Biometrika (2017), **103**, 1, *pp*. 1–13
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 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. In addition, 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 smoothing spline. The smoothing parameters and all variance components are estimated simultaneously using restricted maximum likelihood. We investigate the proposed methodology through simulation. We also illustrate the model by analyzing bivariate longitudinal female hormone data collected daily over multiple consecutive menstrual cycles.

Some key words: Address; Bivariate longitudinal; Cyclic; Gaussian field; Penalized likelihood; Semiparametric.

1. INTRODUCTION

Longitudinal data analysis has wide applications in areas such as medicine, agriculture and so on. One distinctive feature of longitudinal data is that measurements of each subject are collected repeatedly over time; thus, each measurement for the same subject is no longer independent. This induces a correlation structure that needs to be modeled carefully. Many methods have been developed over the years to accommodate this additional structure.

Brumback and Rice **?** used natural cubic splines to model the mean structure of the linear mixed model. They extended the traditional LME model to generalized smoothing spline models for samples of curves stratified by nested and crossed factors and specified the design matrices associated with fixed effects and random effects by bases of functions, as opposed to the usual known covariates matrices. Verbyla et al **?** advocated a similar approach as Brumback & Rice **?**, where data-based determination of the smoothing parameters was advocated in the paper, yet their model specification is slightly different; and the techniques were applied to the analysis of designed experiments.

In addition to modeling the mean structure of using a smoothing spline, some efforts were geared toward modeling complicated within-subject covariance. Taylor et al **?** used a particular stochastic process to model the data in addition to the usual random effect term which induces within-subject correlation. Zhang et al **?** combine both the smoothing spline to measure mean structure and various stationary and nonstationary stochastic processes to model serial correlation into one cohesive model. To further model cyclic responses, Zhang et al **?** extended their previous work and proposed a semiparametric stochastic mixed model for periodic longitudinal data. They used parametric functions to model the covariate effects and a periodic smooth nonparametric function to model the underlying complex periodic time course. The within-subject covariance is modeled using a random intercept and a stochastic process with periodic variance function. Instead of cubic smoothing spline, Welham et al **?** modelled cyclic longitudinal data using mixed model L-splines. Meyer et al **?** proposed a functional data analysis approach to model cyclic

data. Last but not least, Wood **?** modified penalized cubic regression spline to model a cyclic smooth function.

All of the approaches mentioned above are to be applied on univariate longitudinal responses. A growing number of data require techniques to model bivariate, and in more generality, multivariate responses. Liu et al **?** extended the univariate state space model in time series analysis, and proposed a bivariate hierarchical state space model to bivariate longitudinal responses. Each response is modelled by a hierarchical state space model, with both population-average and subject-specific components. The bivariate model is constructed by linking the univariate models based on the hypothesized relationship. Sy, Taylor, and Cumberland **?** employed multivariate stochastic processes to jointly model bivariate longitudinal data.

We extended Zhang et al **? ?** 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. In addition, the between-subject and within-subject correlations are modeled using separate but correlated random effects and a bivariate Gaussian random field, respectively. 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 MODEL

We propose a semiparametric stochastic bivariate mixed model, where the joint model assumes a semiparametric mixed model for each outcome. The univariate models for each outcome are connected through the specification of the correlation structure for the random effects.

2·1. *General bivariate model with joint distribution of random effects*

Denote $\{Y_{1ij}, Y_{2ij}\}$ to be the bivariate response for the ith subject at time point j, $i = 1, \ldots, m$ and $j = 1, \ldots, n_i$. The bivariate model is

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

where $(\boldsymbol{X}_{\mathbf{1}ij}, \boldsymbol{X}_{\mathbf{2}ij})$ are known covariates associated with the fixed effects; $(\boldsymbol{Z}_{\mathbf{1}ij}, \boldsymbol{Z}_{\mathbf{2}ij})$ are known covariates associated with the random effects; $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are $p_1 \times 1$ and $p_2 \times 1$ vectors of regression coefficients, containing the fixed effects, respectively; $\boldsymbol{b}_{\mathbf{1}i}$ and $\boldsymbol{b}_{\mathbf{2}i}$ are $q_1 \times 1$ and $q_2 \times 1$ vectors of random effects, respectively; $f_1(t)$ and $f_2(t)$ are twice-differentiable periodic smooth functions of time with periods T_1 and T_2, respectively; $\{(U_{1i}(t_{ij}), U_{2i}(t_{ij})), t_{ij} \in \{t_{i1}, \ldots, t_{in_i}\}, i = 1, \ldots, m, j = 1, \ldots, n_i\}$ are mean zero bivariate Gaussian field with covariance matrix

$$\begin{aligned} \boldsymbol{C}_i(s,t) &= \begin{pmatrix} E[U_{i1}(s)U_{i1}(t)] & E[U_{i1}(s)U_{i2}(t)] \\ E[U_{i2}(s)U_{i1}(t)] & E[U_{i2}(s)U_{i2}(t)] \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{\xi_1(s)\xi_1(t)}\eta_1(\rho_1; s, t) & \sqrt{\xi_1(s)\xi_2(t)}\eta_3(\rho_3; s, t) \\ \sqrt{\xi_2(s)\xi_1(t)}\eta_3(\rho_3; t, s) & \sqrt{\xi_2(s)\xi_2(t)}\eta_2(\rho_2; s, t) \end{pmatrix} \end{aligned}$$

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 $\rho_1 \in [0, 1]$, $\rho_2 \in [0, 1]$ and $\rho_3 \in [0, 1]$ are correlation parameters; and the measurement errors $(\epsilon_{1ij}, \epsilon_{2ij})^T$ are bivariate normal

$$\begin{pmatrix} \epsilon_{1ij} \\ \epsilon_{2ij} \end{pmatrix} \sim N_2 \left(\mathbf{0}, \begin{pmatrix} \sigma_1^2 & \sigma_{12}^2 \\ \sigma_{12}^2 & \sigma_2^2 \end{pmatrix} \right).$$

We assume that $\boldsymbol{b}_{ki}, k = 1, 2$ to be $(q_1 + q_2)$-dimensional normal with mean zero and covariance matrix $\boldsymbol{D}(\phi)$. These random effects, $\boldsymbol{b}_{\mathbf{1}i}$ and $\boldsymbol{b}_{\mathbf{2}i}$, are assumed to be separate but correlated. Further, we assume that the random effects, the stochastic process and the measurement error to be mutually independent.

Denote

$$\boldsymbol{Y}_{ij} = \begin{pmatrix} Y_{1ij} \\ Y_{2ij} \end{pmatrix} \in \mathbb{R}^2$$

to be the response vector;

$$\boldsymbol{X_{ij}} = \begin{pmatrix} \boldsymbol{X_{1ij}^T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X_{2ij}^T} \end{pmatrix} \in \mathbb{R}^{(p_1+p_2)\times 2}, \quad \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} \in \mathbb{R}^{(p_1+p_2)},$$

to be the matrix of known covariates and the vector of regression coefficients respectively;

$$\boldsymbol{Z_{ij}} = \begin{pmatrix} \boldsymbol{Z_{1ij}^T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z_{2ij}^T} \end{pmatrix} \in \mathbb{R}^{(q_1+q_2)\times 2}, \quad \boldsymbol{b}_i = \begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix} \in \mathbb{R}^{(q_1+q_2)},$$

to be the matrix of known covariates associated with the random effects and the vector of random effects respectively; and finally

$$\boldsymbol{f}(t_{ij}) = \begin{pmatrix} f_1(t_{ij}) \\ f_2(t_{ij}) \end{pmatrix} \in \mathbb{R}^2, \quad \boldsymbol{U}_i(t_{ij}) = \begin{pmatrix} U_{1i}(t_{ij}) \\ U_{2i}(t_{ij}) \end{pmatrix} \in \mathbb{R}^2, \quad \boldsymbol{\epsilon}_{ij} = \begin{pmatrix} \epsilon_{1ij} \\ \epsilon_{2ij} \end{pmatrix} \in \mathbb{R}^2,$$

to be the vectors of the smooth function, the stochastic process, and the measurement error of. 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 **? ?**, 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, we propose to include various stationary and nonstationary Gaussian field 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 **?**); an integrated Wiener process and so on. One particular Gaussian process/field worthy of mentioning is Ornstein-Uhlenbeck (OU) process **?** which has a correlation function that decays exponentially over time 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 inference from the model (2), we will write the model in matrix form - first, in subject level; then, over all subjects. Denote

$$\boldsymbol{Y}_i = \begin{pmatrix} \boldsymbol{Y}_{i1} \\ \vdots \\ \boldsymbol{Y}_{in_i} \end{pmatrix} \in \mathbb{R}^{2n_i},$$

to be the response vector;

$$\boldsymbol{X}_i = \begin{pmatrix} \boldsymbol{X}_{i1}^T \\ \vdots \\ \boldsymbol{X}_{in_i}^T \end{pmatrix} \in \mathbb{R}^{2n_i \times (p_1+p_2)}, \quad \boldsymbol{Z}_i = \begin{pmatrix} \boldsymbol{Z}_{i1}^T \\ \vdots \\ \boldsymbol{Z}_{in_i}^T \end{pmatrix} \in \mathbb{R}^{2n_i \times (q_1+q_2)},$$

to be the corresponding covariate matrix associate with the fixed effects and the random effects respectively; and

$$\boldsymbol{U}_i = \begin{pmatrix} \boldsymbol{U}_i(t_{i1}) \\ \vdots \\ \boldsymbol{U}_i(t_{in_i}) \end{pmatrix} \in \mathbb{R}^{2n_i}, \quad \boldsymbol{\epsilon}_i = \begin{pmatrix} \boldsymbol{\epsilon}_{i1} \\ \vdots \\ \boldsymbol{\epsilon}_{in_i} \end{pmatrix} \in \mathbb{R}^{2n_i},$$

to be the vectors of stochastic process and measurement errors. 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 $\tilde{\boldsymbol{N}}_{\mathbf{1}\boldsymbol{i}}$ 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

$$\tilde{\boldsymbol{N}}_{\mathbf{1}\boldsymbol{i}}[j, \ell] = \begin{cases} 1 & \text{if } t_{1ij}' = t_{1\ell}' \\ 0 & \text{otherwise,} \end{cases}$$

where $\tilde{\boldsymbol{N}}_{\mathbf{1}\boldsymbol{i}}[j, \ell]$ denote the $(j, \ell)^{\text{th}}$ entry of matrix $\tilde{\boldsymbol{N}}_{\mathbf{1}\boldsymbol{i}}$ for $j = 1, \ldots, n_i$ and $\ell = 1, \ldots, r_1$. Similarly, let $\tilde{\boldsymbol{N}}_{\mathbf{2}\boldsymbol{i}}$ 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

$$\tilde{\boldsymbol{N}}_{\mathbf{2}\boldsymbol{i}}[j, \ell] = \begin{cases} 1 & \text{if } t_{2ij}' = t_{2\ell}' \\ 0 & \text{otherwise,} \end{cases}$$

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

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

where

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

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

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

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

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

for subject i, where

$$\boldsymbol{f}_1 = \begin{pmatrix} f_1(t'_{11}) \\ \vdots \\ f_1(t'_{1r_1}) \end{pmatrix} \in \mathbb{R}^{r_1}, \quad \boldsymbol{f}_2 = \begin{pmatrix} f_2(t'_{21}) \\ \vdots \\ f_2(t'_{2r_2}) \end{pmatrix} \in \mathbb{R}^{r_2}.$$

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 $\boldsymbol{D}(\boldsymbol{\phi}) = \mathrm{diag}(\boldsymbol{D}, \dots, \boldsymbol{D})$; $\boldsymbol{\Gamma}(\boldsymbol{\xi}, \rho) = \mathrm{diag}(\boldsymbol{\Gamma}_1(\boldsymbol{t}_1, \boldsymbol{t}_1), \dots, \boldsymbol{\Gamma}_m(\boldsymbol{t}_m, \boldsymbol{t}_m))$ and the $(k, k')^{\mathrm{th}}$ 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*

Let $Y = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{N}_1\boldsymbol{f}_1 + \boldsymbol{N}_2\boldsymbol{f}_2 + \boldsymbol{\epsilon}^*$, where

$$\boldsymbol{\epsilon}^* = \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{U} + \boldsymbol{\epsilon} = \begin{pmatrix} \boldsymbol{Z} & \boldsymbol{I}_{2n\times 2n} & \boldsymbol{I}_{2n\times 2n} \end{pmatrix} \begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{U} \\ \boldsymbol{\epsilon} \end{pmatrix}$$

Then $\boldsymbol{\epsilon}^* \sim N_{2n}(\boldsymbol{0}, \boldsymbol{V})$ where

$$\boldsymbol{V} = A\mathrm{cov} \begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{U} \\ \boldsymbol{\epsilon} \end{pmatrix} A^T = \begin{pmatrix} \boldsymbol{Z} & \boldsymbol{I}_{2n\times 2n} & \boldsymbol{I}_{2n\times 2n} \end{pmatrix} \begin{pmatrix} \boldsymbol{D}(\boldsymbol{\phi}) & \boldsymbol{0} & 0 \\ \boldsymbol{0} & \boldsymbol{\Gamma}(\boldsymbol{\xi}, \rho) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{\Sigma}(\boldsymbol{\sigma}) \end{pmatrix} \begin{pmatrix} \boldsymbol{Z}^T \\ \boldsymbol{I}_{2n\times 2n} \\ \boldsymbol{I}_{2n\times 2n} \end{pmatrix}$$

$$= \begin{pmatrix} \boldsymbol{Z}\boldsymbol{D}(\boldsymbol{\phi}) & \boldsymbol{\Gamma}(\boldsymbol{\xi}, \rho) & \boldsymbol{\Sigma}(\boldsymbol{\sigma}) \end{pmatrix} \begin{pmatrix} \boldsymbol{Z}^T \\ \boldsymbol{I}_{2n\times 2n} \\ \boldsymbol{I}_{2n\times 2n} \end{pmatrix} = \boldsymbol{Z}\boldsymbol{D}\boldsymbol{Z}^T + \boldsymbol{\Gamma} + \boldsymbol{\Sigma}$$

Therefore the proposed model (3) also implies the *marginal model*

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

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

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

We estimate the parameters $\boldsymbol{\beta}$, $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ by maximizing the penalized likelihood **?**:

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

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

Given fixed smoothing parameters and variance parameters, differentiation of (4) with respect to $\boldsymbol{\beta}$, $\boldsymbol{f}_1$, $\boldsymbol{f}_2$ gives the estimators $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{f}}_{\boldsymbol{1}}, \hat{\boldsymbol{f}}_{\boldsymbol{2}})$ that solves

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

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 $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{f}}_{\boldsymbol{1}}$ and $\hat{\boldsymbol{f}}_{\boldsymbol{2}}$

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{Y} \tag{5}$$

$$\hat{\boldsymbol{f}}_{\boldsymbol{1}} = (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{Y} \tag{6}$$

$$\hat{\boldsymbol{f}}_{\boldsymbol{2}} = (\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{Y}, \tag{7}$$

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 *two-level hierarchical model*

$$\boldsymbol{Y}|\boldsymbol{b}, \boldsymbol{U} \sim N_{2n}\left(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{N}_1\boldsymbol{f}_1 + \boldsymbol{N}_2\boldsymbol{f}_2 + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{U}, \boldsymbol{\Sigma}\right) \tag{8}$$

$$\boldsymbol{b} \sim N_{2m}(\boldsymbol{0}, \boldsymbol{D}) \tag{9}$$

$$\boldsymbol{U} \sim N_{2n}(\boldsymbol{0}, \boldsymbol{\Gamma}),$$

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

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

Therefore,

$$E(\boldsymbol{b}|\boldsymbol{Y}) = \boldsymbol{0} + \boldsymbol{D}\boldsymbol{Z}^T\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1\hat{\boldsymbol{f}}_{\boldsymbol{1}} - \boldsymbol{N}_2\hat{\boldsymbol{f}}_{\boldsymbol{2}}) = \boldsymbol{D}\boldsymbol{Z}^T\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1\hat{\boldsymbol{f}}_{\boldsymbol{1}} - \boldsymbol{N}_2\hat{\boldsymbol{f}}_{\boldsymbol{2}})$$

and the estimator or predictor for subject-specific random effects $\boldsymbol{b_i}$ is

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

Similarly, the estimator or predictor for the subject-specific Gaussian field $\boldsymbol{U}_i(\boldsymbol{s}_i)$ is

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

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

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

From closed-form solutions of estimators from equation (5) (6) and (7) 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

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

$$E(\hat{\boldsymbol{f_1}}) - \boldsymbol{f}_1 = (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_2\boldsymbol{f}_2 - \lambda_1\boldsymbol{K}\boldsymbol{f}_1),$$

and

$$E(\hat{\boldsymbol{f_2}}) - \boldsymbol{f}_2 = (\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_1\boldsymbol{f}_1 - \lambda_2\boldsymbol{K}\boldsymbol{f}_2).$$

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 $\lambda_1 \to 0$ and $\lambda_2 \to 0$.

For covariances, simple calculation using (5) (6) and (7) give 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

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

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 $\boldsymbol{\chi} = \begin{pmatrix}\boldsymbol{X} & \boldsymbol{N}_1 & \boldsymbol{N}_2\end{pmatrix}$.

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

By Green (1997) **?**, we can write $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ by a one-to-one linear transformation as

$$\boldsymbol{f}_1 = \boldsymbol{T}_1\boldsymbol{\delta}_1 + \boldsymbol{B}_1\boldsymbol{a}_1$$
$$\boldsymbol{f}_2 = \boldsymbol{T}_2\boldsymbol{\delta}_2 + \boldsymbol{B}_2\boldsymbol{a}_2$$

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,

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{N}_1\boldsymbol{T}_1\boldsymbol{\delta}_1 + \boldsymbol{N}_1\boldsymbol{B}_1\boldsymbol{a}_1 + \boldsymbol{N}_2\boldsymbol{T}_2\boldsymbol{\delta}_2 + \boldsymbol{N}_2\boldsymbol{B}_2\boldsymbol{a}_2 + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{U} + \boldsymbol{\epsilon}, \tag{12}$$

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$(0, \tau_2\boldsymbol{I})$, 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 above modified linear mixed model (12), the REML log-likelihood of $(\tau_1, \tau_2, \boldsymbol{\theta})$ is

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

where $\boldsymbol{X}_* = [\boldsymbol{X}, \boldsymbol{N}_1\boldsymbol{T}_1, \boldsymbol{N}_2\boldsymbol{T}_2]$. Taking derivative with respect to τ_1, τ_1, and $\boldsymbol{\theta}$ and using the identity

$$\boldsymbol{V}_*^{-1}(\boldsymbol{Y} - \boldsymbol{X}_*\hat{\boldsymbol{\beta}}_*) = \boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1\hat{\boldsymbol{f}}_{\boldsymbol{1}} - \boldsymbol{N}_2\hat{\boldsymbol{f}}_{\boldsymbol{2}}),$$

the estimating equation for the smoothing parameters τ_1 τ_2 and variance components $\boldsymbol{\theta}$ can be obtained

$$\frac{\partial\ell_R}{\partial\tau_1} = -\frac{1}{2}\mathrm{tr}(\boldsymbol{P}_*\boldsymbol{B}_{1*}\boldsymbol{B}_{1*}^T) + \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1\hat{\boldsymbol{f}}_{\boldsymbol{1}} - \boldsymbol{N}_2\hat{\boldsymbol{f}}_{\boldsymbol{2}})^T\boldsymbol{V}^{-1}\boldsymbol{B}_{\boldsymbol{1*}}\boldsymbol{B}_{\boldsymbol{1*}}^T\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1\hat{\boldsymbol{f}}_{\boldsymbol{1}} - \boldsymbol{N}_2\hat{\boldsymbol{f}}_{\boldsymbol{2}}), \tag{13}$$

$$\frac{\partial\ell_R}{\partial\tau_2} = -\frac{1}{2}\mathrm{tr}(\boldsymbol{P}_*\boldsymbol{B}_{2*}\boldsymbol{B}_{2*}^T) + \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1\hat{\boldsymbol{f}}_{\boldsymbol{1}} - \boldsymbol{N}_2\hat{\boldsymbol{f}}_{\boldsymbol{2}})^T\boldsymbol{V}^{-1}\boldsymbol{B}_{\boldsymbol{2*}}\boldsymbol{B}_{\boldsymbol{2*}}^T\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1\hat{\boldsymbol{f}}_{\boldsymbol{1}} - \boldsymbol{N}_2\hat{\boldsymbol{f}}_{\boldsymbol{2}}), \tag{14}$$

and

$$\frac{\partial\ell_R}{\partial\theta_j} = -\frac{1}{2}\mathrm{tr}(\boldsymbol{P}_*\frac{\partial\boldsymbol{V}}{\partial\theta_j}) + \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1\hat{\boldsymbol{f}}_{\boldsymbol{1}} - \boldsymbol{N}_2\hat{\boldsymbol{f}}_{\boldsymbol{2}})^T\boldsymbol{V}^{-1}\frac{\partial\boldsymbol{V}}{\partial\theta_j}\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1\hat{\boldsymbol{f}}_{\boldsymbol{1}} - \boldsymbol{N}_2\hat{\boldsymbol{f}}_{\boldsymbol{2}}), \tag{15}$$

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 the smoothing parameters τ_1 τ_2 and variance components $\boldsymbol{\theta}$ can be estimated using Fisher-scoring algorithm, where the Fisher information matrix is obtained using (13), (14) and (15),

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

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. Simulation Study

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

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

where b_{1i} and b_{2i} are independent but correlated random intercepts following a bivariate normal distribution:

$$\begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix} \sim \boldsymbol{N}_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \phi_1 & \phi_2 \\ \phi_2 & \phi_3 \end{pmatrix} \right);$$

U_{1i} and U_{2i} are simulated from mean 0 bivariate NOU fields modeling serial correlation, with variance function 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 $corr(U_{1i}(t), U_{1i}(s)) = \rho_1^{|s-t|}$ $corr(U_{2i}(t), U_{2i}(s)) = \rho_2^{|s-t|}$, 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 0 bivariate normal distribution

$$\begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \sim \boldsymbol{N}_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \right).$$

Further, the nonparametric functions are generated from

$$f_1(t) = 5 \sin\left(\frac{2\pi}{28}\right) t, \quad f_2(t) = 3 \cos\left(\frac{2\pi}{28}\right) t$$

with periods to be 28 days for both responses.

Table 1 recorded 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. 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 minimal for $\hat{f}_2$ which centers around 0, whereas for $\hat{f}_1$ the bias stands a little above 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.

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 96% and 93% for $\hat{f}_1$ and $\hat{f}_2$Our simulate results are largely consistent with those of Zhang et al. (1998) for univariate independent data.

4·1. *Misspecification of Gaussian Fields*

We further conducted simulation studies when the Gaussian fields are incorrectly specified and studied the effect of misspecification of the Gaussian fields. The data are generated as mentioned above using NOU bivariate Gaussian fields, whereas the data are analyzed using OU and Wiener bivariate Gaussian fields and the results are as followed.

We see that OU is relatively robust even if the true Gaussian field is

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

Model parameters	True Value	Parameter estimate	Bias	SE
β_1	1.0000	0.9988	0.0012	0.0012
β_2	0.7500	0.7501	0.0001	0.0013
τ_1	1.0000	0.5577	0.4423	0.0066
τ_2	1.0000	0.5951	0.4049	0.0067
ϕ_1	1.0000	0.9855	0.0145	0.0079
ϕ_2	-0.5000	-0.4992	-0.0016	0.0062
ϕ_3	1.0000	0.9901	0.0099	0.0079
σ_1^2	1.0000	0.9986	0.0014	0.0018
σ_2^2	1.0000	0.9992	0.0008	0.0018
ρ_1	0.1054	0.0822	0.2201	0.0018
a_{10}	-5.0000	-5.0941	-0.0188	0.0555
a_{11}	1.5000	1.5278	0.0185	0.0149
a_{12}	-0.1000	-0.1020	-0.0200	0.0010
ρ_2	0.1054	0.0810	0.2315	0.0017
a_{20}	-5.0000	-5.1081	-0.0216	0.0536
a_{21}	1.5000	1.5317	0.0211	0.0144
a_{22}	-0.1000	-0.1023	-0.0230	0.0009

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

5. SWAN DATA ANALYSIS

The model we proposed was motivated by a dataset from the Study of Women's Health Across the Nation (SWAN), a cohort study of $3,302$ middle-aged women in the United States, collected from seven sites. The women enrolled in the study come from different ethnic backgrounds. 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 **?**. Of these, 338 women collected daily urine samples for at least one

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

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 have been described in detail previously **? ?**.

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 their potential interactions between these two hormones. Thus, we will model jointly for these two responses. For demonstration purposes, we randomly select 50 study participants from the study, with a total of 2714 observations for both responses. Each woman contributes from 16 to 41 observations over a menstrual cycle, resulting an average of 27 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. To make the normality assumption more appropriate, the log transformation was used for both responses.

Figure (**??**) plots the log-transformed progesterone and estrogen levels during a standardized menstrual cycle. Figure (**??**) plots their empirical sample variances calculated at each distinct time points.

Denote $\{(Y_{1ij}, Y_{2ij})\}$ 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:

$$\begin{aligned} Y_{1ij} &= \text{age}_i^T \beta_{11} + \text{BMI}_i^T \beta_{12} + f_1(t_{ij}) + b_{1i} + U_i(t_j) + \epsilon_{1ij} \\ Y_{2ij} &= \text{age}_i^T \beta_{21} + \text{BMI}_i^T \beta_{22} + f_2(t_{ij}) + b_{2i} + U_i(t_j) + \epsilon_{2ij} \\ & i = 1, \ldots, 50; j = 1, \ldots, n_i; t_{ij} \in \{0.5, 1.0, \ldots, 28\} \end{aligned}$$

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

Fig. 3. A graph showing the truth (dot-dash), an estimate (dashes), another estimate (solid), and 95% pointwise confidence limits (small dashes).

$N_2\left(\mathbf{0}, \begin{pmatrix}\phi_1 & \phi_3 \\ \phi_3 & \phi_2\end{pmatrix}\right)$; 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 $\begin{pmatrix}\sigma_1^2 & 0 \\ 0 & \sigma_2^2\end{pmatrix}$.

6. DISCUSSION

This is the concluding part of the paper. It is only needed if it contains new material. It should not repeat the summary or reiterate the contents of the paper.

ACKNOWLEDGEMENT

Acknowledgements should appear after the body of the paper but before any appendices and be as brief as possible subject to politeness. Information, such as contract numbers, of no interest to readers, must be excluded.

SUPPLEMENTARY MATERIAL

Further material such as technical details, extended proofs, code, or additional simulations, figures and examples may appear online, and should be briefly mentioned as Supplementary Material where appropriate. Please submit any such content as a PDF file along with your paper, entitled 'Supplementary material for Title-of-paper'. After the acknowledgements, include a section 'Supplementary material' in your paper, with the sentence 'Supplementary material available at *Biometrika* online includes . . .', giving a brief indication of what is available. However it should be possible to read and understand the paper without reading the supplementary material.

Further instructions will be given when a paper is accepted.

APPENDIX 1

General

Any appendices appear after the acknowledgement but before the references, and have titles. If there is more than one appendix, then they are numbered, as here.

THEOREM A1. *This is a rather dull theorem:*

$$a + b = b + a; \tag{A1}$$

a little equation like this should only be displayed and labelled if it is referred to elsewhere.

APPENDIX 2

Technical details

Often the appendices contain technical details of the main results.

THEOREM B1. *This is another theorem.*

APPENDIX 3

Often the appendices contain technical details of the main results:

$$a + b = c. \tag{C1}$$

Remark C1. This is a remark concerning equations (A1) and (C1).

REFERENCES

BERRENDERO, J. R., C. A. . T. J. L. (2015). On the use of reproducing kernel hilbert spaces in functional classification. *arXiv*: 1507.04398v3.

CLEVELAND, W. S. (1993). *Vizualizing Data*. Summit: Hobart Press.

CLEVELAND, W. S. (1994). *The Elements of Graphing Data*. Summit: Hobart Press, revised ed.

COX, D. R. (1972). Regression models and life tables (with Discussion). *J. R. Statist. Soc.* B **34**, 187–220.

FAN, J. & PENG, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *Ann. Statist.* **32**, 928–61.

HEARD, N. A., HOLMES, C. C. & STEPHENS, D. A. (2006). A quantitative study of gene regulation involved in the immune response of Anopheline mosquitoes: An application of Bayesian hierarchical clustering of curves. *J. Am. Statist. Assoc.* **101**, 18–29.

R DEVELOPMENT CORE TEAM (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, http://www.R-project.org.

TUFTE, E. R. (1983). *The Visual Display of Quantitative Information*. Cheshire: Graphics Press.

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