Estimation and prediction methods for univariate and bivariate cyclic longitudinal data using a semiparametric stochastic mixed effects model

by

Kexin Ji

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics (Biostatistics)

Waterloo, Ontario, Canada, 2018

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Lang Wu
Professor, Dept. of Statistics, University of British Columbia

Supervisor(s): Joel A. Dubin
Professor, Dept. of Statistics and Actuarial Science,
University of Waterloo

Internal Member: Grace Yi
Professor, Dept. of Statistics and Actuarial Science,
University of Waterloo
Leilei Zeng
Professor, Dept. of Statistics and Actuarial Science,
University of Waterloo

Internal-External Member: Ashok Chaurasia
Professor, School of Public Health and Health Systems,
University of Waterloo

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

Abstract

In this thesis, I propose and consider inference for a semiparametric stochastic mixed model for bivariate longitudinal data; and provide a prediction procedure of a future cycle utilizing past cycle information. This thesis is built on the work of Zhang *et al* (1998) [45] and Zhang, Lin & Sowers (2000) [44]. However, the papers are missing big gaps in the theoretical results, are to be applied on univariate longitudinal data, and contain no coverage of prediction of future cycles. We fill in all the gaps in this thesis as well as consider real application of a dataset that contains bivariate longitudinal data. The proposed approach models the mean of outcome variables by parametric fixed effects and a smooth nonparametric function for the underlying time effects, and the relationship across the bivariate responses by a bivariate Gaussian random field and a joint distribution of random effects. The prediction approach is proposed from the frequentist prospective and a prediction density function with predictive intervals will be provided. Simulations studies are performed and a real application of a hormone dataset is considered.

Acknowledgements

I would like to express my deep gratitude towards my advisor, Prof. J. Dubin, for suggesting the problem to me and for his guidance, support and encouragements along the way! I would also like to thank all my committee members, Profs. Lang Wu, Grace Yi, Leilei Zeng and Ashok Chaurasia for their participations and time. I am grateful for all my professors who have taught me both at University of Toronto and University of Waterloo, who set examples for me and inspire me. Special thanks go to Prof. D.A.S Fraser at U of T for taking me on as his research assistant when I was a 3rd year undergraduate student and introduced statistical research to me; to Prof. Ellen Gold at UC-Davis and Prof. Wesley Johnson at UC-Irvine, for their allowance of the use of the hormonal dataset presented in this paper; and to Prof. George A. Heckman for allowing me to participate in a medical research project that partially funded me. I would like to acknowledge that my research was also partially supported by NSERC Discovery Grant (RGPIN-2014-05911), Queen Elizabeth II Graduate Scholarships in Science & Technology, and Department of Statistics at University of Waterloo among others. I treasure and am grateful for the friendship and support from my friends and colleagues at the department who shared my ups and downs during these eventful years. Lastly, I would like to express my utmost gratitude to my dear parents for their unwavering love and support.

To my parents.

Table of Contents

List of Tables

List of Figures

Chapter 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.

Many methods have been developed over the years to accommodate this additional structure, with the emphasis below on models that allow subject-specific predictions of longitudinal trajectories; marginal models will not be emphasized here.

1.1 Major approaches

1.1.1 Linear mixed effects model

The linear mixed effects (LME) model is one of the most popular approaches to modelling continuous longitudinal response data. This model was first proposed by Laird and Ware [19] in 1982 and thus is sometimes called the Laird and Ware Model.

Suppose that there are m subjects in a longitudinal dataset. Let $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^T$ be the set of n_i repeated measurements for the i^{th} subject, then the LME model for subject i is written as

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, m, \tag{1.1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is a vector of fixed effects, $\boldsymbol{b}_i = (b_{i1}, \ldots, b_{iq})^T$ is a vector of random effects assuming to have a multivariate normal distribution with mean zero and covariance matrix $\boldsymbol{D}$, $\boldsymbol{X}_i = (\boldsymbol{X}_{i1}^T, \ldots, \boldsymbol{X}_{in_i}^T)^T$ is an $(n_i \times p)$ matrix of covariates associated with the fixed effects, $\boldsymbol{Z}_i = (\boldsymbol{Z}_{i1}^T, \ldots, \boldsymbol{Z}_{in_i}^T)^T$ is an $(n_i \times q)$ matrix of covariates associated with the random effects, and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})^T$ is a vector of measurement errors assuming to have a multivariate normal distribution with mean zero and covariance matrix $\boldsymbol{\Sigma}_i$, independent of $\boldsymbol{b}_i$.

It is further assumed, initially, that conditional on the random effects, the components of the response $\boldsymbol{Y}_i$ are independent across different measurements for the i^{th} subject, which requires that $\boldsymbol{\Sigma}_i = \sigma^2\boldsymbol{I}_{n_i}$. This is referred to as the *conditional independence assumption*, with equal variance. This assumption implies the *conditional model*:

$$\boldsymbol{Y}_i|\boldsymbol{b}_i \sim N_{n_i}(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i, \boldsymbol{\Sigma}_i);$$

and under the model assumptions, it is straightforward to show

$$\boldsymbol{Y}_i \sim N_{n_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T + \boldsymbol{\Sigma}_i),$$

which is referred as the *marginal model* of $\boldsymbol{Y}_i$. Note that marginally, the Y_{ij} are not independent for a given subject i, with the correlation among the Y_{ij} for subject i being induced by the random effects $\boldsymbol{b}_i$.

The fixed effects are assumed to be the same for all individuals to which they apply in the population (e.g., split up by treatment group). The random effects, on the other hand, have the interpretation how the i^{th} subject effect deviates from those in the population. Under the conditional independence assumption, the introduction of the random effects $\boldsymbol{b}_i$ also induces correlation among the components of $\boldsymbol{Y}_i$, as noted above. The LME model does not require a balanced longitudinal design and each individual can have a unique sequence of measurement times. This makes the LME model well suited for modelling longitudinal data.

The LME model (1.1) can also be formulated to include all subjects into one compact model. Let $n = \sum_{i=1}^{m} n_i$, and denote the vectors

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1 \\ \vdots \\ \boldsymbol{Y}_m \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} \boldsymbol{b}_1 \\ \vdots \\ \boldsymbol{b}_m \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \vdots \\ \boldsymbol{\epsilon}_m \end{pmatrix},$$

of responses, random effects and measurement errors over all subjects with dimension $n \times 1$, $mq \times 1$ and $n \times 1$, respectively; and the design matrices are

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_m \end{pmatrix}, \quad \boldsymbol{Z} = \begin{pmatrix} \boldsymbol{Z}_1 & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{Z}_m \end{pmatrix},$$

of dimension $n \times p$ and $n \times mq$. Then the LME model in the matrix form over all subjects is

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\epsilon}, \tag{1.2}$$

with the distributional assumption

$$\begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{\epsilon} \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{Q} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma} \end{pmatrix} \right),$$

where

$$\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{D} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{D} \end{pmatrix}$$

of dimension $mq \times mq$, and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{\Sigma}_m \end{pmatrix} = \sigma^2 \boldsymbol{I}$$

with $\boldsymbol{I}$ denoting the identity matrix of dimension n.

1.1.2 Nonparametric regression models

Although the LME model is easily implemented and widely used in practice, some complex datasets require more flexible and sophisticated modelling techniques beyond the linear model. Specifically, smoothing techniques can be effectively applied in various scenarios when modelling longitudinal data. We will first review several common smoothing techniques.

Splines, cubic splines, and smoothing splines

A spline is essentially a piecewise polynomial whose different polynomials are joined together such that certain continuity properties are ensured. The points at which the polynomials join are known as the *knots* of the spline. The precise mathematical definition of splines is given in Appendix A.1.

A commonly used spline is called a *cubic spline*. A cubic spline g is a differentiable function defined on some interval $[a, b]$, with h distinct knots such that $a < \tau_1 < \tau_2 < \cdots < \tau_h < b$ and satisfies the following two conditions [17]. First, on each of the intervals $(a, \tau_1), (\tau_1, \tau_2), \ldots, (\tau_h, b)$, g is a cubic polynomial; second, the cubic spline g, its first derivative g' and its second derivative g'' are continuous at each knot $\tau_1, \ldots, \tau_h$, and hence on the whole of $[a, b]$. A cubic spline on an interval $[a, b]$ is said to be a *natural cubic spline* if its second and third derivatives are zero at a and b [17], which implies that g is linear on the two extreme intervals $[a, \tau_1]$ and $[\tau_h, b]$.

Natural cubic spline plays an important role in nonparametric regression. Consider the simplest nonparametric regression model in which y_i are observations with covariate values $t_k, k = 1, \ldots, v$,

$$y_k = g(t_k) + \epsilon_k, \tag{1.3}$$

where $g(t)$ is an unknown smooth function defined on $[a, b]$ and $\epsilon_k, k = 1, \ldots, v$ are distributed as independent and identically Normal with mean 0 and variance σ^2. Let $\mathcal{S}[a, b]$ be the space of functions on some interval $[a, b]$ that have two continuous derivatives[1].

[1] We call a function *smooth* if it is in $\mathcal{S}[a, b]$. [17].

Then under model (1.3), the *smoothing spline estimator* can be obtained by maximizing the following penalized log-likelihood

$$-\frac{1}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\left[-\sum_{i=1}^{v}\{y_k - g(t_k)\}^2 - \lambda\int_a^b \{g''(x)\}^2 dx\right];$$

which is equivalent to minimizing the the following penalized residual sum of squares

$$\sum_{k=1}^{v}\{y_k - g(t_k)\}^2 + \lambda\int_a^b \{g''(x)\}^2 dx, \tag{1.4}$$

where λ is called a smoothing parameter and governs the trade-off between smoothness and goodness-of-fit. As λ goes to infinity, the penalty term $\int_a^b \{g''(x)\}^2 dx$ will be forced to be very small, and thus the fit $\hat{g}$ will approach a linear regression fit; and as λ goes to zero, the main contribution will be the residual sum of squares, and thus the fit $\hat{g}$ will approach the interpolating curve. The resulting estimator is the natural cubic spline and it is the unique minimizer over all functions in $\mathcal{S}[a,b]$[2]. Note that *all* the observed covariate values $\{t_i\}$ are used as knots; i.e., the number of knots is the the number of observations, and the locations of the knots coincide with the locations of the covariate observations $\{t_i\}$. This is the defining feature of *smoothing spline*.

The smoothing spline has close connection with the LME model. It has been shown that the fitted smoothing spline evaluated at the covariate values $(\hat{g}(t_1), \hat{g}(t_2), \ldots, \ldots, \hat{g}(t_v))$ equals the best linear unbiased prediction (BLUP) solution $\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{Z}\hat{\boldsymbol{b}}$ to a linear mixed effect model (1.2). For more details, see [9].

Regression splines

When the number of covariate values become large, computation becomes more difficult and thus the number of knots needs to be reduced. This leads to regression splines and penalized splines.

[2]To be more mathematically rigorous, the minimization is over the space of functions that are differentiable on $[a,b]$ and have absolutely continuous first derivative, which by definition includes all functions that are in $\mathcal{S}[a,b]$ [17].

Regression splines are a basis function-based nonparametric regression method. A *Basis* defines the space of functions (or a close approximation to it) of which f is an element [38], and consists of basis functions, see Appendix A.2 for more details. This approach uses a small number of knots and estimates coefficients as a parametric regression of the basis functions [11].

Let $\{\zeta_1(x), \ldots, \zeta_d(x)\}$ be a set of basis functions where d is the number of knots[3], where d is often small (5 or 6); then one approximates $g(t)$ in (1.3) by

$$g(t) \approx \sum_{i=1}^{d} \alpha_i \zeta_i(t)$$

where $(\alpha_1, \ldots, \alpha_d)$ are unknown parameters and can be estimated by fitting the parametric model

$$y_i = \sum_{i=1}^{d} \alpha_i \zeta_i(t) + \epsilon_i \tag{1.5}$$

via ordinary least squares. Regression splines are computationally easy, as one only needs to run a linear parametric regression. This is one of the key features of the model. However, estimation of $g(t)$ can be sensitive to the choice of number of knots and the locations of the knots. In the statistical context, the knots are often equally spaced or placed at quantiles of the data[11].

Penalized splines

Penalized splines are a hybrid of regression splines and smoothing splines. The estimation proceeds by fitting (1.5) with a quadratic penalty term that is the same to that in smoothing splines, i.e., the integrated square of the second derivative of the fitted curve, scaled by some smoothing parameter.

Typically, the number of knots are much larger than that in regression splines but smaller than the number of observations, as required in smoothing splines. The knot selection is an active research area. Initial attempt at selection of the knots is to use

[3] In general, the number of knots needs *not* to be the same as the number of basis functions.

a *model selection criterion*. With d candidate knots comes 2^d possible models; thus the application of the usual model selection becomes very computational intensive and thus not applicable in many scenarios [27]. A number of other knot selection criterion have been proposed, many of which are based on stepwise regression ideas. For more details, see Section 3.4 and Chapter 5 in [27].

Penalized regression can have various forms of penalties. Eilers & Marx [8] proposed to base the penalty on (higher-order) finite differences of the coefficients of adjacent B-splines, the basis functions used in smoothing:

$$S = \sum_{i=1}^{w} \left\{ y_i - \sum_{j=1}^{u} a_j B_j(x_i) \right\}^2 + \lambda \sum_{j=k+1}^{u} (\Delta^k a_j)^2,$$

where (x_i, y_i) are w data points; $B_j(\cdot)$ is a set of u B-splines; $\sum_{i=1}^{w} \left\{ y_i - \sum_{j=1}^{u} a_j B_j(x_i) \right\}^2$ is the least squares objective function[4]; $\Delta a_j = a_j - a_{j-1}$, and λ is the smoothing parameter governing the goodness-of-fit and smoothness. This is a good discrete approximation of the integrated square of the k^{th} derivative and reduces the dimensionality of the problem to u, the number of B-splines from w, the number of observations.

A number of bases can be used for regression splines and penalized splines. In principle, a change of basis does not change the fit[5]. However, some bases are more numerically stable and allow computation of a fit with greater accuracy. In addition, ease of implementation and interpretation are two other important factors to consider, though the latter may be less important if one is more interested in the fit, not the estimated coefficients [27].

Similarly to bases, there are also other penalties available, see Section 3.8 in [27] for example.

[4] A fitted curve $\hat{y}$ to data (x_i, y_i) is the linear combination $\hat{y}(x) = \sum_{j=1}^{n} \hat{a}_j B_j(x)$.

[5] This is because two bases can be equivalent if they span the same set of functions. The *span* of a set is the set of all possible linear combinations of the basis functions in that set.

1.2 Literature review

1.2.1 Flexible modelling for univariate longitudinal data

There have been many extensions beyond the linear mixed effects models to allow for further flexibility in specifying the mean structure for the modelling of longitudinal data often by using various splines. This comes in two forms - one is to specify design matrices by basis functions [3] [34], and the other is through a nonparametric component in the model [45] [44]. In addition, in some models, various stochastic process are employed to model more flexible within-subject correlation [31] [45].

For example, Brumback and Rice [3] 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. They specify 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*[34] advocated a similar approach as Brumback & Rice [3], where data-based determination of the smoothing parameters was advocated in the paper, yet their model specification is slightly different. The techniques were applied to the analysis of designed experiments, with further details described in [34]. Welham *et al* [37] provided a comparison study of mixed model using splines for curve fitting.

In addition to modeling the mean structure of using a smoothing spline, some efforts were geared toward modelling complicated within-subject covariance. Taylor, Cumberland & Sy [31] used a particular stochastic process to model the data in addition to the usual random effect term which induces within-subject correlation.

1.2.2 Modelling cyclic longitudinal data

One feature in certain longitudinal datasets is that the response is cyclic. This feature happens commonly in the epidemiology studies, particularly related to hormones. For independent data, a periodic cubic smoothing spline has often been used for estimating a periodic function nonparametrically [44]. The smoothing parameter is often estimated by minimizing integrated mean squared error or by cross-validation. However, for correlated data, some of the methods can no longer be accurately applied. Much effort has gone toward solving this additionally data complexity.

To further their efforts on semiparametric stochastic mixed models, Zhang *et al*[44] proposed a semiparametric stochastic mixed model for periodic longitudinal data, where covariate effects are modelled parametrically and the periodic time courses are modelled nonparametrically; and used a stochastic process and random effects to model the within-subject and between-subject correlation respectively.

Instead of cubic smoothing spline, Welham *et al* [36] modelled cyclic longitudinal data using mixed model L-splines. In addition, Meyer *et al*[22] proposed a functional data analysis approach to model cyclic data. They explore modes of variation, displayed as curves, each of which captures some aspect of typical departure from average cyclic behaviour. Last but not least, Wood [38] used penalized cubic regression spline to model a cyclic smooth function. These models, however at this stage to our knowledge, are only applicable to univariate longitudinal analysis.

1.2.3 Modelling multivariate longitudinal data

All of the literature aforementioned reviewed were developed on univariate longitudinal responses. A growing number of data require techniques to model bivariate, or in more generality, multivariate responses.

Some of the efforts of modelling multivariate longitudinal data have emphasized on dimension reduction, such as in Fieuws & Verbeke[10] and Chiou & Muller [5], both of which proposed pairwise modelling approach. Fieuws & Verbeke[10] adapted traditional random effects models to multivariate longitudinal data whereas Chiou & Muller [5] introduced a functional pairwise interaction model from the perspective of functional data analysis. Xiong & Dubin [41] and Xiang *et al*[40] uses nonparametric approaches to model subject-specific curves for multivariate longitudinal data. A review of multivariate longitudinal data analysis can be found in Verbeke *et al*[33]. Further literature review on bivariate longitudinal data analysis will be provided in Chapter 4. Joint models for multivariate longitudinal data and survival data, such as Chi & Ibrahim [4], will not be reviewed here.

1.3 Motivation and the problem

The models proposed in this thesis are motivated by a bivariate longitudinal hormone dataset. One of the response variables is a hormone called progesterone, and the other response variable is another hormone called estrogen, both of which are related to reproductive cycles. Both hormone samples in the dataset are collected daily over several consecutive menstrual cycles from women participated in the study. Since multiple menstrual cycles data are collected, the nature of the data determines that it is intrinsically cyclic. The hormone dataset is described in details in Chapter 2.

Due to the fact that the two hormones may potentially be interactive to each other, it is of interest to model them jointly to explore the correlations between the two hormone levels, as opposed to modelling them separately. We are also interested in modelling the complex time courses of the two hormone levels jointly in a single cycle, as well as some covariate effects, such as age, on the hormones. Furthermore, prediction of a single future observation or of an entire cycle utilizing past cycle observations is meaningful to explore. More often than not, only a single cycle or possibly two consecutive cycles are considered per woman in data analysis in the literature. We will expand upon this focus in the thesis.

To solve this complex problem, we decompose it into three smaller problems and illustrate them in three related chapters in this thesis. In Chapter 3, we briefly review the univariate semiparametric stochastic model in Zhang *et al* [45] and provide derivation details that were absent from the original paper. This chapter will serve as a springboard to Chapter 4 and Chapter 5, where we are building the result to bivariate modelling and it will provide set up for univariate prediction problem in Chapter 5.

In Chapter 4, we extend the univariate model in previous chapter and propose and consider inference for a bivariate semiparametric stochastic mixed model for longitudinal data. The model uses parametric fixed effects to model the covariate effects and smooth nonparametric functions for each of the two underlying time effects. The between-subject correlations are modelled using separate but correlated random effects and the within-subject correlations by a bivariate Gaussian random field.

In Chapter 5, we propose a general prediction procedure of either a single observation or of an entire cycle using past cycle information for periodic univariate longitudinal data from a Frequentist viewpoint. The stochastic process in the univariate model need to have an exponential correlation structure, then by utilizing the Markov property, a prediction density function with prediction intervals can be obtained.

In the last chapter, Chapter 6, we discuss the merits and limitations of findings in previous chapters and potential areas of future work.

Chapter 2

The Bivariate Longitudinal Hormone Dataset

2.1 The hormone dataset for a single menstrual cycle

In the bivariate longitudinal hormone dataset on progesterone and estrogen briefly mentioned in Chapter 1, daily urine samples were collected from 403 employed women aged between 20 to 44 years old, with a completion of a median of five consecutive menstrual cycles each [14]. Among these, 338 women collected at least one complete menstrual cycle daily urine samples with complete covariate information and fewer than three days of missing data in any five-day period, and had no conception in the analyzed cycles. Risk factor data were obtained at baseline in-person interview. The details of the study design and assay methods are described previously Gold *et al* [14] [15].

For analyses that we will later describe and conduct in Chapter 4, one randomly selected cycle is used, and we randomly select 30 study participants from the study due to the heavy computational burden, 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, i.e., day * 28 / max day of observation, based on the assumption that the change of hormone level for each woman depends on the time of the menstrual cycle relative to the cycle length. The standardization generates 56 distinct time points with increments between time points of 1/2 day each. To make the normality assumption more appropriate, a log transformation was used for each of the two hormones.

Figure 2.1 plots the log-transformed progesterone and estrogen levels during a standardized menstrual cycle. Figure 2.2 plots their empirical sample variances calculated at each distinct time point.

2.2 The hormone dataset for multiple menstrual cycles

For analysis that we describe and conduct in Chapter 5, we will use up to 3 consecutive cycles and consider multiple consecutive cycle data for each women, which will lead to a lower number of women in the analysis. Of the 338 women in the study sample, 112 women collected daily urine samples for at least 3 consecutive cycles with complete covariate information and did not have a conception during the analyzed cycles.

Missing data. Considering the 3 consecutive cycles per woman, we expect the ranges of cycles to be of similar lengths; in particular, if the difference of the maximum cycle length and the minimum cycle length per woman is greater than 10 days, then we regard data entries to be unreasonable. For example, for woman X, if the ranges of 3 consecutive cycles considered are 34, 26, 30, respectively, then we find the data to be reasonable; whereas if the ranges of 3 consecutive cycles considered are 53, 26, 27, then we question the validity of the data entries for that particular woman.

The range of the difference of the maximum cycle length and the minimum cycle length length per woman varies between 1 and 37; and of 336 cycles considered (3 cycles for each women), 285 cycles have missing data less than or equal to 10%, see Figure 2.3. Due to the abnormal large range of the difference of the maximum cycle length and the minimum cycle length length for some women and the severity of missing data, we only consider

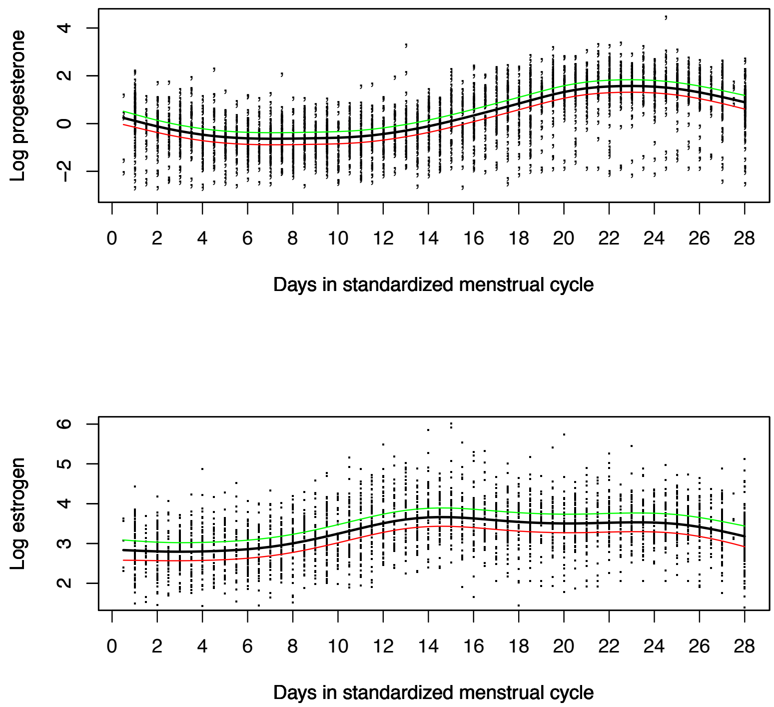


Figure 2.1: Plots of log progesterone and log estrogen levels against days in a standardized menstrual cycle, superimposed by estimated population mean curve $\hat{f}_1$ and $\hat{f}_2$ and their 95% pointwise confidence intervals.

those women that have the range of the difference of the maximum cycle length and the minimum cycle length less than or equal to 10 days and those women that have missing data less than 10%, thus reducing the sample size from 112 to 94.

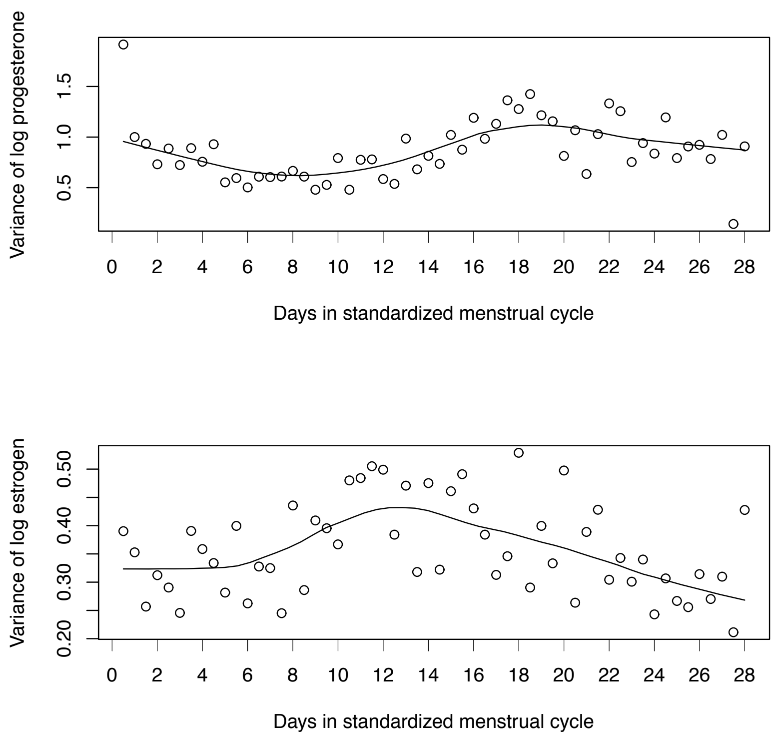


Figure 2.2: Plots of empirical sample variance of log progesterone and log estrogen levels at each distinct time points in a standardized menstrual cycle.

We will conduct analysis in Chapter 5 using this study sample with 3 consecutive cycles. Among the 94 study participants, the total number of observations is 7693*2 for both responses, with each women contributing between 66 and 103 observations. The age range in this new sample is between 23 to 44 years old and the lengths of the menstrual cycles ranges between 19 to 38 days over all women and all cycles, with the average to be

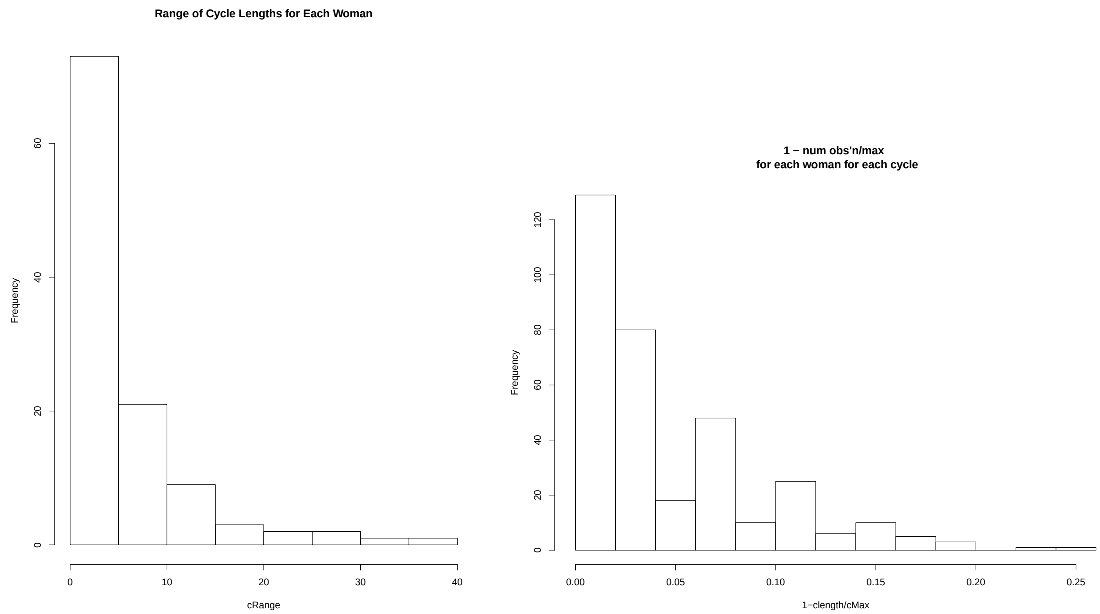


Figure 2.3: Plots of range of cycle lengths and missing data for each cycle for each woman among women that have at least 3 consecutive cycles.

27.3 days. For the same reason as discussed earlier, the cycle length is standardized to a reference 28 days. A log transformation was applied to both of progesterone and estrogen levels to ensure the normality assumptions satisfied as in the single cycle case.

Figure 4.2 displays the log-transformed progesterone and estrogen levels for 3 consecutive cycles which shows that the mean two hormone levels changes over time nonlinearly and periodically.

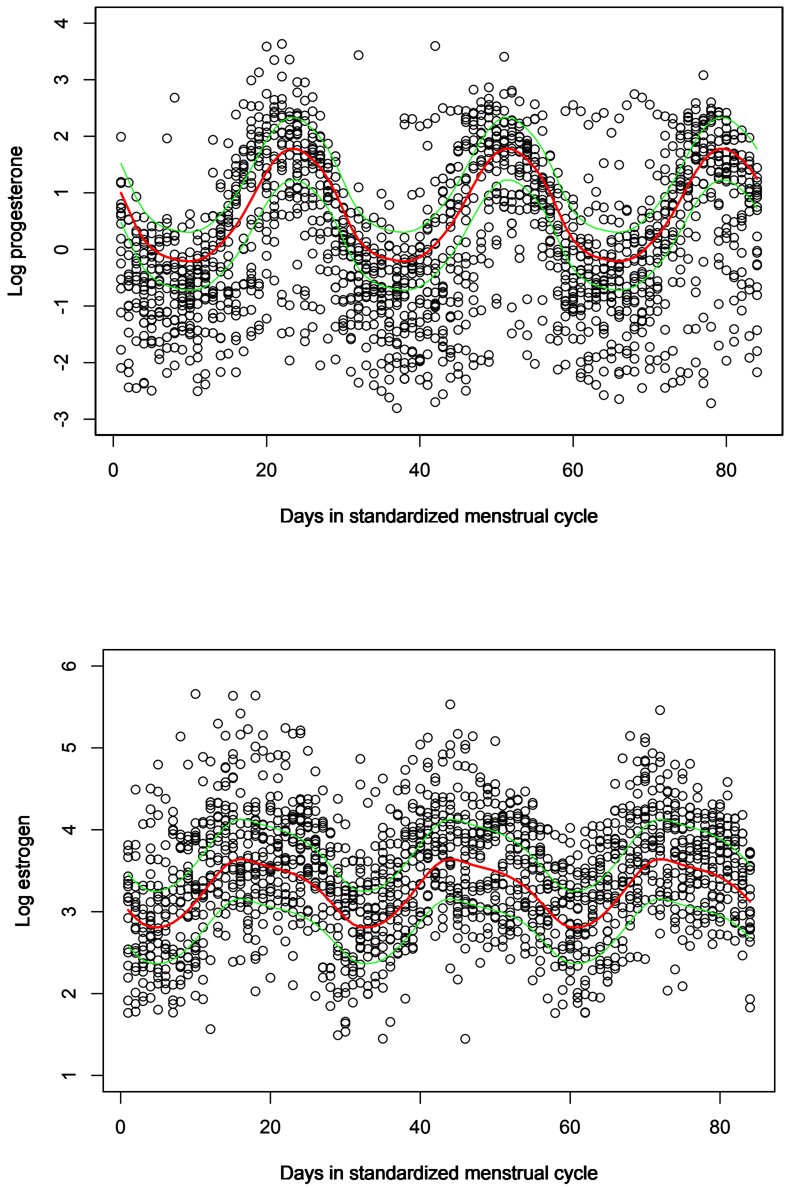


Figure 2.4: Plots of log progesterone and log estrogen levels against days in 3 standardized menstrual cycles, superimposed by estimated population mean curve $\hat{f}_1$ and $\hat{f}_2$ and their 95% pointwise confidence intervals.

Chapter 3

Univariate semiparametric stochastic mixed effects models

3.1 Introduction

In this chapter, I will briefly review the model proposed by Zhang *et al* [45], as the model proposed in the paper would serve as a foundation to the subsequent chapters in the thesis. In addition, non-trivial theoretical details that were absent in the Zhang *et al* [45] are proved in this chapter in details. Furthermore, we will point out some problems in some of the Zhang et al [45] minor yet important results.

The chapter is organized as followed. Section 3.2 outlines the proposed model introduced in Zhang *et al* [45]. Section 3.3 provides theoretical derivations of results given in the paper. Section 3.3.1 provides the derivation of the regression coefficients and nonparametric function. Section 3.3.2 provides the derivation of biases and covariances of the regression coefficients, nonparametric function, random effects, and stochastic processes Section 3.3.3 provides the derivation of the REML estimating equations and the Fisher information Matrix. Section 3.4 provides a summary of the chapter, and discusses challenges and future work.

3.2 Univariate semiparametric mixed effects models

A univariate semiparametric mixed effects model extended the traditional linear mixed effects model by introducing additional nonparametric functions and various stationary and nonstationary stochastic processes to model serial correlation into one cohesive model, see Zhang *et al* [45] and Wu [39]. Specifically, suppose there are m subjects and each subject has n_i measurements, $i = 1, \ldots, m$; let $Y_{ij}, j = 1, \ldots, n_i$ denote the j^{th} measurement for the i^{th} subject, then the model can be written as

$$Y_{ij} = \mathbf{X_{ij}^T}\boldsymbol{\beta} + f(t_{ij}) + \mathbf{Z_{ij}^T b_i} + U_i(t_{ij}) + \epsilon_{ij} \tag{3.1}$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients associated with covariates $\boldsymbol{X}_{ij}$, $f(t)$ is a twice-differentiable smooth function of time, the $\boldsymbol{b}_i$ are independent $q \times 1$ vectors of random effects associated with covariates $\boldsymbol{Z}_{ij}$, the U_i is an independent random processes used to model serial correlation, and ϵ_{ij} are independent measurement errors. The random effects $\mathbf{b_i}$ are assumed to be normally distributed with mean 0 and variance $\boldsymbol{D}(\boldsymbol{\phi})$, where $\boldsymbol{D}$ is assumed to be unstructured and is a positive definite matrix depending on a parameter vector $\boldsymbol{\phi}$, whose components will depend on the dimension of the random effects $\mathbf{b_i}$; $U_i(t)$ is a mean zero Gaussian process[1] with covariance function $\text{cov}(U_i(s), U_i(t)) = \gamma(\boldsymbol{\xi}, \alpha; t, s)$ for some specific parametric function $\gamma(\cdot)$ with a parameter vector $\boldsymbol{\xi}$ and a scalar parameter α, where the components of $\boldsymbol{\xi}$ will depend on the specification of the mean-zero Gaussian process; and ϵ_{ij} is distributed as normal $(0, \sigma^2)$. The $\mathbf{b_i}$, $U_i(t)$ and ϵ_{ij} are assumed to be mutually independent.

Inference of the model can be based on common smoothing methods discussed in Section 1.1.2. In Zhang *et al* [45], smoothing splines was used and the estimators of the regression coefficients $\boldsymbol{\beta}$ and the estimator of the nonparametric function $f(\cdot)$ are obtained using maximum penalized likelihood

$$\ell(\boldsymbol{\beta}, \boldsymbol{f}; \boldsymbol{Y}) - \frac{\lambda}{2}\int [f''(t)]^2 dt \tag{3.2}$$

[1] A *Gaussian process* is a stochastic process $\{X_t, t \in T\}$ (T is a *totally ordered* index set), any finite number of which have a joint Gaussian distribution [26].

where $\ell(\cdot)$ is the log-likelihood function and $\lambda \geq 0$ is the smoothing parameter controlling the balance between the goodness of fit and the roughness of the estimated $f(\cdot)$. The resulting estimators are called the maximum penalized likelihood estimator (MPLE). The penalized likelihood function (3.2) can also be rewritten as

$$\ell(\boldsymbol{\beta}, \boldsymbol{f}; \boldsymbol{Y}) - \frac{\lambda}{2}\boldsymbol{f}^T\boldsymbol{K}\boldsymbol{f}, \tag{3.3}$$

where $\boldsymbol{K}$ is the nonnegative definite smoothing matrix, defined in Equation (2.3) in Green and Silverman [17]. The random effects $\boldsymbol{b}_i$ and the stochastic process U_i are estimated using their conditional means given the data as in LME models.

The use of smoothing spline allows the model to be rewritten as a modified LME model [45], which provides a foundation for the estimation of the smoothing parameter λ and variance parameters. The variance components and the smoothing parameter are estimated using restricted maximum likelihood (REML). The BLUPs from this modified LME model are identical to the MPLEs from (3.2 - 3.3).

3.3 Proof of inference results in the paper

3.3.1 Derivation of model regression coefficients and nonparametric functions

Lemma 1. *Given the log likelihood function of* $(\boldsymbol{\beta}, \boldsymbol{f})$ *from Zhang et al[45]*

$$\ell(\boldsymbol{\beta}, \boldsymbol{f}; \boldsymbol{Y}) = -\frac{1}{2}\log|\boldsymbol{V}| - \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{N}\boldsymbol{f})^T\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{N}\boldsymbol{f}), \tag{3.4}$$

where $\boldsymbol{N}$ *is the incidence matrix defined in from Zhang et al[45] and* $\boldsymbol{\Gamma}_i$ *is the covariance matrix for the Gaussian process* U_i, $\boldsymbol{V} = \mathrm{diag}(V_1, \ldots, V_m)$ *and* $\boldsymbol{V}_i = \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T + \boldsymbol{\Gamma}_i + \boldsymbol{\sigma}^2\boldsymbol{I}_{n_i}, i = 1, \ldots, m$, *show*

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

and

$$\hat{\boldsymbol{f}} = (\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{Y}$$

where $\boldsymbol{W}_x = \boldsymbol{W} - \boldsymbol{W}\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}$, $\boldsymbol{W}_f = \boldsymbol{W} - \boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}$, *and* $\boldsymbol{W} = \boldsymbol{V}^{-1}$.

Proof. Taking derivative of 3.4 with respect to $\boldsymbol{\beta}$, we have [2]

$$\begin{aligned}
\ell_{\boldsymbol{\beta}} &= \left(-\frac{1}{2}\right)(-2)\boldsymbol{X}^T\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{N}\boldsymbol{f}) \\
&= \boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{N}\boldsymbol{f}) \\
&= \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y} - \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}\boldsymbol{f}.
\end{aligned}$$

Set $\ell_{\boldsymbol{\beta}} = 0$, we have

$$\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}\hat{\boldsymbol{f}} = \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y}. \tag{3.5}$$

Thus,

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y} - \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}\hat{\boldsymbol{f}}). \tag{3.6}$$

Similarly, taking derivative of 3.4 with respect to $\boldsymbol{f}$, we have

$$\begin{aligned}
\ell_{\boldsymbol{f}} &= \left(-\frac{1}{2}\right)(-2)\boldsymbol{N}^T\boldsymbol{W}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{N}\boldsymbol{f}) - \frac{\lambda}{2}(2)\boldsymbol{K}\boldsymbol{f} \\
&= \boldsymbol{N}^T\boldsymbol{W}\boldsymbol{Y} - \boldsymbol{N}^T\boldsymbol{W}\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N}\boldsymbol{f} - \lambda\boldsymbol{K}\boldsymbol{f}.
\end{aligned}$$

Setting $\ell_{\boldsymbol{f}} = 0$, we have

$$\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{X}\hat{\boldsymbol{\beta}} + (\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})\hat{\boldsymbol{f}} = \boldsymbol{N}^T\boldsymbol{W}\boldsymbol{Y}, \tag{3.7}$$

which give

$$\hat{\boldsymbol{f}} = (\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{Y} - \boldsymbol{N}^T\boldsymbol{W}\boldsymbol{X}\hat{\boldsymbol{\beta}}). \tag{3.8}$$

Plugging $\hat{\boldsymbol{f}}$ (3.8) into equation (3.5), we have

$$\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}[(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{Y} - \boldsymbol{N}^T\boldsymbol{W}\boldsymbol{X}\hat{\boldsymbol{\beta}})] = \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y}.$$

Expanding and rearranging the equation, we have

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

[2]Note $\frac{\partial}{\partial \boldsymbol{s}}(\boldsymbol{x} - \boldsymbol{A}\boldsymbol{s})^T\boldsymbol{W}(\boldsymbol{x} - \boldsymbol{A}\boldsymbol{s}) = -2\boldsymbol{A}^T\boldsymbol{W}(\boldsymbol{x} - \boldsymbol{A}\boldsymbol{s})$.

which can be reexpressed as

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

or

$$\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{Y}$$

where $\boldsymbol{W}_x = \boldsymbol{W} - \boldsymbol{W}\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}$. Solving for $\hat{\boldsymbol{\beta}}$, we have

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

Similarly, plugging $\hat{\boldsymbol{\beta}}$ (3.6) into equation (3.7), we have

$$\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{X}[(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y} - \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}\hat{\boldsymbol{f}})] + (\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})\hat{\boldsymbol{f}} = \boldsymbol{N}^T\boldsymbol{W}\boldsymbol{Y}.$$

Expanding and rearrangingthe equation, we have

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

which simplifies as

$$(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})\hat{\boldsymbol{f}} = \boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{Y}.$$

We thus obtain the estimator of $\hat{\boldsymbol{f}}$

$$\hat{\boldsymbol{f}} = (\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{Y}$$

where $\boldsymbol{W}_f = \boldsymbol{W} - \boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}$.

Therefore, the maximum penalized likelihood estimators of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{f}}$ from Zhang *et al* [45] are

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

where $\boldsymbol{W}_x = \boldsymbol{W} - \boldsymbol{W}\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}$. and

$$\hat{\boldsymbol{f}} = (\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{Y}$$

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

3.3.2 Derivations of bias and covariance of random effects and stochastic process

Derivations of biases of regression coefficients, nonparametric function, random effects and stochastic process

Lemma 2. *Show*

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

and

$$\mathrm{E}(\hat{\boldsymbol{f}}) - \boldsymbol{f} = -\lambda(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{K}\boldsymbol{f}$$

and these biases go to 0 as λ goes to 0. Also find their covariances.

Proof. For regression coefficient estimator $\hat{\boldsymbol{\beta}}$,

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

Or equivalently, the bias of regression coefficient estimator $\hat{\boldsymbol{\beta}}$ is

$$\mathrm{E}(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta} = (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{N}\boldsymbol{f}. \tag{3.9}$$

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

$$
\begin{aligned}
& \mathrm{E}(\hat{\boldsymbol{f}}) \\
= \; & (\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f E[\boldsymbol{Y}] \\
= \; & (\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f(\boldsymbol{X\beta} + \boldsymbol{N}\boldsymbol{f}) \\
= \; & (\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{X\beta} + (\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N}\boldsymbol{f} \\
= \; & (\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{X\beta} + (\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K} - \lambda\boldsymbol{K})\boldsymbol{f} \\
= \; & (\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{X\beta} + (\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})\boldsymbol{f} \\
& -\lambda(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{K}\boldsymbol{f} \\
= \; & (\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}\boldsymbol{W}_f\boldsymbol{X\beta} + f - \lambda(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{K}\boldsymbol{f} \\
= \; & f - \lambda(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{K}\boldsymbol{f},
\end{aligned}
$$

where the term $(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}\boldsymbol{W}_f\boldsymbol{X\beta}$ vanishes since by plugging $\boldsymbol{W}_f = \boldsymbol{W} - \boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}$, we have

$$
\begin{aligned}
& (\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}\boldsymbol{W}_f\boldsymbol{X\beta} \\
= \; & (\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}\left[\boldsymbol{W} - \boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\right]\boldsymbol{X\beta} \\
= \; & (\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}\boldsymbol{W}\boldsymbol{X\beta} \\
& \qquad -(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X\beta} \\
= \; & (\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}\boldsymbol{W}\boldsymbol{X\beta} \\
& \qquad -(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}\boldsymbol{W}\boldsymbol{X\beta} \\
= \; & \boldsymbol{0}.
\end{aligned}
$$

Plugging $\boldsymbol{W}_x$ into bias (3.9) of $\hat{\boldsymbol{\beta}}$, we have

$$\begin{aligned}
& \mathrm{E}(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta} \\
= & \left[\boldsymbol{X}^T(\boldsymbol{W} - \boldsymbol{W}\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W})\boldsymbol{X}\right]^{-1} \cdot \\
& \qquad \boldsymbol{X}^T(\boldsymbol{W} - \boldsymbol{W}\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W})\boldsymbol{N}\boldsymbol{f} \\
= & \left[\boldsymbol{X}^T(\boldsymbol{W} - \boldsymbol{W}\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W})\boldsymbol{X}\right]^{-1} \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}\boldsymbol{f} \\
& \qquad - \left[\boldsymbol{X}^T(\boldsymbol{W} - \boldsymbol{W}\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W})\boldsymbol{X}\right]^{-1} \cdot \\
& \qquad\qquad \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N}\boldsymbol{f} \\
= & \left[\boldsymbol{X}^T(\boldsymbol{W} - \boldsymbol{W}\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W})\boldsymbol{X}\right]^{-1} \cdot \\
& \qquad \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}\left[\boldsymbol{f} - (\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N}\boldsymbol{f}\right] \\
\to & \; 0 \qquad \text{as } \lambda \to 0
\end{aligned}$$

Note that the last term in square bracket vanishes as $\lambda \to 0$.

The covariance of $\hat{\boldsymbol{\beta}}$ is

$$\begin{aligned}
\mathrm{Cov}(\hat{\boldsymbol{\beta}}) &= (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x Cov(\boldsymbol{Y})\left[(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\right]^T \\
&= (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{V}\boldsymbol{W}_x\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}
\end{aligned}$$

since $\boldsymbol{W}_x^T = \boldsymbol{W}_x$ and $\left[(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\right]^T = \left[(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^T\right]^{-1} = (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}$.

The covariance of $\hat{\boldsymbol{f}}$ is

$$\begin{aligned}
\mathrm{Cov}(\hat{\boldsymbol{f}}) &= (\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f Cov(\boldsymbol{Y})\left[(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f\right]^T \\
&= (\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{V}\boldsymbol{W}_f\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}
\end{aligned}$$

since $\boldsymbol{W}_f^T = \boldsymbol{W}_f$ and

$$\left[(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\right]^T = \left[(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^T\right]^{-1} = (\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}.$$

□

Lemma 3. *Show*

$$E(\hat{\boldsymbol{b}}_i) = \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i\left[\lambda\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{K} - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{N}\right]\boldsymbol{f},$$

and ,

$$E\left[\hat{\boldsymbol{U}}_i(\boldsymbol{s}_i)\right] = \boldsymbol{\Gamma}_i(\boldsymbol{s}_i, \boldsymbol{t}_i)\boldsymbol{W}_i\left[\lambda \boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{K} - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{N}\right]\boldsymbol{f}.$$

and these biases go to 0 as λ goes to 0.

Proof. The estimators of the random effects $\hat{\boldsymbol{b}}_i$ and the stochastic process $\hat{\boldsymbol{U}}_i$ are given by

$$\hat{\boldsymbol{b}}_i = \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i(\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{f}}_i)$$

and

$$\hat{\boldsymbol{U}}_i(\boldsymbol{s}_i) = \boldsymbol{\Gamma}_i(\boldsymbol{s}_i, \boldsymbol{t}_i)\boldsymbol{W}_i(\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{f}}_i).$$

Similarly, we want to find their biases and covariances. For the expected values of these estimators, we first calculate

$$\begin{aligned}
E(\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{f}}_i) &= E(\boldsymbol{Y}_i) - E(\boldsymbol{X}_i\hat{\boldsymbol{\beta}}) - E(\hat{\boldsymbol{f}}_i) \\
&= \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{N}_i\boldsymbol{f} - \boldsymbol{X}_iE(\hat{\boldsymbol{\beta}}) - E(\hat{\boldsymbol{f}}_i) \\
&= \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{N}_i\boldsymbol{f} - \boldsymbol{X}_i\left[\boldsymbol{\beta} + (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{N}\boldsymbol{f}\right] \\
&\qquad - \left[\boldsymbol{N}_i f - \lambda\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{K}\boldsymbol{f}\right] \\
&= \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{N}_i\boldsymbol{f} - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{N}\boldsymbol{f} \\
&\qquad - \boldsymbol{N}_i f + \lambda\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{K}\boldsymbol{f} \\
&= \left[\lambda\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{K} - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{N}\right]\boldsymbol{f}
\end{aligned}$$

where

$$E(\hat{\boldsymbol{f}}_i) = E(\boldsymbol{N}_i\hat{\boldsymbol{f}}) = \boldsymbol{N}_iE(\hat{\boldsymbol{f}}) = \boldsymbol{N}_i f - \lambda\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{K}\boldsymbol{f}.$$

Thus, the expected values are

$$\begin{aligned}
E(\hat{\boldsymbol{b}}_i) &= \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_iE(\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{f}}_i) \\
&= \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i\left[\lambda\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{K} - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{N}\right]\boldsymbol{f},
\end{aligned}$$

and similarly,

$$\begin{aligned}
E\left[\hat{\boldsymbol{U}}_i(\boldsymbol{s}_i)\right] &= \boldsymbol{\Gamma}_i(\boldsymbol{s}_i, \boldsymbol{t}_i)\boldsymbol{W}_iE(\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{f}}_i) \\
&= \boldsymbol{\Gamma}_i(\boldsymbol{s}_i, \boldsymbol{t}_i)\boldsymbol{W}_i\left[\lambda\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{K} - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{N}\right]\boldsymbol{f}.
\end{aligned}$$

Note that as $\lambda \to 0$, 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, since by plugging $\boldsymbol{W}_x = \boldsymbol{W} - \boldsymbol{W}\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}$, we have

$$\begin{aligned}
& \boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{N}\boldsymbol{f} \\
= \;& \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}\boldsymbol{f} - \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N}\boldsymbol{f} \\
= \;& \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}\left[\boldsymbol{f} - (\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N}\boldsymbol{f}\right] \\
\to \;& \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}\left[\boldsymbol{f} - (\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N})^{-1}\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N}\boldsymbol{f}\right] \quad \text{as } \lambda \to 0 \\
= \;& \boldsymbol{0}
\end{aligned}$$

□

Lemma 4. *Let* $\boldsymbol{\chi}_i = \begin{pmatrix}\boldsymbol{X}_i & \boldsymbol{N}_i\end{pmatrix}$ *and* $\boldsymbol{\chi} = \begin{pmatrix}\boldsymbol{X} & \boldsymbol{N}\end{pmatrix}$, *show*

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

Proof. For the covariances, we will first find the covariance of the random effects $\boldsymbol{b}_i$. Since

$$\mathrm{Cov}(\hat{\boldsymbol{b}}_i - \boldsymbol{b}_i) = \mathrm{Var}(\hat{\boldsymbol{b}}_i) + \mathrm{Var}(\boldsymbol{b}_i) - \mathrm{Cov}(\boldsymbol{b}_i, \hat{\boldsymbol{b}}_i) - \mathrm{Cov}(\hat{\boldsymbol{b}}_i, \boldsymbol{b}_i)$$

and $\mathrm{Var}(\boldsymbol{b}_i) = \boldsymbol{D}$, it suffices to find $\mathrm{Var}(\hat{\boldsymbol{b}}_i)$ and $\mathrm{Cov}(\boldsymbol{b}_i, \hat{\boldsymbol{b}}_i)$. To find

$$\mathrm{Var}(\hat{\boldsymbol{b}}_i) = \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i\mathrm{Var}(\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{f}}_i)\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D},$$

denote

$$\boldsymbol{W}_{x_i} = \boldsymbol{W}_i - \boldsymbol{W}_i\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T\boldsymbol{W}_i$$

and

$$\boldsymbol{W}_{f_i} = \boldsymbol{W}_i - \boldsymbol{W}_i\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_i,$$

we have

$$\begin{aligned}
\text{Cov}(\boldsymbol{Y}_i, \boldsymbol{X}_i\hat{\boldsymbol{\beta}}) &= \text{Cov}(\boldsymbol{Y}_i, \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{Y}) \\
&= \text{Cov}\left(\boldsymbol{Y}_i, \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\sum_{j=1}^{m}\boldsymbol{X}_j^T\boldsymbol{W}_{x_j}\boldsymbol{Y}_j\right) \\
&= \text{Cov}\left(\boldsymbol{Y}_i, \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_{x_i}\boldsymbol{Y}_i\right) \\
&= \text{Cov}(\boldsymbol{Y}_i, \boldsymbol{Y}_i)[\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_{x_i}]^T \\
&= \boldsymbol{V}_i\boldsymbol{W}_{x_i}\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T, \\
&= \boldsymbol{V}_i[\boldsymbol{W}_i - \boldsymbol{W}_i\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T\boldsymbol{W}_i]\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T, \\
&= [\boldsymbol{I}_i - \boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T\boldsymbol{W}_i]\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T,
\end{aligned}$$

$$\text{Cov}(\boldsymbol{X}_i\hat{\boldsymbol{\beta}}, \boldsymbol{Y}_i) = [\text{Cov}(\boldsymbol{Y}_i, \boldsymbol{X}_i\hat{\boldsymbol{\beta}})]^T = \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T[\boldsymbol{I}_i - \boldsymbol{W}_i\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T]$$

$$\begin{aligned}
\text{Cov}(\boldsymbol{Y}_i, \hat{\boldsymbol{f}}_i) &= \text{Cov}(\boldsymbol{Y}_i, \boldsymbol{N}_i\hat{\boldsymbol{f}}) \\
&= \text{Cov}(\boldsymbol{Y}_i, \boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{Y}) \\
&= \text{Cov}(\boldsymbol{Y}_i, \boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\sum_{j=1}^{m}\boldsymbol{N}_j^T\boldsymbol{W}_{f_j}\boldsymbol{Y}_j) \\
&= \text{Cov}(\boldsymbol{Y}_i, \boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T\boldsymbol{W}_{f_i}\boldsymbol{Y}_i) \\
&= \text{Cov}(\boldsymbol{Y}_i, \boldsymbol{Y}_i)[\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T\boldsymbol{W}_{f_i}]^T \\
&= \boldsymbol{V}_i\boldsymbol{W}_{f_i}\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T \\
&= \boldsymbol{V}_i[\boldsymbol{W}_i - \boldsymbol{W}_i\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_i]\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T \\
&= [\boldsymbol{I}_i - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_i]\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T
\end{aligned}$$

$$\text{Cov}(\hat{\boldsymbol{f}}_i, \boldsymbol{Y}_i) = [\text{Cov}(\boldsymbol{Y}_i, \hat{\boldsymbol{f}}_i)]^T = \boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T[\boldsymbol{I}_i - \boldsymbol{W}_i\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}_i^T]$$

and[3]

$$\begin{aligned}
\text{Cov}(\boldsymbol{X}_i\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{f}}_i) &= \text{Cov}\left[\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{Y}, \boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{Y}\right] \\
&= \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\text{Cov}(\boldsymbol{Y}, \boldsymbol{Y})[\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f]^T \\
&= \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{V}\boldsymbol{W}_f\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T.
\end{aligned}$$

$$\text{Cov}(\hat{\boldsymbol{f}}_i, \boldsymbol{X}_i\hat{\boldsymbol{\beta}}) = [\text{Cov}(\boldsymbol{X}_i\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{f}}_i)]^T = \boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{V}\boldsymbol{W}_x\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T$$

[3] Recall $\text{Cov}(\boldsymbol{A}\boldsymbol{X}, \boldsymbol{B}^T\boldsymbol{Y}) = \boldsymbol{A}\text{Cov}(\boldsymbol{X}, \boldsymbol{Y})\boldsymbol{B}$

Thus,

$$
\begin{aligned}
& \mathrm{Var}(\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{f}}_i) \\
= \ & \mathrm{Var}(\boldsymbol{Y}_i) + \mathrm{Var}(\boldsymbol{X}_i\hat{\boldsymbol{\beta}}) + \mathrm{Var}(\boldsymbol{N}_i\hat{\boldsymbol{f}}) \\
& \quad -\mathrm{Cov}(\boldsymbol{Y}_i, \boldsymbol{X}_i\hat{\boldsymbol{\beta}}) - \mathrm{Cov}(\boldsymbol{X}_i\hat{\boldsymbol{\beta}}, \boldsymbol{Y}_i) \\
& \quad -\mathrm{Cov}(\boldsymbol{Y}_i, \hat{\boldsymbol{f}}_i) - \mathrm{Cov}(\hat{\boldsymbol{f}}_i, \boldsymbol{Y}_i) \\
& \quad +\mathrm{Cov}(\boldsymbol{X}_i\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{f}}_i) + \mathrm{Cov}(\hat{\boldsymbol{f}}_i, \boldsymbol{X}_i\hat{\boldsymbol{\beta}}) \\
= \ & \boldsymbol{V}_i + \boldsymbol{X}_i(\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}\boldsymbol{X}_i^T \\
& \quad +\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{V}\boldsymbol{W}_f\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T \\
& \quad -[\boldsymbol{I}_i - \boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T\boldsymbol{W}_i]\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T \\
& \quad -\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T[\boldsymbol{I}_i - \boldsymbol{W}_i\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T] \\
& \quad -[\boldsymbol{I}_i - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_i]\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T \\
& \quad -\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T[\boldsymbol{I}_i - \boldsymbol{W}_i\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}_i^T] \\
& \quad +\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{V}\boldsymbol{W}_f\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T \\
& \quad +\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{V}\boldsymbol{W}_x\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T
\end{aligned}
$$

$$
\begin{aligned}
& \mathrm{Var}(\hat{\boldsymbol{b}}_i) \\
= \ & \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{V}_i\boldsymbol{Z}_i\boldsymbol{D} \\
& \quad +\boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i[\boldsymbol{X}_i(\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}\boldsymbol{X}_i^T \\
& \qquad +\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{V}\boldsymbol{W}_f\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T \\
& \qquad -[\boldsymbol{I}_i - \boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T\boldsymbol{W}_i]\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T \\
& \qquad -\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T[\boldsymbol{I}_i - \boldsymbol{W}_i\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T] \\
& \qquad -[\boldsymbol{I}_i - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_i]\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T \\
& \qquad -\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T[\boldsymbol{I}_i - \boldsymbol{W}_i\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}_i^T] \\
& \qquad +\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{V}\boldsymbol{W}_f\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T \\
& \qquad +\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{V}\boldsymbol{W}_x\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T]\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D}
\end{aligned}
$$

We now want to find the covariance between $\hat{\boldsymbol{b}}_i$ and $\boldsymbol{b}_i$. First, we need to find

$$\begin{aligned}
\mathrm{Cov}(\boldsymbol{b}_i, \boldsymbol{Y}_i) &= \mathrm{Cov}[\boldsymbol{b}_i, (\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{N}_i\boldsymbol{f} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{U}_i + \boldsymbol{\epsilon}_i)] \\
&= \mathrm{Cov}(\boldsymbol{b}_i, \boldsymbol{Z}_i\boldsymbol{b}_i) \\
&= \boldsymbol{D}\boldsymbol{Z}_i^T,
\end{aligned}$$

$$\begin{aligned}
\mathrm{Cov}(\boldsymbol{b}_i, \boldsymbol{X}_i\hat{\boldsymbol{\beta}}) &= \mathrm{Cov}[\boldsymbol{b}_i, \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{Y}] \\
&= \mathrm{Cov}\left[\boldsymbol{b}_i, \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\sum_{j=1}^{m}\boldsymbol{X}_j^T\boldsymbol{W}_{x_j}\boldsymbol{Y}_j\right] \\
&= \mathrm{Cov}[\boldsymbol{b}_i, \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_{x_i}\boldsymbol{Y}_i] \\
&= \mathrm{Cov}(\boldsymbol{b}_i, \boldsymbol{Y}_i)[\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_{x_i}]^T \\
&= \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_{x_i}\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T
\end{aligned}$$

$$\begin{aligned}
\mathrm{Cov}(\boldsymbol{b}_i, \hat{\boldsymbol{f}}_i) &= \mathrm{Cov}(\boldsymbol{b}_i, \boldsymbol{N}_i\hat{\boldsymbol{f}}) \\
&= \mathrm{Cov}(\boldsymbol{b}_i, \boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{Y}) \\
&= \mathrm{Cov}\left[\boldsymbol{b}_i, \boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\sum_{j=1}^{m}\boldsymbol{N}_j^T\boldsymbol{W}_{f_j}\boldsymbol{Y}_j\right] \\
&= \mathrm{Cov}\left[\boldsymbol{b}_i, \boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T\boldsymbol{W}_{f_i}\boldsymbol{Y}_i\right] \\
&= \mathrm{Cov}(\boldsymbol{b}_i, \boldsymbol{Y}_i)\left[\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T\boldsymbol{W}_{f_i}\right]^T \\
&= \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_{f_i}^T\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T
\end{aligned}$$

Thus,

$$\begin{aligned}
& \mathrm{Cov}(\boldsymbol{b}_i, \hat{\boldsymbol{b}}_i) \\
= \;& \mathrm{Cov}[\boldsymbol{b}_i, \boldsymbol{D}\boldsymbol{Z}_i^T \boldsymbol{W}_i(\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{f}}_i)] \\
= \;& \mathrm{Cov}(\boldsymbol{b}_i, \boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{f}}_i)\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D} \\
= \;& [\mathrm{Cov}(\boldsymbol{b}_i, \boldsymbol{Y}_i) - \mathrm{Cov}(\boldsymbol{b}_i, \boldsymbol{X}_i\hat{\boldsymbol{\beta}}) - \mathrm{Cov}(\boldsymbol{b}_i, \hat{\boldsymbol{f}}_i)]\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D} \\
= \;& [\boldsymbol{D}\boldsymbol{Z}_i^T - \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_{x_i}\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T \\
& \quad -\boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_{f_i}^T\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T]\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D} \\
= \;& \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D} \\
& \quad -\boldsymbol{D}\boldsymbol{Z}_i^T\left[\boldsymbol{W}_{x_i}\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T + \boldsymbol{W}_{f_i}\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T\right]\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D} \\
= \;& \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D} \\
& \quad -\boldsymbol{D}\boldsymbol{Z}_i^T[(\boldsymbol{W}_i - \boldsymbol{W}_i\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T\boldsymbol{W}_i)\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T \\
& \qquad\qquad +(\boldsymbol{W}_i - \boldsymbol{W}_i\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_i)\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T]\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D} \\
= \;& \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D} \\
& \quad -\boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i[(\boldsymbol{I}_i - \boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T\boldsymbol{W}_i)\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T \\
& \qquad\qquad +(\boldsymbol{I}_i - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_i)\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T]\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D}
\end{aligned}$$

and

$$\begin{aligned}
& \mathrm{Cov}(\hat{\boldsymbol{b}}_i, \boldsymbol{b}_i) \\
= \;& \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D} \\
& \quad -\boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i\left[\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_{x_i} + \boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T\boldsymbol{W}_{f_i}\right]\boldsymbol{Z}_i\boldsymbol{D} \\
= \;& \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D} \\
& \quad -\boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i[\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T(\boldsymbol{W}_i - \boldsymbol{W}_i\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T\boldsymbol{W}_i) \\
& \qquad\qquad +\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T(\boldsymbol{W}_i - \boldsymbol{W}_i\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_i)]\boldsymbol{Z}_i\boldsymbol{D} \\
= \;& \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D} \\
& \quad -\boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i[\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T(\boldsymbol{I}_i - \boldsymbol{W}_i\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T) \\
& \qquad\qquad +\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T(\boldsymbol{I}_i - \boldsymbol{W}_i\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}_i^T)]\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D}
\end{aligned}$$

$$
\begin{aligned}
&\mathrm{Cov}(\hat{\boldsymbol{b}}_i - \boldsymbol{b}_i) \\
= \;& \mathrm{Var}(\hat{\boldsymbol{b}}_i) + \mathrm{Var}(\boldsymbol{b}_i) - \mathrm{Cov}(\boldsymbol{b}_i, \hat{\boldsymbol{b}}_i) - \mathrm{Cov}(\hat{\boldsymbol{b}}_i, \boldsymbol{b}_i) \\
= \;& \boldsymbol{D} + \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{V}_i\boldsymbol{Z}_i\boldsymbol{D} \\
&\quad + \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i[\boldsymbol{X}_i(\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}\boldsymbol{X}_i^T \\
&\qquad + \boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{V}\boldsymbol{W}_f\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T \\
&\qquad - [\boldsymbol{I}_i - \boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T\boldsymbol{W}_i]\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T \\
&\qquad - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T[\boldsymbol{I}_i - \boldsymbol{W}_i\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T] \\
&\qquad - [\boldsymbol{I}_i - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_i]\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T \\
&\qquad - \boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T[\boldsymbol{I}_i - \boldsymbol{W}_i\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}_i^T] \\
&\qquad + \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{V}\boldsymbol{W}_f\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T \\
&\qquad + \boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{V}\boldsymbol{W}_x\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T]\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D} \\
&\quad - \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D} \\
&\qquad + \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i[(\boldsymbol{I}_i - \boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T\boldsymbol{W}_i)\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T \\
&\qquad\quad + (\boldsymbol{I}_i - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_i)\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T]\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D} \\
&\quad - \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D} \\
&\qquad + \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i[\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T(\boldsymbol{I}_i - \boldsymbol{W}_i\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T) \\
&\qquad\quad + \boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T(\boldsymbol{I}_i - \boldsymbol{W}_i\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}_i^T)]\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D} \\
= \;& \boldsymbol{D} - \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{V}_i\boldsymbol{Z}_i\boldsymbol{D} \\
&\quad + \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i[\boldsymbol{X}_i(\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}\boldsymbol{X}_i^T \\
&\qquad + \boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{V}\boldsymbol{W}_f\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T \\
&\qquad + \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{V}\boldsymbol{W}_f\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T \\
&\qquad + \boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{V}\boldsymbol{W}_x\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T]\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D} \qquad (3.10)
\end{aligned}
$$

Denote

$$A = \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x$$

and

$$B = \boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N} + \lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f$$

we have the terms in the square brackets

$$
\begin{aligned}
& \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{V}\boldsymbol{W}_x\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T \\
& \quad +\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N}+\lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{V}\boldsymbol{W}_f\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N}+\lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T \\
& \quad +\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{V}\boldsymbol{W}_f\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N}+\lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T \\
& \quad +\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N}+\lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{V}\boldsymbol{W}_x\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T \\
= \; & AVA^T+BVB^T+AVB^T+BVA^T \\
= \; & AV(A^T+B^T)+BV(B^T+A^T) \\
= \; & (A+B)V(A^T+B^T) \qquad (3.11)
\end{aligned}
$$

where

$$
\begin{aligned}
& A+B \\
= \; & \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x+\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N}+\lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}_f \\
= \; & \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T[\boldsymbol{W}-\boldsymbol{W}\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N}+\lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}] \\
& \quad +\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N}+\lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T[\boldsymbol{W}-\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}] \\
= \; & \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}-\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N}+\lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W} \\
& \quad +\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N}+\lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W} \\
& \quad -\boldsymbol{N}_i(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N}+\lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W} \\
= \; & \begin{pmatrix}\boldsymbol{X}_i & \boldsymbol{N}_i\end{pmatrix}\begin{pmatrix}(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1} & -(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N}+\lambda\boldsymbol{K})^{-1} \\ -(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N}+\lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1} & (\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N}+\lambda\boldsymbol{K})^{-1}\end{pmatrix}\begin{pmatrix}\boldsymbol{X}^T \\ \boldsymbol{N}^T\end{pmatrix}\boldsymbol{W} \\
= \; & \boldsymbol{\chi}_i\boldsymbol{C}^{-1}\boldsymbol{\chi}^T\boldsymbol{W}
\end{aligned}
$$

and

$$
\begin{aligned}
& A^T+B^T \\
= \; & \boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T-\boldsymbol{W}\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N}+\lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T \\
& \quad +\boldsymbol{W}\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N}+\lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T \\
& \quad -\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N}+\lambda\boldsymbol{K})^{-1}\boldsymbol{N}_i^T \\
= \; & \boldsymbol{W}\begin{pmatrix}\boldsymbol{X} & \boldsymbol{N}\end{pmatrix}\begin{pmatrix}(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1} & -(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}(\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{N}+\lambda\boldsymbol{K})^{-1} \\ -(\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N}+\lambda\boldsymbol{K})^{-1}\boldsymbol{N}^T\boldsymbol{W}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1} & (\boldsymbol{N}^T\boldsymbol{W}_f\boldsymbol{N}+\lambda\boldsymbol{K})^{-1}\end{pmatrix}\begin{pmatrix}\boldsymbol{X}_i^T \\ \boldsymbol{N}_i^T\end{pmatrix} \\
= \; & \boldsymbol{W}\boldsymbol{\chi}\boldsymbol{C}^{-1}\boldsymbol{\chi}_i^T
\end{aligned}
$$

Thus, equation (3.11) becomes

$$
\begin{aligned}
& (A+B)\boldsymbol{V}(A^T+B^T) \\
= & \ \boldsymbol{\chi}_i \boldsymbol{C}^{-1} \boldsymbol{\chi}^T \boldsymbol{W} \boldsymbol{V} \boldsymbol{W} \boldsymbol{\chi} \boldsymbol{C}^{-1} \boldsymbol{\chi}_i^T \\
= & \ \boldsymbol{\chi}_i \boldsymbol{C}^{-1} \boldsymbol{\chi}^T \boldsymbol{W} \boldsymbol{\chi} \boldsymbol{C}^{-1} \boldsymbol{\chi}_i^T
\end{aligned}
$$

Therefore, by (3.10)

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

□

3.3.3 Derivation of REML estimating equations and Fisher information matrix

Given in Zhang *et al.* [45], the REML log-likelihood of $(\tau, \boldsymbol{\theta})$ is

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

where $\boldsymbol{X}_* = [\boldsymbol{X}, \boldsymbol{N}\boldsymbol{T}]$, $\boldsymbol{V}_* = \tau \boldsymbol{B}_* \boldsymbol{B}_*^T + \boldsymbol{V}$, and $\boldsymbol{B}_* = \boldsymbol{N}\boldsymbol{B}$ where $\boldsymbol{B} = \boldsymbol{L}(\boldsymbol{L}^T\boldsymbol{L})^{-1}$ and $\boldsymbol{L}$ is $r \times (r-2)$ full-rank matrix satisfying $\boldsymbol{K} = \boldsymbol{L}\boldsymbol{L}^T$ and $\boldsymbol{L}^T\boldsymbol{T} = 0$.

Take derivative with respect to τ, we have

$$
\frac{\partial \ell_R}{\partial \tau} = \left(-\frac{1}{2}\right) \cdot \left[\frac{\partial \log|\boldsymbol{V}_*|}{\partial \tau} + \frac{\partial \log|\boldsymbol{X}_*^T \boldsymbol{V}_*^{-1} \boldsymbol{X}_*|}{\partial \tau} + \frac{\partial (\boldsymbol{Y} - \boldsymbol{X}_*\hat{\boldsymbol{\beta}}_*)^T \boldsymbol{V}_*^{-1}(\boldsymbol{Y} - \boldsymbol{X}_*\hat{\boldsymbol{\beta}}_*)}{\partial \tau}\right] \quad (3.12)
$$

Calculating the terms in the square brackets one by one, the first term gives

$$\begin{aligned}
\frac{\partial \log |\boldsymbol{V}_*|}{\partial \tau} &= \frac{\partial \operatorname{Tr}(\log \boldsymbol{V}_*)}{\partial \tau} \qquad \text{since } \log|X| = \operatorname{Tr}(\log X) \text{ in general} \\
&= \frac{\operatorname{Tr} \partial(\log \boldsymbol{V}_*)}{\partial \tau} \\
&= \operatorname{Tr} \frac{\partial(\log \boldsymbol{V}_*)}{\partial \tau} \\
&= \operatorname{Tr}\left[\frac{\partial(\log \boldsymbol{V}_*)}{\partial \boldsymbol{V}_*}\frac{\partial \boldsymbol{V}_*}{\partial \tau}\right] \\
&= \operatorname{Tr}(\boldsymbol{V}_*^{-1}\boldsymbol{B}_*\boldsymbol{B}_*^T). \qquad (3.13)
\end{aligned}$$

The second term[4] is

$$\begin{aligned}
\frac{\partial \log |\boldsymbol{X}_*^T\boldsymbol{V}_*^{-1}\boldsymbol{X}_*|}{\partial \tau} &= \frac{\partial \operatorname{Tr}\left[\log(\boldsymbol{X}_*^T\boldsymbol{V}_*^{-1}\boldsymbol{X}_*)\right]}{\partial \tau} \\
&= \frac{\operatorname{Tr} \partial \left[\log(\boldsymbol{X}_*^T\boldsymbol{V}_*^{-1}\boldsymbol{X}_*)\right]}{\partial \tau} \\
&= \operatorname{Tr} \frac{\partial \left[\log(\boldsymbol{X}_*^T\boldsymbol{V}_*^{-1}\boldsymbol{X}_*)\right]}{\partial \tau} \\
&= \operatorname{Tr}\left\{\frac{\partial \left[\log(\boldsymbol{X}_*^T\boldsymbol{V}_*^{-1}\boldsymbol{X}_*)\right]}{\boldsymbol{X}_*^T\boldsymbol{V}_*^{-1}\boldsymbol{X}_*}\frac{\partial \boldsymbol{X}_*^T\boldsymbol{V}_*^{-1}\boldsymbol{X}_*}{\partial \boldsymbol{V}_*}\frac{\partial \boldsymbol{V}_*}{\partial \tau}\right\} \\
&= \operatorname{Tr}\left[-\boldsymbol{V}_*^{-1}\boldsymbol{X}_*(\boldsymbol{X}_*^T\boldsymbol{V}_*^{-1}\boldsymbol{X}_*)^{-1}\boldsymbol{X}_*^T\boldsymbol{V}_*^{-1}\boldsymbol{B}_*\boldsymbol{B}_*^T\right]. \qquad (3.14)
\end{aligned}$$

The last term uses identity

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

which, after taking transpose on both sides, is equivalent to

$$(\boldsymbol{Y} - \boldsymbol{X}_*\hat{\boldsymbol{\beta}}_*)^T(\boldsymbol{V}_*^{-1})^T = (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}\hat{\boldsymbol{f}})^T\boldsymbol{V}^{-1}.$$

[4]Recall $\frac{\partial \boldsymbol{a}^T\boldsymbol{X}^{-1}\boldsymbol{b}}{\partial \boldsymbol{X}} = -\boldsymbol{X}^{-1}\boldsymbol{a}\boldsymbol{b}^T\boldsymbol{X}^{-1}$ for some matrix $\boldsymbol{X}$ and vectors $\boldsymbol{a}$, $\boldsymbol{b}$, and $\frac{\partial \boldsymbol{Y}^{-1}}{\partial x} = -\boldsymbol{Y}^{-1}\frac{\partial \boldsymbol{Y}}{\partial x}\boldsymbol{Y}^{-1}$ for some matrix $\boldsymbol{Y}$.

Since $\boldsymbol{V}_* = \tau \boldsymbol{B}_* \boldsymbol{B}_*^T + \boldsymbol{V}$, where $\boldsymbol{B}_*$ is symmetric, and $\boldsymbol{V}$ is diagonal, thus $\boldsymbol{V}_*^T = \boldsymbol{V}_*$ and

$$\begin{aligned} RHS &= (\boldsymbol{Y} - \boldsymbol{X}_*\hat{\boldsymbol{\beta}}_*)^T (\boldsymbol{V}_*^{-1})^T \\ &= (\boldsymbol{Y} - \boldsymbol{X}_*\hat{\boldsymbol{\beta}}_*)^T (\boldsymbol{V}_*^T)^{-1} \\ &= (\boldsymbol{Y} - \boldsymbol{X}_*\hat{\boldsymbol{\beta}}_*)^T \boldsymbol{V}_*^{-1} \end{aligned}$$

$$\Rightarrow \qquad (\boldsymbol{Y} - \boldsymbol{X}_*\hat{\boldsymbol{\beta}}_*)^T \boldsymbol{V}_*^{-1} = (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}\hat{\boldsymbol{f}})^T \boldsymbol{V}^{-1}$$

thus, the partial of the last term is

$$\begin{aligned} & \frac{\partial}{\partial \tau} (\boldsymbol{Y} - \boldsymbol{X}_*\hat{\boldsymbol{\beta}}_*)^T \boldsymbol{V}_*^{-1} (\boldsymbol{Y} - \boldsymbol{X}_*\hat{\boldsymbol{\beta}}_*) \\ = \; & (\boldsymbol{Y} - \boldsymbol{X}_*\hat{\boldsymbol{\beta}}_*)^T \frac{\partial \boldsymbol{V}_*^{-1}}{\partial \tau} (\boldsymbol{Y} - \boldsymbol{X}_*\hat{\boldsymbol{\beta}}_*) \\ = \; & -(\boldsymbol{Y} - \boldsymbol{X}_*\hat{\boldsymbol{\beta}}_*)^T \boldsymbol{V}_*^{-1} \frac{\partial \boldsymbol{V}_*}{\partial \tau} \boldsymbol{V}_*^{-1} (\boldsymbol{Y} - \boldsymbol{X}_*\hat{\boldsymbol{\beta}}_*) \\ = \; & -(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}\hat{\boldsymbol{f}})^T \boldsymbol{V}^{-1} \boldsymbol{B}_* \boldsymbol{B}_*^T \boldsymbol{V}^{-1} (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}\hat{\boldsymbol{f}}) \end{aligned} \tag{3.15}$$

by using the identities. Note that here we didn't take partials of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{f}}$. Plugging results (3.13) (3.14) (3.15) into equation (3.12), we have the score equation for τ as

$$\begin{aligned} \frac{\partial \ell_R}{\partial \tau} &= \left(-\frac{1}{2}\right) \cdot \{ \mathrm{Tr}(\boldsymbol{V}_*^{-1} \boldsymbol{B}_* \boldsymbol{B}_*^T) + \mathrm{Tr}\left[-\boldsymbol{V}_*^{-1} \boldsymbol{X}_* (\boldsymbol{X}_*^T \boldsymbol{V}_*^{-1} \boldsymbol{X}_*)^{-1} \boldsymbol{X}_*^T \boldsymbol{V}_*^{-1} \boldsymbol{B}_* \boldsymbol{B}_*^T \right] \\ & \qquad + (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}\hat{\boldsymbol{f}})^T \boldsymbol{V}^{-1} \boldsymbol{B}_* \boldsymbol{B}_*^T \boldsymbol{V}^{-1} (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}\hat{\boldsymbol{f}}) \} \\ &= \left(-\frac{1}{2}\right) \cdot \{ \mathrm{Tr}\left([\boldsymbol{V}_*^{-1} - \boldsymbol{V}_*^{-1} \boldsymbol{X}_* (\boldsymbol{X}_*^T \boldsymbol{V}_*^{-1} \boldsymbol{X}_*)^{-1} \boldsymbol{X}_*^T \boldsymbol{V}_*^{-1}] \boldsymbol{B}_* \boldsymbol{B}_*^T \right) \\ & \qquad - (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}\hat{\boldsymbol{f}})^T \boldsymbol{V}^{-1} \boldsymbol{B}_* \boldsymbol{B}_*^T \boldsymbol{V}^{-1} (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}\hat{\boldsymbol{f}}) \} \\ &= \left(-\frac{1}{2}\right) \cdot \left\{ \mathrm{Tr}(\boldsymbol{P}_* \boldsymbol{B}_* \boldsymbol{B}_*^T) - (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}\hat{\boldsymbol{f}})^T \boldsymbol{V}^{-1} \boldsymbol{B}_* \boldsymbol{B}_*^T \boldsymbol{V}^{-1} (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}\hat{\boldsymbol{f}}) \right\} \end{aligned}$$

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.

Now, we want to proceed to find Fisher information matrix for $(\tau, \boldsymbol{\theta})$. In other words, we need to find the second-order partial derivatives of ℓ_R with respect to $(\tau, \boldsymbol{\theta})$; and then the expected values of the second-order partial derivative.

$$\begin{aligned}
&\frac{\partial^2 \ell_R}{\partial \tau^2} \\
= \;& \frac{\partial}{\partial \tau}\left(\frac{\partial \ell_R}{\partial \tau}\right) \\
= \;& \frac{\partial}{\partial \tau}\left\{-\frac{1}{2}\operatorname{Tr}(\boldsymbol{P}_*\boldsymbol{B}_*\boldsymbol{B}_*^T) + \frac{1}{2}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{N}\hat{\boldsymbol{f}})^T\boldsymbol{V}^{-1}\boldsymbol{B}_*\boldsymbol{B}_*^T\boldsymbol{V}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{N}\hat{\boldsymbol{f}})\right\} \\
= \;& -\frac{1}{2}\operatorname{Tr}(\frac{\partial}{\partial \tau}\boldsymbol{P}_*\boldsymbol{B}_*\boldsymbol{B}_*^T) + \frac{\partial}{\partial \tau}\left[\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{N}\hat{\boldsymbol{f}})^T\boldsymbol{V}^{-1}\boldsymbol{B}_*\boldsymbol{B}_*^T\boldsymbol{V}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{N}\hat{\boldsymbol{f}})\right] \\
= \;& -\frac{1}{2}\operatorname{Tr}(-\boldsymbol{V}_*^{-1}\frac{\partial \boldsymbol{V}_*}{\partial \tau}\boldsymbol{V}_*^{-1}\boldsymbol{B}_*\boldsymbol{B}_*^T) \\
= \;& \frac{1}{2}\operatorname{Tr}(\boldsymbol{P}_*\boldsymbol{B}_*\boldsymbol{B}_*^T\boldsymbol{P}_*\boldsymbol{B}_*\boldsymbol{B}_*^T)
\end{aligned}$$

and

$$E\left(\frac{\partial^2 \ell_R}{\partial \tau^2}\right) = \frac{1}{2}\operatorname{Tr}(\boldsymbol{P}_*\boldsymbol{B}_*\boldsymbol{B}_*^T\boldsymbol{P}_*\boldsymbol{B}_*\boldsymbol{B}_*^T).$$

$$\begin{aligned}
&\frac{\partial^2 \ell_R}{\partial \tau \partial \theta_j} \\
= \;& \frac{\partial}{\partial \tau}\left(\frac{\partial \ell_R}{\partial \theta_j}\right) \\
= \;& \frac{\partial}{\partial \tau}\left\{-\frac{1}{2}\operatorname{Tr}(\boldsymbol{P}_*\frac{\partial \boldsymbol{V}}{\partial \theta_j}) + \frac{1}{2}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{N}\hat{\boldsymbol{f}})^T\boldsymbol{V}^{-1}\frac{\partial \boldsymbol{V}}{\partial \theta_j}\boldsymbol{V}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{N}\hat{\boldsymbol{f}})\right\} \\
= \;& -\frac{1}{2}\operatorname{Tr}(\frac{\partial}{\partial \tau}\boldsymbol{P}_*\frac{\partial \boldsymbol{V}}{\partial \theta_j}) + \frac{\partial}{\partial \tau}\left[\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{N}\hat{\boldsymbol{f}})^T\boldsymbol{V}^{-1}\frac{\partial \boldsymbol{V}}{\partial \theta_j}\boldsymbol{V}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{N}\hat{\boldsymbol{f}})\right] \\
= \;& -\frac{1}{2}\operatorname{Tr}(-\boldsymbol{V}_*^{-1}\frac{\partial \boldsymbol{V}_*}{\partial \tau}\boldsymbol{V}_*^{-1}\frac{\partial \boldsymbol{V}}{\partial \theta_j}) \\
= \;& \frac{1}{2}\operatorname{Tr}(\boldsymbol{P}_*\boldsymbol{B}_*\boldsymbol{B}_*^T\boldsymbol{P}_*\frac{\partial \boldsymbol{V}}{\partial \theta_j})
\end{aligned}$$

and

$$E\left(\frac{\partial^2 \ell_R}{\partial \tau \partial \theta_j}\right) = \frac{1}{2}\operatorname{Tr}(\boldsymbol{P}_*\boldsymbol{B}_*\boldsymbol{B}_*^T\boldsymbol{P}_*\boldsymbol{B}_*\boldsymbol{B}_*^T).$$

$$\begin{aligned}
&\frac{\partial^2 \ell_R}{\partial \theta_j \partial \theta_k} \\
= \;& \frac{\partial}{\partial \theta_j}\left(\frac{\partial \ell_R}{\partial \theta_k}\right) \\
= \;& \frac{\partial}{\partial \theta_j}\left\{-\frac{1}{2}\operatorname{Tr}(\boldsymbol{P}_*\frac{\partial \boldsymbol{V}}{\partial \theta_k}) + \frac{1}{2}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{N}\hat{\boldsymbol{f}})^T\boldsymbol{V}^{-1}\frac{\partial \boldsymbol{V}}{\partial \theta_k}\boldsymbol{V}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{N}\hat{\boldsymbol{f}})\right\} \\
= \;& -\frac{1}{2}\operatorname{Tr}(\frac{\partial}{\partial \theta_j}\boldsymbol{P}_*\frac{\partial \boldsymbol{V}}{\partial \theta_k}) + \frac{\partial}{\partial \theta_j}\left[\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{N}\hat{\boldsymbol{f}})^T\boldsymbol{V}^{-1}\frac{\partial \boldsymbol{V}}{\partial \theta_k}\boldsymbol{V}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{N}\hat{\boldsymbol{f}})\right] \\
= \;& -\frac{1}{2}\operatorname{Tr}(\boldsymbol{P}_*\frac{\partial^2 \boldsymbol{V}}{\partial \theta_j \partial \theta_k} - \boldsymbol{V}^{-1}\frac{\partial \boldsymbol{V}}{\partial \theta_j}\boldsymbol{V}^{-1}\frac{\partial \boldsymbol{V}}{\partial \theta_k}) \\
&+\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{N}\hat{\boldsymbol{f}})^T[-\boldsymbol{V}^{-1}\frac{\partial \boldsymbol{V}}{\partial \theta_j}\boldsymbol{V}^{-1}\frac{\partial \boldsymbol{V}}{\partial \theta_k}\boldsymbol{V}^{-1} + \boldsymbol{V}^{-1}\frac{\partial^2 \boldsymbol{V}}{\partial \theta_j \partial \theta_k}\boldsymbol{V}^{-1} \\
&\qquad -\boldsymbol{V}^{-1}\frac{\partial \boldsymbol{V}}{\partial \theta_k}\boldsymbol{V}^{-1}\frac{\partial \boldsymbol{V}}{\partial \theta_j}\boldsymbol{V}^{-1}](\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{N}\hat{\boldsymbol{f}}) \\
= \;& -\frac{1}{2}\operatorname{Tr}\left(\boldsymbol{P}_*\frac{\partial^2 \boldsymbol{V}}{\partial \theta_j \partial \theta_k} - \boldsymbol{P}_*\frac{\partial \boldsymbol{V}}{\partial \theta_j}\boldsymbol{P}_*\frac{\partial \boldsymbol{V}}{\partial \theta_k}\right) \\
&-\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{N}\hat{\boldsymbol{f}})^T\left(\boldsymbol{P}_*\frac{\partial \boldsymbol{V}}{\partial \theta_j}\boldsymbol{P}_*\frac{\partial \boldsymbol{V}}{\partial \theta_k} - \boldsymbol{P}_*\frac{\partial^2 \boldsymbol{V}}{\partial \theta_j \partial \theta_k} + \boldsymbol{P}_*\frac{\partial \boldsymbol{V}}{\partial \theta_k}\boldsymbol{P}_*\frac{\partial \boldsymbol{V}}{\partial \theta_j}\right) \\
&\qquad \cdot\boldsymbol{V}^{-1}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{N}\hat{\boldsymbol{f}})
\end{aligned}$$

and

$$\begin{aligned}
E\left(\frac{\partial^2 \ell_R}{\partial \theta_j \partial \theta_k}\right) &= \frac{1}{2}\{\operatorname{Tr}\left(\boldsymbol{P}_*\frac{\partial^2 \boldsymbol{V}}{\partial \theta_j \partial \theta_k} - \boldsymbol{P}_*\frac{\partial \boldsymbol{V}}{\partial \theta_j}\boldsymbol{P}_*\frac{\partial \boldsymbol{V}}{\partial \theta_k}\right) \\
&\quad + \operatorname{Tr}\left(\boldsymbol{P}_*\frac{\partial \boldsymbol{V}}{\partial \theta_j}\boldsymbol{P}_*\frac{\partial \boldsymbol{V}}{\partial \theta_k} - \boldsymbol{P}_*\frac{\partial^2 \boldsymbol{V}}{\partial \theta_j \partial \theta_k} + \boldsymbol{P}_*\frac{\partial \boldsymbol{V}}{\partial \theta_k}\boldsymbol{P}_*\frac{\partial \boldsymbol{V}}{\partial \theta_j}\right)\} \\
&= \frac{1}{2}\operatorname{Tr}\left(\boldsymbol{P}_*\frac{\partial \boldsymbol{V}}{\partial \theta_j}\boldsymbol{P}_*\frac{\partial \boldsymbol{V}}{\partial \theta_k}\right)
\end{aligned}$$

as desired.

3.4 Discussion and future work

I provide essential theoretical proofs and derivations for results from Zhang *et al* [45], which pave the way for the theoretical development of the next chapter, Chapter 4. No simulation results is shown here since the paper already contains it, though I wrote the R program myself and have replicated both the simulation and the analysis of data application. The R code is available upon request.

In addition, I also updated the findings from Zhang *et al* [45] and have shown in Lemma 3 that as the smoothing parameter λ goes to 0, the biases in the estimators of the random effects $\hat{\boldsymbol{b}}_i$ and the stochastic process $\hat{\boldsymbol{U}}_i$ also goes to zero.

However, I would like to point out some other issues with that paper. First, there is a discrepancy between the covariances of the estimators of the random effects $\hat{\boldsymbol{b}}_i$ stated in the paper and the results that I derive in Lemma 4. In the paper,

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

whereas, what I find is

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

as shown in Lemma 4. The discrepancy is shown in colour red.

Second, an important omission in the simulation study is identified. Zhang *et al* [45] provided an estimate of τ, which is the inverse of the smoothing parameter, in the data analysis; however, mysteriously, they left out the estimate of τ in the simulation studies. Having run a set of simulation study myself (results are not shown here), I can speculate the reason why the estimates of τ was not presented in their simulation results is due to the poor quality of its estimation; specifically, very poor relative bias. We would need to do further research to determine why τ's performance is poor in my simulation study; and very likely in their simulation and possibly to improve upon its estimation.

In spite of the issues mentioned above, the model proposed in Zhang *et al* [45] provide a flexible yet easy-to-implement framework to model complex univariate longitudinal data. It incorporated a nonparametric term into the model as part of the population mean structure to better model the complex time course of the hormone levels. The incorporation of the stochastic process gives more structure to model the within-subject covariance. The theoretical derivation shown in this chapter serve as a springboard to Chapter 4 where a semiparametric stochastic mixed effects model is proposed to model the bivariate longitudinal data; and Chapter 5 where a prediction procedure of a future cycle is presented under the univariate semiparametric stochastic mixed model framework.

Chapter 4

Semiparametric Stochastic Mixed Effects Models for Bivariate Longitudinal Data

4.1 Introduction

In this chapter, I am interested in modelling time courses for the estrogen and progesterone metabolite profiles for a single cycle, the effects of covariates on the hormone excretion, and the potential correlation between the two hormones. Joint modeling of the hormone profiles of estrogen and progesterone is challenging. First, the time courses of the univariate hormones profiles is complex such that to model using a simple parametric function, such as linear mixed effects model, is insufficient; see Figure 4.1. Second, multiple layers of correlation structures, say within-subject correlation between the bivariate hormones at different time points, also present a challenge.

Some of the univariate techniques discussed in Chapter 1 have been extended to the bivariate case. For example, Sy, Taylor & Cumberland[30] employed multivariate stochastic processes to jointly model bivariate longitudinal data. Funatogawa, Funatogawa & Ohashi [12] proposed a bivariate autoregressive linear mixed effects model for longitudi-

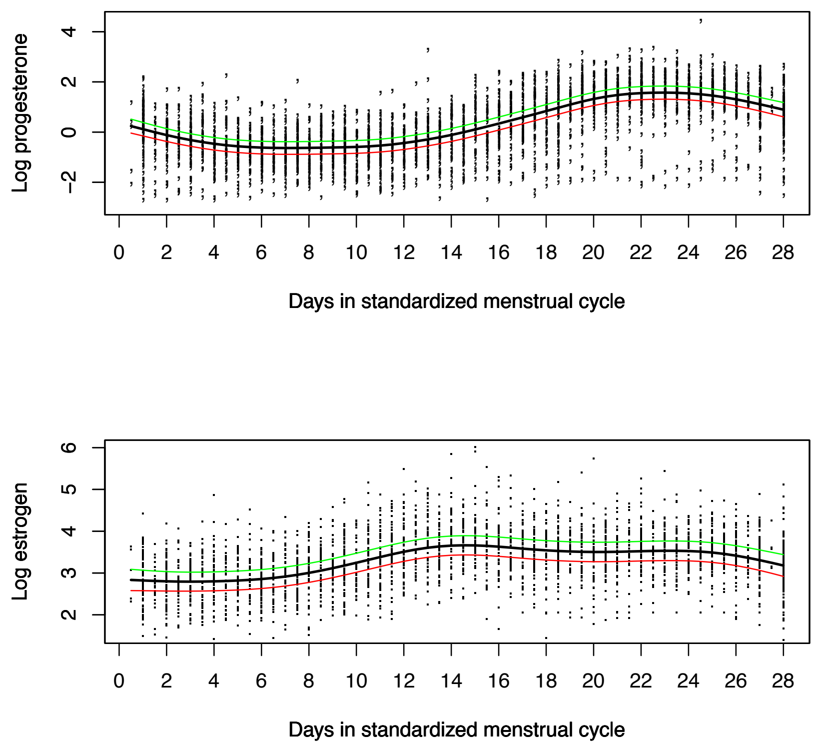


Figure 4.1: Plots of log progesterone and log estrogen levels against days in a standardized menstrual cycle, superimposed by estimated population mean curve $\hat{f}_1$ and $\hat{f}_2$.

nal data. And shared random effects is another approach for jointly modeling outcomes [33]. More recently, Raffa & Dubin[25] modelled bivariate longitudinal responses, from outcomes of different data types, via a mixed effects hidden Markov modeling approach.

And, most relevant in terms of the type of data focused upon in this paper, Liu *et al*[21] extended the univariate state space model in time series analysis, and proposed a bivariate hierarchical state space model to bivariate circadian rhythmic longitudinal responses; each response is modelled by a hierarchical state space model, with both population-average and subject-specific components; and the bivariate model is constructed by linking the univariate models based on the hypothesized relationship between the responses.

It is worth noting that some Bayesian semiparametric modelling was proposed in the literature. For example, Ghosh & Hanson [13] extended the work of Zhang *et al*[45] and Zeger & Diggle [43] to the multivariate longitudinal data from a Bayesian prospective. A multivariate mixture of Polya trees prior distribution was used to model the multivariate random effects. Das *et al* [6] proposed a Bayesian semiparametric model for bivariate sparse longitudinal data.

In this chapter, we extend Zhang *et al* [45] and propose a bivariate semiparametric stochastic mixed model for bivariate repeated measures data. The bivariate model uses parametric fixed effects and smooth nonparametric functions for each of the two underlying time effects. The between-subject correlations are modelled using separate but correlated random effects and the within-subject correlations by a bivariate Gaussian field. The model allows us to investigate the relationship of the two responses through the correlation of the random effects and the bivariate Gaussian field, which can not only describe the concurrent relationship of the two responses but also allows for characterizations of the relationship across time points. We derive maximum penalized likelihood estimators for both the fixed effects regression coefficients and the nonparametric time functions. The smoothing parameters and all variance components are estimated simultaneously using restricted maximum likelihood.

The chapter is organized as followed. Section 4.2 specifies the proposed model with assumptions. Section 4.3 provides estimation and inference procedures. Specifically, Section 4.3.1 gives estimation procedures for the model parameters, the nonparametric components, random effects and the Gaussian fields. Section 4.3.2 specifies the biases and covariances for all the estimators given in Section 4.3.1; and Section 4.3.3 concludes this section by providing the estimation procedures of the smoothing parameters and variance components. Section 4.4 extends the model proposed in Section 4.2 to accommodate bivariate periodic longitudinal data with multiple cycles. Section 4.5 investigates the proposed methodology through a simulation study. Section 4.6 illustrates the model by analyzing bivariate longitudinal female hormone data collected daily over a single menstrual cycle, and finally, Section 4.7 provides a summary of our proposed model, and discusses challenges and future work. Technical details are included in Appendix A.4.

4.2 The Bivariate Semiparametric Stochastic Mixed Effects Model

4.2.1 The Proposed Model Specifications and Assumptions

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

$$\boldsymbol{Y}_{ij} = \boldsymbol{X}_{ij}^T\boldsymbol{\beta} + \boldsymbol{f}(t_{ij}) + \boldsymbol{Z}_{ij}^T\boldsymbol{b}_i + \boldsymbol{U}_i(t_{ij}) + \boldsymbol{\epsilon}_{ij}, \tag{4.1}$$

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

$$\boldsymbol{C}_i(s,t) = \begin{pmatrix} \sqrt{\xi_1(s)\xi_1(t)}\eta_1(\rho_1; s,t) & \sqrt{\xi_1(s)\xi_2(t)}\eta_3(\rho_3; s,t) \\ \sqrt{\xi_2(s)\xi_1(t)}\eta_3(\rho_3; t,s) & \sqrt{\xi_2(s)\xi_2(t)}\eta_2(\rho_2; s,t) \end{pmatrix} \tag{4.2}$$

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

This model (4.1) is an extension to the model proposed in Zhang *et al* [45], where a univariate semiparametric stochastic mixed model for longitudinal data was proposed. The challenge here is that we are modeling a bivariate longitudinal response model, which is achieved by modeling a joint distribution of the random effects and the bivariate Gaussian field. The two random effects are assumed separate but correlated. This specification of random effect structure is preferred as opposed to the shared random effects due to the flexibility that the structure allows over imposing a common random effect. The distributions of the two random effects can be potentially distinct, with different distributions or the same distribution with different parameters.

4.2.2 Matrix Notation

To make inferences from the model (4.1), we will write the model in matrix form - first, in subject level; then, over all subjects.

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

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

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

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

Then, the proposed bivariate semiparametric stochastic mixed model (4.1) can be written as

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

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

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

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{N}_1\boldsymbol{f}_1 + \boldsymbol{N}_2\boldsymbol{f}_2 + \boldsymbol{Zb} + \boldsymbol{U} + \boldsymbol{\epsilon}, \tag{4.3}$$

with assumptions

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

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

4.2.3 Covariance Structures

(a) Mean and Covariance of the Proposed Model

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

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

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

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

The mean response for a specific subject is

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

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

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

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

(b) Modelling the Relationship of the Bivariate Responses

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

$$\begin{aligned}
& \text{cov}(Y_{1ij}, Y_{2ik}) \\
= \;& \text{cov}(\boldsymbol{X}_{\boldsymbol{1ij}}^T\boldsymbol{\beta_1} + f_1(t_{ij}) + \boldsymbol{Z}_{\boldsymbol{1ij}}^T\boldsymbol{b_{1i}} + U_{1i}(t_{ij}) + \epsilon_{1ij}, \boldsymbol{X}_{\boldsymbol{2ik}}^T\boldsymbol{\beta_2} + f_2(t_{ik}) + \boldsymbol{Z}_{\boldsymbol{2ik}}^T\boldsymbol{b_{2i}} + U_{2i}(t_{ik}) + \epsilon_{2ik}) \\
= \;& \text{cov}(\boldsymbol{Z}_{\boldsymbol{1ij}}^T\boldsymbol{b_{1i}} + U_{1i}(t_{ij}) + \epsilon_{1ij}, \boldsymbol{Z}_{\boldsymbol{2ik}}^T\boldsymbol{b_{2i}} + U_{2i}(t_{ik}) + \epsilon_{2ik}) \\
= \;& \boldsymbol{Z}_{\boldsymbol{1ij}}^T\text{cov}(\boldsymbol{b_{1i}}, \boldsymbol{b_{2i}})\boldsymbol{Z_{2ik}} + \text{cov}(U_{1i}(t_{ij}), U_{2i}(t_{ik})) \\
= \;& \boldsymbol{Z}_{\boldsymbol{1ij}}^T\boldsymbol{G}'\boldsymbol{Z_{2ik}} + \sqrt{\xi_1(t_{ij})\xi_2(t_{ik})}\eta_3(\rho_3; t_{ij}, t_{ik}),
\end{aligned} \tag{4.4}$$

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

4.2.4 The Gaussian Field Specification

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

There are potentially many choices available: Wiener process or Brownian motion [31]; an integrated Wiener process and so on. One particular Gaussian process/field worthy of mentioning is the Ornstein-Uhlenbeck (OU) process [18] which has a correlation function that decays exponentially over time $\text{corr}(U_i(t), U_i(s)) = \exp\{-\alpha|s-t|\}$. The variance function for OU process $\xi(t) = \sigma^2/2a$ is a constant, thus the process is strictly stationary. When $\xi(t)$ varies over time, then the process becomes nonhomogenesous (NOU) and, for example, we can assume $\xi(t) = \exp(a_0 + a_1 t + a_1 t^2)$.

4.3 Estimation and Inference

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

The proposed model (4.3) implies the *marginal model*

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

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

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

for given fixed variance parameters. We estimate the parameters $\boldsymbol{\beta}$, $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ by maximizing the penalized likelihood [35]:

$$\ell(\boldsymbol{\beta}, \boldsymbol{f}_1, \boldsymbol{f}_2; \boldsymbol{Y}) - \lambda_1 \int_a^b [f_1''(t)]^2 dt - \lambda_2 \int_a^b [f_2''(t)]^2 dt = \ell(\boldsymbol{\beta}, \boldsymbol{f}_1, \boldsymbol{f}_2; \boldsymbol{Y}) - \lambda_1 \boldsymbol{f}_1^T \boldsymbol{K} \boldsymbol{f}_1 - \lambda_2 \boldsymbol{f}_2^T \boldsymbol{K} \boldsymbol{f}_2 \tag{4.5}$$

where λ_1 and λ_2 are smoothing parameters; a and b is the range of time t; and $\boldsymbol{K}$ is the nonnegative definite smoothing matrix, defined in Equation (2.3) in Green & Silverman[17]. Since observation time points t_{ij} are assumed to be the same for both responses, the smoothing matrix $\boldsymbol{K}$, which is determined by time increments, is also the same. The resulting estimators for the nonparametric functions are the natural cubic spline estimators of $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$.

Differentiation of (4.5) 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

$$\begin{pmatrix} \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}_1 & \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{N}_2 \\ \boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{X} & \boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K} & \boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_2 \\ \boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{X} & \boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_1 & \boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{f}_1 \\ \boldsymbol{f}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y} \\ \boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{Y} \\ \boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{Y} \end{pmatrix}, \tag{4.6}$$

where $\boldsymbol{W} = \boldsymbol{V}^{-1}$, for details see Appendix A.4.1. To study the theoretical properties of the estimates, such as bias and covariance, we derive the closed-form solutions for $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{f}}_1$ and $\hat{\boldsymbol{f}}_2$

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{Y} \tag{4.7}$$

$$\hat{\boldsymbol{f}}_1 = (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{Y} \tag{4.8}$$

$$\hat{\boldsymbol{f}}_2 = (\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{Y}, \tag{4.9}$$

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 $\boldsymbol{W}_2 = \boldsymbol{W} - \boldsymbol{W}\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}$. Empirically, all inverses $(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}$, $(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}$, and $(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}$ exist. Note that they are the corresponding block-diagnoal elements of the inverse of the coefficient matrix in equation (4.6).

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

$$\hat{\boldsymbol{b}}_i = E(\boldsymbol{b}|\boldsymbol{Y}) = \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{f}}_{1i} - \hat{\boldsymbol{f}}_{2i}) \tag{4.10}$$

and similarly,

$$\hat{\boldsymbol{U}}_i(\boldsymbol{s}_i) = \boldsymbol{\Gamma}(\boldsymbol{s}_i, \boldsymbol{t}_i)\boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{f}}_{1i} - \hat{\boldsymbol{f}}_{2i}) \tag{4.11}$$

where $\hat{\boldsymbol{f}}_{1i} = \boldsymbol{N}_{1i}\hat{\boldsymbol{f}}_1$ and $\hat{\boldsymbol{f}}_{2i} = \boldsymbol{N}_{2i}\hat{\boldsymbol{f}}_2$. Technical details for this subsection is included in the Appendix A.4.2.

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

From closed-form solutions of estimators from equation (4.7) (4.8) and (4.9) in Section 4.3.1, the biases of the estimators $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{f}}_1$ and $\hat{\boldsymbol{f}}_2$ can be easily calculated, see Appendix A.4.3, and we have

$$E(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta} = (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x(\boldsymbol{N}_1\boldsymbol{f}_1 + \boldsymbol{N}_2\boldsymbol{f}_2) \tag{4.12}$$

$$E(\hat{\boldsymbol{f}}_1) - \boldsymbol{f}_1 = (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_2\boldsymbol{f}_2 - \lambda_1\boldsymbol{K}\boldsymbol{f}_1) \tag{4.13}$$

$$E(\hat{\boldsymbol{f}}_2) - \boldsymbol{f}_2 = (\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_1\boldsymbol{f}_1 - \lambda_2\boldsymbol{K}\boldsymbol{f}_2). \tag{4.14}$$

Similarly, the expected values of the estimators in (4.10) and (4.11) 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 $\mathbf{0}$ as both smoothing parameters $\lambda_1 \to 0$ and $\lambda_2 \to 0$, see Lemma 5 in Appendix A.4.3.

For covariances, simple calculation using (4.7) (4.8) and (4.9) gives the covariance of $\hat{\boldsymbol{\beta}}$

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

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

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

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

The covariances of the estimators in (4.10) and (4.11) for the subject-specific random effects $\boldsymbol{b}_i$ and for the subject-specific Gaussian field $\boldsymbol{U}_i(\boldsymbol{s}_i)$ are

$$\text{cov}(\hat{\boldsymbol{b}}_i - \boldsymbol{b}_i) = \boldsymbol{D} - \boldsymbol{D}\boldsymbol{Z}_i^T \boldsymbol{W}_i \boldsymbol{Z}_i \boldsymbol{D} + \boldsymbol{D}\boldsymbol{Z}_i^T \boldsymbol{W}_i \boldsymbol{\chi}_i \boldsymbol{C}^{-1} \boldsymbol{\chi}^T \boldsymbol{W} \boldsymbol{\chi} \boldsymbol{C}^{-1} \boldsymbol{\chi}_i^T \boldsymbol{W}_i \boldsymbol{Z}_i \boldsymbol{D} \qquad (4.15)$$

and

$$\begin{aligned} & \text{cov}(\hat{\boldsymbol{U}}_i(\boldsymbol{s}_i) - \boldsymbol{U}_i(\boldsymbol{s}_i)) \\ = \; & \boldsymbol{\Gamma}(\boldsymbol{s}_i, \boldsymbol{s}_i) - \boldsymbol{\Gamma}(\boldsymbol{s}_i, \boldsymbol{t}_i) \boldsymbol{W}_i \boldsymbol{\Gamma}(\boldsymbol{s}_i, \boldsymbol{t}_i)^T + \boldsymbol{\Gamma}(\boldsymbol{s}_i, \boldsymbol{t}_i) \boldsymbol{W}_i \boldsymbol{\chi}_i \boldsymbol{C}^{-1} \boldsymbol{\chi}^T \boldsymbol{W} \boldsymbol{\chi} \boldsymbol{C}^{-1} \boldsymbol{\chi}_i^T \boldsymbol{W}_i \boldsymbol{\Gamma}(\boldsymbol{s}_i, \boldsymbol{t}_i)^T, \end{aligned}$$

where $\boldsymbol{\chi}_i = \begin{pmatrix} \boldsymbol{X}_i & \boldsymbol{N}_{1i} & \boldsymbol{N}_{2i} \end{pmatrix}$ and $\boldsymbol{\chi} = \begin{pmatrix} \boldsymbol{X} & \boldsymbol{N}_1 & \boldsymbol{N}_2 \end{pmatrix}$. The calculation of the results is non-trivial and is provided in Appendix A.4.4.

4.3.3 Estimation of the Smoothing Parameters and Variance Parameters

To estimate the smoothing parameters and variance components jointly using the restricted maximum likelihood (REML), we rewrite the proposed semiparametric model as a modified linear mixed model. Specifically, by Green [16], the nonparametric functions $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ under a one-to-one linear transformation are

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

where δ_1 and δ_2 are vectors of dimensions 2; $a_1$and a_2 are of dimensions $r-2$; $\boldsymbol{B} = \boldsymbol{L}(\boldsymbol{L}^T\boldsymbol{L})^{-1}$ and $\boldsymbol{L}$ is $r \times (r-2)$ full-rank matrix satisfying $\boldsymbol{K} = \boldsymbol{L}\boldsymbol{L}^T$ and $\boldsymbol{L}^T\boldsymbol{T} = 0$. Thus the proposed semiparametric mixed model (4.3) can be rewritten as a modified linear mixed model [45],

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{N}_1\boldsymbol{T}\boldsymbol{\delta}_1 + \boldsymbol{N}_1\boldsymbol{B}\boldsymbol{a}_1 + \boldsymbol{N}_2\boldsymbol{T}\boldsymbol{\delta}_2 + \boldsymbol{N}_2\boldsymbol{B}\boldsymbol{a}_2 + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{U} + \boldsymbol{\epsilon}, \qquad (4.16)$$

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

Under the modified linear mixed model (4.16), 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 $\boldsymbol{X}_* = [\boldsymbol{X}, \boldsymbol{N}_1 \boldsymbol{T}, \boldsymbol{N}_2 \boldsymbol{T}]$. Taking the derivative of ℓ_R with respect to τ_1, τ_1, and $\boldsymbol{\theta}$ and using the identity $\boldsymbol{V}_*^{-1}(\boldsymbol{Y} - \boldsymbol{X}_* \hat{\boldsymbol{\beta}}_*) = \boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1 \hat{\boldsymbol{f}}_1 - \boldsymbol{N}_2 \hat{\boldsymbol{f}}_2)$, the estimating equations for smoothing parameters τ_1, τ_2 and variance components $\boldsymbol{\theta}$ can be obtained:

$$\frac{\partial \ell_R}{\partial \tau_1} = -\frac{1}{2}\text{tr}(\boldsymbol{P}_* \boldsymbol{B}_{1*} \boldsymbol{B}_{1*}^T) + \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1 \hat{\boldsymbol{f}}_1 - \boldsymbol{N}_2 \hat{\boldsymbol{f}}_2)^T \boldsymbol{V}^{-1} \boldsymbol{B}_{1*} \boldsymbol{B}_{1*}^T \boldsymbol{V}^{-1} (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1 \hat{\boldsymbol{f}}_1 - \boldsymbol{N}_2 \hat{\boldsymbol{f}}_2), \tag{4.17}$$

$$\frac{\partial \ell_R}{\partial \tau_2} = -\frac{1}{2}\text{tr}(\boldsymbol{P}_* \boldsymbol{B}_{2*} \boldsymbol{B}_{2*}^T) + \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1 \hat{\boldsymbol{f}}_1 - \boldsymbol{N}_2 \hat{\boldsymbol{f}}_2)^T \boldsymbol{V}^{-1} \boldsymbol{B}_{2*} \boldsymbol{B}_{2*}^T \boldsymbol{V}^{-1} (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1 \hat{\boldsymbol{f}}_1 - \boldsymbol{N}_2 \hat{\boldsymbol{f}}_2), \tag{4.18}$$

and

$$\frac{\partial \ell_R}{\partial \theta_j} = -\frac{1}{2}\text{tr}(\boldsymbol{P}_* \frac{\partial \boldsymbol{V}}{\partial \theta_j}) + \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1 \hat{\boldsymbol{f}}_1 - \boldsymbol{N}_2 \hat{\boldsymbol{f}}_2)^T \boldsymbol{V}^{-1} \frac{\partial \boldsymbol{V}}{\partial \theta_j} \boldsymbol{V}^{-1} (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{N}_1 \hat{\boldsymbol{f}}_1 - \boldsymbol{N}_2 \hat{\boldsymbol{f}}_2), \tag{4.19}$$

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 (4.17), (4.18) and (4.19),

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

where

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

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

and

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

4.4 Bivariate semiparametric mixed effects model for perodic longitudinal data from multiple cycles

We extend methods from Section 4.2 to fit bivariate periodic longitudinal data with multiple cycles. The model was motivated by the dataset described in Chapter 2, where the hormone data changes periodically from one cycle to another, and both the mean and the variance demonstrate periodic features, see Figure 4.2. Specifically, Denote $\boldsymbol{Y}_{ij\ell} = (Y_{ij1}, Y_{ij2})^T$ to be the bivariate responses for the ith subject at time point $t_{ij\ell}$, $i = 1, \ldots, m$, $j = 1, \ldots, n_i$ and $\ell = 1, \ldots, l$. The proposed model for bivariate periodic longitudinal responses is

$$\boldsymbol{Y}_{ij\ell} = \boldsymbol{X}_{ij}^T\boldsymbol{\beta} + \boldsymbol{f}(t_{ij\ell}) + \boldsymbol{Z}_{ij}^T\boldsymbol{b}_i + \boldsymbol{U}_i(t_{ij\ell}) + \boldsymbol{\epsilon}_{ij\ell}, \tag{4.20}$$

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

$$\boldsymbol{C}_i(s,t) = \begin{pmatrix} \sqrt{\xi_1(s)\xi_1(t)}\eta_1(\rho_1; s,t) & \sqrt{\xi_1(s)\xi_2(t)}\eta_3(\rho_3; s,t) \\ \sqrt{\xi_2(s)\xi_1(t)}\eta_3(\rho_3; t,s) & \sqrt{\xi_2(s)\xi_2(t)}\eta_2(\rho_2; s,t) \end{pmatrix} \tag{4.21}$$

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 $\epsilon_{ij} = (\epsilon_{1ij}, \epsilon_{2ij})^T$ are bivariate normal with mean $\boldsymbol{0}$ and variance $\mathrm{diag}(\sigma_1^2, \sigma_2^2)$. We assume $\boldsymbol{b}_i$ to be $2q$-dimensional normal with mean zero and unstructured covariance matrix $\boldsymbol{G}(\boldsymbol{\phi})$, and that the random effects, the stochastic process and the measurement error to be mutually independent; and that the fixed effects $\boldsymbol{\beta}$ is across all cycles.

This model (4.20) is an extension to the model (4.1) and the model proposed in Zhang *et al* [44]), where a univariate semiparametric stochastic mixed model for periodic longitudinal data was proposed. A key difference from model (4.1) in Section 4.2 is that both the nonparametric function and the variance of the Gaussian fields are constrained to be periodic, thus resulting in the specification of the smoothing matrix $\boldsymbol{K}$ and the variance of the Gaussian fields to be different. Also by taking advantage of the periodicity of the data, the incidence matrix can be largely reduced by only estimating the nonparametric function for one period. Preliminary application to the hormone shows that the method works well, see Figure 4.2, where the smooth curve is the estimated nonparametric function and its pointwise 95% confidence intervals. This is still ongoing research and more will be touched upon in the discussion section.

4.5 Simulations

4.5.1 A Simulation Study using NOU

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

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

where b_{1i} and b_{2i} are separate but correlated random intercepts following a bivariate normal distribution with mean $\mathbf{0}$ and unstructured covariance matrix $\boldsymbol{D}(\phi_1, \phi_{1,2}, \phi_2)$; U_{1i} and U_{2i} are simulated from mean $\mathbf{0}$ bivariate NOU fields modeling serial correlation, with variance function $\text{var}(U_{1i}(t)) = \exp\{a_{10} + a_{11}t + a_{12}t^2\}$, var $(U_{2i}(t)) = \exp\{a_{20} + a_{21}t + a_{22}t^2\}$ and

$\mathrm{corr}(U_{1i}(t), U_{1i}(s)) = \rho_1^{|s-t|}$ $\mathrm{corr}(U_{2i}(t), U_{2i}(s)) = \rho_2^{|s-t|}$, i.e. the covariance function for the bivariate NOU field is

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

ϵ_{1ij} and ϵ_{2ij} are simulated from a mean $\mathbf{0}$ bivariate normal distribution with covariance $\mathrm{diag}(\sigma_1^2, \sigma_2^2)$; and the nonparametric smooth functions are generated from $f_1(t) = 5\sin(2\pi/28)\,t$ and $f_2(t) = 3\cos(2\pi/28)\,t$. Covariate age is randomly generated between 20 and 44 years old by R.

Table 4.1: Estimates of regression coefficients, variance parameter and smoothing parameter for the progesterone and estrogen data.

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

Table 4.1 records the simulation results for estimates of model parameters based on 500 simulation replicates and 30 subjects, where the model-based SE is yet to be computed for

the smoothing parameters and the variance components. 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 4.3. Figure 4.4 shows that model standard errors of estimates of $\hat{f}_1$ and $\hat{f}_2$ agree quite well with the empirical standard errors.

Figure 4.5 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.

4.5.2 Misspecification of Gaussian Fields

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

Based on 400 simulations results for each choice of Gaussian field, the estimates of regression coefficients and random intercepts are fairly robust with bias close to zero even when the bivariate Gaussian field is misspecified as bivariate OU or Wiener field. The estimates for the smoothing parameters is much more biased for both bivariate OU or Wiener field, though misspecification in OU field would lead to less bias than that in Wiener. The estimates for variance of the measurement error are almost unbiased with the misspecification of bivariate Wiener field; whereas it is 20% more biased in the case of misspecification of bivariate OU field. In conclusion, misspecification of Gaussian field does not have a major influence if more emphasis is placed on the estimates of regression coefficients, yet the estimates of smoothing parameters and some variance components can vary significantly from the true values in the presence of misspecification of Gaussian field.

4.6 Bivariate Longitudinal Hormone Data Analysis

Denoting $\{(Y_{1ij}, Y_{2ij})\}$ the j^{th} log-transformed progesterone and estrogen values measured at standardized day t_{ij} since menstruation for the i^{th} woman, we consider the following bivariate semiparametric stochastic mixed model for the hormone dataset described in Section 2.1:

$$\begin{aligned} Y_{1ij} &= \text{age}_i^T\beta_{11} + \text{underWeight}_i^T\beta_{12} + \text{overWeight}_i^T\beta_{13} + f_1(t_{ij}) + b_{1i} + U_{1i}(t_{ij}) + \epsilon_{1ij} \\ Y_{2ij} &= \text{age}_i^T\beta_{21} + \text{underWeight}_i^T\beta_{22} + \text{overWeight}_i^T\beta_{23} + f_2(t_{ij}) + b_{2i} + U_{2i}(t_{ij}) + \epsilon_{2ij} \\ & \quad i = 1, \ldots, 30; j = 1, \ldots, n_i; t_{ij} \in \{0.5, 1.0, \ldots, 28\} \end{aligned}$$

where the specifications of random intercepts b_{1i} and b_{2i}, the bivariate Gaussian field U_{1i} and U_{2i}, the measurement errors ϵ_{1ij} and ϵ_{2ij}, and the nonparametric smooth functions are the same as those in the simulation study. Covariates underWeight and overWeight are indicator variables, which is characterized by Body Mass Index (BMI) where if BMI is less than 19.0 then the person is categorized as underWeight whereas if BMI is greater than 25.7, then overWeight. For future work, more potential covariates will be considered. 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 $f_2(t)$ represent the progesterone and estrogen curves, respectively, for women of 33 years old with normal weight.

Table 4.2 records the results of estimates of regression coefficients, smoothing parameters and variance components. The standard errors of our fixed effects parameter estimates are sufficiently large such that none of the fixed effects parameter estimates are statistically significant. That said, in terms of point estimates, we find a negative association on both responses with age, and both overweight and underweight (compared to regular BMI) are also negatively associated with progesterone only. The estimated correlation of

Table 4.2: Estimates of regression coefficients, variance parameter and smoothing parameter for the progesterone and estrogen data.

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

the bivariate responses can be calculated from variance estimates from Table 4.2:

$$\text{corr}(Y_{1ij}, Y_{2ik}) = \frac{\text{cov}(Y_{1ij}, Y_{2ik})}{\sqrt{\text{var}(Y_{1ij})}\sqrt{\text{var}(Y_{2ik})}}$$

$$= \frac{\phi_{1,2}}{\sqrt{\sigma_1^2 + \phi_1 + \rho_1^0 \exp\{a_{10} + a_{11}t_{ij} + a_{12}t_{ij}^2\}}\sqrt{\sigma_2^2 + \phi_2 + \rho_2^0 \exp\{a_{20} + a_{21}t_{ik} + a_{22}t_{ik}^2\}}}$$

$$= \frac{-0.2755}{\sqrt{0.6499 + 0.6455 + \exp\{-0.7699 + 0.2894t_{ij} - 0.1673t_{ij}^2\}}} \cdot \frac{1}{\sqrt{0.6019 + 0.6208 + \exp\{-1.7431 + 0.5172t_{ik} - 0.0800t_{ik}^2\}}},$$

since $\text{cov}(Y_{1ij}, Y_{2ij}) = \text{cov}(b_1, b_2)$ in this case. For example, if $t_{ij} = 5$ and $t_{ik} = 6$, then $\text{corr}(Y_{1ij}, Y_{2ik}) = -0.2343$, which indicates that that the two hormones are negatively correlated when progesterone is at $t_{ij} = 5$ and estrogen is at $t_{ik} = 6$. The estimates of nonparametric function $\hat{f}_1$ and $\hat{f}_2$ and their 95% confidence interval are superimposed over the log responses in Figure 4.1, respectively, which accurately captures the underlying trends of the bivariate longitudinal responses.

4.7 Discussion

We propose and build a model for analysis of bivariate cyclic longitudinal data and provide inference procedures. The model is proposed in the likelihood framework and the regression parameters and nonparametric functions are estimated by maximizing a penalized likelihood function. The smoothing parameter and variance components are numerically estimated using the Fisher-scoring algorithm based on restricted maximum likelihood. Modelling the time effect nonparametrically gives more flexibility in specifying the response mean structure, and the Gaussian field allows for additional flexibility in specifying the within-subject correlation structure, including possibly non-stationarity. The correlation of the two responses were explained only through the correlation of the random effects in the data analysis in Section 4.6, though the proposed model 4.1 can accommodate more complicated correlation structure of the responses through the covariance matrix (4.21) of the bivariate Gaussian field, as illustrated in subsection 4.2.3. Simulation results show that inference procedure performs well in all estimation results.

The bivariate semiparametric stochastic longitudinal model we proposed can be readily extended to multivariate longitudinal data. Dimensionality can pose as a challenge during the extension however. In the bivariate studies, we employed both C++ and parallel computing in the simulation study. Despite the effort, there is still computational burden on estimation procedure. Also, due to the high dimensionality of the model and the estimation problem, the algorithm is not always converging for all parameters; however, it does very well for the estimations of all parameters. Considering the high dimensional feature, the algorithm actually performs very well, specifically, about 90% to 95% of the time the algorighm converges in the simulation study and always converges in the real dataset applications for the parameter estimations.

Parameter initializations of the bivariate Gaussian fields need to be properly chosen as some initialization of Gaussian field parameters may lead to infinity in some entries of variance-covariance matrix, thus causing the matrix degenerate. This said, in the analysis of the real dataset, we tried three very different initializations of the Gaussian field parameters from one another, and all estimates of the regression parameters, the variance components, and smoothing parameters are qualitatively the same, which is reassuring. Initialization of the smoothing parameter in the Fisher-Scoring algorithm could also be part of consideration for future work. Note in the simulation study and the data analysis, both smoothing parameters are initialized to be 1 and it may not apply in the general setting. In our proofs, we allow the smoothing parameters λ_1 and λ_2 to go to zero, and this have implication on variance, which can be further investigated.

In this chapter, we consider modelling the bivariate longitudinal responses jointly via separate but correlated random effects and bivariate Gaussian field. For future work, the bivariate longitudinal responses can also be modelled separately for comparison purposes in order to compare the efficiency gains from modelling the data jointly. The specification of the random effect structure allows flexibility over imposing a shared random effects.

We would like to further explore sensitivity/robustness to the model assumptions. We have investigated the impact of Gaussian field misspecification in the simulation studies, which show that the choice of Gaussian field has little impact on the fixed effect parameters of interest. However, if we were interested in the underlying biological process, a deeper understanding of the choice of the Gaussian field is needed, including for the covariate structure in equation (4.4). For example, residual plots may be considered as a potential tool for model diagnostics and also for automatic choices for parameter selections; such as a procedure that makes sense for initialing values to better avoid any local maximization or lack of convergence for parameters to be estimated in the algorithm. Also, non-Guassian models can be generalized under the proposed framework if needed. In spite of further work to consider, this is a flexible and informative method for modeling bivariate longitudinal response data and we look forward to further extensions of this work in the above and possibly other directions.

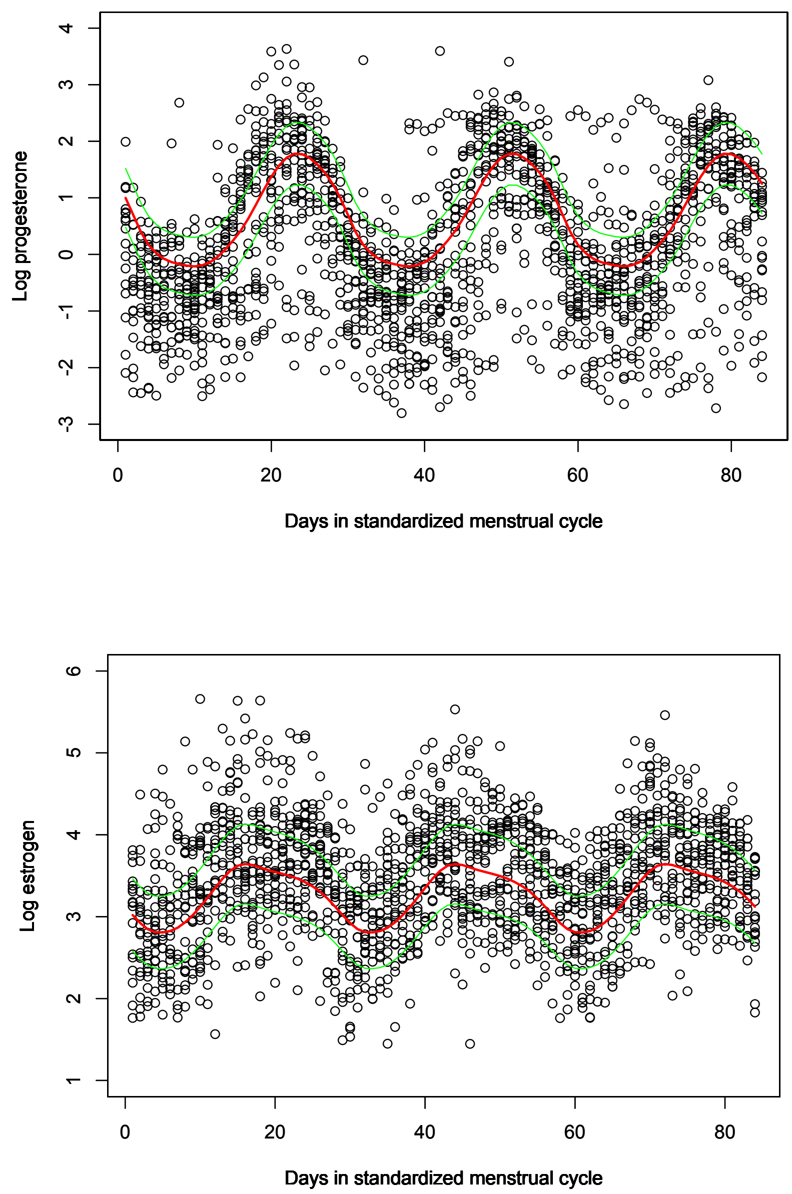


Figure 4.2: Plots of log progesterone and log estrogen levels against days in 3 standardized menstrual cycles, superimposed by estimated population mean curve $\hat{f}_1$ and $\hat{f}_2$ and their 95% pointwise confidence intervals.

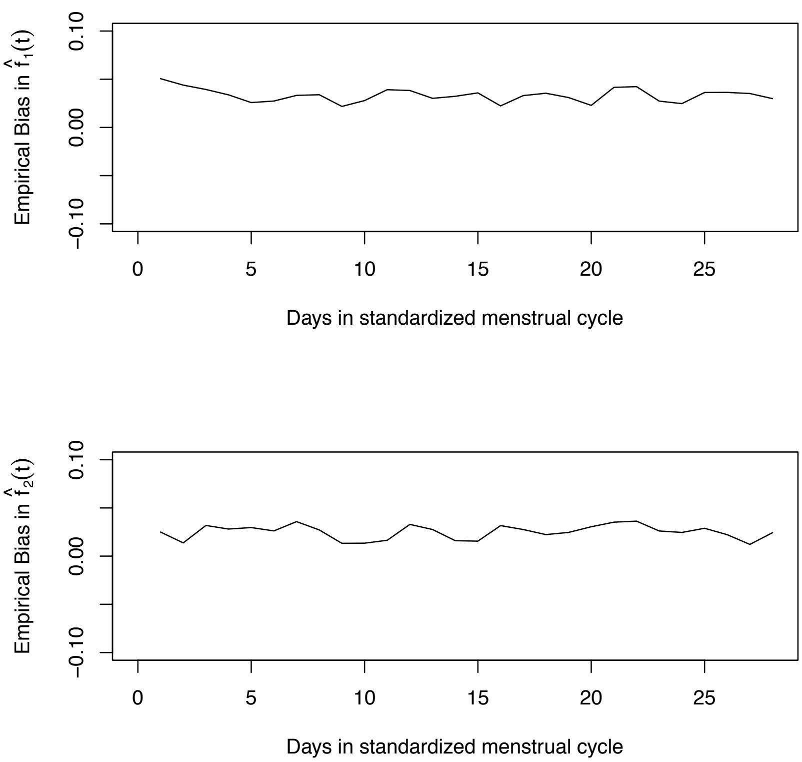


Figure 4.3: Empirical Bias in estimated nonparametric functions $\hat{f}_1$ and $\hat{f}_2$ based on 500 simulation replications.

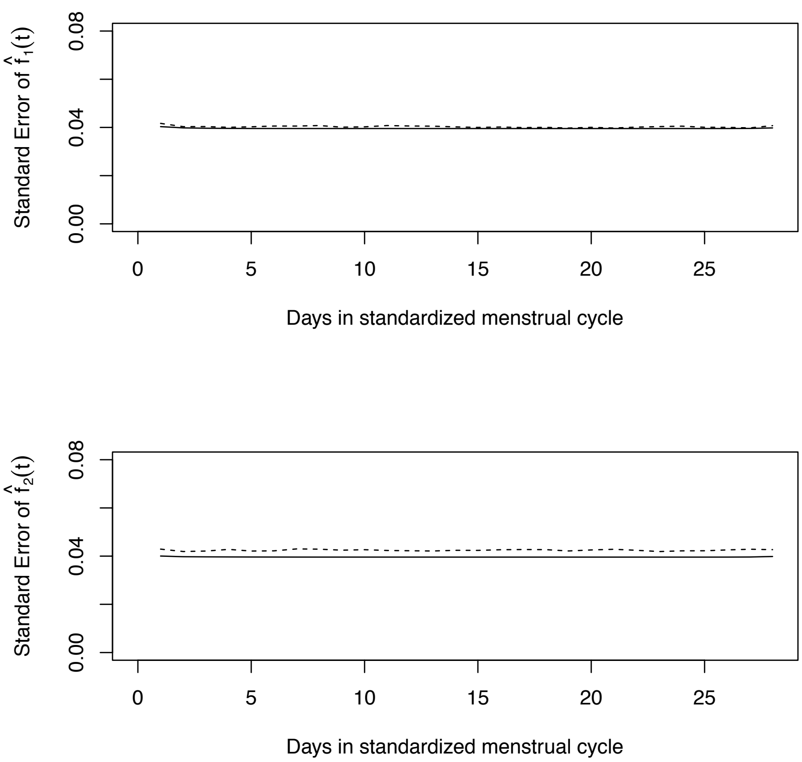


Figure 4.4: Pointwise empirical (dashes) and frequentist (solid) standard errors of the estimated nonparametric functions $\hat{f}_1$ and $\hat{f}_2$ based on 500 simulation replications.

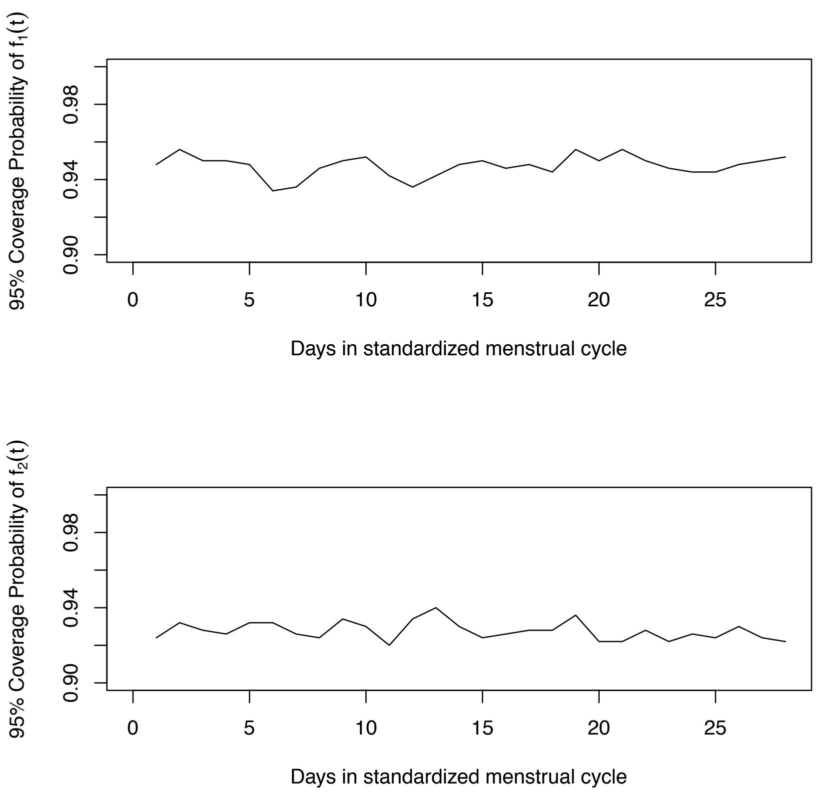
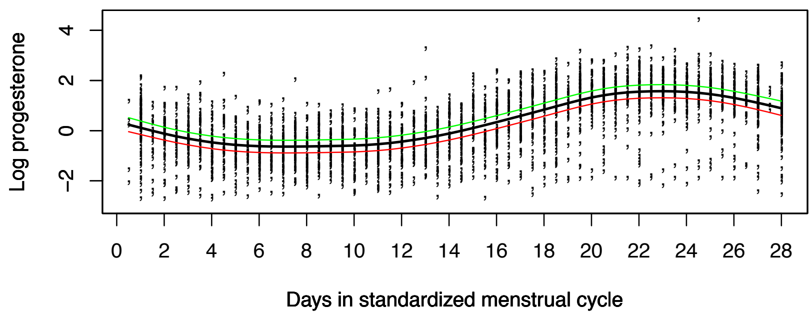


Figure 4.5: A graph showing the estimated 95% coverage probabilities of the true non-parametric functions f_1 and f_2 based on 500 simulation replications.



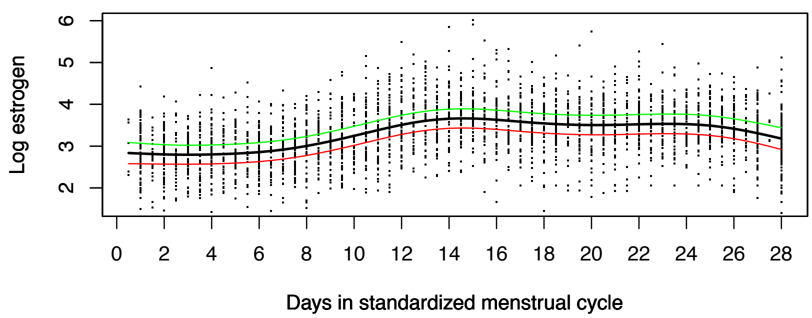


Figure 4.6: Plots of empirical sample variance of log progesterone and log estrogen levels at each distinct time points in a standardized menstrual cycle.

Chapter 5

Prediction

5.1 Introduction and literature review

Prediction is one of the fundamental problems in the statistical discipline. Usually due to finite resources or other constraints, only a small portion of the data are available for making reasonable predictions of new/future observations, thereby making prediction a challenging problem. In the context of periodic longitudinal data, I would like to predict future cycles for a single response from past cycles, to detect any potential signs of abnormalities or even illnesses.

In the literature, most prediction problems were approached from the Bayesian prospective, such as Taylor *et al* [32], Yu, Taylor & Sandler [42], and Proust-Lima & Taylor [24]. These three papers are concerned with prediction of prostate-specific antigen(PSA) using joint longitudinal and survival models. Serrat *et al* [29] studied the same problem using both frequentist and Bayesian approaches. Proust-Lima *et al* [23] provide a review on the methods. Lawless & Fredette [20] propose a frequentist approach to obtain prediction intervals and predictive distributions under a general parametric framework. The key to the solution is via exact or approximate pivotal quantity.

In this chapter, I propose a prediction procedure from the frequentist prospective. Instead of approaching the problem using exact or approximate pivotal quantities, underlying covariance structure of the Gaussian process in the univariate semiparametric model introduced in Zhang *et al* [44] is utilized to help solve the problem. The method can be generally applied when the stochastic process in the semiparametric model has exponential correlation functions, which implicitly induces the Markov property. The prediction approach will be illustrated through simulation studies and application to the hormone dataset. Univariate prediction only will be considered in this chapter in the methodological development; however, we will discuss the prediction problem in the bivariate context in the discussion section.

The chapter is organized as followed. Section 5.2 proposes a general prediction procedure of a single future observation for univariate longitudinal data. Specifically, Section 5.2.1 provides a prediction density function and prediction intervals for a single future observation. Section 5.2.2 illustrates the proposed method by an example; and Section 5.2.3 concludes this section by presenting methods and challenges when predicting an entire future cycle. Section 5.3 investigates the proposed methodology through a simulation study. Section 5.4 illustrates the model by analyzing bivariate longitudinal female hormone data collected daily over a single menstrual cycle, and finally, Section 5.5 provides a summary of our proposed model, and discusses challenges and future work.

5.2 Predictive Distribution and Predictive Intervals

5.2.1 General Method

We present in this section a general prediction procedure of a single future observation for univariate longitudinal data. Consider the periodic semiparametric stochastic mixed model introduced in Zhang *et al* [44], which is the univaraiate version of the model proposed in Section 4.4. Denote $\{Y_{ij}\}$ to be the j^{th} measurement for the i^{th} subject and suppose there are m subjects and each subject has n_i measurements; then the model is written as

$$Y_{ij} = \boldsymbol{X}_{ij}^T\boldsymbol{\beta} + f(t_{ij}) + b_i + U_i(t_{ij}) + \epsilon_{ij}, \tag{5.1}$$

where $\boldsymbol{\beta}$ is regression coefficients associated with covariates $\boldsymbol{X}_{ij}$; $f(t)$ is a twice-differentiable periodic function with period length to be T; the b_i are independent subject-specific random intercepts assumed to be normally distributed with mean 0 and variance ϕ; the $U_i(t)$ are independent mean zero Gaussian process with periodic covariance function $\xi(t)$ and correlation function $\mathrm{corr}(U_i(s), U_i(t)) = \eta(\rho; s, t)$ where ρ is correlation coefficient and between 0 and 1, and ϵ_{ij} is distributed as independent normal $(0, \sigma_\epsilon^2)$. The b_i, $U_i(t)$ and ϵ_{ij} are assumed to be mutually independent.

The aim of the prediction procedure is to predict a single future observation $Y_{i,j}$ given past observations $Y_{i1}, \ldots, Y_{i,j-1}$ and the associated covariates $\boldsymbol{X_{ij}}$. The proposed method take advantage of the fact that past observations $Y_{i1}, \ldots, Y_{i,j-1}$ explicitly influence the present observation $Y_{i,j}$; and thus past can be treated as additional predictor variables. If Gaussian process U_i's have exponential correlation structure

$$\mathrm{cov}(U_i(s), U_i(t)) = \sigma^2 \exp(\rho|s - t|),$$

then the conditional distribution of $U_i(t_{ij})$ given all past observations $U_i(t_{i1}), \ldots, U_i(t_{i,j-1})$ depends only on the previous observation value $U_i(t_{i,j-1})$, and model (5.1) can be reinterpreted as a *transition Markov* model by rewriting it as

$$Y_{ij} = \boldsymbol{X_{ij}^T \beta} + f(t_{ij}) + b_i + U_i(t_{ij}) + \epsilon_{ij}$$

where

$$U_i(t_{ij}) = \alpha(t_{ij}) U_i(t_{ij-1}) + H(t_{ij}), \tag{5.2}$$

$\alpha(t_{ij}) = \exp(\rho|t_{ij} - t_{i,j-1}|)$, and the $H(t_{ij})$ are mutually independent $N(0, G)$ random variables where $G = \sigma^2[1 - \alpha(t_{ij})^2]$. By substituting $U_i(t_{ij}) = Y_{ij} - \boldsymbol{X_{ij}^T \beta} - f(t_{ij}) - b_i - \epsilon_{ij}$ to Equation 5.2,

$$\begin{aligned} & Y_{ij} - \boldsymbol{X_{ij}^T \beta} - f(t_{ij}) - b_i - \epsilon_{ij} \\ = \quad & \alpha(t_{ij})(Y_{i,j-1} - \boldsymbol{X_{i,j-1}^T \beta} - f(t_{i,j-1}) - b_i - \epsilon_{i,j-1}) + Z(t_{ij}), \end{aligned}$$

and rearragning gives

$$\begin{aligned} Y_{ij} \quad = \quad & \boldsymbol{X_{ij}^T \beta} + f(t_{ij}) + b_i + \epsilon_{ij} \\ & + \alpha(t_{ij})(Y_{i,j-1} - \boldsymbol{X_{i,j-1}^T \beta} - f(t_{i,j-1}) - b_i - \epsilon_{i,j-1}) + Z(t_{ij}). \end{aligned}$$

Therefore, the predictive density function is

$$Y_{ij}|Y_{i,j-1}, \boldsymbol{X_{ij}}, \boldsymbol{X_{i,j-1}} \sim N(\mu_{cond}, \sigma_{cond}),$$

where

$$E(Y_{ij}|Y_{i,j-1}, \boldsymbol{X_{ij}}, \boldsymbol{X_{i,j-1}}) = \mu_{cond} = \boldsymbol{X_{ij}^T\beta} + f(t_{ij}) + \alpha(t_{ij})(Y_{i,j-1} - \boldsymbol{X_{i,j-1}^T\beta} - f(t_{i,j-1}))$$

and

$$\begin{aligned} Var(Y_{ij}|Y_{i,j-1}, \boldsymbol{X_{ij}}, \boldsymbol{X_{i,j-1}}) &= \sigma_{cond} = Var[b_i - \alpha(t_{ij})(b_i + \epsilon_{i,j-1}) + H(t_{ij}) + \epsilon_{ij})] \\ &= [1 + \alpha(t_{ij})^2]Var(b_i) + [1 + \alpha(t_{ij})^2]Var(\epsilon_{i,j-1}) + Var[H(t_{ij})] \\ &= [1 + \alpha(t_{ij})^2]\phi + [1 + \alpha(t_{ij})^2]\sigma_\epsilon^2 + \sigma^2[1 - \alpha(t_{ij})^2]. \end{aligned}$$

This is an extension of the transition Markov models [7] to the periodic semiparametric model (5.1). The appeal of this method is that it takes advantage of the properties of Gaussian process and can be applied as long as U_i's have the exponential correlation structure. This model can also be applied when times of measurements are not common to all subjects, which does not always hold for other correlation models [7].

5.2.2 An illustrative example

Suppose the stochastic process $U_i(t)$ in the model follows mean-zero non-homogeneous Ornstein-Uhlenbeck (NOU) process with covariance structure

$$\text{cov}(U_i(s), U_i(t)) = \exp\left(\frac{\xi(t) + \xi(s)}{2}\right)\exp(\log\rho|s - t|)$$

where $\xi(t) = \xi_0 + \xi_1 s_1(t) + \xi_2 s_2(t)$ is the periodic cubic spline [44], where $s_j(t) = a_j(t) + b_j(t^2) + c_j t^3 + (t - t_j)_+^3$, with coefficients a_j, b_j, and c_j to be

$$\begin{aligned} a_j &= -\frac{T(T - t_j)}{2} + \frac{3(T - t_j)^2}{2} - \frac{(T - t_j)^3}{T} \\ b_j &= \frac{3(T - t_j)}{2} - \frac{3(T - t_j)^3}{2T} \\ c_j &= -\frac{T - t_j}{T}. \end{aligned}$$

Then by the Markov properties of the NOU process, the conditional distribution of $U_i(t_{ij})$ given all its predecessors depends only on the previous value $U_i(t_{i,j-1})$, and

$$U_i(t_{ij})|U_i(t_{i,j-1}) \sim N\left(\alpha(t_{ij})U_i(t_{i,j-1}), \exp\left(\frac{\xi(t_{ij})+\xi(t_{i,j-1})}{2}\right)(1-\alpha(t_{ij})^2)\right)$$

where

$$\alpha(t_{ij}) = \exp(\log\rho|t_{ij} - t_{i,j-1}|)$$

and the process is initiated by

$$U_i(t_{i1}) \sim N(0, \exp(\xi(t_{i1}))).$$

Equivalently, with the same initial conditions,

$$U_i(t_{ij}) = \alpha(t_{ij})U_i(t_{i,j-1}) + H(t_{ij}) \tag{5.3}$$

where $H(t_{ij})$ are mutually independent $N(0, G)$ random variables where $G = \exp((\xi(t_{ij}) + \xi(t_{i,j-1}))/2)(1-\alpha(t_{ij})^2)$. Substitute equation (5.3) into model (5.1) and rearranging, we have

$$\begin{aligned}
Y_{ij} &= \boldsymbol{X_{ij}^T\beta} + f(t_{ij}) + b_i + \alpha(t_{ij})U_i(t_{i,j-1}) + H(t_{ij}) + \epsilon_{ij} \\
&= \boldsymbol{X_{ij}^T\beta} + f(t_{ij}) + b_i \\
&\quad +\alpha(t_{ij})(Y_{i,j-1} - \boldsymbol{X_{i,j-1}^T\beta} - f(t_{i,j-1}) - b_i - \epsilon_{i,j-1}) + H(t_{ij}) + \epsilon_{ij}.
\end{aligned}$$

Therefore,

$$Y_{ij}|Y_{i,j-1}, \boldsymbol{X_{ij}}, \boldsymbol{X_{i,j-1}} \sim N(\mu_{cond}, \sigma_{cond}).$$

where

$$E(Y_{ij}|Y_{i,j-1}, \boldsymbol{X_{ij}}, \boldsymbol{X_{i,j-1}}) = \mu_{cond} = \boldsymbol{X_{ij}^T\beta} + f(t_{ij}) + \alpha(t_{ij})(Y_{i,j-1} - \boldsymbol{X_{i,j-1}^T\beta} - f(t_{i,j-1}))$$

and

$$\begin{aligned}
& Var(Y_{ij}|Y_{i,j-1}, \boldsymbol{X_{ij}}, \boldsymbol{X_{i,j-1}}) \\
= \; & \sigma_{cond} = Var[b_i - \alpha(t_{ij})(b_i + \epsilon_{i,j-1}) + H(t_{ij}) + \epsilon_{ij})] \\
= \; & [1+\alpha(t_{ij})^2]Var(b_i) + [1+\alpha(t_{ij})^2]Var(\epsilon_{i,j-1}) + Var(H(t_{ij})) \\
= \; & [1+\alpha(t_{ij})^2]\phi + [1+\alpha(t_{ij})^2]\sigma_\epsilon^2 + \exp\left[\frac{\xi(t_{ij})+\xi(t_{i,j-1})}{2}\right][1-\alpha(t_{ij})^2].
\end{aligned}$$

5.2.3 Prediction of an entire cycle

To predict an entire cycle given the past cycles, we have to come up with ways to best approximate previous observations $Y_{i,j-1}, j = 2, \ldots, n_i$ for the entire cycle, which is unknown to us. There are several ways to approach it.

One way is that we can take advantage of the cyclic nature of the data, by using $(j-1)^{th}$ observation value from previous cycle, denoting $Y_{i,j-1,pc}$, to approximate the previous observation $Y_{i,j-1}$. Another way is to use the predictive value $Y_{i,j-1,pre}$ as a proxy for the previous observation $Y_{i,j-1}$. A third way is weighted version of the previous two methods, with some weights w upon the two proxies, e.g. $w * Y_{i,j-1,pc} + (1-w) * Y_{i,j-1,pre}, 0 < w < 1$. We will demonstrate the effectiveness of these methods in Section 5.3.

In the application of hormone data, due to the measurements to be unequally spaced, it is possible that $Y_{i,j-1,pc}$ is not available for that particular time point $j-1$ from previous cycle. In this case, we can either interpolate the past cycle to obtain the measurement or we can rely on the observations two cycles away, denoting $Y_{i,j-1,pc2}$, as a proxy. We will touch upon this issue in more detail in Section 5.4.

5.3 Simulation Studies

A simulation study is conducted to evaluate the performance of the prediction method. Univariate periodic longitudinal data are generated ($m = 50$) according to the following model:

$$\begin{aligned} Y_{ij} &= \text{age}_i^T \beta_1 + f(t_{ij}) + b_i + U_i(t_{ij}) + \epsilon_{ij} \\ & \qquad i = 1, \ldots, 50;\ j = 1, \ldots, 28;\ t_{ij} \in \{1, \ldots, 28\} \end{aligned}$$

where b_i are mutually independent $N(0, \phi)$; U_i are simulated from zero-mean NOU process modeling serial correlation, with variance function $\xi(t) = \xi_0 + \xi_1 s_1(t) + \xi_2 s_2(t)$ as described in the last section, and $\text{corr}(U_i(t), U_i(s)) = \rho^{|s-t|}$. ϵ_{ij} are simulated from a mutually independent $N(0, \sigma^2)$; and the nonparametric periodic smooth function is generated from

$f_1(t) = 5\sin(2\pi/28)\,t$. Our goal is to predict the longitudinal response values of the second cycle, using the observations from the first cycle.

There are, in general, two ways to apply the prediction method in the simulation study. Since the true values of β and f are known in simulation, they can directly be used to obtain the prediction result; whereas in real life scenario when the true values of β and f are unknown, $\hat{\beta}$ and $\hat{f}$ from the estimating procedure can be used to best approximate the true values. I will demonstrate the latter method to better mimic what I will do for the application dataset in the next section. The performance of different methods for each simulation is evaluated by the *average predictive squared error* (PSE) defined

$$PSE = \frac{1}{m}\sum_{i=1}^{m}(Y_{ij} - \mu_{cond})^2$$

where Y_{ij} is the true simulated response and μ_{cond} is the mean value from the conditional distribution.

Table 5.1 records PSE values for different proxies for $Y_{i,j-1}$ based on 100 simulation replicates and 50 subjects. pno means $Y_{i,j}$ conditional on the true simulated data $Y_{i,j-1}$; pc conditional on the $Y_{i,j-1,pc}$ from previous cycle; pre conditional on the $Y_{i,j-1,pre}$ from predictive prior observation; and the rest are weighted version of pc and pre. We see that of three main methods, pno, pc and pre, pno method performs the best as expected with the lowest PSE value, with pc follows closely behind; and pre performs the worst, especially towards the latter part of the cycle, see Figure 5.2. In the interest of not overcrowding, not all of the weighted versions are added to the plots. What is interesting is that both of first two weighted versions perform better than the pno version, i.e. μ_{cond} are computed with more or equal weights are put into conditional on the $Y_{i,j-1,pre}$ from predictive prior observation. Conditional on $Y_{i,j-1,pre}$ takes more account of present cycle while conditional on $Y_{i,j-1,pc}$ takes only account of previous cycle; resulting the hybrid version with more weights on $Y_{i,j-1,pre}$ in more desirable results.

Note that it is unsurprising that using the estimates instead of the true values takes much more computational time since parameters have to be estimated for each simulation before the prediction method takes place.

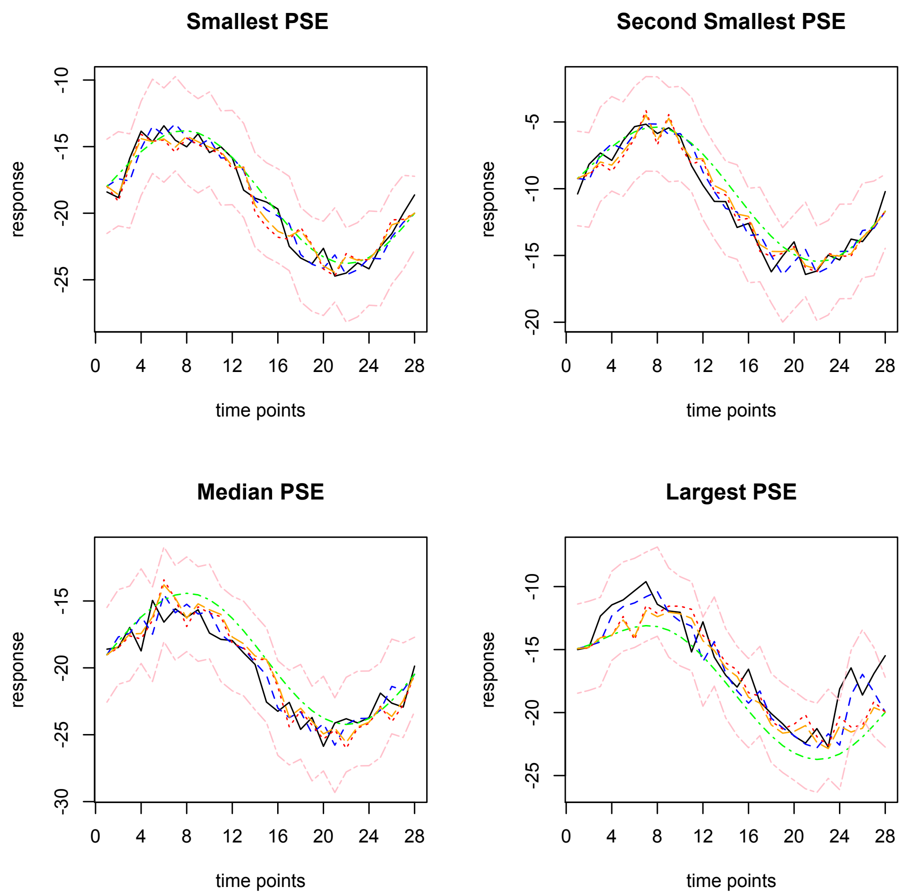


Figure 5.1: Plots of simulated subjects with the lowest, second lowest, median and largest PSE values with respect to conditional on $Y_{i,j-1,pc}$ based on 100 simulations, with pointwise 95% confidence interval. True data: black; pno: blue; pc: red; pre: green; 0.25*pre + 0.75*pc: orange; 95% CI: pink.

Table 5.1: PSE values for different proxies for $Y_{i,j-1}$ based on 100 simulations.

	pno	pc	pre	0.25*pre + 0.75*pc	0.5*pre + 0.5*pc	0.75*pre + 0.25*pc
PSE	1.7735	1.7774	2.3246	1.6642	1.7177	1.9379

5.4 Application to the univariate cyclic longitudinal hormone dataset

We apply the prediction method to the hormone dataset with 3 consecutive cycles described in Chapter 2. Denoting Y_{ij} the j^{th} log-transformed progesterone values measured at standardized day t_{ij} since menstruation for the i^{th} woman, we consider the following univariate semiparametric stochastic mixed model:

$$\begin{aligned} Y_{ij} &= \text{age}_i^T\beta_1 + \text{underWeight}_i^T\beta_2 + \text{overWeight}_i^T\beta_3 + f(t_{ij}) + b_i + U_i(t_{ij}) + \epsilon_{ij} \\ & i = 1, \ldots, 50; j = 1, \ldots, n_i; t_{ij} \in \{1, \ldots, 28\} \end{aligned}$$

where the specifications of random intercepts b_i, the Gaussian process U_i, the measurement errors ϵ_{ij} and the nonparametric smooth functions are the same as those in the simulation study. Covariates underWeight and overWeight are indicator variables, which is characterized by Body Mass Index (BMI) where if BMI is less than 19.0 then the person is categorized as underWeight whereas if BMI is greater than 25.7, then overWeight; this means the comparison group is normal weight, i.e., between 19.0 and 25.7 BMI. For computational stability, standardized days were centered at the median day 14 and divided by 10; covariate age is also centered at median 36 years old and divided by 50. Thus, $f(t)$ represents the progesterone curve for women of 36 years old with normal weight.

Our goal is to predict the progesterone level for the third cycle, using the previous two cycles. Similar to the simulation study, there are several proxies for the previous observed progesterone level $Y_{i,j-1}$ in Cycle 3, which is assumed to be unknown. One key problem in the hormone data analysis is that one of the proxies, $Y_{i,j-1,pc}$ may not be always available from previous cycle, i.e., Cycle 2 in this case; since the response observations are unequally spaced and the menstrual cycle lengths vary from cycle to cycle. This problem is solved by linear interpolation of Cycle 2 using all available observed hormone levels from the same cycle. However, in the case where the end points are unavailable, e.g. progesterone observation is missing at day 1 in Cycle 2 but is needed as a proxy for $Y_{i,1,pc}$; interpolation, either linear interpolation or cubic spline interpolation, would not produce the interpolated end points. In this case, I propose to interpolate the end points by using observations from

Cycle 1 and combined with other available interpolated values from Cycle 2 to use as proxies for $Y_{i,j-1}$ in Cycle 3.

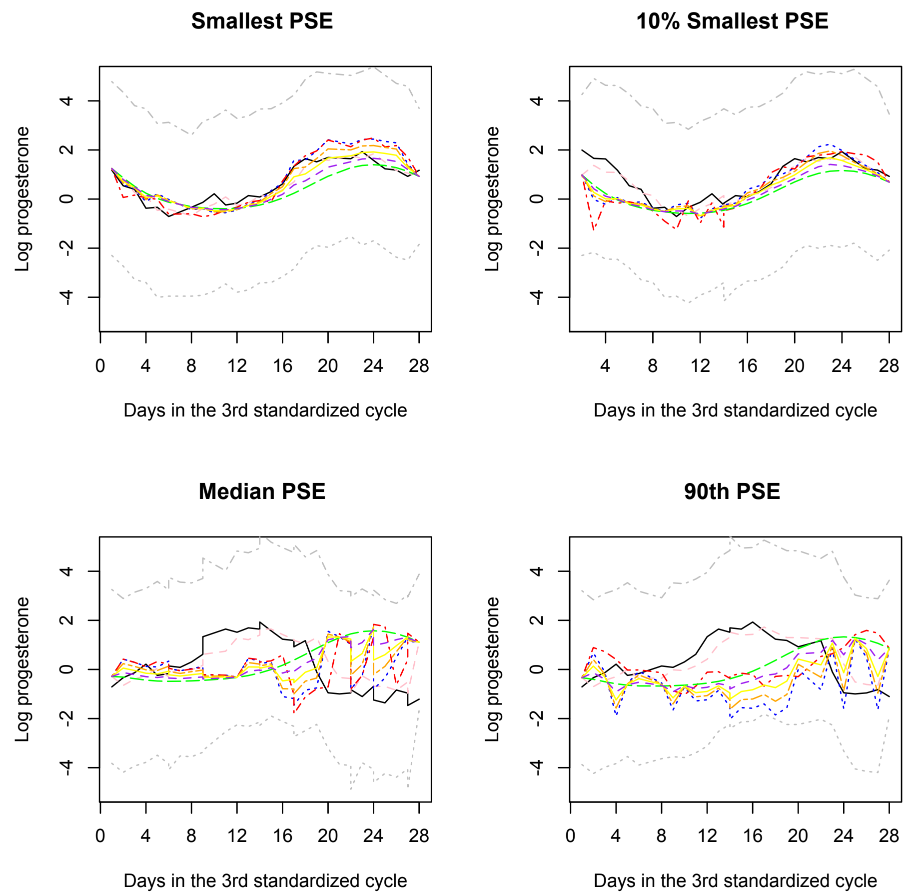


Figure 5.2: Plots of progesterone levels for women with the lowest, 10th quantile lowest, median and 90th quantile PSE values with respect to conditional on $Y_{i,j-1,pc2}$, with pointwise 95% confidence interval. True data: black; pno: pink; pc2: blue; pc1: red; pre: green; 0.25 * pre + 0.75 * pc: orange; 0.5 * pre + 0.5 * pc: yellow; 0.75 * pre + 0.25 * pc: purple; 95% CI:gray.

Table 5.2 records PSE values for different proxies for $Y_{i,j-1}$ for the aforementioned hormone dataset of 3 consecutive cycles. For demonstration purposes, we randomly select

50 women from the dataset to perform the analysis. Method pno means $Y_{i,j}$ conditional on the true previous observation $Y_{i,j-1}$; pc1, pc2 conditional on the $Y_{i,j-1,pc1}$ and $Y_{i,j-1,pc2}$ from Cycle 1 and Cycle 2, respectively, either the from true previous observations or from interpolations; pre conditional on the $Y_{i,j-1,pre}$ from predictive prior observation; and the rest are weighted version of pc2 and pre. We see that of four main methods, pno, pc2, pc1 and pre, it comes as no surprise that pno method performs much better than the rest of 3 proxies with the lowest PSE value; the other three proxies produce similar PSE values, with pre performs the better than the other two pc's, see Figure 5.2, which indicates that none of the observations from either cycles, Cycle 1 or Cycle 2, are good proxies for $Y_{i,j-1}$. However, since the estimated nonparametric functions are estimates of the overall time trend for the entire sample, it is not suitable for some study participants.

Table 5.2: PSE values for different proxies for $Y_{i,j-1}$ based on the hormone dataset with 3 consecutive cycless.

	pno	pc2	pc1	pre	pre/4 + 3*pc2/4	pre/2 + pc2/2	3*pre/4 + pc2/4
PSE	0.2716	2.4998	2.2335	1.5324	2.1081	1.8163	1.6243

Since results from conditional on predictive prior observations $Y_{i,j-1,pre}$ is better, the weighted version puts weights between pre and pc2. Suppose pc1 and pc2 perform much better, weighted version can be changed to $w * pc1 + (1-w) * pc2, 1 < w < 0$. As expected, the weighted versions perform better than either the pc versions, especially when more weights are put on $Y_{i,j-1,pre}$.

5.5 Discussion and future work

In conclusion, we provide a novel prediction procedure from a frequentist viewpoint, while most of other prediction methods in the literature are from the Bayesian standpoint. The method requires that the stochastic process in model (5.1) has exponential correlation structure, which implicitly implies Markov property; resulting in the model to be able to be rewritten in the form of autoregressive model of order one for the method to work.

For prediction of a single future observation, a direct application of this method produces the prediction density function and prediction intervals of the single future observation of interest, given the previous longitudinal measurement is observed. The method is straightforward to apply and the performance was shown to be fairly good in both a simulation study and in a real application dataset. Also note that the univariate longitudinal data does not have to be periodic in this case.

For prediction of an entire future cycle, iterative application of this method needs to be applied; and due to the unavailability of the previous longitudinal observations, proxies of the previous longitudinal observations need to be utilized. The choice of proxies and the variations between cycles affect the accuracies of the prediction results. If there is little variation from cycle to cycle, then the method performs well; otherwise, due to great variations of cycles, proxies using past cycle information provide a poor substitute for the true previous longitudinal observation that is unavailable; resulting in less than satisfactory results.

Several areas of future work can be considered. First, extension of the univariate prediction method to the bivariate case can be explored. I would like to utilize the correlation between the bivariate responses to help with the prediction of a given response, beyond just using the past response or past cycle information when predicting a future response in a univariate longitudinal context. Secondly, when predicting a future cycle, the longitudinal observations of a cycle can be viewed as realizations of a multivariate stochastic/Gaussian process and thus treating observations of a cycle as one unit in the vector form. Then, the properties of the process may give rise to a more general prediction procedure for prediction of a future cycle. Third, the proposed prediction procedure is performed under the standardized time scale. It would be interesting to explore how the prediction procedure would work under the original time scale; or given the predictions under the standardized time scale, one could back transform to go to the original scale for an individual person, if of interest. Lastly, I would like to investigate further the robustness of the accuracy of prediction under the violations of our model assumptions.

Chapter 6

Conclusion and Future Work

6.1 Summary

In Chapter 3, I briefly reviewed the proposed model from Zhang *et al* [45] and presented and updated detailed derivations for the findings contained in the paper; also two minor discrepancies were identified and further simulation studies need to be explored to ensure the accuracy of the smoothing parameter estimate. The chapter provides a solid foundation for me to extend it to the bivariate longitudinal model proposed in the next Chapter, and also gives the set-up needed for the prediction problem in Chapter 5.

In Chapter 4, I propose and develop a model for analysis of bivariate longitudinal data and provide inference procedures, building off of specification and procedures for univariate longitudinal data, inspired by earlier work from Zhang *et al*[45] among other researchers having proposed semiparametric models for univariate longitudinal response data. The bivariate model is proposed in the likelihood framework and the regression parameters and nonparametric functions are estimated by maximizing a penalized likelihood function. The smoothing parameter and variance components are numerically estimated using the Fisher-scoring algorithm based on restricted maximum likelihood. Modelling the time effect nonparametrically gives more flexibility in specifying the response mean structure, and the Gaussian field allows for additional flexibility in specifying the within-subject correlation structure, including possibly non-stationarity. The correlation of the two responses is specified only through the correlation of the random effects in the simulation study, though the proposed model can accommodate more complicated correlation structure of the responses through the covariance matrix of the bivariate Gaussian field. Simulation results show that the inference procedure performs well across a variety of simulation settings.

In Chapter 5, we proposed a prediction procedure of either a single observation or of an entire cycle under the frequentist framework, which is not often done in the literature for longitudinal data. Under Bayesian modelling, dependence on the prior, complicated MCMC algorithm and robustness of the prior specification pose as a challenge in modelling, but the predictive distribution follows naturally from the posterior distribution assuming the prior distribution assumptions are reasonable. Frequentist methods, on the other hand, can be more challenging as there is no prior to help augment observed data but it is more computationally efficient. Prediction of a single future observation is straightforward and ready to use. Both simulation and real data analysis produce satisfactory results. Prediction of an entire cycle poses more of a challenge than prediction of a single observation. I overcome the challenge by finding proper proxies for previous observed data point. If the data behaves cyclic in a very uniform way, i.e., the pattens of cycles do not vary much from cycle to cycle for all units under consideration, then the proxy using past cycle performs well, as demonstrated in the simulation study. However, if the cycle varies greatly from cycle to cycle and from unit to unit under consideration, then the proxy using past cycle to predict the present cycle can be inaccurate, as shown in the real data analysis.

6.2 Future work

In spite of the advances conveyed in previous chapters, I look forward to extend in following subject areas.

6.2.1 Robustness of model assumptions

All models proposed in this thesis are investigated under Gaussian framework, which includes the bivariate Gaussian fields, random effects and measurement errors. We would like to further explore sensitivity/robustness to this assumption. When or if the model assumptions are violated, how much would the violation affect the estimations of the parameters of interest? Some theoretical explorations can be explored and simulation studies can be performed.

Also, non-Guassian models can be considered under the proposed framework in the future to allow for more generality. However, under non-Gaussian models, it can pose some challenge on the estimating procedure. For example, if the random effect is no longer assumed to be normally distributed, then the penalized likelihood would be difficult to specify, which is the key to the entire proposed estimation procedure.

Further, only joint models of the bivariate longitudinal data are considered in the thesis, it is of interest to model the bivariate longitudinal data separately to justify the efficiency gains by modelling the bivariate longitudinal data jointly.

6.2.2 Model diagnostics

It is very challenging to identify or justify which Gaussian field to be chosen to appropriately model the underlying biological or other process. It is also of interest to identify the most reasonable covariance structure for the bivariate responses in equation (4.4); and to justify the use of bivariate Gaussian field to model the complicated within-subject correlations, or can we get away with simpler specifications of serial correlations, such as in ϵ term in a more standard linear mixed effects model. Although extensive empirical simulations can be performed to help solve the issue, there is not a clear or easy-to-use method for this. As discussed in Chapter 4, we have investigated the impact of Gaussian field misspecification in the simulation studies, which show that the choice of Gaussian field has little impact on the fixed effect parameters of interest; and perhaps residual plots could be utilized as a tool to help solve the problem. However, if we were interested in the underlying biological process or any of issues mentioned above, a deeper understanding is needed.

6.2.3 Missing data

In the current real data applications, we excluded women with more than 10% missing data and with larger than normal range of the difference of maximum cycle length and minimum cycle length. Even if the exclusions are reasonable under the current setting, it might be of interest to include women with more missing data to obtain a more comprehensive sample to avoid selection bias.

6.2.4 Extension to multivariate longitudinal data with one or multiple cycles

The model proposed in Chapter 4 can be readily extended to more general multivariate cyclic longitudinal data. Specifically, extension to model bivariate longitudinal data with multiple cycles is proposed and preliminarily tested by using a small sample ($m = 20$) from the hormone study in Section 4.4. However, as the number of subjects increases in the sample size and as the number of cycles increases, computation time becomes extended. As discussed in Chapter 4, even with both C++ and parallel computing were utilized in the simulation study, the computation time is extended; modeling of a high-dimensional single dataset may also lead to some degree of computational burden for the proposed methodology.

6.2.5 Prediction for the bivariate model

An extension of the prediction problem discussed in Chapter 5 could be made to the bivariate case. However, there is an additional layer of dependence between the two responses due to their correlation. In the bivariate case, the correlation between the bivariate responses can be utilized to help with the prediction of a given response, beyond just using the past response or past cycle information when predicting a future response in a univariate longitudinal context.

References

[1] R. J. Adler and Jonathan E. Taylor. *Random Fields and Geometry.* Springer, New York, NY, 2007.

[2] Carl de Boor. *A Practical Guide to Splines.* Springer, New York, 1978.

[3] Babette A. Brumback and John A. Rice. Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of the American Statistical Association*, 93(443):961–976, 1998.

[4] Yueh-Yun Chi and Joseph G Ibrahim. Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*, 62(2):432–445, 2006.

[5] Jeng-Min Chiou and Hans-Georg Mller. A pairwise interaction model for multivariate functional and longitudinal data. *Biometrika*, 103(2):377–396, 2016.

[6] Kiranmoy Das, Runze Li, Subhajit Sengupta, and Rongling Wu. A bayesian semi-parametric model for bivariate sparse longitudinal data. *Statistics in medicine*, 32(22):3899–3910, 2013.

[7] Peter Diggle. *Analysis of longitudinal data.* Oxford University Press, 2002.

[8] Paul H. C. Eilers and Brian D. Marx. Flexible smoothing with b-splines and penalties. *Statistical Science*, 11(2):89–121, 1996.

[9] Randall L. Eubank. *Nonparametric Regression and Spline Smoothing.* Marcel Dekker, New York, second edition, 1999.

[10] S. Fieuws and G. Verbeke. Pairwise fitting of mixed models for the joint modelling of multivariate longitudinal profiles. *Biometrics*, 62(2):424–431, 2006.

[11] Garrett Fitzmaurice, Marie Davidian, Geert Verbeke, and Geert Molenberghs. *Longitudinal Data Analysis*. Chapman and Hall/CRC, Boca Raton, Florida, 2008.

[12] Ikuko Funatogawa, Takashi Funatogawa, and Yasuo Ohashi. A bivariate autoregressive linear mixed effects model for the analysis of longitudinal data. *Statistics in medicine*, 27(30):6367–6378, 2008.

[13] Pulak Ghosh and Timothy Hanson. A semiparametric bayesian approach to multivariate longitudinal data. *Australian & New Zealand journal of statistics*, 52(3):275–288, 2010.

[14] Eilen B Gold, Brenda Eskenazi, Bill L Lasley, Steven J Samuels, Marianne O'Neill Rasor, James W Overstreet, and Marc B Schenker. Epidemiologic methods for prospective assessment of menstrual cycle and reproductive characteristics in female semiconductor workers. *American journal of industrial medicine*, 28(6):783–797, 1995.

[15] Ellen B Gold, Brenda Eskenazi, S Katharine Hammond, Bill L LaSley, Steven J Samuels, Marianne O'Neill Rasor, Cynthia J Hines, James W Overstreet, and Marc B Schenker. Prospectively assessed menstrual cycle characteristics in female wafer-fabrication and nonfabrication semiconductor employees. *American journal of industrial medicine*, 28(6):799–815, 1995.

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

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

[18] Leonid B. Koralov and Yakov G. Sinai. *Theory of Probability and Random Processes*. Springer, New York, NY, second edition, 2007.

[19] Nan M. Laird and James H. Ware. Random-effects models for longitudinal data. *Biometrics,*, 38(4):963–974, 1982.

[20] JF Lawless and Marc Fredette. Frequentist prediction intervals and predictive distributions. *Biometrika*, 92(3):529–542, 2005.

[21] Ziyue Liu, Anne R. Cappola, Leslie J. Crofford, and Wensheng Guo. Modeling bivariate longitudinal hormone profiles by hierarchical state space models. *Journal of the American Statistical Association*, 109(505):108–118, 2014.

[22] P. M. Meyer, S. L. Zeger, Sioban D. Harlow, M. Sowers, S. Crawford, J. L. Luborsky, I. Janssen, D. S. McConnell, J. F. Randolph, and G. Weiss. Characterizing daily urinary hormone profiles for women at midlife using functional data analysis. *American Journal of Epidemiology*, 165(8):936–945, 2007.

[23] Cécile Proust-Lima, Mbéry Séne, Jeremy MG Taylor, and Hélène Jacqmin-Gadda. Joint latent class models for longitudinal and time-to-event data: A review. *Statistical methods in medical research*, 23(1):74–90, 2014.

[24] Cécile Proust-Lima and Jeremy MG Taylor. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment psa: a joint modeling approach. *Biostatistics*, 10(3):535–549, 2009.

[25] Jesse D Raffa and Joel A Dubin. Multivariate longitudinal data analysis with mixed effects hidden markov models. *Biometrics*, 71(3):821–831, 2015.

[26] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, Massachusetts, 2006.

[27] David Ruppert, M. P. Wand, and R. J. Carroll. *Semiparametric Regression*. Cambridge University Press, New York, 2003.

[28] Larry Schumaker. *Spline Functions: Basic Theory*. Cambridge University Press, Cambridge, UK, third edition, 2007.

[29] Carles Serrat, Montserrat Rué, Carmen Armero, Xavier Piulachs, Hèctor Perpiñán, Anabel Forte, Álvaro Páez, and Guadalupe Gómez. Frequentist and bayesian approaches for a joint model for prostate cancer risk and longitudinal prostate-specific antigen data. *Journal of Applied Statistics*, 42(6):1223–1239, 2015.

[30] J. P. Sy, J. M. G. Taylor, and W. G. Cumberland. A stochastic model for the analysis of bivariate longitudinal aids data. *Biometrics*, 53(2):542–555, 1997.

[31] J. M. G. Taylor, W. G. Cumberland, and J. P. Sy. A stochastic model for analysis of longitudinal aids data. *Journal of the American Statistical Association*, 89:727–736, 1994.

[32] Jeremy MG Taylor, Yongseok Park, Donna P Ankerst, Cecile Proust-Lima, Scott Williams, Larry Kestin, Kyoungwha Bae, Tom Pickles, and Howard Sandler. Real-time individual predictions of prostate cancer recurrence using joint models. *Biometrics*, 69(1):206–213, 2013.

[33] Geert Verbeke, Steffen Fieuws, Geert Molenberghs, and Marie Davidian. The analysis of multivariate longitudinal data: A review. *Statistical methods in medical research*, 23(1):42–59, 2014.

[34] Arunas P. Verbyla, Brian R. Cullis, Michael G. Kenward, and Sue J. Welham. The analysis of designed experiments and longitudinal data by using smoothing splines. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 48(3):269–311, 1999.

[35] Yuedong Wang, Wensheng Guo, and Morton B Brown. Spline smoothing for bivariate data with applications to association between hormones. *Statistica Sinica*, pages 377–397, 2000.

[36] Sue J. Welham, Brian R. Cullis, Michael G. Kenward, and Robin Thompson. The analysis of longitudinal data using mixed model l-splines. *Biometrics*, 62(2):392–401, 2006.

[37] Sue J. Welham, Brian R. Cullis, Michael G. Kenward, and Robin Thompson. A comparison of mixed model splines for curve fitting. *Australian & New Zealand Journal of Statistics*, 49(1):1–23, 2007.

[38] Simon Wood. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, Boca Raton, Florida, 2006.

[39] Lang Wu. *Mixed effects models for complex data*. CRC Press, 2009.

[40] Dongdong Xiang, Peihua Qiu, and Xiaolong Pu. Nonparametric regression analysis of multivariate longitudinal data. *Statistica Sinica*, pages 769–789, 2013.

[41] Xiaoqin Xiong and Joel A Dubin. A binning method for analyzing mixed longitudinal data measured at distinct time points. *Statistics in medicine*, 29(18):1919–1931, 2010.

[42] Menggang Yu, Jeremy M G Taylor, and Howard M Sandler. Individual prediction in prostate cancer studies using a joint longitudinal survival–cure model. *Journal of the American Statistical Association*, 103(481):178–187, 2008.

[43] Scott L Zeger and Peter J Diggle. Semiparametric models for longitudinal data with application to cd4 cell numbers in hiv seroconverters. *Biometrics*, pages 689–699, 1994.

[44] D. Zhang, X. Lin, and M. Sowers. Semiparametric regression for periodic longitudinal hormone data from multiple menstrual cycles. *Biometrics*, 56(1):31–39, 2000.

[45] Daowen Zhang, Xihong Lin, Jonathan Raz, and MaryFran Sowers. Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association*, 93(442):710–719, 1998.

APPENDICES

Appendix A

In this thesis, a number of technical definitions and results have been given. This appendix provides some details on them. Further details can be found in corresponding text books, some of which are listed in references.

A.1 Polynomials & splines

In this section, we give precise definitions [28] related to polynomials and splines.

Polynomials of Order m. The space

$$\mathcal{P}_m = \left\{ p(x) : p(x) = \sum_{i=1}^{m} c_i x^{i-1}, \quad c_1, \ldots, c_m, x \in \mathcal{R} \right\}$$

of *polynomials of order m* has the following attractive features:

1. $\mathcal{P}_m$ is a finite dimensional linear spaces with a convenient basis;
2. Polynomials are smooth functions;
3. Polynomials are easy to store, manipulate, and evaluate on a digital computer;
4. The derivatives and antiderivatives of polynomials are again polynomials whose coefficients can be found algebraically;

5. The number of zeros of a polynomial of order m cannot exceed $m-1$;
6. Various matrices (arising in interpolation and approximation by polynomials) are always nonsignular, ad they have strong sign-regulatiry properties;
7. The sign structure and shape of a polynomial are intimately related to the sign structure of its set of coefficients;
8. Given any continuous function on the interval $[a,b]$, there exists a polynomial which is uniformly close to it;
9. Precise rates of convergence can be given for approximation of smooth functions by polynomials
10. May approximation processes involving polynomials tend to produce polynomial approximations that oscillate wildly.

Piecewise Polynomials. The main drawback of the space $\mathcal{P}_m$ of polynomials for approximation purposes is that the class is relatively inflexible [28]. Thus we introduce the concept of *piecewise polynomials*.

Let $a = x_0 < x_1 < \cdots < x_k < x_{k+1} = b$, and write $\Delta = \{x_i\}_0^{k+1}$. The set Δ partitions the interval $[a,b]$ into $k+1$ subinterval, $I_i = [x_i, x_{i+1}), i = 0, 1, \ldots, k-1$ and $I_k = [x_i, x_{k+1}]$. Given a positive integer m, let

$$\begin{aligned}\mathcal{PP}_m(\Delta) = \{f : \text{there exists polynomials } p_0, p_1, \ldots, p_k \text{ in } \mathcal{P}_m \\ \text{with } f(x) = p_i(x) \text{ for } x \in I_i, i = 0, 1, \ldots, k\} \quad \text{(A.1)}\end{aligned}$$

We call $\mathcal{PP}_m(\Delta)$ the *space of piecewise polynomials of order m with knots $x_1, \ldots, x_k$.*

Note piecewise polynomial functions are not necessarily smooth and can be discontinuous.

Polynomial Splines With Simple Knots. Let Δ be a partition of the interval $[a,b]$ as in definition of piecewise polynomials, and let m be a positive integer. Let

$$\mathcal{S}_m(\Delta) = \mathcal{PP}_m(\Delta) \cap C^{m-2}[a,b],$$

where $\mathcal{PP}_m(\Delta)$ is the space of pieceswise polynomials. We call $\mathcal{S}_m(\Delta)$ the space of *polynomial splines of order m with simple knots at the points* $x_1, \ldots, x_k$.

Polynomial splines possess the following attractive features:

1. Polynomial spline spaces are finite dimensional linear spaces with very convenient bases;

2. Polynomial splines are relatively smooth functions;

3. Polynomial splines are easy to store, manipulate, and evaluate on a digital computer;

4. The derivatives and antiderivatives of polynomial splines are again polynomial splines whose expansions can be found on a computer;

5. Polynomial splines possess nice zero properties analogous to those for polynomials;

6. Various matrices arising naturally in the use of splines in approximation theory and numerical analysis have convenient sign and determinantal properties;

7. The sign structure and shape of a polynomial spline can be related to the sign structure of its coefficients;

8. Every continuous function on the interval $[a, b]$ can be approximated arbitrarily well by polynomial splines *with the order m fixed*, provided a sufficient number of knots are allowed;

9. Precise rates of convergence can be given for approximation of smooth functions by splines - not only are the functions themselves approximated to high order, but their derivatives are *simultaneously* approximated well;

10. Low-order splines are very flexible, and do not exhibit the oscillations usually associated with polynomials.

A.2 Basis functions

There are two main popular bases - truncated power functions and the B-spline basis. Using the truncated power functions, the *truncated power basis of degree p* is

$$1, t, t^2, \ldots, t^p, (t-\kappa_1)_+^p, \ldots, (t-\kappa_r)_+^p$$

where $\{\kappa_1, \ldots, \kappa_r\}$ is the set of knots and r is the number of knots and $(t-\kappa)_+^p$ is defined to be

$$(t-\kappa)_+^p = \begin{cases} (t-\kappa)_+^p & t > \kappa \\ 0 & \text{otherwise} \end{cases}$$

For example, a cubic spline can be written as

$$g(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \sum_{k=1}^{r} b_k (t-\kappa_k)_+^3.$$

Another popular basis is the B-spline basis. Compared to the truncated power basis, the B-spline basis is more numerically stable, since the former is far from being orthogonal. A B-spline consists of specially connected polynomial pieces. A B-spline of degree q has the following general properties [8]:

- It consists of $q+1$ polynomial pieces, each of degree q;
- the polynomial pieces join at q inner knots;
- at the joining points, derivatives up to order $q-1$ are continuous;
- the B-spline is positive on a domain spanned by $q+2$ knots; zero everywhere else;
- at a given t, $q+1$ B-splines are nonzero.

For example, let $B_{j,4}(t)$ be the jth B-spline basis function, $j = -3, \ldots, e$ of order 4 (degree 3) for knots $\{\kappa_1, \ldots, \kappa_e\}$. The cubic B-splines can be defined recursively [2, 11] in terms of lower-order B-splines

$$B_{j,1}(t) = \begin{cases} 1 & \kappa_j \le t < \kappa_{j+1}, \\ 0 & \text{otherwise} \end{cases}$$

$$B_{j,k+1}(t) = \frac{t-\kappa_j}{\kappa_{j+k}-\kappa_j} B_{j,k}(t) + \frac{\kappa_{j+k+1}-t}{\kappa_{j+k+1}-\kappa_{j+1}} B_{j+1,k}(t), \qquad k > 0.$$

In addition, there are other basis functions available, such as polynomial basis, cubic spline basis and radial basis functions. For more details, see Section 3.2 in [38] and Section 3.7 in [27].

A.3 Stochastic processes and random fields

Stochastic process and random fields

Given a parameter space T, a *stochastic process* f over T is a collection of random variables

$$\{f(t) : t \in T\}.$$

If T is a set of dimension N, and the random variables $f(t)$ are all vector valued of dimension d, then we call the vector valued *random field* f a (N, d) random field [1].

A stochastic process $\{Y_t : t \geq 0\}$ is said to be

- **stationary** if, for all $t_1 < \cdots < t_n$ and $h > 0$, the random n-vectors $(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_n})$ and $(Y_{t1+h}, Y_{t2+h}, \ldots, Y_{tn+h})$ are identically distributed; that is, time shifts leave joint probabilities unchanged.
- **Markovian** if, for all $t_1 < t_2 < \cdots < t_n$, $P(Y_{t_n} \leq y | Y_{t_1}, Y_{t_2}, \ldots, Y_{t_{n-1}}) = P(Y_{t_n} \leq y | Y_{t_{n-1}})$; that is, the future is determined only by the present and not the past.

Gaussian process and Gaussian fields

A real-valued *Gaussian (random) field* [1] or *Gaussian (random) process* is defined to be a random field f on a parameter set T for which the finite distributions of $(f_{t_1}, f_{t_2}, \ldots, f_{t_k})$ is multivariate Gaussian for each $1 \leq k < \infty$ and each $(t_1, \ldots, t_k) \in T^k$.

Since multivariate Gaussian distributions are determined by means and covariances, Gaussian random fields are also determined by their mean and covariance functions, given by

$$m(t) = \mathbb{E}[f(t)]$$

and

$$C(s,t) = \mathbb{E}[(f(s) - m(s))(f(t) - m(t))].$$

Multivariate Gaussian fields taking values in $\mathbb{R}^d$ are fields for which $\langle \alpha, f_t \rangle$ are a real valued Gaussian field for every $\alpha \in \mathbb{R}^d$. The mean function $m(t)$ now takes values in $\mathbb{R}^d$ and the covariance is non-negative definite $d \times d$ matrices:

$$C(s,t) = \mathbb{E}[(f(s) - m(s))'(f(t) - m(t))]$$

with (i,j)th element being

$$C_{ij}(s,t) = \mathbb{E}[(f_i(s) - m_i(s))'(f_j(t) - m_j(t))].$$

Ornstein-Hulenbeck (OU) process

An Ornstein-Hulenbeck process X_t [18] with parameters (a, σ) and initial condition ξ satisfies the following stochastic differential equation:

$$dX_t = -aX_t dt + \sigma dW_t,$$

whose solution is

$$X_t = e^{-at}\xi + \sigma \int_0^t e^{-a(t-s)} dW_s.$$

Its expectation and covariance are

$$E(X_t) = e^{-at} E(\xi),$$

and

$$\begin{aligned} E(X_s X_t) &= e^{-as} e^{-at} E(\xi^2) + \sigma^2 \int_0^{s \wedge t} e^{-a(s-u)-a(t-u)} du \\ &= e^{-a(s+t)} \left(E(\xi^2) + \sigma^2 \frac{e^{2as \wedge t} - 1}{2a} \right). \end{aligned}$$

If ξ is Gaussian, then X_t is a Gaussian process. In particular, if ξ is Gaussian with mean zero and $E(\xi^2) = \frac{\sigma^2}{2a}$, then

$$E(X_s X_t) = \frac{\sigma^2 e^{-a|s-t|}}{2a}.$$

The OU process is stationary, Gaussian, and Markovian, and is the only nontrivial process that satisfies these three conditions.

A.4 Theoretical derivations for Chapter 4

A.4.1 Proof of the normal matrix (4.6)

Proof of (4.6). Taking derivative of the log-likelihood function (4.5) with respect to $\boldsymbol{\beta}$, $\boldsymbol{f}_1$, and $\boldsymbol{f}_1$, we have

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

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

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

$$\boldsymbol{N}_1^T\boldsymbol{W}(\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{N}_1\hat{\boldsymbol{f}}_1+\boldsymbol{N}_2\hat{\boldsymbol{f}}_2)+\lambda_1\boldsymbol{K}\hat{\boldsymbol{f}}_1 = \boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{Y} \tag{A.3}$$

$$\boldsymbol{N}_2^T\boldsymbol{W}(\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{N}_1\hat{\boldsymbol{f}}_1+\boldsymbol{N}_2\hat{\boldsymbol{f}}_2)+\lambda_2\boldsymbol{K}\hat{\boldsymbol{f}}_2 = \boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{Y}, \tag{A.4}$$

which can be rewritten as (4.6). □

A.4.2 Proof of (4.7), (4.8) and (4.9)

Proof of (4.7), (4.8) and (4.9). From equations (A.2), (A.3) and (A.4), we can reexpress the parameter estimators

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

$$\hat{\boldsymbol{f}}_1 = (\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1+\lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{N}_2\hat{\boldsymbol{f}}_2) \tag{A.5}$$

$$\hat{\boldsymbol{f}}_2 = (\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2+\lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{N}_1\hat{\boldsymbol{f}}_1). \tag{A.6}$$

To solve explicitly for $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{f}}_1$ and $\hat{\boldsymbol{f}}_2$, we first plug $\hat{\boldsymbol{f}}_1$ (A.5) into equations (A.2) and (A.4), rearrange and obtain

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

and

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

respectively; which can be rewritten as

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

and

$$\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{X}\hat{\boldsymbol{\beta}} + (\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2 + \boldsymbol{\lambda}_2\boldsymbol{K})\hat{\boldsymbol{f}}_2 = \boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{Y} \tag{A.8}$$

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

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

and

$$\hat{\boldsymbol{f}}_2 = (\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_1(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) \tag{A.10}$$

respectively. Then plugging (A.9) into (A.8) and (A.10) into (A.7) and rearrange, we have

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

and

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

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

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

and

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

where $\boldsymbol{W}_{f_2} = \boldsymbol{W}_1 - \boldsymbol{W}_1\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1$, and $\boldsymbol{W}_x = \boldsymbol{W}_1 - \boldsymbol{W}_1\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_1$. Similarly, to obtain the closed-form solution for $\hat{\boldsymbol{f}}_{\boldsymbol{1}}$, we plug (A.6) into equation (A.3) and obtain

$$\boldsymbol{N}_1^T\boldsymbol{W}_2\boldsymbol{X}\hat{\boldsymbol{\beta}} + (\boldsymbol{N}_1^T\boldsymbol{W}_2\boldsymbol{N}_1 + \boldsymbol{\lambda}_1\boldsymbol{K})\hat{\boldsymbol{f}}_1 = \boldsymbol{N}_1^T\boldsymbol{W}_2\boldsymbol{Y}, \tag{A.11}$$

where $\boldsymbol{W}_2 = \boldsymbol{W} - \boldsymbol{W}\boldsymbol{N}_2(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}$. Plugging $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_2(\boldsymbol{Y} - \boldsymbol{N}_1\hat{\boldsymbol{f}}_1)$ into (A.11), the closed-form solution for $\hat{\boldsymbol{f}}_1$ is

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

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

Proof of (4.10) and (4.11). From model assumptions, it follows that

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

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

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

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

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

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

A.4.3 Proof of biases of (4.12) and (4.13) and a lemma

Proof of (4.12) and (4.13). 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

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

□

Remark. The bias of nonparametric function estimator $\hat{\boldsymbol{f}}_2$ in (4.14) can be derived similarly as that of $\hat{\boldsymbol{f}}_1$.

Lemma 5. *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

$$\boldsymbol{W}_x \to \boldsymbol{W}_1 - \boldsymbol{W}_1 \boldsymbol{N}_2 (\boldsymbol{N}_2^T \boldsymbol{W}_1 \boldsymbol{N}_2)^{-1} \boldsymbol{N}_2^T \boldsymbol{W}_1, \tag{A.12}$$

where

$$\boldsymbol{W}_1 \to \boldsymbol{W} - \boldsymbol{W} \boldsymbol{N}_1 (\boldsymbol{N}_1^T \boldsymbol{W} \boldsymbol{N}_1)^{-1} \boldsymbol{N}_1^T \boldsymbol{W} \tag{A.13}$$

Plugging $\boldsymbol{W}_x$ in (A.12) into bias of $\hat{\boldsymbol{\beta}}$ in (4.12), we have

$$
\begin{aligned}
& \mathrm{E}(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta} \\
= \; & (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\left[\boldsymbol{W}_1 - \boldsymbol{W}_1\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_1\right](\boldsymbol{N}_1\boldsymbol{f}_1 + \boldsymbol{N}_2\boldsymbol{f}_2) \\
= \; & (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1(\boldsymbol{N}_1\boldsymbol{f}_1 + \boldsymbol{N}_2\boldsymbol{f}_2) \\
& \qquad - (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_1(\boldsymbol{N}_1\boldsymbol{f}_1 + \boldsymbol{N}_2\boldsymbol{f}_2) \\
= \; & (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{N}_1\boldsymbol{f}_1 \\
& + (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{N}_2\left[\boldsymbol{f}_2 - (\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2\boldsymbol{f}_2\right] \\
& - (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_1\boldsymbol{f}_1 \\
= \; & (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{N}_1\boldsymbol{f}_1 \\
& \qquad - (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_1\boldsymbol{f}_1
\end{aligned} \tag{A.14}
$$

Further, plugging $\boldsymbol{W}_1$ in (A.13) into(A.14), we have

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

$$
\mathrm{E}(\hat{\boldsymbol{f}}_{\mathbf{1}}) - \boldsymbol{f}_{\mathbf{1}} \to (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_2\boldsymbol{f}_2 \tag{A.15}
$$

where

$$\boldsymbol{W}_{f_1} = \boldsymbol{W}_2 - \boldsymbol{W}_2\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_2 \tag{A.16}$$

and

$$\boldsymbol{W}_2 \to \boldsymbol{W} - \boldsymbol{W}\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W} \tag{A.17}$$

Plugging $\boldsymbol{W}_{f_1}$ in (A.16) into bias of $\hat{\boldsymbol{f}}_{\boldsymbol{1}}$ in (A.15) and plugging $\boldsymbol{W}_2$ in (A.17) into $\boldsymbol{W}_{f_1}$ in (A.16), we have

$$\begin{aligned}
& \mathrm{E}(\hat{\boldsymbol{f}}_{\boldsymbol{1}}) - \boldsymbol{f}_{\boldsymbol{1}} \\
= \; & (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_2\boldsymbol{N}_2\boldsymbol{f}_2 - (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_2\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{N}_2\boldsymbol{f}_2 \\
= \; & (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_2\boldsymbol{f}_2 - (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2\boldsymbol{f}_2 \\
& -(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1)^{-1}\boldsymbol{N}_1^T\left[\boldsymbol{W} - \boldsymbol{W}\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2)^{-1}\boldsymbol{N}_2^T\boldsymbol{W}\right]\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}^T \cdot \\
& \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 \cdot \\
& \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. □

A.4.4 Proof of covariance of random effects (4.15)

Proof.

$$\text{Cov}(\hat{\boldsymbol{b}}_i - \boldsymbol{b}_i) = \text{Var}(\hat{\boldsymbol{b}}_i) + \text{Var}(\boldsymbol{b}_i) - \text{Cov}(\boldsymbol{b}_i, \hat{\boldsymbol{b}}_i) - \text{Cov}(\hat{\boldsymbol{b}}_i, \boldsymbol{b}_i).$$

Since $\text{Var}(\boldsymbol{b}_i) = \boldsymbol{D}$, it suffices to find $\text{Var}(\hat{\boldsymbol{b}}_i)$ and $\text{Cov}(\boldsymbol{b}_i, \hat{\boldsymbol{b}}_i)$. To find

$$\text{Var}(\hat{\boldsymbol{b}}_i) = \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i\text{Var}(\boldsymbol{Y}_i - \boldsymbol{X_i}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{f}}_{\boldsymbol{1i}} - \hat{\boldsymbol{f}}_{\boldsymbol{2i}})\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D},$$

we will first find

$$\begin{aligned}
& \text{Var}(\boldsymbol{Y}_i - \boldsymbol{X_i}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{f}}_{\boldsymbol{1i}} - \hat{\boldsymbol{f}}_{\boldsymbol{2i}}) \\
= \; & \text{Var}(\boldsymbol{Y}_i) + \text{Var}(\boldsymbol{X}_i\hat{\boldsymbol{\beta}}) + \text{Var}(\boldsymbol{N}_{1i}\hat{\boldsymbol{f}}_{\boldsymbol{1}}) + \text{Var}(\boldsymbol{N}_{2i}\hat{\boldsymbol{f}}_{\boldsymbol{2}}) \\
& -\text{Cov}(\boldsymbol{Y}_i, \boldsymbol{X}_i\hat{\boldsymbol{\beta}}) - \text{Cov}(\boldsymbol{X}_i\hat{\boldsymbol{\beta}}, \boldsymbol{Y}_i) \\
& -\text{Cov}(\boldsymbol{Y}_i, \hat{\boldsymbol{f}}_{1i}) - \text{Cov}(\hat{\boldsymbol{f}}_{1i}, \boldsymbol{Y}_i) - \text{Cov}(\boldsymbol{Y}_i, \hat{\boldsymbol{f}}_{2i}) - \text{Cov}(\hat{\boldsymbol{f}}_{2i}, \boldsymbol{Y}_i) \\
& +\text{Cov}(\boldsymbol{X}_i\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{f}}_{1i}) + \text{Cov}(\hat{\boldsymbol{f}}_{1i}, \boldsymbol{X}_i\hat{\boldsymbol{\beta}}) + \text{Cov}(\boldsymbol{X}_i\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{f}}_{2i}) + \text{Cov}(\hat{\boldsymbol{f}}_{2i}, \boldsymbol{X}_i\hat{\boldsymbol{\beta}}) \\
& +\text{Cov}(\hat{\boldsymbol{f}}_{1i}, \hat{\boldsymbol{f}}_{2i}) + \text{Cov}(\hat{\boldsymbol{f}}_{2i}, \hat{\boldsymbol{f}}_{1i})
\end{aligned}$$

Denote

$$\begin{aligned}
\boldsymbol{W}_{x_i} &= \boldsymbol{W}_{1i} - \boldsymbol{W}_{1i}\boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T\boldsymbol{W}_{1i} \\
\boldsymbol{W}_{f_{1i}} &= \boldsymbol{W}_{2i} - \boldsymbol{W}_{2i}\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_{2i} \\
\boldsymbol{W}_{f_{2i}} &= \boldsymbol{W}_{1i} - \boldsymbol{W}_{1i}\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_{1i}
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{W}_{1i} &= \boldsymbol{W}_i - \boldsymbol{W}_i\boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T\boldsymbol{W}_i \\
\boldsymbol{W}_{2i} &= \boldsymbol{W}_i - \boldsymbol{W}_i\boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T\boldsymbol{W}_i
\end{aligned}$$

we have

$$
\begin{aligned}
\mathrm{Cov}(\boldsymbol{Y}_i, \boldsymbol{X}_i\hat{\boldsymbol{\beta}}) &= \mathrm{Cov}(\boldsymbol{Y}_i, \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{Y}) \\
&= \mathrm{Cov}\left(\boldsymbol{Y}_i, \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\sum_{j=1}^{m}\boldsymbol{X}_j^T\boldsymbol{W}_{x_j}\boldsymbol{Y}_j\right) \\
&= \mathrm{Cov}\left(\boldsymbol{Y}_i, \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_{x_i}\boldsymbol{Y}_i\right) \\
&= \mathrm{Cov}(\boldsymbol{Y}_i, \boldsymbol{Y}_i)[\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_{x_i}]^T \\
&= \boldsymbol{V}_i\boldsymbol{W}_{x_i}\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T \\
&= \boldsymbol{V}_i[\boldsymbol{W}_{1i} - \boldsymbol{W}_{1i}\boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T\boldsymbol{W}_{1i}]\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T \\
&= \boldsymbol{V}_i\boldsymbol{W}_{1i}[\boldsymbol{I}_i - \boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T\boldsymbol{W}_{1i}]\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T \\
&= \boldsymbol{V}_i[\boldsymbol{W}_i - \boldsymbol{W}_i\boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T\boldsymbol{W}_i] \\
&\qquad [\boldsymbol{I}_i - \boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T\boldsymbol{W}_{1i}]\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T \\
&= [\boldsymbol{I}_i - \boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T\boldsymbol{W}_i] \\
&\qquad [\boldsymbol{I}_i - \boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T\boldsymbol{W}_{1i}]\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Cov}(\boldsymbol{X}_i\hat{\boldsymbol{\beta}}, \boldsymbol{Y}_i) \\
=\ & [\mathrm{Cov}(\boldsymbol{Y}_i, \boldsymbol{X}_i\hat{\boldsymbol{\beta}})]^T \\
=\ & \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T[\boldsymbol{I}_i - \boldsymbol{W}_{1i}\boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T]\cdot \\
&\qquad [\boldsymbol{I}_i - \boldsymbol{W}_i\boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T]
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Cov}(\boldsymbol{Y}_i, \hat{\boldsymbol{f}}_{1i}) &= \mathrm{Cov}(\boldsymbol{Y}_i, \boldsymbol{N}_{1i}\hat{\boldsymbol{f}}_i) \\
&= \mathrm{Cov}(\boldsymbol{Y}_i, \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{Y}) \\
&= \mathrm{Cov}(\boldsymbol{Y}_i, \boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\sum_{j=1}^{m}\boldsymbol{N}_{1j}^T\boldsymbol{W}_{f_{1j}}\boldsymbol{Y}_j) \\
&= \mathrm{Cov}(\boldsymbol{Y}_i, \boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T\boldsymbol{W}_{f_{1i}}\boldsymbol{Y}_i) \\
&= \mathrm{Cov}(\boldsymbol{Y}_i, \boldsymbol{Y}_i)[\boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T\boldsymbol{W}_{f_{1i}}]^T \\
&= \boldsymbol{V}_i\boldsymbol{W}_{f_{1i}}\boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T \\
&= \boldsymbol{V}_i[\boldsymbol{W}_{2i} - \boldsymbol{W}_{2i}\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_{2i}] \cdot \\
&\qquad \boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T \\
&= \boldsymbol{V}_i\boldsymbol{W}_{2i}[\boldsymbol{I}_i - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_{2i}]\boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T \\
&= \boldsymbol{V}_i[\boldsymbol{W}_i - \boldsymbol{W}_i\boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T\boldsymbol{W}_i] \\
&\qquad [\boldsymbol{I}_i - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_{2i}]\boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T \\
&= [\boldsymbol{I}_i - \boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T\boldsymbol{W}_i] \\
&\qquad [\boldsymbol{I}_i - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_{2i}]\boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Cov}(\hat{\boldsymbol{f}}_i, \boldsymbol{Y}_i) &= [\mathrm{Cov}(\boldsymbol{Y}_i, \hat{\boldsymbol{f}}_i)]^T \\
&= \boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T[\boldsymbol{I}_i - \boldsymbol{W}_{2i}\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}_i^T] \\
&\qquad [\boldsymbol{I}_i - \boldsymbol{W}_i\boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T]
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Cov}(\boldsymbol{Y}_i, \hat{\boldsymbol{f}}_{2i}) &= \mathrm{Cov}(\boldsymbol{Y}_i, \boldsymbol{N}_{2i}\hat{\boldsymbol{f}}_i) \\
&= \mathrm{Cov}(\boldsymbol{Y}_i, \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{Y}) \\
&= \mathrm{Cov}(\boldsymbol{Y}_i, \boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\sum_{j=1}^{m}\boldsymbol{N}_{2j}^T\boldsymbol{W}_{f_{2j}}\boldsymbol{Y}_j) \\
&= \mathrm{Cov}(\boldsymbol{Y}_i, \boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T\boldsymbol{W}_{f_{2i}}\boldsymbol{Y}_i) \\
&= \mathrm{Cov}(\boldsymbol{Y}_i, \boldsymbol{Y}_i)[\boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T\boldsymbol{W}_{f_{2i}}]^T \\
&= \boldsymbol{V}_i\boldsymbol{W}_{f_{2i}}\boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T \\
&= \boldsymbol{V}_i[\boldsymbol{W}_{1i} - \boldsymbol{W}_{1i}\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_{1i}] \cdot \\
&\qquad \boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T \\
&= \boldsymbol{V}_i\boldsymbol{W}_{1i}[\boldsymbol{I}_i - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_{1i}]\boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T \\
&= \boldsymbol{V}_i[\boldsymbol{W}_i - \boldsymbol{W}_i\boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T\boldsymbol{W}_i] \cdot \\
&\qquad [\boldsymbol{I}_i - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_{1i}]\boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T \\
&= [\boldsymbol{I}_i - \boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T\boldsymbol{W}_i] \cdot \\
&\qquad [\boldsymbol{I}_i - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_{1i}]\boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Cov}(\hat{\boldsymbol{f}}_i, \boldsymbol{Y}_i) &= [\mathrm{Cov}(\boldsymbol{Y}_i, \hat{\boldsymbol{f}}_i)]^T \\
&= \boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T[\boldsymbol{I}_i - \boldsymbol{W}_{1i}\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}_i^T] \\
&\qquad [\boldsymbol{I}_i - \boldsymbol{W}_i\boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T]
\end{aligned}
$$

where $I_i = I_{n_i}$, and[1]

$$
\begin{aligned}
&\mathrm{Cov}(\boldsymbol{X}_i\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{f}}_{1i}) \\
= \;& \mathrm{Cov}\left[\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{Y}, \boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{Y}\right] \\
= \;& \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\mathrm{Cov}(\boldsymbol{Y},\boldsymbol{Y})[\boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}]^T \\
= \;& \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{V}\boldsymbol{W}_{f_1}\boldsymbol{N}_1(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Cov}(\hat{\boldsymbol{f}}_{1i}, \boldsymbol{X}_i\hat{\boldsymbol{\beta}}) &= [\mathrm{Cov}(\boldsymbol{X}_i\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{f}}_i)]^T \\
&= \boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{V}\boldsymbol{W}_x\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Cov}(\boldsymbol{X}_i\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{f}}_{2i}) \\
= \;& \mathrm{Cov}\left[\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{Y}, \boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{Y}\right] \\
= \;& \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\mathrm{Cov}(\boldsymbol{Y},\boldsymbol{Y})[\boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}]^T \\
= \;& \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{V}\boldsymbol{W}_{f_2}\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Cov}(\hat{\boldsymbol{f}}_{2i}, \boldsymbol{X}_i\hat{\boldsymbol{\beta}}) &= [\mathrm{Cov}(\boldsymbol{X}_i\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{f}}_i)]^T \\
&= \boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{V}\boldsymbol{W}_x\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathrm{Cov}(\hat{\boldsymbol{f}}_{1i}, \hat{\boldsymbol{f}}_{2i}) \\
= \;& \mathrm{Cov}\left[\boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{Y}, \boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{Y}\right] \\
= \;& \boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\mathrm{Cov}(\boldsymbol{Y},\boldsymbol{Y})[\boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}]^T \\
= \;& \boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{V}\boldsymbol{W}_{f_2}\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{Cov}(\hat{\boldsymbol{f}}_{2i}, \hat{\boldsymbol{f}}_{1i}) \\
= \;& [\mathrm{Cov}(\boldsymbol{X}_i\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{f}}_i)]^T \\
= \;& \boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{V}\boldsymbol{W}_{f_1}\boldsymbol{N}_1(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T.
\end{aligned}
$$

[1]Recall $\mathrm{Cov}(\boldsymbol{A}\boldsymbol{X}, \boldsymbol{B}^T\boldsymbol{Y}) = \boldsymbol{A}\mathrm{Cov}(\boldsymbol{X},\boldsymbol{Y})\boldsymbol{B}$

Thus,

$$
\begin{aligned}
& \operatorname{Var}(\boldsymbol{Y}_i - \boldsymbol{X_i}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{f}}_{\boldsymbol{1i}} - \hat{\boldsymbol{f}}_{\boldsymbol{2i}}) \\
= \; & \operatorname{Var}(\boldsymbol{Y}_i) + \operatorname{Var}(\boldsymbol{X}_i\hat{\boldsymbol{\beta}}) + \operatorname{Var}(\boldsymbol{N}_{1i}\hat{\boldsymbol{f}}_{\boldsymbol{1}}) + \operatorname{Var}(\boldsymbol{N}_{2i}\hat{\boldsymbol{f}}_{\boldsymbol{2}}) \\
& \quad -\operatorname{Cov}(\boldsymbol{Y}_i, \boldsymbol{X}_i\hat{\boldsymbol{\beta}}) - \operatorname{Cov}(\boldsymbol{X}_i\hat{\boldsymbol{\beta}}, \boldsymbol{Y}_i) \\
& \quad -\operatorname{Cov}(\boldsymbol{Y}_i, \hat{\boldsymbol{f}}_{1i}) - \operatorname{Cov}(\hat{\boldsymbol{f}}_{1i}, \boldsymbol{Y}_i) \\
& \quad -\operatorname{Cov}(\boldsymbol{Y}_i, \hat{\boldsymbol{f}}_{2i}) - \operatorname{Cov}(\hat{\boldsymbol{f}}_{2i}, \boldsymbol{Y}_i) \\
& \quad +\operatorname{Cov}(\boldsymbol{X}_i\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{f}}_{1i}) + \operatorname{Cov}(\hat{\boldsymbol{f}}_{1i}, \boldsymbol{X}_i\hat{\boldsymbol{\beta}}) \\
& \quad +\operatorname{Cov}(\boldsymbol{X}_i\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{f}}_{2i}) + \operatorname{Cov}(\hat{\boldsymbol{f}}_{2i}, \boldsymbol{X}_i\hat{\boldsymbol{\beta}}) \\
& \quad +\operatorname{Cov}(\hat{\boldsymbol{f}}_{1i}, \hat{\boldsymbol{f}}_{2i}) + \operatorname{Cov}(\hat{\boldsymbol{f}}_{2i}, \hat{\boldsymbol{f}}_{1i}) \\
= \; & \boldsymbol{V}_i + \boldsymbol{X}_i(\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}\boldsymbol{X}_i^T \\
& \quad +\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{V}\boldsymbol{W}_{f_1}\boldsymbol{N}_1(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T \\
& \quad +\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{V}\boldsymbol{W}_{f_2}\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T \\
& \quad -[\boldsymbol{I}_i - \boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T\boldsymbol{W}_i] \\
& \qquad [\boldsymbol{I}_i - \boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T\boldsymbol{W}_{1i}]\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T \\
& \quad -\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T[\boldsymbol{I}_i - \boldsymbol{W}_{1i}\boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T] \\
& \qquad [\boldsymbol{I}_i - \boldsymbol{W}_i\boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T] \\
& \quad -[\boldsymbol{I}_i - \boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T\boldsymbol{W}_i][\boldsymbol{I}_i - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_{2i}] \\
& \qquad \boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T \\
& \quad -\boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T[\boldsymbol{I}_i - \boldsymbol{W}_{2i}\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}_i^T] \\
& \qquad [\boldsymbol{I}_i - \boldsymbol{W}_i\boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T] \\
& \quad -[\boldsymbol{I}_i - \boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T\boldsymbol{W}_i][\boldsymbol{I}_i - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_{1i}] \\
& \qquad \boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T \\
& \quad -\boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T[\boldsymbol{I}_i - \boldsymbol{W}_{1i}\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}_i^T] \\
& \qquad [\boldsymbol{I}_i - \boldsymbol{W}_i\boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T] \\
& \quad +\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\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}\boldsymbol{N}_{1i}^T \\
& \quad +\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{V}\boldsymbol{W}_x\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T \\
& \quad +\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\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}\boldsymbol{N}_{2i}^T \\
& \quad +\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{V}\boldsymbol{W}_x\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T \\
& \quad +\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{V}\boldsymbol{W}_{f_2}\boldsymbol{N}_2(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T \\
& \quad +\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{V}\boldsymbol{W}_{f_1}\boldsymbol{N}_1(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T.
\end{aligned}
$$

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

Now, we want to find $\text{Cov}(\hat{\boldsymbol{b}}_i, \boldsymbol{b}_i)$

$$\begin{aligned}
\text{Cov}(\boldsymbol{b}_i, \hat{\boldsymbol{b}}_i) &= \text{Cov}[\boldsymbol{b}_i, \boldsymbol{D}\boldsymbol{Z}_i^T \boldsymbol{W}_i(\boldsymbol{Y}_i - \boldsymbol{X_i}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{f}}_{\boldsymbol{1i}} - \hat{\boldsymbol{f}}_{\boldsymbol{2i}})] \\
&= \text{Cov}(\boldsymbol{b}_i, \boldsymbol{Y}_i - \boldsymbol{X_i}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{f}}_{\boldsymbol{1i}} - \hat{\boldsymbol{f}}_{\boldsymbol{2i}})\boldsymbol{W}_i \boldsymbol{Z}_i \boldsymbol{D} \\
&= [\text{Cov}(\boldsymbol{b}_i, \boldsymbol{Y}_i) - \text{Cov}(\boldsymbol{b}_i, \boldsymbol{X}_i\hat{\boldsymbol{\beta}}) - \text{Cov}(\boldsymbol{b}_i, \hat{\boldsymbol{f}}_{1i}) - \text{Cov}(\boldsymbol{b}_i, \hat{\boldsymbol{f}}_{2i})]\boldsymbol{W}_i \boldsymbol{Z}_i \boldsymbol{D}
\end{aligned}$$

where

$$\begin{aligned}
\text{Cov}(\boldsymbol{b}_i, \boldsymbol{Y}_i) &= \text{Cov}(\boldsymbol{b}_i, \boldsymbol{X_i}\boldsymbol{\beta} + \boldsymbol{N}_{1i}\boldsymbol{f}_1 + \boldsymbol{N}_{2i}\boldsymbol{f}_2 + \boldsymbol{Z_i}\boldsymbol{b_i} + \boldsymbol{U}_i + \boldsymbol{\epsilon}_i) \\
&= \text{Cov}(\boldsymbol{b}_i, \boldsymbol{Z}_i\boldsymbol{b}_i) \\
&= \boldsymbol{D}\boldsymbol{Z}_i^T,
\end{aligned}$$

$$\begin{aligned}
\text{Cov}(\boldsymbol{b}_i, \boldsymbol{X}_i\hat{\boldsymbol{\beta}}) &= \text{Cov}[\boldsymbol{b}_i, \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{Y}] \\
&= \text{Cov}\left[\boldsymbol{b}_i, \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\sum_{j=1}^{m}\boldsymbol{X}_j^T\boldsymbol{W}_{x_j}\boldsymbol{Y}_j\right] \\
&= \text{Cov}[\boldsymbol{b}_i, \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_{x_i}\boldsymbol{Y}_i] \\
&= \text{Cov}(\boldsymbol{b}_i, \boldsymbol{Y}_i)[\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_{x_i}]^T \\
&= \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_{x_i}\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T \\
&= \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_{1i}[\boldsymbol{I}_i - \boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T\boldsymbol{W}_{1i}] \cdot \\
&\qquad \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T \\
&= \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i[\boldsymbol{I}_i - \boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T\boldsymbol{W}_i] \\
&\qquad [\boldsymbol{I}_i - \boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_1\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T\boldsymbol{W}_{1i}]\boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}_i^T
\end{aligned}$$

$$
\begin{aligned}
\mathrm{Cov}(\boldsymbol{b}_i, \hat{\boldsymbol{f}}_{1i}) &= \mathrm{Cov}(\boldsymbol{b}_i, \boldsymbol{N}_{1i}\hat{\boldsymbol{f}}_i) \\
&= \mathrm{Cov}[\boldsymbol{b}_i, \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{Y}] \\
&= \mathrm{Cov}\left[\boldsymbol{b}_i, \boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\sum_{j=1}^{m}\boldsymbol{N}_{1j}^T\boldsymbol{W}_{f_{1j}}\boldsymbol{Y}_j\right] \\
&= \mathrm{Cov}\left[\boldsymbol{b}_i, \boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T\boldsymbol{W}_{f_{1i}}\boldsymbol{Y}_i\right] \\
&= \mathrm{Cov}(\boldsymbol{b}_i, \boldsymbol{Y}_i)\left[\boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T\boldsymbol{W}_{f_{1i}}\right]^T \\
&= \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_{f_{1i}}\boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T \\
&= \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_{2i}[\boldsymbol{I}_i - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_{2i}]\cdot \\
&\qquad\qquad \boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T \\
&= \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i[\boldsymbol{I}_i - \boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T\boldsymbol{W}_i] \\
&\quad [\boldsymbol{I}_i - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_2\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_{2i}]\boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Cov}(\boldsymbol{b}_i, \hat{\boldsymbol{f}}_{2i}) &= \mathrm{Cov}(\boldsymbol{b}_i, \boldsymbol{N}_{2i}\hat{\boldsymbol{f}}_i) \\
&= \mathrm{Cov}(\boldsymbol{b}_i, \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{Y}) \\
&= \mathrm{Cov}\left[\boldsymbol{b}_i, \boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\sum_{j=1}^{m}\boldsymbol{N}_{2j}^T\boldsymbol{W}_{f_{2j}}\boldsymbol{Y}_j\right] \\
&= \mathrm{Cov}\left[\boldsymbol{b}_i, \boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T\boldsymbol{W}_{f_{2i}}\boldsymbol{Y}_i\right] \\
&= \mathrm{Cov}(\boldsymbol{b}_i, \boldsymbol{Y}_i)\left[\boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T\boldsymbol{W}_{f_{2i}}\right]^T \\
&= \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_{f_{2i}}\boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T \\
&= \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_{1i}[\boldsymbol{I}_i - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_{1i}]\cdot \\
&\qquad\qquad \boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T \\
&= \boldsymbol{D}\boldsymbol{Z}_i^T\boldsymbol{W}_i[\boldsymbol{I}_i - \boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_{1i}^T\boldsymbol{W}_i] \\
&\quad [\boldsymbol{I}_i - \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_1\boldsymbol{X})^{-1}\boldsymbol{X}_i^T\boldsymbol{W}_{1i}]\boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_{2i}^T
\end{aligned}
$$

Thus,

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

and

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

Therefore, after cancelling out identical terms, we have

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

$\boldsymbol{W}_i\boldsymbol{Z}_i\boldsymbol{D}$.

Denote

$$
\begin{aligned}
A &= \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x \\
B &= \boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1} \\
C &= \boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}
\end{aligned}
$$

the terms in the square brackets become

$$
\begin{aligned}
& AVA^T + BVB^T + CVC^T + AVB^T + BVA^T + AVC^T + CVA^T + BVC^T + CVB^T \\
= \; & AV(A^T + B^T + C^T) + BV(B^T + A^T + C^T) + CV(C^T + A^T + B^T) \\
= \; & (A + B + C)V(A^T + B^T + C^T) \qquad \text{(A.18)}
\end{aligned}
$$

Recall from the estimation procedure,

$$
\begin{pmatrix} (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{Y} \\ (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1 + \lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{Y} \\ (\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2 + \lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{Y} \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{f}}_1 \\ \hat{\boldsymbol{f}}_2 \end{pmatrix} = \boldsymbol{C}^{-1} \begin{pmatrix} \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y} \\ \boldsymbol{N}_1^T\boldsymbol{W}\boldsymbol{Y} \\ \boldsymbol{N}_2^T\boldsymbol{W}\boldsymbol{Y} \end{pmatrix}.
$$

Dividing $\boldsymbol{Y}$ from both sides and take out the common term $\boldsymbol{W}$, we have

$$\Rightarrow \quad \begin{pmatrix} (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x \\ (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1+\lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1} \\ (\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2+\lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_{f_2} \end{pmatrix} = \boldsymbol{C}^{-1}\begin{pmatrix} \boldsymbol{X}^T \\ \boldsymbol{N}_1^T \\ \boldsymbol{N}_2^T \end{pmatrix}\boldsymbol{W}. \tag{A.19}$$

Thus,

$$\begin{aligned} A+B+C &= \boldsymbol{X}_i(\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x + \boldsymbol{N}_{1i}(\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1+\lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1} \\ &\quad +\boldsymbol{N}_{2i}(\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2+\lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_{f_2} \\ &= \begin{pmatrix} \boldsymbol{X}_i & \boldsymbol{N}_{1i} & \boldsymbol{N}_{2i} \end{pmatrix} \begin{pmatrix} (\boldsymbol{X}^T\boldsymbol{W}_x\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}_x \\ (\boldsymbol{N}_1^T\boldsymbol{W}_{f_1}\boldsymbol{N}_1+\lambda_1\boldsymbol{K})^{-1}\boldsymbol{N}_1^T\boldsymbol{W}_{f_1} \\ (\boldsymbol{N}_2^T\boldsymbol{W}_{f_2}\boldsymbol{N}_2+\lambda_2\boldsymbol{K})^{-1}\boldsymbol{N}_2^T\boldsymbol{W}_{f_2} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{X}_i & \boldsymbol{N}_{1i} & \boldsymbol{N}_{2i} \end{pmatrix} \boldsymbol{C}^{-1} \begin{pmatrix} \boldsymbol{X}^T \\ \boldsymbol{N}_1^T \\ \boldsymbol{N}_2^T \end{pmatrix} \boldsymbol{W} \\ &= \boldsymbol{\chi}_i\boldsymbol{C}^{-1}\boldsymbol{\chi}^T\boldsymbol{W} \end{aligned}$$

by plugging equation (A.19), and

$$A^T+B^T+C^T = (A+B+C)^T = \boldsymbol{W}\boldsymbol{\chi}\boldsymbol{C}^{-1}\boldsymbol{\chi}_i^T$$

which by equation (A.18) gives

$$\begin{aligned} &(A+B+C)\boldsymbol{V}(A^T+B^T+C^T) \\ =\ & \boldsymbol{\chi}_i\boldsymbol{C}^{-1}\boldsymbol{\chi}^T\boldsymbol{W}\boldsymbol{V}\boldsymbol{W}\boldsymbol{\chi}\boldsymbol{C}^{-1}\boldsymbol{\chi}_i^T \\ =\ & \boldsymbol{\chi}_i\boldsymbol{C}^{-1}\boldsymbol{\chi}^T\boldsymbol{W}\boldsymbol{\chi}\boldsymbol{C}^{-1}\boldsymbol{\chi}_i^T. \end{aligned}$$

Therefore, $\mathrm{Cov}(\hat{\boldsymbol{b}}_i - \boldsymbol{b}_i)$ becomes

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

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

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

where $\boldsymbol{\chi}_i$ and $\boldsymbol{\chi}$ are defined the same as before. □