathbf{0}$ and variance $\begin{pmatrix} \sigma_1^2 & \sigma_{12}^2 \\ \sigma_{12}^2 & \sigma_2^2 \end{pmatrix}$. We assume that $\boldsymbol{b}_{ki}, k = 1,2$ to be $(q_1 +$

q_2)-dimensional normal with mean zero and covariance matrix $\boldsymbol{D}(\boldsymbol{\phi})$, and that the random effects, the stochastic process and the measurement error to be mutually independent.

Denote $\boldsymbol{Y}_{ij} = (Y_{1ij}, Y_{2ij})^T$ and similarly for $\boldsymbol{\beta}$, $\boldsymbol{b}_i$, $\boldsymbol{f}(t_{ij})$, $\boldsymbol{U}_i(t_{ij})$ and $\boldsymbol{\epsilon}_{ij}$, and let $\boldsymbol{X}_{ij} = \mathrm{diag}(\boldsymbol{X}_{1ij}^T, \boldsymbol{X}_{2ij}^T)$ and similarly for $\boldsymbol{Z}_{ij}$. Then model (1) can also be rewritten as

$$\boldsymbol{Y}_{ij} = \boldsymbol{X}_{ij}^T\boldsymbol{\beta} + \boldsymbol{f}(t_{ij}) + \boldsymbol{Z}_{ij}^T\boldsymbol{b}_i + \boldsymbol{U}_i(t_{ij}) + \boldsymbol{\epsilon}_{ij}, \tag{2}$$

with the same model assumptions.

This model (1) is an extension to the model proposed in Zhang et al. (1998) and Zhang et al. (2000), where a univariate semiparametric stochastic mixed model for (periodic) 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. The distributions of the two random effects can be potentially distinct, with different distributions or the same distribution with different parameters; but, as mentioned above, the two random effects are assumed separate but correlated.

2·2. The Gaussian Field Specification

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

There are potentially many choices available: Wiener process or Brownian motion (Taylor et al., 1994); an integrated Wiener process and so on. One particular Gaussian process/field worthy of mentioning is the Ornstein-Uhlenbeck (OU) process (Koralov & Sinai, 2007) 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 become nonhomogenesous (NOU) and, for example, we can assume $\xi(t) = \exp(a_0 + a_1 t)$.

3. Estimation and Inference

3·1. Matrix Notation

To make inferences from the model (2), 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 $\boldsymbol{\epsilon}_i$. Assume $t_{ij} > 0$ and $\min\{t_{ij}\} = 0$. Since $f_1(t)$ and $f_2(t)$ are periodic functions with periods T_1 and T_2, we only need to estimate $f_1(t)$ for $t \in [0, T_1)$ and $f_2(t)$ for $t \in [0, T_2)$. Let $\boldsymbol{t}'_1 = (t'_{11}, \ldots, t'_{1r_1})$ to be a vector of ordered distinct values of $t'_{1ij} = \mod (t_{ij}, T_1)$ for $i = 1, \ldots, m$ and $j = 1 \ldots n_i$, and let $\boldsymbol{t}'_2 = (t'_{21}, \ldots, t'_{2r_2})$ to be a vector of ordered distinct values of $t'_{2ij} = \mod (t_{ij}, T_2)$ for $i = 1, \ldots, m$ and $j = 1 \ldots n_i$, thus $t'_{1k} \in [0, T_1)$ for $k = 1, \ldots, r_1$ and $t'_{2k} \in [0, T_2)$ for $k = 1, \ldots, r_2$. Then, let $\tilde{\boldsymbol{N}}_{1i}$ be the $n_i \times r_1$ incidence matrix for the i^{th} subject for the first response connecting $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{in_i})^T$ and $\boldsymbol{t}'_1$ such that

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

where $\tilde{\boldsymbol{N}}_{1i}[j, \ell]$ denote the $(j, \ell)^{\text{th}}$ entry of matrix $\tilde{\boldsymbol{N}}_{1i}$ for $j = 1, \ldots, n_i$ and $\ell = 1, \ldots, r_1$. Similarly, let $\tilde{\boldsymbol{N}}_{2i}$ be the $n_i \times r_2$ incidence matrix for the i^{th} subject for the second response connecting $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{in_i})^T$ and $\boldsymbol{t}'_2$ such that

$$\tilde{\boldsymbol{N}}_{2i}[j, \ell] = \begin{cases} 1 & \text{if } t'_{2ij} = t'_{2\ell} \\ 0 & \text{otherwise,} \end{cases}$$

where $\tilde{\boldsymbol{N}}_{2i}[j,\ell]$ denote the $(j,\ell)^{\text{th}}$ entry of matrix $\tilde{\boldsymbol{N}}_{2i}$ for $j=1,\dots,n_i$ and $\ell=1,\dots,r_2$. Further, we need to refine the incidence matrix $\tilde{\boldsymbol{N}}_{1i}$ to make it correspond to the first response such that

$$\boldsymbol{N}_{1i} = \boldsymbol{A}_{1i}\tilde{\boldsymbol{N}}_{1i}$$

where

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

thus the refined incidence matrix $\boldsymbol{N}_{1i}$ is of dimension $2n_i \times r_1$. Similarly, the refined incidence matrix $\boldsymbol{N}_{2i}$ of dimension $2n_i \times r_2$ for the second response is

$$\boldsymbol{N}_{2i} = \boldsymbol{A}_{2i}\tilde{\boldsymbol{N}}_{2i}$$

where

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

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

$$\boldsymbol{Y}_i = \boldsymbol{X}_{\boldsymbol{i}}\boldsymbol{\beta} + \boldsymbol{N}_{1i}\boldsymbol{f}_1 + \boldsymbol{N}_{2i}\boldsymbol{f}_2 + \boldsymbol{Z}_{\boldsymbol{i}}\boldsymbol{b}_{\boldsymbol{i}} + \boldsymbol{U}_i + \boldsymbol{\epsilon}_i$$

for subject i, where $\boldsymbol{f}_1 = (f_1(t'_{11},\dots,f_1(t'_{1r_1})^T$ and $\boldsymbol{f}_2 = (f_2(t'_{21}),\dots,f_2(t'_{2r_2})^T$.

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 let $n=\sum_{i=1}^m n_i$, then the bivariate semiparametric stochastic mixed effects model over all subjects is written as

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

where

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

with $\boldsymbol{D}(\boldsymbol{\phi}) = \operatorname{diag}(\boldsymbol{D},\dots,\boldsymbol{D})$; $\boldsymbol{\Gamma}(\boldsymbol{\xi},\rho) = \operatorname{diag}(\boldsymbol{\Gamma}_1(\boldsymbol{t}_1,\boldsymbol{t}_1),\dots,\boldsymbol{\Gamma}_m(\boldsymbol{t}_m,\boldsymbol{t}_m))$ and the $(k,k')^{\text{th}}$ entry of $\boldsymbol{\Gamma}_i(\boldsymbol{t}_i,\boldsymbol{t}_i)$ is $\boldsymbol{C}_i(k,k')$; and $\boldsymbol{\Sigma}(\boldsymbol{\sigma}^2) = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \boldsymbol{I}_n$.

3·2. Estimation of Model Coefficients, Nonparametric Function, Random Effects and Gaussian Fields

The proposed model (3) also implies the marginal model

$$Y = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{N}_1\boldsymbol{f}_1 + \boldsymbol{N}_2\boldsymbol{f}_2 + \boldsymbol{\epsilon}^*, \boldsymbol{\epsilon}^* \sim N_{2n}(\boldsymbol{0},\boldsymbol{V}) \tag{4}$$

where $\boldsymbol{V} = \boldsymbol{Z}\boldsymbol{D}\boldsymbol{Z}^T + \boldsymbol{\Gamma} + \boldsymbol{\Sigma}$. Thus, the log-likelihood function for $(\boldsymbol{\beta},\boldsymbol{f}_1,\boldsymbol{f}_2)$ by (4) is :

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

We estimate the parameters $\boldsymbol{\beta}$, $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ by maximizing the penalized likelihood (Wang, Guo & Brown, 2000):

$$\ell(\boldsymbol{\beta}, \boldsymbol{f}_1, \boldsymbol{f}_2; \boldsymbol{Y}) - \lambda_1 \int_a^b [f_1''(t)]^2 dt - \lambda_2 \int_a^b [f_2''(t)]^2 dt = \ell(\boldsymbol{\beta}, \boldsymbol{f}_1, \boldsymbol{f}_2; \boldsymbol{Y}) - \lambda_1 \boldsymbol{f}_1^T \boldsymbol{K} \boldsymbol{f}_1 - \lambda_2 \boldsymbol{f}_2^T \boldsymbol{K} \boldsymbol{f}_2 \tag{5}$$

where $\boldsymbol{K}$ is the nonnegative definite smoothing matrix, defined in Equation (2.3) in Green & Silverman (1994). And the resulting estimators for the nonparametric functions are the natural cubic spline estimators of $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$. 125

Given fixed smoothing parameters and variance parameters, differentiation of (5) 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

$$\begin{pmatrix} \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}_1 & \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}_2 \\ \boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{X} & \boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K} & \boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_2 \\ \boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{X} & \boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_1 & \boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{f}_1 \\ \boldsymbol{f}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y} \\ \boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{Y} \\ \boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{Y} \end{pmatrix}, \tag{6}$$

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 $\hat{\boldsymbol{f}}_2$ 130

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{Y} \tag{7}$$
$$\hat{\boldsymbol{f}}_1 = (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{Y} \tag{8}$$
$$\hat{\boldsymbol{f}}_2 = (\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{Y}, \tag{9}$$

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 $\boldsymbol{Y}$. Note that the proposed model (3) can also be rewritten as a two-level hierarchical model 135

$$\begin{aligned} \boldsymbol{Y}|\boldsymbol{b}, \boldsymbol{U} &\sim N_{2n}\left(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{N}_1\boldsymbol{f}_1 + \boldsymbol{N}_2\boldsymbol{f}_2 + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{U}, \boldsymbol{\Sigma}\right) \\ \boldsymbol{b} &\sim N_{2m}(\boldsymbol{0}, \boldsymbol{D}) \\ \boldsymbol{U} &\sim N_{2n}(\boldsymbol{0}, \boldsymbol{\Gamma}); \end{aligned}$$

then by the property of normality, we have

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

since $\mathrm{cov}(\boldsymbol{Y}, \boldsymbol{b}) = ZD$ after some simple calculation. Therefore,

$$\hat{\boldsymbol{b}}_i = E(\boldsymbol{b}|\boldsymbol{Y}) = \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{f}}_{1i} - \hat{\boldsymbol{f}}_{2i}) \tag{10}$$

and similarly, 140

$$\hat{\boldsymbol{U}}_i(\boldsymbol{s}_i) = \boldsymbol{\Gamma}(\boldsymbol{s}_i, \boldsymbol{t}_i)\boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{f}}_{1i} - \hat{\boldsymbol{f}}_{2i}) \tag{11}$$

where $\hat{\boldsymbol{f}}_{1i} = \boldsymbol{N}_{1i}\hat{\boldsymbol{f}}_1$ and $\hat{\boldsymbol{f}}_{2i} = \boldsymbol{N}_{2i}\hat{\boldsymbol{f}}_2$.

3·3. Biases and Covariances of Model Coefficients, Nonparametric Function, Random Effects and Gaussian Fields

From closed-form solutions of estimators from equation (7) (8) and (9) in Section 3·2, the biases of the estimators $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{f}}_1$ and $\hat{\boldsymbol{f}}_2$ can be easily calculated and we have

$$\mathrm{E}(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta} = (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x(\boldsymbol{N}_1\boldsymbol{f}_1 + \boldsymbol{N}_2\boldsymbol{f}_2) \tag{12}$$
$$E(\hat{\boldsymbol{f}}_1) - \boldsymbol{f}_1 = (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_2\boldsymbol{f}_2 - \lambda_1\boldsymbol{K}\boldsymbol{f}_1) \tag{13}$$
$$E(\hat{\boldsymbol{f}}_2) - \boldsymbol{f}_2 = (\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_1\boldsymbol{f}_1 - \lambda_2\boldsymbol{K}\boldsymbol{f}_2). \tag{14}$$

Similarly, the expected values of the estimators in (10) and (11) for the subject-specific random effects $\boldsymbol{b}_i$ and for the subject-specific Gaussian field $\boldsymbol{U}_i(\boldsymbol{s}_i)$ are

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

and

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

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 $\lambda_2 \to 0$.

For covariances, simple calculation using (7) (8) and (9) 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 covariance 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 (10) and (11) for the subject-specific random effects $\boldsymbol{b}_i$ and for the subject-specific Gaussian field $\boldsymbol{U}_i(\boldsymbol{s}_i)$ are

$$\mathrm{cov}(\hat{\boldsymbol{b}}_i - \boldsymbol{b}_i) = \boldsymbol{D} - \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D} + \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i\boldsymbol{\chi}_i\boldsymbol{C}^{-1}\boldsymbol{\chi}^T\boldsymbol{W}\boldsymbol{\chi}\boldsymbol{C}^{-1}\boldsymbol{\chi}_i^T\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D} \tag{15}$$

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 $\boldsymbol{\chi}_i = \left(\boldsymbol{X}_i\ \boldsymbol{N}_{1i}\ \boldsymbol{N}_{2i}\right)$ and $\boldsymbol{\chi} = \left(\boldsymbol{X}\ \boldsymbol{N}_1\ \boldsymbol{N}_2\right)$.

3·4. Estimation of the Smoothing Parameters and Variance Parameters

By Green (1987), we can write $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ by a one-to-one linear transformation as

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

where δ_1 and a_1 are of dimensions 2 and $r_1 - 2$ and δ_2 and a_2 are of dimensions 2 and $r_2 - 2$. $\boldsymbol{B}_1 = \boldsymbol{L}_1(\boldsymbol{L}_1^T\boldsymbol{L}_1)^{-1}$ and $\boldsymbol{L}_1$ is $r_1 \times (r_1 - 2)$ full-rank matrix satisfying $\boldsymbol{K}_1 = \boldsymbol{L}_1\boldsymbol{L}_1^T$ and $\boldsymbol{L}_1^T\boldsymbol{T}_1 = 0$.

$\boldsymbol{B}_2 = \boldsymbol{L}_2(\boldsymbol{L}_2^T\boldsymbol{L}_2)^{-1}$ and $\boldsymbol{L}_2$ is $r_2 \times (r_2-2)$ full-rank matrix satisfying $\boldsymbol{K}_2 = \boldsymbol{L}_2\boldsymbol{L}_2^T$ and $\boldsymbol{L}_2^T\boldsymbol{T}_2 = 0$. Thus the proposed semiparametric mixed model (3) can be rewritten as a modified linear mixed model,

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{N}_1\boldsymbol{T}_1\boldsymbol{\delta}_1 + \boldsymbol{N}_1\boldsymbol{B}_1\boldsymbol{a}_1 + \boldsymbol{N}_2\boldsymbol{T}_2\boldsymbol{\delta}_2 + \boldsymbol{N}_2\boldsymbol{B}_2\boldsymbol{a}_2 + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{U} + \boldsymbol{\epsilon}, \tag{16}$$

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 a_1 distributed as normal $(0, \tau_1\boldsymbol{I})$, 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 $\boldsymbol{B}_{1*} = \boldsymbol{N}_1\boldsymbol{B}_1$ and $\boldsymbol{B}_{2*} = \boldsymbol{N}_2\boldsymbol{B}_2$.

Under the modified linear mixed model (16), 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 $\boldsymbol{X}_* = [\boldsymbol{X}, \boldsymbol{N}_1\boldsymbol{T}_1, \boldsymbol{N}_2\boldsymbol{T}_2]$. Taking the derivative 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 equation for the smoothing parameters τ_1, τ_2 and variance components $\boldsymbol{\theta}$ can be obtained:

$$\frac{\partial \ell_R}{\partial \tau_1} = -\frac{1}{2}\mathrm{tr}(\boldsymbol{P}_*\boldsymbol{B}_{1*}\boldsymbol{B}_{1*}^T) + \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1\hat{\boldsymbol{f}}_1 - \boldsymbol{N}_2\hat{\boldsymbol{f}}_2)^T\boldsymbol{V}^{-1}\boldsymbol{B}_{1*}\boldsymbol{B}_{1*}^T\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1\hat{\boldsymbol{f}}_1 - \boldsymbol{N}_2\hat{\boldsymbol{f}}_2), \tag{17}$$

$$\frac{\partial \ell_R}{\partial \tau_2} = -\frac{1}{2}\mathrm{tr}(\boldsymbol{P}_*\boldsymbol{B}_{2*}\boldsymbol{B}_{2*}^T) + \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1\hat{\boldsymbol{f}}_1 - \boldsymbol{N}_2\hat{\boldsymbol{f}}_2)^T\boldsymbol{V}^{-1}\boldsymbol{B}_{2*}\boldsymbol{B}_{2*}^T\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1\hat{\boldsymbol{f}}_1 - \boldsymbol{N}_2\hat{\boldsymbol{f}}_2), \tag{18}$$

and

$$\frac{\partial \ell_R}{\partial \theta_j} = -\frac{1}{2}\mathrm{tr}(\boldsymbol{P}_*\frac{\partial \boldsymbol{V}}{\partial \theta_j}) + \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1\hat{\boldsymbol{f}}_1 - \boldsymbol{N}_2\hat{\boldsymbol{f}}_2)^T\boldsymbol{V}^{-1}\frac{\partial \boldsymbol{V}}{\partial \theta_j}\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1\hat{\boldsymbol{f}}_1 - \boldsymbol{N}_2\hat{\boldsymbol{f}}_2), \tag{19}$$

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 (17), (18) and (19),

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

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. 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 for the smooth-

Table 1. Simulation results for estimates of model parameters based on 500 simulation replicates.

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	
ϕ_2	-0.50	-0.5019	-0.0038	0.0730	
ϕ_3	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	

ing parameters and the variance parameters. Bivariate cyclic longitudinal data are generated according to the following model:

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

where b_{1i} and b_{2i} are independent but correlated random intercepts following a bivariate normal distribution with mean $\mathbf{0}$ and covariance $\begin{pmatrix} \phi_1 & \phi_2 \\ \phi_2 & \phi_3 \end{pmatrix}$. U_{1i} and U_{2i} are simulated from mean 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

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

lastly, ϵ_{1ij} and ϵ_{2ij} are simulated from a mean $\mathbf{0}$ bivariate normal distribution with covariance $\begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$. Further, the nonparametric cyclic functions are generated from $f_1(t) = 5\sin(2\pi/28)\,t$ and $f_2(t) = 3\cos(2\pi/28)\,t$ with periods to be 28 days for both responses.

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 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 1. Figure 2 shows that model standard errors of estimates of $\hat{f}_1$ and $\hat{f}_2$ agree quite well with the empirical standard errors.

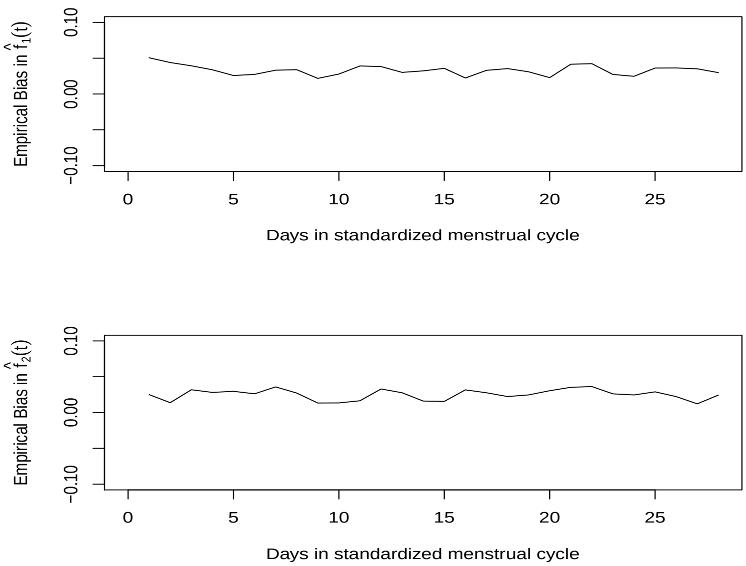

Fig. 1. Empirical Bias in estimated nonparametric functions $\hat{f}_1$ and $\hat{f}_2$ based on 500 simulation replications.

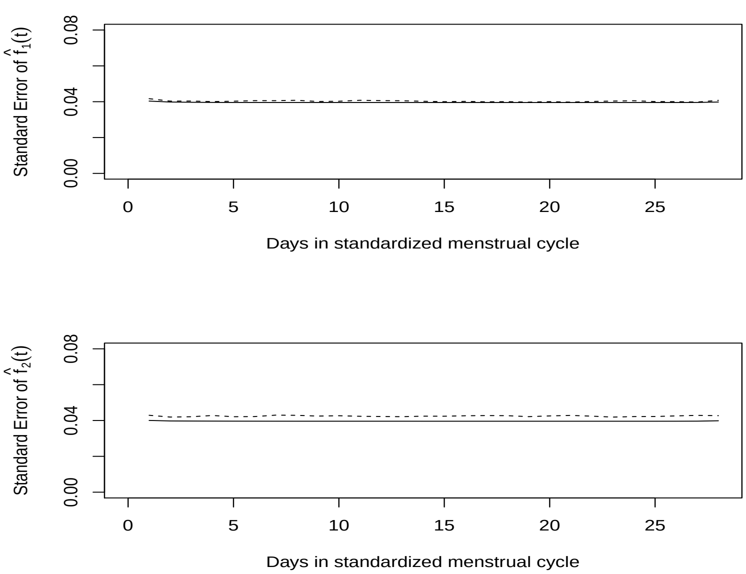

Fig. 2. Pointwise empirical (dashes) and frequentist (solid) standard errors of the estimated nonparametric functions $\hat{f}_1$ and $\hat{f}_2$ based on 500 simulation replications.

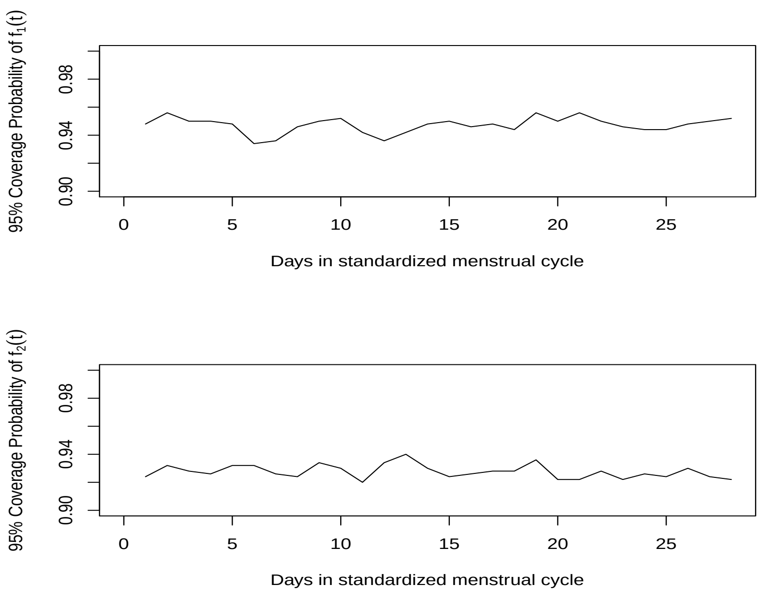

Fig. 3. A graph showing the estimated 95% coverage probabilities of the true nonparametric functions f_1 and f_2 based on 500 simulation replications.

Figure 3 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 conducted simulation studies when the Gaussian fields are incorrectly specified and studied the effect of this misspecification. Specifically, we used 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 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 is 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.

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 Gold et al. (1995bs). Of these, 338 women collected daily urine samples for at least one complete menstrual

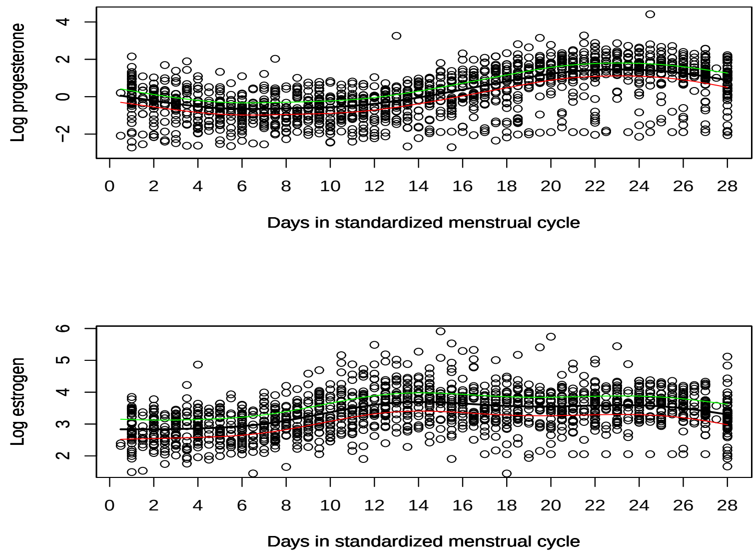

Fig. 4. Plots of log progesterone and log estrogen levels against days in a standardized menstrual cycle, superimposed by estimated population mean curve $\hat{f}_1$ and $\hat{f}_2$.

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 Gold et al. (1995bs) and Gold et al. (1995a).

We are interested in modelling the mean curve for women's daily urinary estrogen (EIC) and progesterone (PdG) metabolite profiles, and their relationships to demographic and lifestyle factors over a 28-day reference menstrual cycle. Also, we are interested in the potential association between these two hormones; thus, we jointly model these two responses. 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 responses.

Figure 4 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.

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

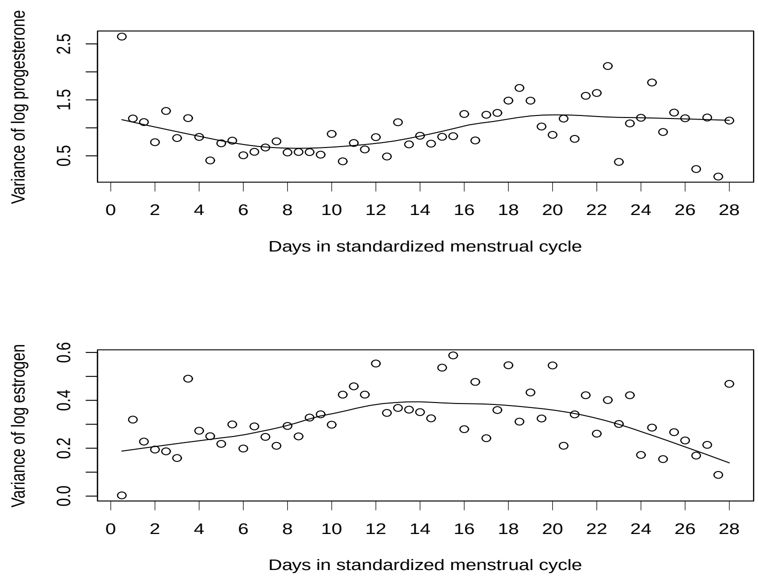

Fig. 5. Plots of empirical sample variance of log progesterone and log estrogen levels at each distinct time points in a standardized menstrual cycle.

semiparametric stochastic mixed model:

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

where b_{1i} and b_{2i} are the random intercepts that are correlated between the two hormone response but independent between subjects following a bivariate normal distribution with mean zero and variance $\begin{pmatrix} \phi_1 & \phi_3 \\ \phi_3 & \phi_2 \end{pmatrix}$; the U_i are mean 0 bivariate Gaussian field modeling serial correlation, and the ϵ_{1ij} and ϵ_{2ij} are independent mean zero measurement errors following bivariate normal with variance $\begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$. Covariates underWeight and overWeight are indicator variables. 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 curve for women of 33 years old with normal weight.

Table 2 records the results of estimates of regression coefficients, smoothing parameters and variance components. It comes as no surprise that the estimate of age has negative impact on both responses and almost all overweight or underweight also affect both responses negatively; however, noting the associated standard errors of each estimate, possibly larger than might be expected based on the relatively small sample size of this demonstration analysis, these effects are not statistically meaningful. Figure 4 superimposed the estimates of nonparametric function $\hat{f}_1$ and $\hat{f}_2$ and their 95% confidence interval, respectively, which accurately captured the underlying trends of the bivariate longitudinal responses.

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

Model parameters	Parameter estimate	SE
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	
ϕ_2	-0.2755	
ϕ_3	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_{20}	0.5172	
a_{20}	-0.0800	

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 and the Gaussian field allows for additional flexibility in within-subject correlation structure. Simulation results show that inference procedure performs well in all estimation results.

The model we proposed can also be readily extended to multivariate cyclic 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 much 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 estimates of parameter of interest. However, if we were interested in the underlying biological process, a deeper understanding of the choice in the Gaussian field is needed. In spite of further work to consider, this is a flexible and informative method for modeling bivariate cyclic longitudinal response data and we look forward to further extensions of this work in the above and possibly other directions.

Appendix

Technical details of Section 3·2

Proof of (6). Taking derivative of the log-likelihood function (5) with respect to $\boldsymbol{\beta}$, $\boldsymbol{f}_1$, and $\boldsymbol{f}_1$, we have

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

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

$$\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{N}_1\hat{\boldsymbol{f}}_1+\boldsymbol{N}_2\hat{\boldsymbol{f}}_2) = \boldsymbol{X}^T\boldsymbol{WY} \tag{B1}$$
$$\boldsymbol{N}_1^T\boldsymbol{W}(\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{N}_1\hat{\boldsymbol{f}}_1+\boldsymbol{N}_2\hat{\boldsymbol{f}}_2)+\lambda_1\boldsymbol{K}\hat{\boldsymbol{f}}_1 = \boldsymbol{N}_1^T\boldsymbol{WY} \tag{B2}$$
$$\boldsymbol{N}_2^T\boldsymbol{W}(\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{N}_1\hat{\boldsymbol{f}}_1+\boldsymbol{N}_2\hat{\boldsymbol{f}}_2)+\lambda_2\boldsymbol{K}\hat{\boldsymbol{f}}_2 = \boldsymbol{N}_2^T\boldsymbol{WY}, \tag{B3}$$

which can be rewritten as (6). □

Proof of (7), (8) and (9). From equations (B1), (B2) and (B3), we can reexpress the parameter estimators

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{WX})^{-1}\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{N}_1\hat{\boldsymbol{f}}_1-\boldsymbol{N}_2\hat{\boldsymbol{f}}_2)$$
$$\hat{\boldsymbol{f}}_1 = (\boldsymbol{N}_1^T\boldsymbol{WN}_1+\lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{N}_2\hat{\boldsymbol{f}}_2) \tag{B4}$$
$$\hat{\boldsymbol{f}}_2 = (\boldsymbol{N}_2^T\boldsymbol{WN}_2+\lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{N}_1\hat{\boldsymbol{f}}_1). \tag{B5}$$

To solve explicitly for $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{f}}_1$ and $\hat{\boldsymbol{f}}_2$, we first plug $\hat{\boldsymbol{f}}_1$ (B4) into equations (B1) and (B3), rearrange and obtain

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

and

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

respectively; which can be rewritten as

$$\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{N}_2\hat{\boldsymbol{f}}_2 = \boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{Y} \tag{B6}$$

and

$$\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{X}\hat{\boldsymbol{\beta}}+(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2+\boldsymbol{\lambda}_2\boldsymbol{K})\hat{\boldsymbol{f}}_2 = \boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{Y} \tag{B7}$$

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

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1(\boldsymbol{Y}-\boldsymbol{N}_2\hat{\boldsymbol{f}}_2), \tag{B8}$$

and

$$\hat{\boldsymbol{f}}_2 = (\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2+\lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_1(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}) \tag{B9}$$

respectively. Then plugging (B8) into (B7) and (B9) into (B6) 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

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

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 (B5) into equation (B2) and obtain

$$\boldsymbol{N}_1^T\boldsymbol{W}_2\boldsymbol{X}\hat{\boldsymbol{\beta}} + (\boldsymbol{N}_1^T\boldsymbol{W}_2\boldsymbol{N}_1 + \boldsymbol{\lambda}_1\boldsymbol{K})\hat{\boldsymbol{f}}_1 = \boldsymbol{N}_1^T\boldsymbol{W}_2\boldsymbol{Y}, \tag{B10}$$

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 (B10), 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 $\boldsymbol{W}_{f_1} = \boldsymbol{W}_2 - \boldsymbol{W}_2\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_2$. □

Technical details of Section 3·3

Proof of (12) and (13). For regression coefficient estimator $\hat{\boldsymbol{\beta}}$,

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

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

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

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

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

□

Remark B1. The bias of nonparametric function estimator $\hat{\boldsymbol{f}}_2$ in (14) can be derived similarly as that of $\hat{\boldsymbol{f}}_1$.

Lemma B1. 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

$$\boldsymbol{W}_x \to \boldsymbol{W}_1 - \boldsymbol{W}_1\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_1, \tag{B11}$$

where

$$\boldsymbol{W}_1 \to \boldsymbol{W} - \boldsymbol{W}\boldsymbol{N}_1(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\boldsymbol{W} \tag{B12}$$

Plugging $\boldsymbol{W}_x$ in (B11) into bias of $\hat{\boldsymbol{\beta}}$ in (12), we have

$$\begin{aligned}
&\mathrm{E}(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta}\\
&= (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\left[\boldsymbol{W}_1 - \boldsymbol{W}_1\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_1\right](\boldsymbol{N}_1\boldsymbol{f}_1 + \boldsymbol{N}_2\boldsymbol{f}_2)\\
&= (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1(\boldsymbol{N}_1\boldsymbol{f}_1 + \boldsymbol{N}_2\boldsymbol{f}_2)\\
&\qquad - (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_1(\boldsymbol{N}_1\boldsymbol{f}_1 + \boldsymbol{N}_2\boldsymbol{f}_2)\\
&= (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{N}_1\boldsymbol{f}_1\\
&\quad + (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{N}_2\left[\boldsymbol{f}_2 - (\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2\boldsymbol{f}_2\right]\\
&\quad - (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_1\boldsymbol{f}_1\\
&= (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{N}_1\boldsymbol{f}_1 - (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_1\boldsymbol{f}_1
\end{aligned} \tag{B13}$$

Further, plugging $\boldsymbol{W}_1$ in (B12) into(B13), 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\\
&\quad - (\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\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\\
&\quad - (\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\\
&\quad + (\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\\
&\quad + (\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\\
&\quad - (\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\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 $\hat{\boldsymbol{f}}_{\boldsymbol{1}}$

$$\mathrm{E}(\hat{\boldsymbol{f}}_{\boldsymbol{1}}) - \boldsymbol{f}_{\boldsymbol{1}} \to (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_2\boldsymbol{f}_2 \tag{B14}$$

where

$$\boldsymbol{W}_{f_1} = \boldsymbol{W}_2 - \boldsymbol{W}_2\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_2 \tag{B15}$$

and

$$\boldsymbol{W}_2 \to \boldsymbol{W} - \boldsymbol{W}\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W} \tag{B16}$$

Plugging $\boldsymbol{W}_{f_1}$ in (B15) into bias of $\hat{\boldsymbol{f}}_1$ in (B14) and plugging $\boldsymbol{W}_2$ in (B16) into $\boldsymbol{W}_{f_1}$ in (B15), we have

$$
\begin{aligned}
& \mathrm{E}(\hat{\boldsymbol{f}}_1) - \boldsymbol{f}_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 \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 \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 B2. The covariance result for the random effect in (15) is non-trivial and long, and is available upon request.

References

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

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. (1995). Prospectively assessed menstrual cycle characteristics in female wafer-fabrication and nonfabrication semiconductor employees. Am. J. Industrial Medicine 28, 799–815.

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

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

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

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

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.

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

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.

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.

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.
Wang, Y., Guo, W. & Brown, M. B. (2000). Spline smoothing for bivariate data with applications to association between hormones. Statistica Sinica. 377–497.
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.
Wood, S. (2006). Generalized Additive Models: An Introduction with R. Boca Raton, Florida: Chapman and Hall/CRC.
Zhang, D., Lin, X., Raz, J. & Sowers, M. (1998). Semiparametric stochastic mixed models for longitudinal data. J. Am. Statist. Assoc. 93, 710–719.
Zhang, D., Lin, X. & Sowers, M. (2000). Semiparametric regression for periodic longitudinal hormone data from multiple menstrual cycles. Biometrics 56, 31–39.

[Received 2 *January* 2017. Editorial decision on 1 *April* 2017]