ESTIMATION OF COVARIANCE STRUCTURES IN FUNCTIONAL MIXED MODELS WITH APPLICATION TO HERITABILITY ESTIMATION

by

Mohammad Shaha Alam Patwary

APPROVED BY SUPERVISORY COMMITTEE:

Pankaj K. Choudhary, Chair

Swati Biswas

Min Chen

Sam Efromovich

Copyright © 2020

Mohammad Shaha Alam Patwary

All rights reserved

To the departed soul of my beloved father,

my beloved mother,

beautiful wife,

and

lovely daughter.

ESTIMATION OF COVARIANCE STRUCTURES IN FUNCTIONAL MIXED MODELS WITH APPLICATION TO HERITABILITY ESTIMATION

by

MOHAMMAD SHAHA ALAM PATWARY, BS, MS

DISSERTATION

Presented to the Faculty of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY IN

STATISTICS

THE UNIVERSITY OF TEXAS AT DALLAS

August 2020

ACKNOWLEDGMENTS

I would like to start by thanking my advisor, Professor Pankaj Choudhary, for all his scholastic guidance, generosity with his time, and for offering me his mentorship, fatherly love, and care. Under his guidance, I successfully overcame many difficulties and learnt a lot. His deep knowledge of statistics, enthusiasm for the subject, and patience were invaluable not only in writing this dissertation but also in every step of my graduate career at The University of Texas at Dallas. I could not have imagined having a better advisor for my PhD study.

I am also very grateful to the members of my committee, Professor Swati Biswas, Professor Min Chen, and Professor Sam Efromovich, for their valuable advice and suggestions. I would also like to express my gratitude to the Mathematical Sciences Department at UTD for providing me an opportunity to pursue my PhD by offering me admission and financial support as a Teaching Assistant and Teaching Associate. These opportunities led me to develop myself for my future career. My sincere thanks go to everyone at the department for the warm and friendly atmosphere they provided. Thank you, all my friends, for your consistent support and being there whenever I needed a peer.

My deepest gratitude goes to my parents and family members for all the support and love they have given me throughout my life.

Last but not least, I thank my wife, Aklima, and our daughter, Umaija, for their love and patience. Thank you, Aklima, for your boundless support and for always being there with me in every step of this journey. I would not be here without your love and generosity.

June 2020

ESTIMATION OF COVARIANCE STRUCTURES IN FUNCTIONAL MIXED MODELS WITH APPLICATION TO HERITABILITY ESTIMATION

Mohammad Shaha Alam Patwary, PhD
The University of Texas at Dallas, 2020

Supervising Professor: Pankaj K. Choudhary, Chair

When the response on a subject can be naturally viewed as a smooth curve or function, it is said to be a functional response. The response may be observed at a set of discrete times, possibly with noise. Functional data arise when observations of a functional response are available from a sample of subjects. Thus, the functional data essentially consist of a sample of curves. One example of such data is the usual longitudinal data where a variable of interest is measured over time on a sample of subjects. Functional data arise in a variety of disciplines, including economics, environmental science, public health, medicine, and genetics. Analysis of such data is currently an active area of statistical research.

Functional data are often analyzed by modeling them as a functional mixed model. This model commonly assumes that the within-subject errors are homoscedastic and uncorrelated. But this assumption is often violated in practice, which may sometimes lead to potentially misleading inferences.This is especially an issue if the object of inference is a function of both random effect and error autocovariance functions. One such quantity is heritability function, defined as the proportion of variance explained by the random effect. In genetics, the random effect can be interpreted as the additive genetic component of a quantitative trait. This way, heritability is the ratio of additive genetic variance to the total phenotypic variance of the trait. It measures the extent to which individuals' phenotypes are determined

by the genes transmitted from the parents. This makes heritability a fundamental quantity of interest in genetics.

This dissertation makes three contributions toward the issue of estimating both random effect and error covariance structures in a functional mixed model. First, it develops a methodology for modeling functional data from independent subjects that incorporates parametric models for error covariance structure. The methodology is evaluated using a simulation study. Its application is illustrated by analyzing a growth curve data. Next, this methodology is extended for family data where the subjects may be grouped into families and subjects from the same family are dependent. This methodology is also evaluated using simulation. Finally, it introduces the novel notion of a singular mixed model, whose further development in future may allow modeling the error covariance structure nonparametrically, enhancing the flexibility of functional mixed models.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Motivation and problem formulation

Functional data analysis[1,2] is currently an active area of statistical research. Functional data arise when the response measured on a subject is a smooth curve, which may be observed possibly with noise at a set of discrete time points, rather than a scalar or vector quantity. Thus, a functional response consists of repeated measurements generally indexed by time but the index may also be something else such as a spatial location. Examples of functional response include longitudinal measurements on an individual taken over a period of time such as the growth curve, electrical activity of a heart measured by a wearable device every minute over the course of a day, and gene expression profile of an individual[3]. The functional data consist of one observed curve per subject. The data may be considered dense or sparse depending upon the regularity and closeness of the observation times.

A common approach to functional data analysis involves representing the response curve as a functional mixed model[2,4,5]. To be specific, let $y(t)$, $t \in \mathcal{T}$ be a random function denoting the functional response of a randomly selected subject from the population. The functional mixed model assumes that

$$y(t) = f(t) + g(t) + \epsilon(t),\ t \in \mathcal{T}, \tag{1.1}$$

where $f(t)$ is the (fixed) mean function in the population; $g(t)$ is a random effect function that captures the deviation from the population mean for the subject, with mean zero and autocovariance function $\psi(s,t)$; and $\epsilon(t)$ represents the random error, distributed independently of $g(t)$, with mean zero and autocovariance function $\sigma(s,t) = \sigma^2 I(s=t)$, where $I(\cdot)$ denotes an indicator function. This model postulates that the function $f(t) + g(t)$ is the true response of the subject which is observed with error as $y(t)$. Both f and g functions are

usually modeled through basis expansions[4–8]. Marginally, the response has mean $f(t)$ and autocovariance function $v(s,t) = \psi(s,t) + \sigma^2 I(s = t)$.

The assumption in the above model that the errors are homoscedastic and uncorrelated is rather strong and is often violated in practice. Indeed, examples of longitudinal data, which are a special case of functional data[9,10], abound where it is necessary to allow heteroscedasticity and autocorrelation in errors to adequately capture the true covariance structure of the response[11]. The issue of adequately capturing both the random effect and error covariance structures is especially important if the object of inference is

$$h^2(t) = \frac{\text{var}(g(t))}{\text{var}(y(t))} = \frac{\psi(t,t)}{\psi(t,t) + \sigma(t,t)}, \tag{1.2}$$

a quantity called the *heritability function*, because it involves both the functions. Heritability of a genetic trait is a fundamental quantity in statistical genetics[12,13]. Its concept goes back to the seminal works of Fisher[14] and Wright[15]. It is defined as the ratio of additive genetic variance to the total phenotypic variance of the trait. It measures the extent to which an individual's phenotype is determined by the genes transmitted from parents[12]. Although, in this dissertation, we will not have an opportunity to work with genetic data, we simply take the heritability function to mean the proportion of total variation explained by the random effect, but it does serve as the object of primary interest. Besides, some of the methodology developed in this work can be readily adapted to analyze functional genetic trait data.

The task of modeling both random effect and error covariance functions is complicated by the fact that there is an identifiability issue in the model (1.1). If parametric assumptions are not made for at least one of the two, depending upon the type of data, it may not be possible for us to separately estimate both the functions. Specifically, if the subjects are independent, the random effect is confounded with the error. So the two cannot be separated. On the other hand, if the subjects come from a type of family data, we may be able to separate them in principle, but in practice, we may run into the problem of near-unidentifiability.

Therefore, in this dissertation, only the random effect is modeled nonparametrically and parametric models are considered for the error covariance structure.

This dissertation makes the following three contributions to the literature. First, we develop a methodology for modeling functional data from independent subjects that incorporates parametric models for error covariance structure and considers inference on the heritability function. This is the topic of Chapter 2. Next, we extend this methodology for family data where the subjects may be grouped in families and subjects from the same family are dependent. This is the topic of Chapter 3. Some additional development of this approach in future will allow us to estimate heritability from functional genetic trait data. In Chapter 4, we introduce the novel notion of a singular mixed model and study it briefly. Further development of this model in future may allow the error covariance structured to be modeled nonparametrically, extending the flexibility of functional mixed models. The remainder of this chapter is developed to reviewing mixed models for scalar data and a common approach for construction of confidence intervals and bands, and introducing a motivating dataset that will be analyzed later in the dissertation.

1.2 Classical mixed model

Let $\mathbf{Y}_i$ be the response vector regarding subject i. A classical linear mixed model decompose the outcome as a sum of overall (fixed) mean effects, subject-specific random effects, and measurement error terms. Therefore, for the ith subject, classical formulation of linear mixed model can be presented presented as[11,16]

$$\mathbf{Y}_i := \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i; i = 1, \ldots, n, \tag{1.3}$$

where

$\mathbf{Y}_i \in \mathbb{R}^{M_i \times 1}$ is a vector of response.

$\mathbf{X}_i \in \mathbb{R}^{M_i \times p}$ is a design matrix for fixed effects.

$\boldsymbol{\beta} \in \mathbb{R}^{p\times 1}$ is a vector of fixed effects.

$\mathbf{Z}_i \in \mathbb{R}^{M_i\times q}$ is a design matrix for random effects.

$\mathbf{b}_i \in \mathbb{R}^{q\times 1}$ is a vector of random effects.

$\boldsymbol{\epsilon}_i \in \mathbb{R}^{M_i\times 1}$ is a vector of random errors.

Assumptions regarding (1.3) are as follows:

$\mathbf{b}_i \sim \mathcal{N}_q(\mathbf{0}, \mathbf{G})$,

$\mathbf{Z}_i\mathbf{b}_i \sim \mathcal{N}_{M_i}(\mathbf{0}, \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i')$,

$\boldsymbol{\epsilon}_i \overset{i.d}{\sim} \mathcal{N}_{M_i}(\mathbf{0}, \sigma^2\mathbf{I}_{M_i})$, and

$\mathbf{b}_i$ and $\boldsymbol{\epsilon}_i$ are independent, implying $\text{cov}(\mathbf{b}_i, \boldsymbol{\epsilon}_i) = \mathbf{0},\ i = 1, \ldots, n.$

Therefore, $\mathbf{Y}_i \sim \mathcal{N}_{M_i}(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i' + \sigma^2\mathbf{I}_{M_i})$

Thus, for all n subjects, the complete data model can be written in matrix form as[11,16,17]

$$\mathbf{Y} := \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}, \qquad \mathbf{b} \sim \mathcal{N}_{nq}(\mathbf{0}, \mathbf{I}_n \otimes \mathbf{G}),\ \boldsymbol{\epsilon} \sim \mathcal{N}_N(\mathbf{0}, \sigma^2\mathbf{I}). \tag{1.4}$$

where

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_n \end{bmatrix} \in \mathbb{R}^{N\times 1}, \quad N := \textstyle\sum_{i=1}^n M_i\ , \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{bmatrix} \in \mathbb{R}^{N\times p}\ , \quad \boldsymbol{\beta} \in \mathbb{R}^{p\times 1},$$

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \ldots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{Z}_n \end{bmatrix} = \oplus_{i=1}^n \mathbf{Z}_i = \text{diag}\{\mathbf{Z}_1, \ldots, \mathbf{Z}_n\} \in \mathbb{R}^{N\times nq},$$

$$\mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{bmatrix} \in \mathbb{R}^{nq\times 1},\ \boldsymbol{\epsilon} = \begin{bmatrix} \boldsymbol{\epsilon_1} \\ \boldsymbol{\epsilon_2} \\ \vdots \\ \boldsymbol{\epsilon_n} \end{bmatrix} \in \mathbb{R}^{N\times 1},$$

$\mathbf{X}$ denotes the design matrix for mean function with associated parameter vector $\boldsymbol{\beta}$, and

$\mathbf{Z}$ is the design matrix for the random effects with associated parameter vector $\mathbf{b}$.

However, in modeling data through the model (1.3), it assumes that within-subject errors are homoscedastic and uncorrelated, that is $\boldsymbol{\epsilon}_i \sim \mathcal{N}_{M_i}(\mathbf{0}, \sigma^2\mathbf{I})$. However, it is often not plausible to assume that within-subject errors are homoscedastic and uncorrelated. Thus, with the assumption of heteroscedastic and correlated within-subject errors, we can re-write the classical mixed model presented in (1.3) and (1.4) as follows[11,16]:
For the ith subject, :

$$\mathbf{Y}_i := \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i; i = 1, \ldots, n, \tag{1.5}$$

where, $\boldsymbol{\epsilon}_i \sim \mathcal{N}_{M_i}(\mathbf{0}, \sigma^2\mathbf{R}_i)$, $\mathbf{R}_i \in \mathbb{R}^{M_i \times M_i}$.
Therefore, $\mathbf{Y}_i \sim \mathcal{N}_{M_i}(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i' + \sigma^2\mathbf{R}_i)$

Thus, all n, subjects the model can be written as[11,16,17]:

$$\mathbf{Y} := \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \epsilon, \qquad \mathbf{b} \sim \mathcal{N}_{nq}(\mathbf{0}, \mathbf{I}_n \otimes \mathbf{G}),\ \boldsymbol{\epsilon} \sim \mathcal{N}_N(\mathbf{0}, \sigma^2\mathbf{R}). \tag{1.6}$$

where $\mathbf{R} = \oplus_{i=1}^{n}\mathbf{R}_i = \text{diag}\{\mathbf{R}_1, \ldots, \mathbf{R}_n\} \in \mathbb{R}^{N \times N}$ Therefore, $\mathbf{Y} \sim \mathcal{N}_N(\mathbf{X}\boldsymbol{\beta}, \mathbf{Z}(\mathbf{I}_n \otimes \mathbf{G})\mathbf{Z}' + \sigma^2\mathbf{R})$
Note that, notations and all other assumptions for (1.5) and (1.6) remain same as we presented in the models (1.3) and (1.4), respectively.
In the following subsections of this section, we briefly present different structures and modeling capacities related to the matrices $\mathbf{G}$ and $\mathbf{R}$.

1.2.1 Covariance structures of random effects parameters

Based on practical situation, goal of research, and nature of data, one can specify the covariance matrix $\mathbf{G}$ of random effects parameters from quite a few choices of structures such as, block-diagonal matrix, compound symmetry structure matrix, diagonal matrix, multiple of

an identity matrix, and general positive-definite matrix in model fitting. A nice description of such structures can be found in[11]. Among all these choices, general positive-definite matrix is the most complex one. Therefore, both the model defined in (1.4) and (1.6), allows considerable flexibility in specifying the random effects covariance structure.

1.2.2 Decomposition of within-subject error covariance structure

Mixed model (1.4) restricts within-subject errors to be identically and independently distributed random variables with mean vector $\mathbf{0}$ and covariance matrix $\sigma^2\mathbf{I}$. In contrast, the model defined in (1.6) allows considerable flexibility in specifying within-subject errors covariance structure parametrically by generalizing the identical and independent assumption. Nevertheless, there are many applications involving grouped data which allows the within-subject errors are homoscedastics and correlated ($\mathbf{R}_i = \sigma^2\boldsymbol{\rho}(s,t)$), heteroscedastic and uncorrelated ($\mathbf{R}_i = \sigma^2(t)\mathbf{I}$), and heteroscedastic and correlated ($\mathbf{R}_i = \sigma^2(t)\boldsymbol{\rho}(s,t)$). $\boldsymbol{\rho}(s,t)$ represents within-subject errors correlation structure.

Moreover, the within-subject covariance matrix $\mathbf{R}_i$ can always be decomposed into a product of two relatively simpler matrices[18] as

$$\mathbf{R}_i = \mathbf{D}_i\mathbf{C}_i\mathbf{D}_i$$

where $\mathbf{D}_i$ is a within-subject dispersion matrix and $\mathbf{C}_i$ is a within-subject correlation matrix. This decomposition of $\mathbf{R}_i$ into a variance structure component $\mathbf{D}_i$ and a correlation structure $\mathbf{C}_i$ allows us flexibility and convenience in both theoretical development and to avail computational facilities. We present comprehensive summary of variance function structures for $\mathbf{D}_i$ in Section 1.2.3 and correlation structures for $\mathbf{C}_i$ in Section 1.2.4.

1.2.3 Modeling heteroscedastic variance of within-subject errors

If all the diagonal elements of $\mathbf{D}_i$ is equal to 1 then $Var(\epsilon_{ij}) = \sigma^2$ which reduces the case to a very simpler case of homoscedastic within-subject error variance as we presented in the model (1.4). But, in reality, if all the diagonal elements of $\mathbf{D}_i$ is not equal to 1 then the situation of heteroscedastic error variance arises. To address this issue, variance functions are used to model the heteroscedastic variance structure of within-subject errors $\boldsymbol{\epsilon}_i$ using covariates. For this kind of modeling, a more detailed outline can be found in the context of extended linear models[19] and in the context of mixed model given in (1.6)[11,20]. However, a more comprehensive detail and implementation procedures are given in[11]. One may choose one of the variance function model from the list of fixed variance, different variances per stratum, power of covariate, exponential of covariate, constant plus power of covariate, and combination of variance functions[11,21] depending of modeling facility and point of research interest.

1.2.4 Modeling dependence among within-subject errors

In this section we present a summary of different correlation structures to represent the correlation matrix $\mathbf{C}_i$. In the context of mixed model, correlation structures are used for modeling dependence among within-subject errors. Moreover, in statistical and economic literature, correlation structures are mainly developed and used for two broad classes of data, such as, time series data and spatial data[11]. Historically, time series data are associated with observations recorded or reported at an integer-valued time variable. In contrast, the spatial data mainly refers to the observations recorded according to a two-dimensional spatial coordinate vector on a plane.

Time series correlation structures are often known as serial correlation structures. A detail study on time series correlation structures for linear models without random effects can be found in[22]. However, in the context of linear mixed models, detail description of time series

correlation structures can be found in[23]. Time series correlation structures typically assumes that the data are observed at discrete and integer time points and it not easy to generalize to continuous time points. This is the main limitation of these structures. A comprehensive study of most commonly used time series correlation structures[11] are outlined below :

Compound symmetry correlation structure:

This correlation structure is simplest among all correlation structures found in literature. It assumes equal correlation among all within-subject errors in the same group and the corresponding correlation model is give by

$$h(d, \rho) = \rho, \qquad d = 1, 2, \ldots,$$

where the correlation parameter ρ is commonly known as intraclass correlation coefficient. However, for overlay simplistic nature of compound symmetry correlation model, it has very limited practical use in the application of time series data.

General correlation structure:

In general correlation structure, each lag-1 correlation in within-subject error is represented by a different parameter according to the following correlation function

$$h(d, \boldsymbol{\rho}) = \rho_d, \qquad d = 1, 2, \ldots.$$

Thus, the number of parameter increases quadratically with the number of observations within each group.

Autoregressive moving average correlation structure:

This family of correlation structures is a mixture of two linear stationary models: autoregressive (AR) models and moving average (MA) models, commonly known as ARMA models[22].

The distance (or lag) between two conservators at time t and s is given by $|t-s|$. Thus, before summarizing ARMA models, we present summary of AR and MA models below:

AR correlation structure:

These models express the relationship between current observation and previous observation, such that the current observation in within-subject errors as a liner function of past observations along with an independent white noise term, u_t. The order of an AR model is denoted by p is defined as the number of past observations included in the model. Therefore, an AR(p) is given by the following form:

$$\epsilon_t = \sum_{i=1}^{p} \phi_i \epsilon_{t-i} + u_t \tag{1.7}$$

When the order of the model is greater than 1 ($p > 1$), the form of the correlation function is quite involved but can be obtained recursively through the following equation[22]

$$h(d, \boldsymbol{\phi}) = \sum_{i=1}^{p} \phi_i h(|d-i|, \boldsymbol{\phi}), \quad d = 1, 2, \ldots \tag{1.8}$$

However, the simplest and most useful model in AR family is AR of order 1, called AR(1). The variance of above AR(1) model is given by

$$\gamma(0) = \text{var}(\epsilon_t) = \frac{\sigma^2}{1-\phi^2}$$

Correlation function of AR(1) model declines exponentially with the absolute value of lag d and is given by

$$h(d, \phi) = \phi^d, \quad d = 0, 1, 2, \ldots, \tag{1.9}$$

where the parameter ϕ represents lag-1 correlation and $-1 \leq \phi \leq 1$.

Continuous time autoregressive correlation structure:

Due to the simplicity in its form, AR(1) model can be easily generalized to continuous time

measurements. Thus, the correlation function of continuous time AR(1) or CAR(1) is given as

$$h(s, \phi) = \phi^s, \quad s \geq 0, \quad \phi \geq 0. \tag{1.10}$$

Unlike the correlation function of AR(1), the correlation parameter in the correlation function of CAR(1) must be non-negative. In contrast to AR(1) model, CAR(1) model allows unevenly spaced continuous time points. This is a remarkable flexibility of CAR(1).

MA correlation Structure:

MA correlation models represent the current observation of within-subject errors as a linear function of independent and identically distributed white noise terms, u_t. The order q of MA is defined as the number of white noise terms added in the model. The form of MA models of order q, referred as MA(q) is given by

$$\epsilon_t = u_t + \sum_{j=1}^{q} \theta_j u_{t-j} \tag{1.11}$$

Therefore, the correlation function of MA(q) given in 1.11 is

$$h(d, \boldsymbol{\theta}) = \begin{cases} \dfrac{\theta_d + \theta_1\theta_{d-1} + \ldots + \theta_{d-q}\theta_q}{1 + \theta_1^2 + \ldots + \theta_q^2}, & d = 1, \ldots, q \\ 0, & d = q+1, q+2, \ldots \end{cases} \tag{1.12}$$

It is clearly observed from 1.12 that the observations more than q time points apart are indeed uncorrelated, because they do not share any common white noise terms u_t.

The variance of MA(1) model is obtained as

$$\gamma(0) = \text{var}(\epsilon_t) = \sigma^2(1 + \theta^2)$$

Clearly, the correlation function of MA(1) model is given by

$$h(d, \theta) = \begin{cases} \dfrac{\theta}{1 + \theta^2}, & d = 1 \\ 0, & d = 2, 3, \ldots \end{cases} \tag{1.13}$$

Surprisingly, MA(1) has only lag-1 correlation and correlations are zero for lag-2 and onward resulting sudden decrease in correlation function.

Finally, by combining an AR(p) and MA(q) together we get autoregressive moving average model of order (p, q), called an ARMA(p, q) model. The model contains p autoregressive parameters $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$ and q moving average parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)$ corresponding to MA(q). Therefore, the model contains total of $p + q$ parameters. The functional form of an ARMA(p, q) is given as

$$\epsilon_t = \sum_{i=1}^{p} \phi_i \epsilon_{t-i} + u_t + \sum_{j=1}^{q} \theta_j u_{t-j} \tag{1.14}$$

Conventionally, $ARMA(p, 0) \equiv AR(p)$ and $ARMA(0, q) \equiv MA(q)$. Hence, both AR models and MA models are special cases of ARMA models.

By using the recursive relations, we obtained the correlation function of ARMA(p, q) as

$$h(d, \boldsymbol{\rho}) = \begin{cases} \phi_1 h(|d-1|, \boldsymbol{\rho}) + \ldots + \phi_p h(|d-p|, \boldsymbol{\rho}) + \\ \qquad \theta_1 \psi(d-1, \boldsymbol{\rho}) + \ldots + \theta_q \psi(d-q, \boldsymbol{\rho}), & d = 1, \ldots, q \\ \phi_1 h(|d-1|, \boldsymbol{\rho}) + \ldots + \phi_p h(|d-p|, \boldsymbol{\rho}), & d = q+1, q+2, \ldots, \end{cases} \tag{1.15}$$

where $\psi(k, \boldsymbol{\phi}, \boldsymbol{\theta}) = \mathbb{E}[\epsilon_{t-k} u_t]/Var(\epsilon_t)$.

For lags between 1 and q, the correlation function of ARMA(p, q) model seems like an AR(p) correlation function plus a term related to the MA part of the model. For, lags greater than q, the correlation function of ARMA(p, q) seems like the correlation function of an AR(p).

The variance of ARMA(1,1) is given by

$$\gamma(0) = \text{var}(\epsilon_t) = \frac{1 + 2\phi\theta + \theta^2}{1 - \phi^2}\sigma^2$$

However, the correlation function for ARMA(1,1) model with $p = 1$ and $q = 1$ is obtained as

$$h(d, \phi, \theta) = \begin{cases} \dfrac{(\phi + \theta)(1 + \phi\theta)}{1 + 2\phi\theta + \theta^2}, & d = 1 \\ \phi h(d-1, \phi, \theta), & d = 2, 3, \ldots \end{cases} \tag{1.16}$$

To this end, we present a very concise summary of different spatial correlation structures to model within-subject errors correlation matrix $\mathbf{C}_i$. Spatial correlation structures were originally proposed to model dependence in data indexed by continuous two-dimensional position vectors, such as geostatistical data, lattice data, and point patterns[11]. But, in practice, they can be used with discrete time data, such as time series data[11,18,21]. Detail studies on spatial correlation structures in the context of linear models without random effects[24] and mixed models[11,25] can be found in statistical literatures.

Thus, to model within-subject errors correlation structure by using spatial correlation structures, some popular choices of variogram models are presented below[11]:

Exponential model:

$$\gamma(s, \rho) = 1 - exp(-s/\rho)$$

Gaussian model:

$$\gamma(s, \rho) = 1 - exp\left[-(s/\rho)^2\right]$$

Linear model:

$$\gamma(s, \rho) = 1 - (1 - s/\rho) I(s < \rho)$$

Rational quadratic model:

$$\gamma(s, \rho) = (s/\rho)^2 / \left[1 + (s/\rho)^2\right]$$

Spherical model:

$$\gamma(s, \rho) = 1 - \left[1 - 1.5(s/\rho) + 0.5(s/\rho)^3\right] I(s < \rho)$$

We can deduce the correlation function $h(s, \rho)$ by using a simple relation $h(s, \rho) = 1 - \gamma(s, \rho)$.

1.3 Pointwise and simultaneous confidence intervals

Suppose we are interested in $\psi \equiv \psi(\boldsymbol{\theta})$, a function of model parameters $\boldsymbol{\theta}$. $\boldsymbol{\theta}$ may be a scalar quantity but in many applications it is a function of time. Example of scalar ψ includes but not limited to σ^2. The mean function, random effects variance function, and heritability function are the examples of time dependent parametric functions, denoted by $\psi(t)$. We focus on constructing two-sided confidence bands for ψ and $\psi(t)$.

An approximate $100(1-\alpha)\%$ two-sided pointwise confidence band for $\psi(t)$ can be computed[26] as

$$\text{two-sided band: } \hat{\psi}(t_l) - b_l \pm z_{1-\alpha/2}\sqrt{s_{ll}}, \quad l = 1, \ldots, M, \tag{1.17}$$

where $\hat{\psi}(t_l)$ is the estimate of the parameter at the time point l, $l = 1, \ldots, M$, b_l is the lth element of bias vector $\mathbf{b} = (b_1, \ldots, b_M)^T$, s_{ll} is the lth diagonal element of the covariance matrix $\mathbf{S}$ of the estimators, and z_α is the 100αth percentile of a $\mathcal{N}_1(0, 1)$ distribution.

If we replace z_α in (1.17) by an appropriate percentile detailed in Chapter 3 of[27], we can get a 100(1-α)% simultaneous confidence interval as given below:

$$\text{two sided band: } \hat{\psi}(t_l) - b_l \pm c_{1-\alpha}\sqrt{s_{ll}}, \quad l = 1, \ldots, M, \tag{1.18}$$

where, the notations $\hat{\psi}(t_l)$, b_l, and S_{ll} retain same meaning from (1.17). Additionally, $c_{1-\alpha} =$ $100(1-\alpha)$th percentile of $|Z|_{max}$, $|Z|_{max} = max_{l=1,\ldots,M}|Z_l|$, $Z_l = (\hat{\psi}_l - \psi_l)/SE(\hat{\psi}_l) \overset{approx}{\sim}$ $\mathcal{N}_1(0, 1)$ with $P\left(|Z|_{max} \leq c_{1-\alpha}\right) \approx 1 - \alpha$.

This computation of 100(1-α)% simultaneous intervals can be adapted using `multcomp` package of[28] in `R` or via Monte Carlo simulation[26].

We implement Monte Carlo simulation method along with Bootstrap methodology to compute coverage probabilities of 100(1-α)% confidence intervals and bounds for different parameters and parameter functions of interest. To compute the bias vector $\mathbf{b}$ and covariance

matrix $\mathbf{S}$ of parameter estimates, we use the bootstrap methodology presented in[26]. A concise summary of the methodology is presented below:

1. From n subjects with indices $1, \ldots, n$, we sample n indices of subjects with replacement and take all the observations associated with the sampled subjects indices. The new data would be a resample of original data.

2. We apply the computation methods to estimate $\boldsymbol{\theta}$ from the bootstrap data to get $\hat{\boldsymbol{\theta}}^*$.

3. By using $\hat{\boldsymbol{\theta}}^*$ from step 2, we estimate $\boldsymbol{\psi}(\mathbf{t})$ as $\hat{\boldsymbol{\psi}}^*(\mathbf{t})$. Thus, we can consider $\hat{\boldsymbol{\psi}}^*(\mathbf{t})$ as a resample of $\hat{\boldsymbol{\psi}}(\mathbf{t})$.

4. We then repeat the steps (1)–(3) Q times to get the bootstrap resamples $\hat{\boldsymbol{\psi}}_q^*(\mathbf{t}),\ q = 1, \ldots, Q$ of $\hat{\boldsymbol{\psi}}(\mathbf{t})$.

5. Now, we compute the bias vector $\mathbf{b}$ using $\sum_{q=1}^{Q} \hat{\boldsymbol{\psi}}_q^*(\mathbf{t})/Q - \hat{\boldsymbol{\psi}}(\mathbf{t})$, and the covariance matrix $\mathbf{S}$ of the resamples $\hat{\boldsymbol{\psi}}_q^*(\mathbf{t})$ of $\hat{\boldsymbol{\psi}}(\mathbf{t})$,$q = 1, \ldots, Q$.

In many cases, a parameter function ψ does not depend on time t, that is, ψ is a scalar quantity like σ^2. In that case, the whole procedure can easily be adapted for computing confidence intervals for such types of parameters. Indeed, it is a very special case of the bootstrap methodology outlined in this section with bias vector $\mathbf{b}$ of length 1 and covariance matrix $\mathbf{S}$ of order 1×1. However, the term b_l in (1.17) and (1.18) can be omitted if there is an evidence that the bias correction is not necessary.

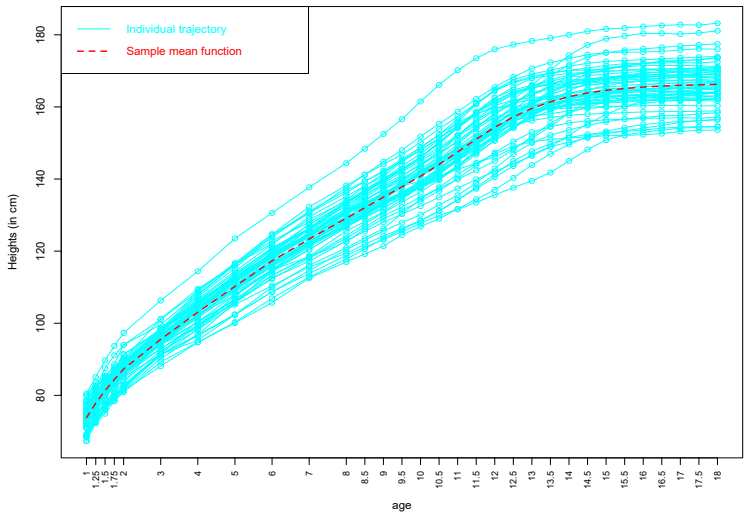

Figure 1.1. Individual trajectory plot and sample mean of growth data.

1.4 Dataset

For illustration of our methodologies devoted in subsequent chapters, we consider a dataset, called Berkeley Growth Study[29–31] data. This dataset consists of heights measured in centimeter (cm) of 39 boys and 54 girls from California at 31 different ages starting from age 1 year up to age 18 years. The ages are not equally spaced; four measurements were take from age 1 to 2 years; annual measurements from age 2 years onward until 8 years, biannual measurements from age 8 years to 18 years. Since girls' list of data has higher number subjects, we choose to use only girls data throughout this dissertation. Figure 1.1 shows individual trajectories of heights of 54 California girls measured at 31 different ages and additionally a superimposed sample mean function. It is clear that, heights are increasing function of ages up to a certain age. From exploratory analysis, we found that the data is balanced and dense.

CHAPTER 2

A FUNCTIONAL MIXED MODEL FOR DATA FROM INDEPENDENT SUBJECTS

2.1 Introduction

In this chapter, we study a functional mixed model for data from independent subjects that allows the within-subject errors to be heteroscedastic and correlated. The chapter is organized as follows. Section 2.2 describes the model and discusses its fitting. Section 2.3 presents results of a simulation study to evaluate performance of the methodology. An illustration of the methodology is presented in Section 2.4. Section 2.5 concludes with a summary and mentions some ongoing and future work.

2.2 Modeling data

Suppose the observed data consist of functional responses of n subjects, indexed as $i = 1, \ldots, n$, observed at discrete times $t_{ij} \in \mathcal{T}$, $j = 1, \ldots, M_i$, $i = 1, \ldots, n$. Here the domain $\mathcal{T} = [a,\ b]$, $a < b \in \mathbb{R}$ is bounded. Specifically, let $y_i(t_{ij})$ denote the response of subject i at time t_{ij}. Thus, subject i contributes M_i observations. The actual observation times and their numbers need not be the same for all subjects, but the set of all observation times in the data is assumed to be dense in $\mathcal{T}$. Let $N = \sum_{i=1}^{n} M_i$ be the total number of observations in the data. We assume that the subjects involved in the study are independent in that their responses are statistically independent.

2.2.1 Population model

To get a model for the data, let us first describe a model for the population from which the observed response curves are drawn. Let $y(t)$, $t \in \mathcal{T}$ be a random function denoting the

functional response of a randomly selected subject from the population. The model for $y(t)$ is given by the functional mixed model (1.1), which is written as

$$y(t) = f(t) + g(t) + \epsilon(t),\ t \in \mathcal{T}, \tag{2.1}$$

where $f(t)$ is the (fixed) population mean function; $g(t)$ is a random effect function that captures the deviation from the population mean for the subject, with mean zero and covariance function $\psi(s,t)$; and $\epsilon(t)$ represents the random error, distributed independently of $g(t)$, with mean zero and covariance function $\sigma(s,t)$, and the errors are allowed to be heteroscedastic and correlated. From (2.1), it follows that

$$\text{cov}(y(s), y(t)) = \psi(s,t) + \sigma(s,t), \tag{2.2}$$

and the heritability function is

$$h^2(t) = \frac{\text{var}(g(t))}{\text{var}(y(t))} = \frac{\psi(t,t)}{\psi(t,t) + \sigma(t,t)}. \tag{2.3}$$

2.2.2 Data model

The population model (2.1) implies the following model for the observed data:

$$y_i(t_{ij}) = f(t_{ij}) + g_i(t_{ij}) + \epsilon_i(t_{ij}),\ j = 1, \ldots, M_i,\ i = 1, \ldots, n, \tag{2.4}$$

where $f(t)$ is the population mean function, $g_i(t)$ is the random effect function for subject i, and $\epsilon_i(t)$ is the random error function, distributed independently of $g_i(t)$. Here the $g_i(t)$ are independent realizations of the function $g(t)$ defined in (2.1) which are observed at the sites t_{ij}. Thus, $g_i(t)$ has mean zero and its covariance function is $\psi(s,t)$. Further, the $\epsilon_i(t)$ are independent realizations of the error function $\epsilon(t)$ from (2.1) which are observed at the sites t_{ij}. Therefore, the errors have mean zero. Moreover, the errors associated with the same subject have covariance function $\sigma(s,t)$, known as within-subject covariance function, whereas those associated with different subjects are independent.

2.2.3 Basis expansions for fixed and random effect functions

The functions f and g in the population model (2.1) are represented as linear combinations of specified basis functions. The basis systems for two functions need not be the same. Let p and q respectively denote the number of basis functions used to represent f and g. Specifically,

$$f(t) = \sum_{k=1}^{p} \beta_k \phi_{fk}(t) = \mathbf{\Phi}_f^T(t)\boldsymbol{\beta}, \quad g(t) = \sum_{k=1}^{q} b_k \phi_k(t) = \mathbf{\Phi}^T(t)\mathbf{b}, \tag{2.5}$$

where $\mathbf{\Phi}_f(t) = (\phi_{f1}(t), \ldots, \phi_{fp}(t))^T$ and $\mathbf{\Phi}(t) = (\phi_1(t), \ldots, \phi_q(t))^T$ are $p \times 1$ and $q \times 1$ vectors of basis functions with $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ and $\mathbf{b} = (b_1, \ldots, b_q)^T$ as $p \times 1$ and $q \times 1$ vectors of their respective coefficients. The coefficient vector $\boldsymbol{\beta}$ for f is a fixed, unknown quantity, whereas the coefficient vector $\mathbf{b}$ for g is a random quantity. It assumed that $\mathbf{b} \sim (\mathbf{0}, \mathbf{G})$. The matrix $\mathbf{G}$ is an arbitrary, unknown positive-definite matrix. The resulting random effect covariance function is

$$\psi(s,t) = \mathbf{\Phi}^T(s)\mathbf{G}\mathbf{\Phi}(t). \tag{2.6}$$

From the expansion (2.5), we have

$$g_i(t) = \sum_{k=1}^{q} b_{ki} \phi_k(t) = \mathbf{\Phi}^T(t)\mathbf{b}_i, \tag{2.7}$$

where the $q \times 1$ vector $\mathbf{b}_i = (b_{1i}, \ldots, b_{qi})^T$ of coefficients for subject i is an independent draw from the distribution of $\mathbf{b}$. Upon replacing f and g_i functions in (2.4) with their basis expansions from (2.5) and (2.7), the model becomes

$$y_i(t_{ij}) = \mathbf{\Phi}_f^T(t_{ij})\boldsymbol{\beta} + \mathbf{\Phi}^T(t_{ij})\mathbf{b}_i + \epsilon_i(t_{ij}), \quad j = 1, \ldots, M_i, \quad i = 1, \ldots, n \tag{2.8}$$

2.2.4 Representation as a classical mixed model

Let $\mathbf{t}_i = (t_{i1}, \ldots, t_{iM_i})^T$ be the $M_i \times 1$ vector of observation locations for the ith subject. Define the following quantities:

$$\mathbf{Y}_i = \mathbf{Y}_i(\mathbf{t}_i) = \begin{pmatrix} y_i(t_{i1}) \\ \vdots \\ y_i(t_{iM_i}) \end{pmatrix}, \ \mathbf{X}_i = \mathbf{X}_i(\mathbf{t}_i) = \begin{pmatrix} \boldsymbol{\Phi}_f^T(t_{i1}) \\ \vdots \\ \boldsymbol{\Phi}_f^T(t_{iM_i}) \end{pmatrix}, \ \boldsymbol{\epsilon}_i = \boldsymbol{\epsilon}_i(\mathbf{t}_i) = \begin{pmatrix} \epsilon_i(t_{i1}) \\ \vdots \\ \epsilon_i(t_{iM_i}) \end{pmatrix}. \quad (2.9)$$

Here $\mathbf{Y}_i$ is the vector of observations on subject i; $\mathbf{X}_i$ is the fixed effect design matrix; and $\boldsymbol{\epsilon}_i$ is the vector of random errors. Next, define

$$\mathbf{b}_i = \begin{pmatrix} b_{1i} \\ \vdots \\ b_{qi} \end{pmatrix}, \ \mathbf{Z}_i = \mathbf{Z}_i(\mathbf{t}_i) = \begin{pmatrix} \boldsymbol{\Phi}^T(t_{i1}) \\ \vdots \\ \boldsymbol{\Phi}^T(t_{iM_i}) \end{pmatrix}.$$

Here $\mathbf{b}_i$ is the $q \times 1$ vector of random coefficients and $\mathbf{Z}_i(\mathbf{t}_i)$ is the associated $M_i \times q$ design matrix. Further, define $\mathbf{R}_i = \mathbf{R}_i(\mathbf{t}_i) = \text{var}(\boldsymbol{\epsilon}_i) = (\sigma(t_{ij}, t_{iq}))$ as the $M_i \times M_i$ covariance matrix of the within-subject errors. Next, let the notation $\mathbf{x} \sim (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ mean that the random vector $\mathbf{x}$ has mean $\boldsymbol{\mu}$ and variance matrix $\boldsymbol{\Sigma}$. We can now write the functional mixed model (2.8) for data from subject i in the form of a classical mixed model from Section 1.5 as

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \ \mathbf{b}_i \sim (\mathbf{0}, \mathbf{G}), \ \boldsymbol{\epsilon}_i \sim (\mathbf{0}, \mathbf{R}_i), \quad i = 1, \ldots, n, \quad (2.10)$$

with $\text{cov}(\mathbf{b}_i, \mathbf{b}_l) = \mathbf{0}$ and $\text{cov}(\mathbf{b}_i, \boldsymbol{\epsilon}_l) = \mathbf{0}$ for all i and l, and $\text{cov}(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_l) = \mathbf{0}$ for $i \neq l$.

Next, we would like to write the functional mixed model for all the data together in the classical mixed model form. For this, define $N \times 1$ vectors $\mathbf{Y}$ and $\boldsymbol{\epsilon}$, $\mathbf{N} \times p$ matrix $\mathbf{X}$, and a $nq \times 1$ vector $\mathbf{b}$ as

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_n \end{pmatrix}, \ \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_n \end{pmatrix}, \ \boldsymbol{\epsilon} = \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \vdots \\ \boldsymbol{\epsilon}_n \end{pmatrix}, \ \mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_n \end{pmatrix},$$

and a $N \times nq$ matrix $\mathbf{Z}$ and a $N \times N$ matrix $\mathbf{R}$ as block diagonal matrices

$$\mathbf{Z} = \text{diag}\{\mathbf{Z}_1, \ldots, \mathbf{Z}_n\}, \quad \mathbf{R} = \text{diag}\{\mathbf{R}_1, \ldots, \boldsymbol{R}_n\}.$$

We can now write the model (2.8) for all the data together as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}, \quad \mathbf{b} \sim (\mathbf{0}, \mathbf{I} \otimes \mathbf{G}), \ \boldsymbol{\epsilon} \sim (\mathbf{0}, \mathbf{R}), \ \text{cov}(\mathbf{b}, \boldsymbol{\epsilon}) = \mathbf{0}, \tag{2.11}$$

where the symbol '$\otimes$' denotes a Kronecker product. All the quantities in (2.11) except $\mathbf{b}$ and its covariance matrix $\mathbf{G}$ depend on the observation times in the data as well. But this dependence is suppressed for notational simplicity. The model (2.11) is in the form of a classical mixed model. It implies

$$E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}, \quad \text{var}(\mathbf{Y}) = \mathbf{Z}(\mathbf{I} \otimes \mathbf{G})\mathbf{Z}^T + \mathbf{R}. \tag{2.12}$$

In practice, the number of basis functions q should be moderately large to adequately capture the true underlying random effect covariance function $\psi(s,t)$ via (2.6). However, this may make the estimation of the covariance matrix $\mathbf{G}$—an unstructured positive-definite matrix—a computationally challenging task. As in Wang[32], this issue can be addressed by applying a transformation due to Chen and Dunson[33] based on modified Cholesky decomposition. The matrix $\mathbf{G}$ is decomposed as $\mathbf{G} = \boldsymbol{\Omega}\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T\boldsymbol{\Omega}$, where $\boldsymbol{\Omega} = \text{diag}\{\omega_1, \ldots, \omega_q\}$ is a diagonal matrix and $\boldsymbol{\Gamma} = (\gamma_{km})$ is a unit lower triangular matrix with ones on the diagonal. Here, $\omega_k > 0$, and $\gamma_{kk} = 1$ for $k = 1, \ldots, q$; $\gamma_{km} = 0$ for $m = k+1, \ldots, q$; and the remaining elements of $\boldsymbol{\Gamma}$ are unconstrained. Thus, the diagonal elements of $\boldsymbol{\Omega}$ are proportional to the standard deviations of the basis coefficients and the unconstrained elements of $\boldsymbol{\Gamma}$ are related to the correlations among the coefficients. These correlations do not involve $\boldsymbol{\omega}$. With this decomposition, we can write $\mathbf{b}_i = \boldsymbol{\Omega}\boldsymbol{\Gamma}\mathbf{u}_i$, allowing us to reparameterize the model (2.10) as

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\Omega}\boldsymbol{\Gamma}\mathbf{u}_i + \boldsymbol{\epsilon}_i, \ \mathbf{u}_i \sim (\mathbf{0}, \mathbf{I}_q), \ \boldsymbol{\epsilon}_i \sim (\mathbf{0}, \mathbf{R}_i), \ i = 1, \ldots, n, \tag{2.13}$$

with $\text{cov}(\mathbf{u}_i, \mathbf{u}_l) = \mathbf{I}_q$. In this formulation, the elements of covariance matrix $\mathbf{G}$ are essentially converted into "mean" related parameters, making the task of model fitting easier. We can proceed as before to write this model also in the form of one joint classical mixed model for all the data together as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}, \quad \mathbf{u} \sim (\mathbf{0}, \mathbf{I}_n \otimes \mathbf{I}_q), \ \boldsymbol{\epsilon} \sim (\mathbf{0}, \mathbf{R}), \ \text{cov}(\mathbf{u}, \boldsymbol{\epsilon}) = \mathbf{0}, \tag{2.14}$$

where the notation is same as in (2.11) with the exception that $\mathbf{Z} = \text{diag}\{\mathbf{Z}_1\boldsymbol{\Omega}\boldsymbol{\Gamma}, \ldots, \mathbf{Z}_n\boldsymbol{\Omega}\boldsymbol{\Gamma}\}$, which now depends on unknown model parameters.

2.2.5 Models for error covariance function

The heritability function, given by (2.3), depends on the random effects covariance function $\psi(s,t)$ and the error covariance function $\sigma(s,t)$. Given that estimation of heritability is the primary objective of proposed modeling effort, it is imperative that both the covariance structures are modeled adequately. For specifying a structure for the covariance matrix $\mathbf{G}$ of the basis coefficients, one has quite a few choices, including a diagonal matrix, block-diagonal matrix, compound symmetric matrix, multiple of an identity matrix, and an unstructured general positive-definite matrix. Among all these choices, the last one provides maximum flexibility in capturing the true structure of $\psi(s,t)$. This is why we leave the matrix $\mathbf{G}$ unstructured. By the same token, we also need flexibility in capturing the true structure of $\sigma(s,t)$. Although, once $\psi(s,t)$ is adequately modeled, we may expect $\sigma(s,t)$ to be sparse, but imposing structure such as homoscedasticity and uncorrelated errors or even stationarity of the covariance function may be restrictive. On the other hand, we can not leave $\sigma(s,t)$ completely unspecified, otherwise we may run into the issue of identifiability of two covariance functions. Therefore, in practice, we propose exploring a variety of parametric models for the error covariance function and let the data point to the appropriate models.

Here we specifically mention three models for $\sigma(s,t)$ or equivalently for the $M_i \times M_i$ error covariance matrix $\mathbf{R}_i = \text{var}(\boldsymbol{\epsilon}_i) = (\sigma(t_{ij}, t_{im}))$ in (2.10). The first is $\sigma(s,t) = \sigma^2 I(s = t)$,

implying homoscedastic and uncorrelated within-subject errors, which, as mentioned in Section 1.2, is also a standard assumption in functional mixed models. The second is $\sigma(s,t) = \sigma^2(t)I(s=t)$, heteroscedastic and uncorrelated within-subject errors. For the variance function $\sigma^2(t)$, both parametric models such as polynomial, exponential, and power functions of t; and a nonparametric model specified using the basis expansion, $\log\{\sigma^2(t)\} = \sum_{k=1}^{r} \nu_k \phi_k(t)$ as in (2.5) can be incorporated. The third is $\sigma(s,t) = \sigma^2(t)\rho(s,t)$, where the variance function $\sigma^2(t)$ is modeled as above and the correlation function $\rho(s,t)$ is modeled using either stationary serial correlation models from the time series analysis such as autoregressive (AR), moving average (MA), and autoregressive moving average (ARMA) models; or isotropic variogram models for spatial correlation from spatial statistics such as exponential, Gaussian, Linear, and spherical models[11,23,34,35] (see Section 1.2.4). Some have also been considered for functional mixed models[32,36].

2.2.6 Model fitting and inference

For model fitting, we additionally make the normality assumption for basis coefficients and errors in (2.8) or equivalently (2.10) so that $\mathbf{b}_i \sim \mathcal{N}_q(\mathbf{0}, \mathbf{G})$ and $\boldsymbol{\epsilon}_i \sim \mathcal{N}_{M_i}(\mathbf{0}, \mathbf{R}_i)$, and use the likelihood-based methods. In the equivalent, alternate formulation (2.13), the former assumption amounts to $\mathbf{u}_i \sim \mathcal{N}_q(\mathbf{0}, \mathbf{I})$. Due to equivalence between the functional mixed model (2.8) and the classical mixed model (2.11), the former can be fit using any software package for fitting the latter, such as `nlme`[37] in `R` and `proc mixed` in `SAS`, which are widely popular. These packages also implement the parametric models for error autovariance function $\sigma(s,t)$ that are specifically mentioned in Section 2.2.5 and some other models as well. Once the model parameters are estimated, the other parametric function, including the heritability function $h^2(t)$ given by (2.3), can be estimated by plug-in. Further, large-sample theory of likelihood-based estimators and bootstrap can used to construct the confidence intervals and bands (see Section 1.3). The number and placement of knots can be chosen according to standard guidelines developed for semiparametric regression[38].

2.3 A simulation study

In this section, we perform Monte Carlo simulation studies to evaluate performance of interval estimators of the following parameters and parametric functions: mean function $f(t)$, variance function $\psi(t,t)$ of the random effect function, error variance σ^2, and heritability function $h^2(t)$. We take coverage probability of an interval estimator as the measure of its accuracy. Two scenarios are considered for the error covariance structure in the model. In the first, the errors are uncorrelated, whereas, in the second, they are correlated. In both scenarios, however, they are assumed to be homoscedastic.

Let us consider the first scenario. The data are simulated from (2.11) on a grid of $M = 30$ equally-spaced time points as $t \in \mathcal{T} = \{1, 2, \ldots, 30\}$ under the following settings and assumptions: $f(t) = 74 + 6.5t - 0.1t^2$; $g(t)$ is a linear combination of $q = 4$ cubic spline basis functions with coefficient vectors distributed as $\mathcal{N}_4(\mathbf{0}, \text{diag}\{1, 10, 15, 50\})$; $\sigma(s,t) = \sigma^2 I(s = t)$ with $\sigma^2 = 5$; $n \in \{100, 200, 300\}$; and $1 - \alpha = 0.95$ for both pointwise and simultaneous confidence intervals.

The model is fit by restricted maximum likelihood using a B-spline basis[29,39] with $p = 14$ basis functions for the fixed effect function and $q = 4$ basis functions for the random effect function. Further, bootstrap as described in Section 1.3 is used to construct confidence intervals with $Q = 50$ replications, both with and without a bias correction. For higher accuracy, the confidence intervals for σ^2 and $\psi(t,t)$ are first constructed on natural log scale and then exponentiated. Likewise, the confidence interval for heritablity function is first constructed on logit scale and then the inverse logit transformation is applied to get the results on the original scale. The whole process of simulating data, model fitting, and constructing confidence intervals is repeated 300 times to estimate the coverage probabilities.

The estimated coverage probabilities for σ^2 are presented in Table 2.1. They are close to the nominal level in all cases and it does not seem to matter whether the bias correction

Table 2.1. Estimated coverage probabilities (in %) of 95% confidence intervals for error variance σ^2 assuming uncorrelated errors. Settings (a) and (b) respectively represent with and without bias correction.

n	(a)	(b)
100	93.0	92.8
200	93.8	92.8
300	93.5	94.0

Table 2.2. Average estimated pointwise coverage probabilities (in %) of 95% confidence intervals for mean function $f(t)$, random effects variance function $\psi(t,t)$, and heritability function $h^2(t)$ assuming uncorrelated errors. Settings (a) and (b) respectively represent with and without bias correction.

Parameter	n	(a)	(b)
$f(t)$	100	94.6	94.9
	200	93.6	94.1
	300	95.6	96.1
$\psi(t,t)$	100	93.3	94.4
	200	95.1	95.4
	300	94.4	94.5
$h^2(t)$	100	93.5	94.2
	200	95.0	95.2
	300	94.3	94.8

is made. The average estimated pointwise coverage probabilities for $f(t)$, $\psi(t,t)$, and $h^2(t)$ presented in Table 2.2 are close to nominal level. Figure 2.1 displays the pointwise coverage probabilities for $h^2(t)$ without the bias correction. The estimates for $n \geq 200$ are close to the nominal level for all t.

Table 2.3 presents estimated simultaneous coverage probabilities for the three parameter functions. Here also we see little effect of bias correction. Some of the entries for $f(t)$ are close to the nominal level in all cases. However, the same cannot be said for $\psi(t,t)$ and $h^2(t)$, increase of n has minimal effect on performance of intervals.

On the whole, it is clear that the bias correction does not help in increasing accuracy of the interval estimators. Also, $n \geq 100$ seems to provide acceptable accuracy for pointwise intervals in all cases and for simultaneous interval for $f(t)$, $\psi(t,t)$ and $h^2(t)$ functions.

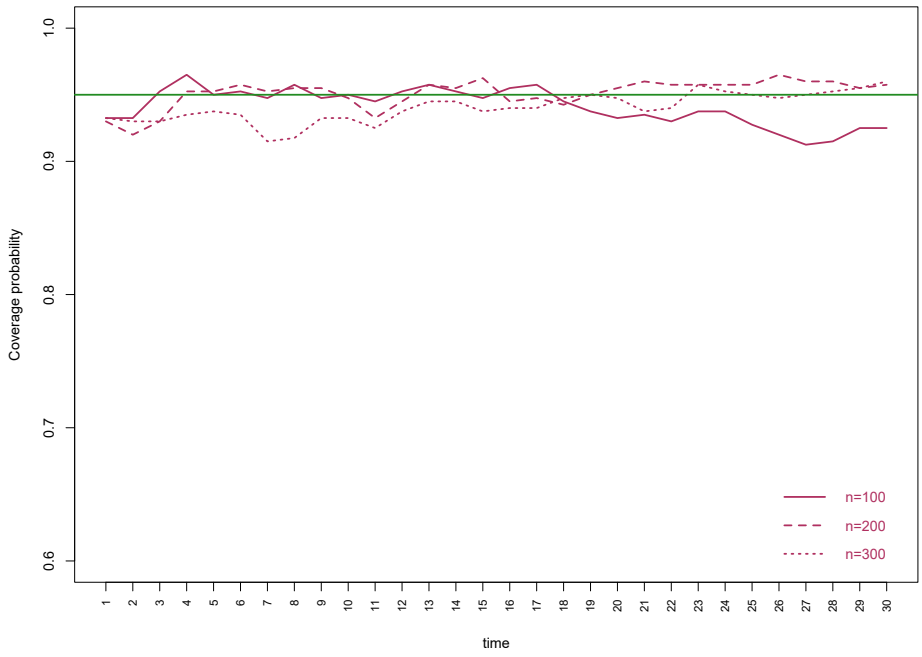

Figure 2.1. Estimated pointwise coverage probability of 95% confidence intervals for heritability function $h^2(t)$ assuming uncorrelated errors. The horizontal solid dark green line represents the nominal 0.95 level.

Table 2.3. Estimated simultaneous coverage probabilities (in %) of 95% confidence intervals for mean function $f(t)$, random effects variance function $\psi(t,t)$, and heritability function $h^2(t)$ assuming uncorrelated errors. Settings (a) and (b) respectively represent with and without bias correction.

Parameter	n	(a)	(b)
$f(t)$	100	91.5	90.5
	200	93.0	92.5
	300	94.0	95.0
$\psi(t,t)$	100	91.8	92.5
	200	92.0	92.5
	300	92.1	92.3
$h^2(t)$	100	89.8	92.5
	200	91.5	92.5
	300	92.5	92.2

Table 2.4. Estimated coverage probabilities (in %) of 95% confidence intervals for error variance σ^2 under model with AR(1) error correlation structure with parameter ϕ. Settings (a) and (b) respectively represent with and without bias correction.

	$\phi = 0.10$		$\phi = 0.70$		$\phi = 0.90$	
n	(a)	(b)	(a)	(b)	(a)	(b)
100	95.3	94.3	91.8	85.8	80.3	80.5
200	94.5	94.5	88.8	89.8	81.8	81.0
300	93.8	94.0	92.5	92.8	84.3	85.0

Next, we consider the second scenario where the errors are assumed to be correlated. Specifically, we consider three time series models for error correlation from Section 1.2.4: AR(1) with parameter $\phi \in \{0.1, 0.7, 0.9\}$; MA(1) with parameter $\theta \in \{0.1, 0.3, 0.8\}$; and ARMA(1,1) with $(\phi, \theta) \in \{(0.05, 0.05), (0.3, 0.3), (0.7, 0.8)\}$. The innovation variance in the time series models is scaled in such a way so that the resulting error variance equals σ^2. The parameter settings are the same as in the previous scenario and we proceed in the same way to simulate data, fit model, and compute interval estimates. The correct error covariance structure is assumed when fitting the model.

The results for the AR(1) models are presented in Tables 2.4, 2.5, and 2.6 and Figure 2.2. We may make the following general conclusions based on them: bias correction makes little difference; the intervals appear liberal in that their coverage probabilities are less than the nominal level; for $\phi = 0.7$ and $\phi = 0.9$, their accuracy tends to get worse as ϕ increases; and the accuracy increases with n. On the whole, the intervals may be considered to have acceptable accuracy only when $\phi = 0.1$ and $n \geq 100$; except for the simultaneous intervals.

The results for the MA(1) models are presented in Tables 2.7, 2.8, and 2.9 and Figure 2.3. These results look much better than those for the AR(1) models and allow the following general conclusions: bias correction makes little difference; the intervals appear liberal but the simultaneous intervals appear more liberal than their pointwise counterparts; the accuracy seem to be increased as θ increases; for $f(t)$, $\psi(t,t)$, and $h^2(t)$. Overall, the pointwise and simultaneous intervals may be considered to have acceptable accuracy for $n \geq 100$, for $f(t)$,

Table 2.5. Average estimated pointwise coverage probabilities (in %) of 95% confidence intervals for mean function $f(t)$, random effects variance function $\psi(t,t)$, and heritability function $h^2(t)$ under model with AR(1) error correlation structure with parameter ϕ. Settings (a) and (b) respectively represent with and without bias correction.

		$\phi = 0.10$		$\phi = 0.70$		$\phi = 0.90$	
Parameter	n	(a)	(b)	(a)	(b)	(a)	(b)
$f(t)$	100	93.7	94.0	93.4	93.5	93.8	93.8
	200	94.9	95.1	94.4	94.5	94.2	94.4
	300	94.3	94.6	92.7	92.8	92.4	93.0
$\psi(t,t)$	100	92.5	93.1	90.1	91.2	81.5	83.6
	200	92.9	93.3	93.0	93.4	81.6	83.6
	300	93.7	93.9	91.4	92.7	82.7	86.8
$h^2(t)$	100	92.6	93.4	89.2	90.0	74.4	78.8
	200	93.3	93.7	90.9	91.8	78.5	79.6
	300	94.3	94.4	90.9	92.2	78.4	82.0

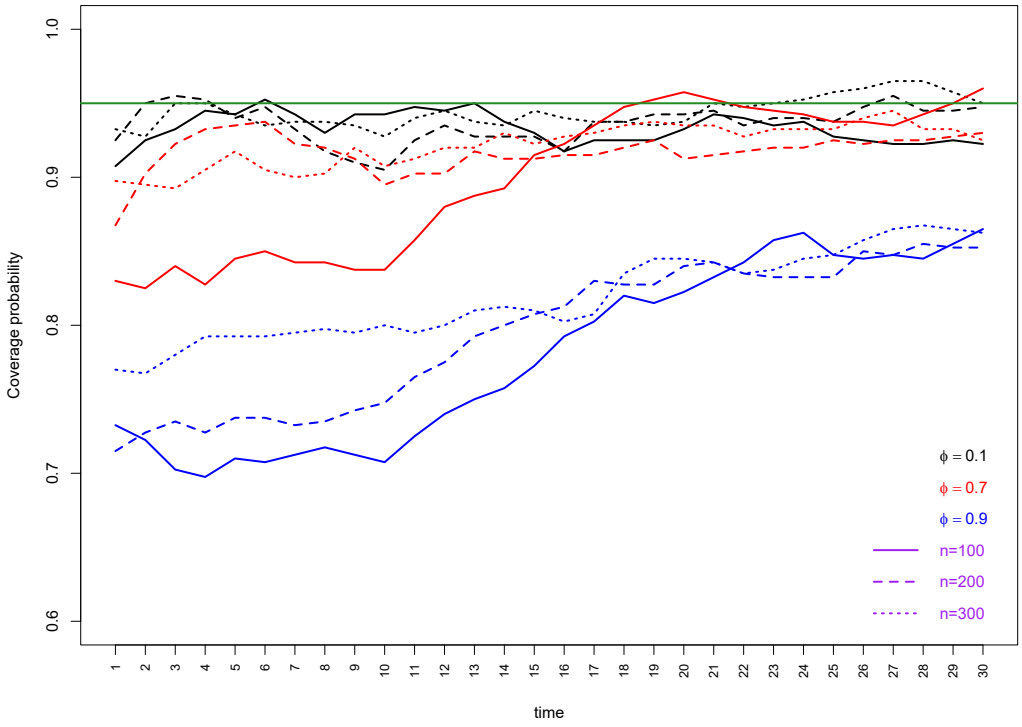

Figure 2.2. Estimated pointwise coverage probability of 95% confidence intervals for heritability function $h^2(t)$ under model with AR(1) error correlation structure with parameter ϕ. The horizontal solid dark green line represents the nominal 0.95 confidence level.

Table 2.6. Estimated simultaneous coverage probabilities (in %) of 95% confidence intervals for mean function $f(t)$, random effects variance function $\psi(t,t)$, and heritability function $h^2(t)$ under model with AR(1) error correlation structure with parameter ϕ. Settings (a) and (b) respectively represent with and without bias correction.

		$\phi = 0.10$		$\phi = 0.70$		$\phi = 0.90$	
Parameter	n	(a)	(b)	(a)	(b)	(a)	(b)
$f(t)$	100	88.0	88.5	91.3	92.0	92.3	93.3
	200	91.3	91.8	92.5	91.5	92.5	93.0
	300	92.8	94.0	88.3	88.3	89.3	89.5
$\psi(t,t)$	100	87.3	88.3	80.5	82.0	64.0	70.8
	200	89.3	90.3	85.0	86.8	66.8	69.9
	300	90.5	92.0	87.0	90.0	70.0	74.3
$h^2(t)$	100	86.3	88.8	79.0	79.5	64.5	69.3
	200	90.3	90.3	84.8	85.0	68.0	70.0
	300	89.0	91.5	86.8	88.8	69.8	75.3

Table 2.7. Estimated coverage probabilities (in %) of 95% confidence intervals for error variance σ^2 under model with MA(1) correlation structure with parameter θ. Settings (a) and (b) respectively represent with and without bias correction.

	$\theta = 0.10$		$\theta = 0.30$		$\theta = 0.80$	
n	(a)	(b)	(a)	(b)	(a)	(b)
100	93.3	93.8	93.8	94.1	94.5	95.1
200	93.1	93.8	94.2	94.2	94.3	95.3

$\psi(t,t)$ and $h^2(t)$ with $\theta = 0.8$. Nonetheless, higher sample size n with higher value of θ may provide higher performance. But with higher value of θ, performance for all sample sizes considered here may be competitive.

The results for the ARMA(1, 1) models are presented in Tables 2.10, 2.11, and 2.12 and Figure 2.4. These results appear intermediate between those for AR(1) and MA(1) models. We may make the following general conclusions based on them: bias correction is not improving the performance of the interval estimators; the pointwise intervals are acceptably close to the nominal level but the simultaneous intervals appear more liberal than their pointwise counterparts; the accuracy of both the intervals for σ^2, $\psi(t,t)$ and $h^2(t)$ does seem to depend on the AR(1) and MA(1) parameter choices; and the accuracy improves with n. Overall, the pointwise and simultaneous intervals may be considered to

Table 2.8. Average estimated pointwise coverage probabilities (in %) of 95% confidence intervals for mean function $f(t)$, random effects variance function $\psi(t,t)$, and heritability function $h^2(t)$ under model with MA(1) correlation structure with parameter θ. Settings (a) and (b) respectively represent with and without bias correction.

		$\theta = 0.10$		$\theta = 0.30$		$\theta = 0.80$	
Parameter	n	(a)	(b)	(a)	(b)	(a)	(b)
$f(t)$	100	91.6	92.0	93.1	92.8	94.6	95.2
	200	92.5	92.1	92.3	93.3	95.1	95.6
$\psi(t,t)$	100	93.5	93.0	93.5	94.1	94.8	95.7
	200	94.1	94.0	94.3	94.5	95.6	96.1
$h^2(t)$	100	93.3	93.1	93.7	94.1	94.2	94.6
	200	95.3	95.1	95.3	95.4	95.7	96.3

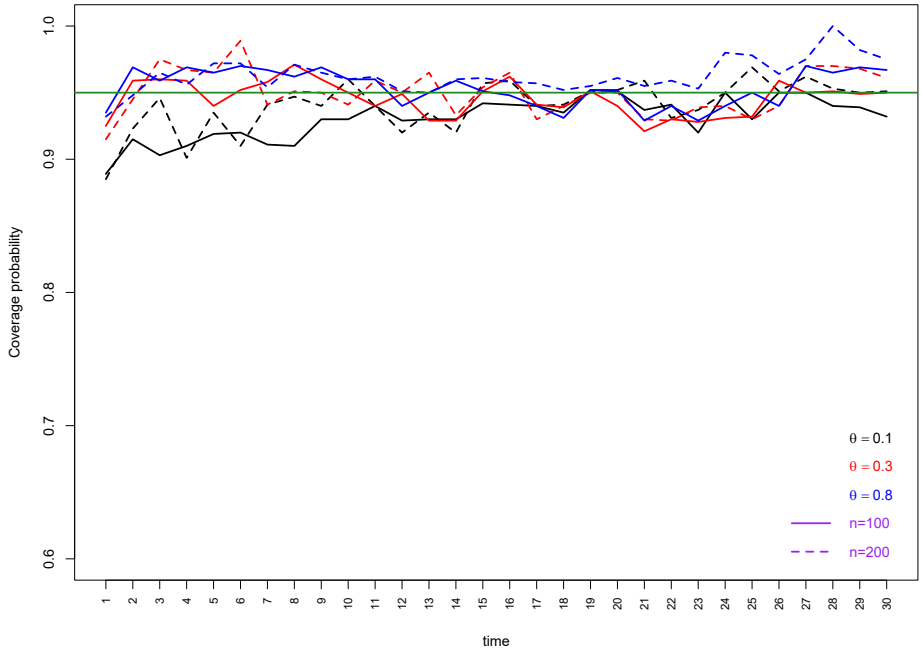

Figure 2.3. Estimated pointwise coverage probability of 95% confidence intervals for heritability function $h^2(t)$ under model with MA(1) error correlation structure with parameter θ.

Table 2.9. Estimated simultaneous coverage probabilities (in %) of 95% confidence intervals for mean function $f(t)$, random effects variance function $\psi(t,t)$, and heritability function $h^2(t)$ under model with MA(1) correlation structure with parameter θ. Settings (a) and (b) respectively represent with and without bias correction. .

		$\theta = 0.10$		$\theta = 0.30$		$\theta = 0.80$	
Parameter	n	(a)	(b)	(a)	(b)	(a)	(b)
$f(t)$	100	90.1	91.1	92.3	92.5	93.6	93.5
	200	91.2	91.3	93.2	93.1	93.3	94.1
$\psi(t,t)$	100	91.5	91.3	91.8	91.7	92.8	92.7
	200	92.1	92.0	92.3	92.7	93.1	94.6
$h^2(t)$	100	90.3	90.8	91.3	91.2	92.6	93.3
	200	90.5	91.1	92.3	92.1	93.7	94.1

Table 2.10. Estimated coverage probabilities (in %) of 95% confidence intervals for error variance σ^2 under model with ARMA(1, 1) correlation structure with three combinations of parameters (ϕ, θ): (1) $(0.05, 0.05)$, (2) $(0.30, 0.30)$, and (3) $(0.70, 0.80)$. Settings (a) and (b) respectively represent with and without bias correction.

	(1)		(2)		(3)	
n	(a)	(b)	(a)	(b)	(a)	(b)
100	93.5	94.3	93.3	90.5	90.3	89.8
200	97.3	96.3	94.8	92.0	92.0	90.2

Table 2.11. Average estimated pointwise coverage probability (in %) of 95% confidence intervals for mean function $f(t)$, random effects variance function $\psi(t,t)$, and heritability function $h^2(t)$ under model with ARMA(1, 1) correlation structure with three combinations of parameters (ϕ, θ): (1) $(0.05, 0.05)$, (2) $(0.30, 0.30)$, and (3) $(0.70, 0.80)$. Settings (a) and (b) respectively represent with and without bias correction.

		(1)		(2)		(3)	
Parameter	n	(a)	(b)	(a)	(b)	(a)	(b)
$f(t)$	100	95.1	95.3	94.9	95.0	94.3	94.6
	200	95.9	95.6	93.6	93.9	94.2	94.5
$\psi(t,t)$	100	93.3	94.0	90.7	91.0	91.2	92.1
	200	94.7	94.7	93.2	93.7	93.5	94.4
$h^2(t)$	100	93.5	94.2	90.2	90.7	90.8	90.4
	200	94.9	95.1	92.9	93.8	93.3	93.1

have acceptable accuracy for $n \geq 100$ in case of all parameters and parameter functions for lower values of AR(1) and MA(1) parameters.

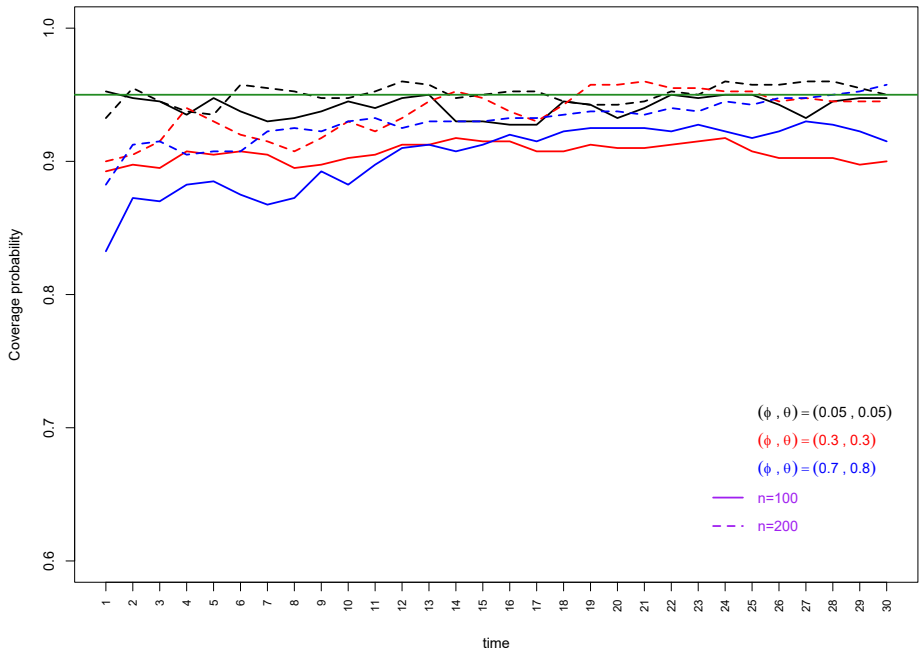

Figure 2.4. Estimated pointwise coverage probability of 95% confidence intervals for heritability function $h^2(t)$ under model with ARMA(1, 1) correlation structure with parameters (ϕ, θ). The horizontal solid dark green line represents the nominal 0.95 confidence level.

Table 2.12. Estimated simultaneous coverage probabilities (in %) of 95% confidence intervals for mean function $f(t)$, random effects variance function $\psi(t,t)$, and heritability function $h^2(t)$ under model with ARMA(1, 1) correlation structure with three combinations of parameters (ϕ, θ): (1) $(0.05, 0.05)$, (2) $(0.30, 0.30)$, and (3) $(0.70, 0.80)$. Settings (a) and (b) respectively represent with and without bias correction.

		(1)		(2)		(3)	
Parameter	n	(a)	(b)	(a)	(b)	(a)	(b)
$f(t)$	100	94.0	94.5	93.5	93.3	91.8	91.5
	200	93.5	94.5	91.3	91.8	93.5	93.8
$\psi(t,t)$	100	89.3	90.5	82.8	84.5	84.0	86.5
	200	91.5	92.5	87.3	88.5	86.8	90.3
$h^2(t)$	100	89.8	91.5	84.0	86.5	85.0	86.0
	200	93.3	92.8	87.0	89.0	88.3	88.0

Taken together, the simulation results indicate that the bias correction is not necessary; the pointwise intervals often have acceptable accuracy for $n \geq 100$; the simultaneous intervals tend to be accurate only when the error correlations are small and they often need $n \geq 100$ for acceptable accuracy.

2.4 Illustration

In this section, we illustrate the methodology developed in this chapter by analyzing the Berkeley growth data introduced in Section 1.4. As mentioned therein, these data consist of growth curves height (in cm) of $n = 54$ girls measured at the same $M = 31$ ages for each girl. The ages, however, are not equally spaced as four measurements are available between 1 and 2 years; annual measurements between 2 and 8 years; and biannual measurements between 8 and 18 years. Figure 1.1 displays the individual trajectories of the growth curves superimposed with the sample mean function. It is clear that the heights increase up to around age 14 and then stabilize. This is consistent with what we expect.

Next, we fit the standard functional mixed model (2.8) that assumes homoscedastic and uncorrelated errors. The fixed and random effect functions are represent using the cubic B-spline basis[29,39] with $p = 5$ and $q = 4$ basis functions, respectively. The model is fit by restricted maximum likelihood (REML) method using the `nlme` package in `R`. The resulting estimated mean function is also displayed in Figure 2.5. We see that it close to the sample mean function, which is expected due to the balanced nature of the data.

To check adequacy of the homoscedastic and uncorrelated assumption for the errors, we compute the sample correlation matrix of the normalized residuals from the model and display a heat map of it in Figure 2.6. If the assumption holds, the sample correlation matrix should resemble an identity matrix, making its heat map look essentially whitish below the diagonal. This is clearly not the case in the figure because there is a band structure with four prominent bands. It seems that as lag increases, the autocorrelation initially decreases from

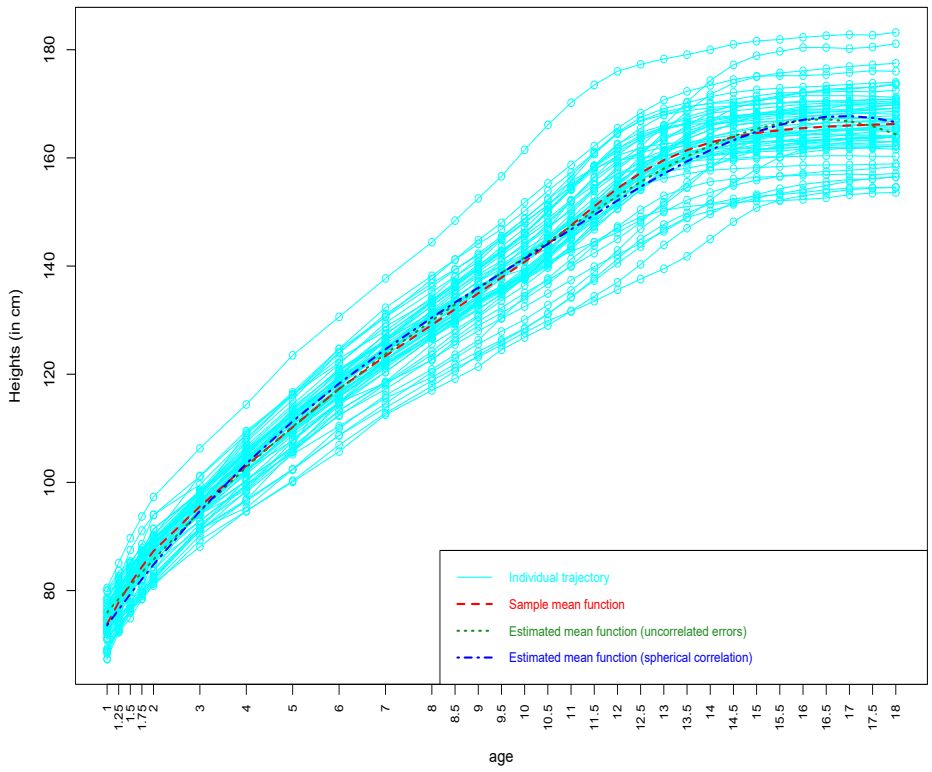

Figure 2.5. Individual trajectories of height from the Berkeley growth curve data superimposed with the sample mean function, the estimated mean function assuming uncorrelated errors, and the estimated mean function assuming spherical correlation structure in the errors. All the estimated mean functions essentially close to each other.

large positive to large negative values, then it increases to moderately positive values, and decreases again to negative or near-zero values. Thus, the assumption of uncorrelated and homoscedastic errors is not tenable and there is a clear need to the model the autocorrelation in the errors. However, it is difficult to capture the observed band structure through standard parametric models for autocorrelation.

Among the available error autocorrelation models in the `nlme` package, the time series models described in Section 1.2.4 are out of consideration because the observation times (i.e., the ages) are not equally-spaced. However, we can employ spatial correlation structures described in Section 1.2.4 based on variogram models because they allow unevenly

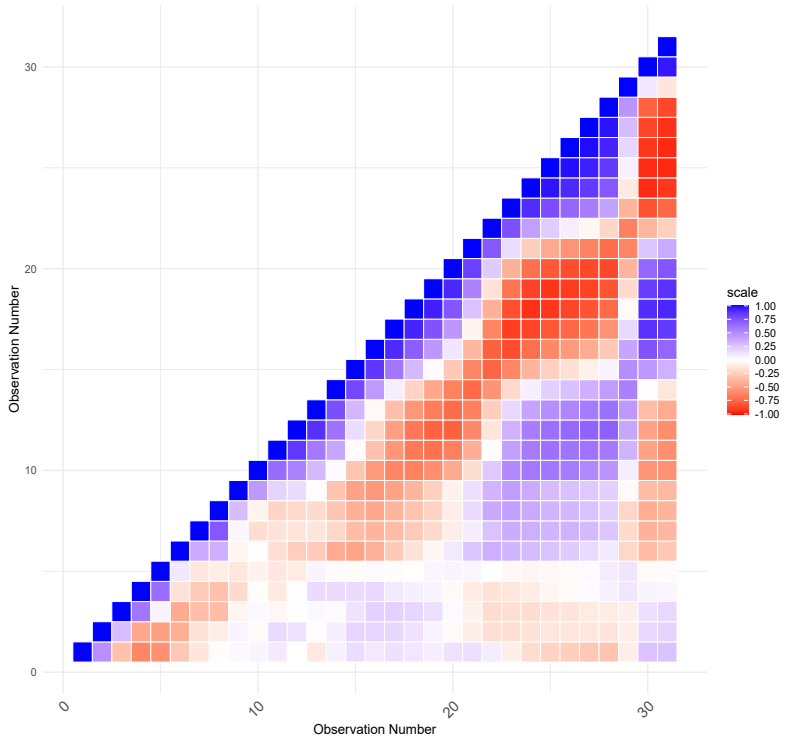

Figure 2.6. Heat map of sample correlation matrix of normalized residuals from the model assuming uncorrelated errors.

spaced continuous observation times. We refit the proposed model using four different spatial correlation structures to the errors, namely, exponential, rational quadratic, spherical, and Gaussian models. AIC, BIC, and the maximum log-likelihood values for the resulting models are presented in Table 2.13. The spherical correlation model is appears the best, followed by the exponential correlation model (which is equivalent to the continuous AR(1) model). The heat map of sample correlation matrix of normalized residuals from this model is shown in Figure 2.7. Although the blocks below the diagonal are not perfectly whitish, but they are comparatively whiter than those in Figure 2.6, showing a marked improvement over the uncorrelated errors model.

The estimated mean function for the spherical correlation model is also displayed in Figure 2.5. Due to the balanced nature of the data, this mean function also close to the other

Table 2.13. Model selection criteria for different spatial correlation structures.

Model	Structure	AIC	BIC	log-likelihood
1	Exponential	5089.721	5181.860	-2527.860
2	Rational Quadratic	5500.498	5592.638	-2733.249
3	Spherical	4976.840	5068.979	-2471.420
4	Gaussian	6114.226	6206.366	-3040.113

estimated mean functions displayed in that figure. Next, Figures 2.8 and 2.9 respectively present the estimated random effect variance function $\hat{\psi}(t,t)$ and heritability function $\hat{h}^2(t)$ along with superimposed 95% simultaneous confidence bands under both uncorrelated and correlated error models. We see that not taking into account of autocorrelation structure in the errors leads to a slightly higher random effect variance throughout the age interval. This in turn leads to a slightly higher estimate of heritability function over all the ages. However, the difference is not substantial. Thus, in these data, the difference between the heritability estimate with and without the correlation structure is not substantial. This may be partly due to the fact that the error variance in these data is small relative to the between-subject variation. We may expect a larger difference when the error variance is substantial relative to the between-subject variation.

2.5 Summary and future work

In this chapter, we present a functional mixed model that allows the within-subject errors to be heteroscedastic and correlated. Although the random effect covariance structure is modeled nonparametrically via splines, parametric models are assumed for error covariance structure to avoid identifiability issues. The following additional research is planned for the near future:

- Develop a cross-validation type method for selecting the number of basis functions.
- Incorporate penalized estimation of mean function in the methodology.

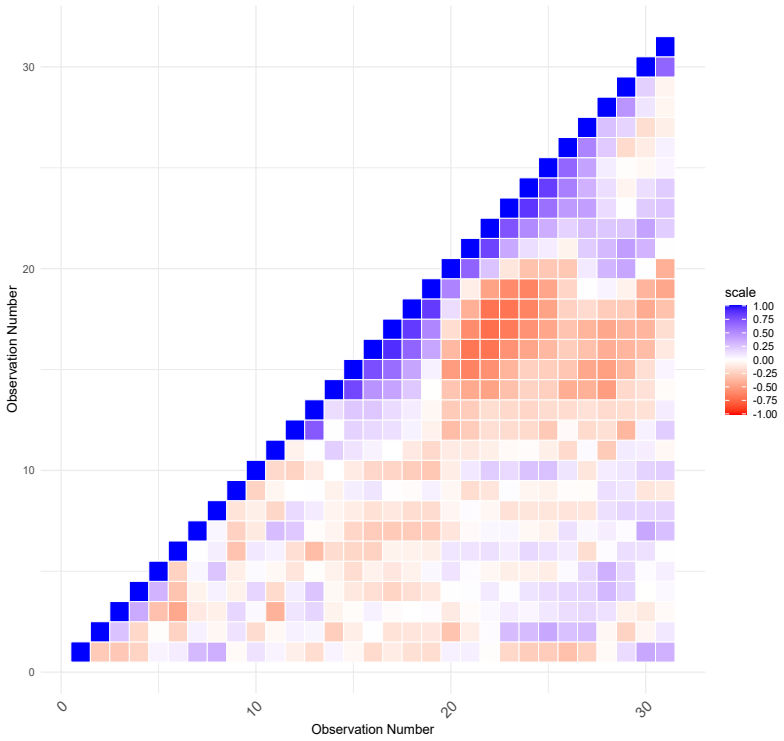

Figure 2.7. Heat map of sample correlation matrix of normalized residuals from the model assuming a spherical autocorrelation structure for the errors.

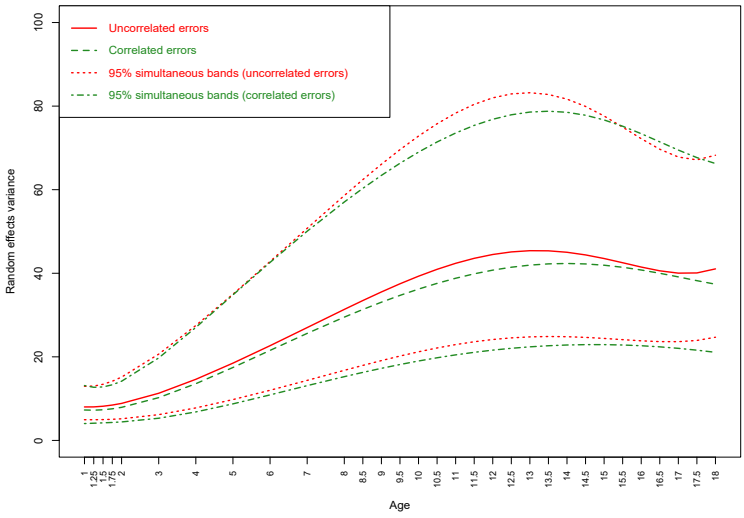

Figure 2.8. Estimate of random effects variance function $\hat{\psi}(t,t)$ assuming uncorrelated errors and a spherical correlation structure in the errors.

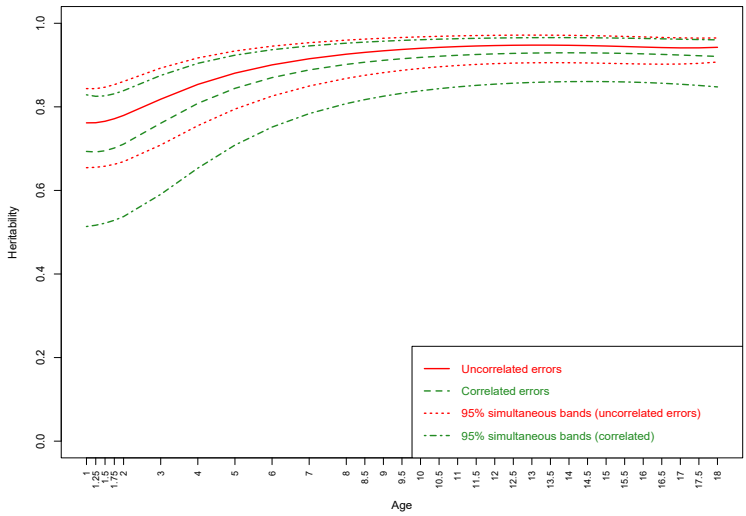

Figure 2.9. Estimate of heritability function $\hat{h}^2(t)$ assuming uncorrelated errors and a spherical correlation structure in the errors.

- Evaluate via simulation the impact of a model selection step to select the error correlation structure.
- Develop an alternative for case when the errors are highly correlated as the simultaneous interval for $h^2(t)$ does not work well in that case.
- Evaluate spatial correlation structures via simulation.
- Develop an `R` package to implement the methodology and make it publicly available.

CHAPTER 3

A FUNCTIONAL MIXED MODEL FOR FAMILY DATA

3.1 Introduction

The data considered in this chapter have the same structure as in Chapter 2 in that they consist of functional responses of n subjects but with one distinction: the subjects are not mutually independent anymore. Instead, they are grouped into families. This distinction is important because members of the same family are genetically related, inducing dependence in their responses, while members of different families can be assumed to provide independent responses. To get correct inference, modeling and analysis of family data must take into to account of the familial dependence among the subjects and hence the dependence in their responses. Our goal in this chapter is to adapt the functional mixed model methodology of the previous chapter for family data.

Family data are quite common in statistical genetics. As mentioned in Chapter 1, they are used to estimate heritability of a quantitative trait. Developing methods for estimation of heritability has been a mainstay of quantitative genetics. Over time, a number of methods have become available for estimating heritability from a variety of data types, including univariate traits[40–43], multivariate traits[44,45], and longitudinal traits[32,36,46–50]. These methods generally assume a mixed model (or often a special case of it called a variance components model) for the data, where the mean is a fixed effect and the genetic effect is an additive random effect, and use likelihood-based methods for estimating the unknown parameters. In the case of longitudinal data, the fixed mean effect and the random genetic effect are functions of time. These effects can also be specified nonparametrically using splines[32,36,49,50].

As the individuals in a family are correlated, their correlation is incorporated into the analysis in the form of a correlation matrix[46,48]. Traditionally, the methods use pedigree information to infer this relatedness[42,46]. However, if the pedigree information is partially or

wholly unknown, genetic markers[51–53] or the observed genotype data[54–59] can be used to infer relatedness between pairs of individuals. In the situation where neither familial relationships nor genotypes are available, a recent work[60] has explored the use of methylation data for estimating the relatedness. For model estimation, although it is possible to use specialized software, it is more common nowadays to adapt a mainstream statistical software for fitting mixed models such as `nlme`[37] and `lme4`[61] packages in the open-source system `R`[62] as they are powerful, flexible, and efficient. Adaptations such as `lme4qtl`[63] are necessary because the standard software assume independence across individuals, which is not true for family data. This also the approach we adopt here. An alternative to this, applicable for certain types of family data, is to first perform a reparameterization of the genetic random effect[64] and then use a standard software.

The rest of this chapter is organized as follows. In Section 3.2, we adapt the functional mixed model of Chapter 2 for family data by essentially following the same development. The new functional mixed model generalizes many existing quantitative trait models from statistical genetics. In Section 3.3, we report results of a simulation study to evaluate properties of the proposed estimation procedures. Section 3.5 concludes with summary and mentions some ongoing and future work.

3.2 Modeling family data

As mentioned in Section 3.1, the family data have the same structure as the data from Chapter 2 with the exception that the subjects may have familial dependence. This dependence is incorporated through a $n \times n$ *known* correlation matrix

$$\mathbf{A} = (r_{il}). \tag{3.1}$$

Here r_{il} can be computed from the pedigree data using the known relationship between the individuals i and l as twice their kinship coefficient[65]. Alternatively, this matrix can be computed using genetic markers[51–53] or observed genotype information[54–59].

3.2.1 Population and data models

Borrowing notation from the previous chapter, our functional mixed model for the population is identical to that of (2.1), i.e.,

$$y(t) = f(t) + g(t) + \epsilon(t),\ t \in \mathcal{T}, \tag{3.2}$$

where $y(t)$, $t \in \mathcal{T}$ denotes the functional response of a randomly selected subject from the population; $f(t)$ is the population mean function; $g(t)$ is a random effect function that captures the deviation from the population mean for the subject, with mean zero and covariance function $\psi(s,t)$; $\epsilon(t)$ represents the random error, distributed independently of $g(t)$, with mean zero and covariance function $\sigma(s,t)$; and the errors are allowed to be heteroscedastic and correlated.

The observed data consist of $y_i(t_{ij})$, $j = 1, \ldots, M_i$, $i = 1, \ldots, n$, where $y_i(t)$ is the functional response of subject i, discretely observed at the observation times t_{ij}. Following (2.4), the model for these family data can also be written as

$$y_i(t_{ij}) = f(t_{ij}) + g_i(t_{ij}) + \epsilon_i(t_{ij}),\ j = 1, \ldots, M_i,\ i = 1, \ldots, n, \tag{3.3}$$

where $f(t)$ is the population mean function, $g_i(t)$ is the random effect function of individual i, and $\epsilon_i(t)$ is the random error function, distributed independently of $g_i(t)$. Here it is assumed that the $\epsilon_i(t)$ are independent realizations of the error function $\epsilon(t)$ from (3.2) which are observed at the sites t_{ij}. Therefore, the errors have mean zero. Moreover, the errors associated with the same individual have covariance function $\sigma(s,t)$, whereas those associated with different individuals are independent. Further, the $g_i(t)$ are realizations of the function $g(t)$ defined in (3.2) which are observed at the sites t_{ij}. Thus, $g_i(t)$ has mean zero and its covariance function is $\psi(s,t)$. However, unlike (2.4), the $g_i(t)$ now may be dependent across subjects. Specifically, it is now assumed that the covariance between random effect functions of subjects i and l is

$$\text{cov}(g_i(s), g_l(t)) = r_{il}\psi(s,t), \quad s,t \in \mathcal{T},\ i,l = 1, \ldots, n, \tag{3.4}$$

where r_{il} is given by (3.1). It follows from (3.4) that $\text{corr}(g_i(t), g_l(t)) = r_{il}$, $t \in \mathcal{T}$. Thus, we may interpret r_{il} as the correlation between genetic effects of individuals i and l at the same t. This correlation does not depend on t. As in (2.3), the heritability function is

$$h^2(t) = \frac{\text{var}(g(t))}{\text{var}(y(t))} = \frac{\psi(t,t)}{\psi(t,t) + \sigma(t,t)}, \quad t \in \mathcal{T}. \tag{3.5}$$

3.2.2 A natural generalization of quantitative trait models

If we think of the response y as a quantitative trait and assume only an additive genotypic effect and no additive major gene effects or shared environmental effects, then several mixed models from quantitative genetics may be seen as special cases of the proposed model (3.3). Examples include the classical model for one genotype[12], polygenic model[65], and the multivariate trait model[45,65]. To see the last model, suppose there are M traits, indexed as $j = 1, \ldots, M$. The traits may be polygenic but there is no need be specific about the number of loci involved. The model considered in Sec. 8.4 of Lange[65] can be written as $y_{ij} = f_j + g_{ij} + \epsilon_{ij}$, $j = 1, \ldots, M$, $i = 1, \ldots, n$, where y_{ij} is the phenotypic value of jth trait for individual i, g_{ij} is the genotypic effect acting on the jth trait of individual i, and ϵ_{ij} is the error term. The g_{ij} have mean zero and $\text{cov}(g_{ij}, g_{lq}) = r_{il}\psi_{jq}$, $i, l = 1, \ldots, n$, $j, q = 1, \ldots, M$; and the ϵ_{ij} have mean zero and $\text{cov}(\epsilon_{ij}, \epsilon_{iq}) = \sigma_{jq}$. Here ψ_{jj} and ψ_{jq} are the additive variances and covariances of the genetic effects and σ_{jj} and σ_{jq} are the variances and covariances of the errors associated with the same individual. By a suitable choice of the quantities involved in (3.3), we can see that this multivariate trait model is a special case of it.

Likewise, using appropriate notation, the longitudinal model[46] for a single trait can also be seen as a special case of (3.3). The variance function $\psi(t,t)$ in the proposed model generalizes the notion of *additive variance* in quantitative trait models by letting it depend on t. Moreover, for covariance between genetic effects acting on individuals i and l, these models assume the structure $\text{cov}(g_i, g_l) = r_{il}\psi$ in case of a univariate trait[12,65] and $\text{cov}(g_{ij}, g_{lq}) =$

$r_{il}\psi_{jq}$ in case of a multivariate trait[65]. The covariance between the functional genetic effects g_i and g_l given by (3.4) is a natural generalization of this structure. Thus, the proposed functional mixed model (3.3) may be justifiably seen as a generalization of the commonly used mixed models in quantitative genetics that only include additive genetic effects and random errors.

3.2.3 Basis expansion and representation as a classical mixed model

Proceeding along the lines of Chapter 2, the data model (3.3) can be written using the basis expansion (2.5) to get the analog of (2.8) as

$$y_i(t_{ij}) = \boldsymbol{\Phi}_f^T(t_{ij})\boldsymbol{\beta} + \boldsymbol{\Phi}^T(t_{ij})\mathbf{b}_i + \epsilon_i(t_{ij}), \quad j = 1, \ldots, M_i, \quad i = 1, \ldots, n. \tag{3.6}$$

But here, unlike in (2.8), $\mathbf{b}_i$ are not assumed to be mutually independent. Instead, it is assumed that

$$\text{cov}(\mathbf{b}_i, \mathbf{b}_l) = r_{il}\mathbf{G}, \tag{3.7}$$

so that we have $\text{cov}(g_i(s), g_l(t)) = \boldsymbol{\Phi}^T(s)\text{cov}(\mathbf{b}_i, \mathbf{b}_l)\boldsymbol{\Phi}(t) = r_{il}\boldsymbol{\Phi}^T(s)\mathbf{G}\boldsymbol{\Phi}(t) = r_{il}\psi(s,t)$, which is the desired covariance structure from (3.4). This also shows that with family data, the random effect covariance structure can be separated from the error covariance structure, resolving the identifiability issue that afflicts the model from the previous chapter where the subjects are independent.

Now the model (3.6) also be represented in form of a classical mixed model to get the analog of (2.10) as

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \ \mathbf{b}_i \sim (\mathbf{0}, \mathbf{G}), \ \boldsymbol{\epsilon}_i \sim (\mathbf{0}, \mathbf{R}_i), \quad i = 1, \ldots, n, \tag{3.8}$$

with $\text{cov}(\mathbf{b}_i, \mathbf{b}_l) = r_{il}\mathbf{G}$ and $\text{cov}(\mathbf{b}_i, \boldsymbol{\epsilon}_l) = \mathbf{0}$ for all i and l, and $\text{cov}(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_l) = \mathbf{0}$ for $i \neq l$; and to get the analog of (2.13) as

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\Omega}\boldsymbol{\Gamma}\mathbf{u}_i + \boldsymbol{\epsilon}_i, \ \mathbf{u}_i \sim (\mathbf{0}, \mathbf{I}_q), \ \boldsymbol{\epsilon}_i \sim (\mathbf{0}, \mathbf{R}_i), \ i = 1, \ldots, n, \tag{3.9}$$

where $\text{cov}(\mathbf{u}_i, \mathbf{u}_l) = r_{il}\mathbf{I}_q$. Next, the model for all the data together than can also represented in the form of a joint classical mixed model to get the analog of (2.11) as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}, \quad \mathbf{b} \sim (\mathbf{0}, \mathbf{A} \otimes \mathbf{G}), \ \boldsymbol{\epsilon} \sim (\mathbf{0}, \mathbf{R}), \ \text{cov}(\mathbf{b}, \boldsymbol{\epsilon}) = \mathbf{0}, \tag{3.10}$$

where $\mathbf{A}$ is the family correlation matrix given by (3.1). We can also get the analog of (2.14) as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}, \quad \mathbf{u} \sim (\mathbf{0}, \mathbf{A} \otimes \mathbf{I}_q), \ \boldsymbol{\epsilon} \sim (\mathbf{0}, \mathbf{R}), \ \text{cov}(\mathbf{u}, \boldsymbol{\epsilon}) = \mathbf{0}, \tag{3.11}$$

At this point, note that, in the models (3.3) and (3.6) we did not use specific index for family information because family information is being taken into account in correlation matrix given in (3.1). The model (2.11) or equivalently (2.14) for independent subjects is obtained by setting $\mathbf{A} = \mathbf{I}_n$ in (3.10) or (3.11), respectively.

The fact that the individual random effects vector $\mathbf{b}_i$ in (3.8) or $\mathbf{u}_i$ in (3.9) are correlated, not independent which is a standard assumption in mixed models, precludes the use of common software for fitting mixed models. To address this, we extend an idea used in `lme4qtl` package[63] and apply a transformation to convert the model in the standard form where the random effects vector has a covariance matrix of the form $\mathbf{I}_n \otimes \mathbf{I}_q$. To this end, note that the vector $\mathbf{u}$ in (3.11) is written as n stacks of $q \times 1$ vectors, where each stack consists of q coefficients for a particular individual. Define a $nq \times 1$ vector by rearranging the elements of $\mathbf{u}$ so that it is written as q stacks of $n \times 1$ vectors, where each stack consists of values of a particular coefficient on n individuals. Specifically, let $\tilde{\mathbf{u}}_k = (u_{k1}, \ldots, u_{kn})^T$ be the $n \times 1$ vector of values of the kth coefficient and $\tilde{\mathbf{u}} = (\tilde{\mathbf{u}}_1^T, \ldots, \tilde{\mathbf{u}}_q^T)^T$ be the $nq \times 1$ vector of all the coefficients. We have, $\tilde{\mathbf{u}}_k \sim (\mathbf{0}, \mathbf{A})$ and $\tilde{\mathbf{u}} \sim (\mathbf{0}, \mathbf{I}_q \otimes \mathbf{A})$. Let $\tilde{\mathbf{Z}}$ be the $N \times nq$ design matrix associated with the random effects vector $\tilde{\mathbf{u}}$. It is a rearrangement of columns of $\mathbf{Z}$ and has the form $\tilde{\mathbf{Z}} = [\tilde{\mathbf{Z}}_1, \ldots, \tilde{\mathbf{Z}}_q]$, where each $\tilde{\mathbf{Z}}_k$ is a $N \times n$ matrix.

Let the $n \times n$ matrix $\mathbf{U}$ denote the upper triangular factor of the Cholesky decomposition of the correlation matrix $\mathbf{A}$ so that $\mathbf{U}^T\mathbf{U} = \mathbf{A}$. By construction,

$$\mathbf{Zu} = \tilde{\mathbf{Z}}\tilde{\mathbf{u}} = \sum_{k=1}^{q} \tilde{\mathbf{Z}}_k \tilde{\mathbf{u}}_k = \sum_{k=1}^{q} \tilde{\mathbf{Z}}_k \mathbf{U}^T (\mathbf{U}^T)^{-1} \tilde{\mathbf{u}}_k = \sum_{k=1}^{q} \tilde{\mathbf{Z}}_k^* \tilde{\mathbf{u}}_k^* = \tilde{\mathbf{Z}}^* \tilde{\mathbf{u}}^* = \mathbf{Z}^* \mathbf{u}^*, \tag{3.12}$$

where $\tilde{\mathbf{Z}}_k^* = \tilde{\mathbf{Z}}_k \mathbf{U}^T$, $\tilde{\mathbf{u}}_k^* = (\mathbf{U}^T)^{-1}\tilde{\mathbf{u}}_k$, $\tilde{\mathbf{Z}}^* = [\tilde{\mathbf{Z}}_1^*, \ldots, \tilde{\mathbf{Z}}_q^*]$, $\tilde{\mathbf{u}}^* = (\tilde{\mathbf{u}}_1^{*T}, \ldots, \tilde{\mathbf{u}}_q^{*T})^T$, and $\mathbf{Z}^*$ is the $N \times nq$ design matrix associated with $\mathbf{u}^*$. Here $(\mathbf{u}^*, \mathbf{Z}^*)$ is obtained from $(\tilde{\mathbf{u}}^*, \tilde{\mathbf{Z}}^*)$ by applying the reverse of the operations that led to $(\tilde{\mathbf{u}}, \tilde{\mathbf{Z}})$ from $(\mathbf{u}, \mathbf{Z})$. It follows that $\tilde{\mathbf{u}}_k^* \sim (\mathbf{0}, \mathbf{I}_n)$, $\tilde{\mathbf{u}}^* \sim (\mathbf{0}, \mathbf{I}_q \otimes \mathbf{I}_n)$, $\mathbf{u}_i^* \sim (\mathbf{0}, \mathbf{I}_q)$, and $\mathbf{u}^* \sim (\mathbf{0}, \mathbf{I}_n \otimes \mathbf{I}_q)$. Thus, an equivalent representation of model (3.11) is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}^*\mathbf{u}^* + \boldsymbol{\epsilon}, \quad \mathbf{u}^* \sim (\mathbf{0}, \mathbf{I}_n \otimes \mathbf{I}_q), \ \boldsymbol{\epsilon} \sim (\mathbf{0}, \mathbf{R}), \ \mathrm{cov}(\mathbf{u}^*, \boldsymbol{\epsilon}) = \mathbf{0}. \tag{3.13}$$

In this formulation, the individual-specific components of the transformed vector $\mathbf{u}^*$ are independent draws from a q-variate distribution with mean zero and covariance matrix $\mathbf{I}_q$. Due to (3.12), $\mathrm{var}(\mathbf{Zu}) = \mathrm{var}(\mathbf{Z}^*\mathbf{u}^*)$, implying that $\mathbf{Z}(\mathbf{A} \otimes \mathbf{I}_q)\mathbf{Z}^T = \mathbf{Z}^*(\mathbf{I}_n \otimes \mathbf{I}_q)\mathbf{Z}^{*T}$. Unlike $\mathbf{Z}$ in (3.11), $\mathbf{Z}^*$ in (3.13) is not a block diagonal matrix with individual-specific blocks on the diagonal. The formulation (3.13) can be used for fitting the model. We can proceed along similar lines to get an equivalent representation of (3.10) as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}^*\mathbf{b}^* + \boldsymbol{\epsilon}, \quad \mathbf{b}^* \sim (\mathbf{0}, \mathbf{I}_n \otimes \mathbf{G}), \ \boldsymbol{\epsilon} \sim (\mathbf{0}, \mathbf{R}), \ \mathrm{cov}(\mathbf{b}^*, \boldsymbol{\epsilon}) = \mathbf{0}. \tag{3.14}$$

3.2.4 Models for error covariance function

As in Section 2.2.5, we let the covariance matrix $\mathbf{G}$ be unstructured and consider only parametric models for error covariance function $\sigma(s, t)$.

3.2.5 Model fitting and inference

As in Section 2.2.6, for model fitting, we additionally make the normality assumption, and in principle, the model can be fit using likelihood-based methods and large-sample theory

together with bootstrap can be used for inference. It would appear that once the functional mixed model (3.3) is represented as a classical mixed model (3.14), one should be able to use a software package for fitting the latter model to fit the former model. However, due to the quirks of the design of the mixed model fitting packages, challenges remain in incorporating familial dependence among the subjects. Thus far, we have only been able to implement model fitting under the assumption of uncorrelated and homoscedastic errors, i.e., $\sigma(s,t) = \sigma^2 I(s=t)$, using the `R` package `lme4`[61].

3.3 A simulation study

In this section, we use Monte Carlo simulation to evaluate performance of interval estimators of the following quantities: mean function $f(t)$, variance function $\psi(t,t)$ of the random effect function, error variance σ^2, and heritability function $h^2(t)$. Coverage probability for an interval estimator is taken as the measure of accuracy. Both pointwise and simultaneous coverage probabilities are considered for parameter function that depend on t. This evaluation only focuses on the model with uncorrelated and homoscedastic errors, i.e., $\mathbf{R} = \sigma^2\mathbf{I}$, for which we have been able to implement the fitting.

For computational convenience, we assume that the number of subjects n is a multiple of 4 and the subjects are grouped into $n/4$ families, each consisting of 4 members: father, mother, and two children. The correlation structure of the family members is as follows: father and mother are uncorrelated, meaning $r_{12} = 0$; parents and children are correlated with correlation $r_{13} = r_{14} = r_{23} = r_{24} = 0.5$; and the two siblings are also correlated with correlation $r_{34} = 0.5$. Thus, the correlation matrix for a family is

$$\mathbf{A}_0 = \begin{bmatrix} 1 & 0 & 0.5 & 0.5 \\ 0 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 \end{bmatrix}.$$

Table 3.1. Estimated coverage probabilities (in %) of 95% confidence intervals for error variance σ^2. Settings (a) and (b) respectively represent with and without bias correction.

# Families	n	(a)	(b)
25	100	94.5	96.0
50	200	93.3	92.3
75	300	97.8	99.0

Thus, the familial correlation matrix $\mathbf{A}$ for n subjects will have a diagonal structure with $\mathbf{A}_0$ repeated on the diagonal $n/4$ times. In other words, it can be written as $\mathbf{A} = \mathbf{I}_{n/4} \otimes \mathbf{A}_0$.

The data are simulated from (3.14) on a grid of $M = 30$ equally-spaced time points as $t \in \mathcal{T} = \{1, 2, \ldots, 30\}$ under the following settings and assumptions: $f(t) = 74.0 + 6.5t - 0.10t^2$; $g(t)$ is a linear combination of $q = 4$ cubic spline basis functions with coefficient vectors distributed as $\mathcal{N}_4(\mathbf{0}, \mathbf{G} = \text{diag}\{1, 10, 15, 50\})$; $\mathbf{R} = \sigma^2\mathbf{I}$ with $\sigma^2 = 5$; $n \in \{100, 200, 300\}$; and $1 - \alpha = 0.95$ for both pointwise and simultaneous confidence intervals.

As in Section 2.3, the model is fit by restricted maximum likelihood using a B-spline basis[29,39] with $p = 14$ basis functions for the fixed effect function and $q = 4$ basis functions for the random effect function. Further, bootstrap as described in Section 1.3 is used to construct confidence intervals $Q = 50$ replications, both with and without a bias correction. For higher accuracy, the confidence intervals for σ^2 and $\psi(t, t)$ are first constructed on natural log scale and then exponentiated. Likewise, the confidence interval for heritablity function is first constructed on logit scale and then the inverse logit transformation is applied to get the results on the original scale. The whole process of simulating data, model fitting, and constructing confidence intervals is repeated 300 times to estimate the coverage probabilities.

The estimated coverage probabilities for σ^2 are presented in Table 3.1. We observe that confidence intervals without bias correction have higher coverage than those with it except when $n = 200$. There is indication that bias correction may be leading to a wider confidence interval and hence higher coverage.

Table 3.2 presents average estimated pointwise coverage probability for three parameter functions of time: mean function $f(t)$, random effects variance function $\psi(t, t)$, and heri-

Table 3.2. Average estimated pointwise coverage probability (in %) of 95% confidence intervals for mean function $f(t)$, random effects variance function $\psi(t,t)$, and heritability function $h^2(t)$. Settings (a) and (b) respectively represent with and without bias correction.

Parameter	# Families	n	(a)	(b)
$f(t)$	25	100	92.5	92.9
	50	200	94.1	94.2
	75	300	92.8	92.7
$\psi(t,t)$	25	100	92.7	93.4
	50	200	94.1	94.3
	75	300	93.3	93.5
$h^2(t)$	25	100	92.5	93.5
	50	200	94.1	94.4
	75	300	94.1	94.4

tability function $h^2(t)$. In either case, the bias correction does not seem to matter as the probabilities are more or less similar with and without the correction. The probabilities tend to increase with n and become closer to the nominal level. A graph of estimated pointwise coverage probability for $h^2(t)$ against t is shown in Figure 3.1. On the whole, the pointwise intervals may be considered to have acceptable accuracy with $n \geq 200$.

Table 3.3 presents estimated simultaneous coverage probabilities for the three parameter functions mentioned before. We see that intervals without the bias correction tend to have slightly higher coverage probability which are closer to the nominal level than those with the correction. There is indication that $n \geq 200$ may be enough to provide acceptable accuracy.

3.4 Application to estimation of DNA methylation heritability

Although genetic changes take many generations of natural selection, according to Darwinism, epigenetic changes, on the other hand, can happen within a short period of time. Powerful environmental conditions may leave an imprint on genetic material, which can lead to passage of new traits in a single generation through the epigenetic process. Such a process is known as DNA methylation, a biological process through which methyl groups are added to the DNA molecules.

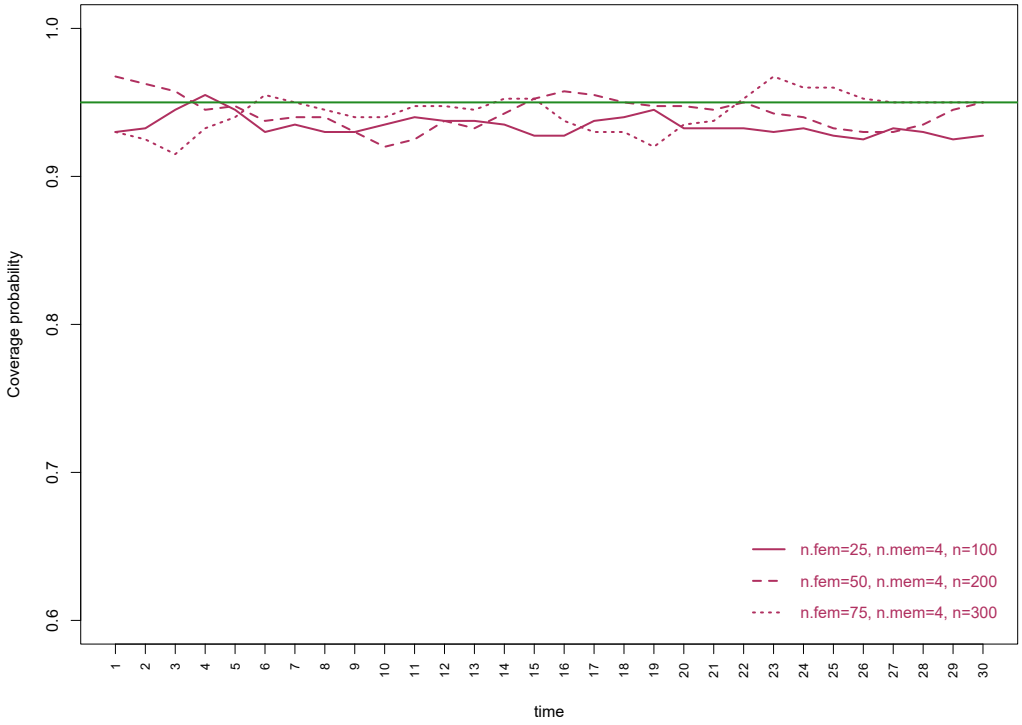

Figure 3.1. Estimated pointwise coverage probability of 95% confidence intervals for heritability function $h^2(t)$. The horizontal solid dark green line represents the nominal 0.95 confidence level.

Table 3.3. Estimated simultaneous coverage probabilities (in %) of 95% confidence intervals for mean function $f(t)$, random effects variance function $\psi(t,t)$, and heritability function $h^2(t)$. Settings (a) and (b) respectively represent with and without bias correction.

Parameter	# Families	n	(a)	(b)
$f(t)$	25	100	88.8	89.0
	50	200	93.0	93.8
	75	300	90.5	90.5
$\psi(t,t)$	25	100	88.5	90.3
	50	200	91.3	92.5
	75	300	89.5	90.8
$h^2(t)$	25	100	88.8	90.5
	50	200	92.8	94.0
	75	300	90.0	91.3

In particular, DNA methylation, the cornerstone of epigenetic changes, has been widely studied to understand its role in many biological traits and cell development, including embryonic development, cell differentiation, genomic imprinting, X-chromosomal inactivation, and autoimmunity[66–72]. In recent years, its role in marking the biological clock has also been studied[73,74], and it has been hypothesized that DNA methylation is heritable from generation to generation[75–81]. There are several types of DNA methylation, with the most common (almost exclusively) one in humans being in the context of CG dinucleotides — also referred to as CpG or 5-mC — where the cytosine (C) may be methylated. This is the context we consider here. To study methylation heritability, most of the existing methods estimate it separately at each CpG site[77,82]. However, it is known that methylation levels are correlated between sites that are separated by less than 1000 base pairs (bp)[83–85], and there is evidence that methylation patterns in a genomic region have greater association with gene expression than individual CpG sites[86–90]. As such, it would be more appropriate to understand the entire landscape of DNA methylation heritability by studying CpG regions and by taking correlation into consideration rather than studying individual sites. By doing so, one may begin to characterize DNA methylation heritability patterns within and across genomes.

As elaborated above, statistical methods for estimating heritability of quantitative traits based on mixed models are well developed. A number of approaches are available for modeling methylation data as well but they are primarily directed at identifying differentially methylated cytosines or differentially methylated regions[89–96]. For estimating methylation heritability, the most common approach is to treat the methylation level at each single site throughout the genome as a univariate trait and apply methods based on mixed models *separately* for each site[60,77,80,82,97,98]. This site-wise approach is naive in that it completely ignores the correlation in the methylation levels of nearby sites, which are known to be highly correlated[83–85,92,95,99,100]. Further, by not borrowing information across the neighboring sites within the region, it makes a suboptimal and potentially wasteful use of data. As a result, the

site-wise heritability estimates are likely to be inefficient and have high variability. This is a potentially serious problem as high variability may mask clear patterns that may be present in the true heritability profiles. These shortcomings of a site-wise approach are statistical in nature; however, its shortcoming of a biological nature poses a more serious concern. As articulated in a number of works[86–88], the working unit of DNA methylation is perhaps a region rather than a site; therefore, it would be much more biologically meaningful to profile the heritability of CpG regions rather than individual CpG sites.

The shortcomings of a site-wise approach can be corrected by taking a region-wise approach that *jointly* models the entire methylation profile of a given CpG region. We can directly apply our model (3.14) to jointly model such data. This approach involves treating the observed methylation profiles of individuals as draws from a smooth random curve (function) that are measured discretely at the sites in the region. Thus, the data on an individual are treated as one observed curve rather than a sequence of observations. The curves for different individuals are dependent due to the relatedness of the individuals in the observed sample (which could be from known familial relationship or simply representing cryptic relatedness). Each individual methylation curve is decomposed into a fixed population mean function, a random genetic effect function, and a random error function representing the environmental effect. The population mean and random effect functions are specified nonparametrically through basis expansions. This way no a priori assumptions are made regarding their shapes and they are estimated from the data. The within-individual errors are allowed to be heteroscedastic and correlated by letting their autocovariance function depend on site.

One specific dataset we are trying to obtain is Framingham Heart Study (FHS) data for application to our proposed methodology. In this family data set, along with age and gender of participants, peripheral blood cell DNA methylation levels are also recorded on 48 sites belonging to the chr6:29894140-29895117 CpG island. The data consists of 2541 individuals, 2369 of whom are known to be related. A small number of subjects have missing values on a

number of methylation sites, but that posed no problem as our method can handle missing data.

3.5 Summary and future work

In this chapter, we present a functional mixed model for modeling and analysis of family data. The proposed model generalizes various quantitative trait models from statistical genetics for scalar, vector, and longitudinal responses to the case of functional response. The computations for fitting the model have only been implemented for the case when errors are uncorrelated and homoscedastic. Once fully implemented, the model can be used to estimate heritability of functional traits. Having family data allows separation of covariance structure due to random effects and errors, resolving the non-identifiability issue. This would also open door for incorporating nonparametric models for error covariance function, enhancing the flexibility of the methodology. Additional simulation studies are underway to get a more clear conclusion regarding performance of interval estimators of model parameter and parameter functions. Further work is needed in future to make this methodology viable. We specifically plan to explore the following in the near future:

- Implement computations for fitting the model (3.14) for assuming parametric models for error covariance structure, as in Chapter 2.
- Perform a comprehensive simulation study to evaluate properties of the estimators.
- Develop a cross-validation type method for selecting the number of basis functions.
- Illustrate the methodology by estimating methylation heritability from FHS data introduced in Section 3.4.
- Incorporate penalized estimation of mean function in the methodology.

- Extend the methodology to allow nonparametric models for error covariance, test it on real and simulated data, and compare it with alternative competing methods.
- Develop an `R` package to implement the methodology and make it publicly available.

CHAPTER 4

A SINGULAR MIXED MODEL

4.1 Introduction

As we have seen in previous chapters, classical mixed models provide a flexible framework for modeling and analysis of functional data. These models, however, assume that the error covariance matrix is nonsingular. Moreover, when the error covariance structure is modeled, we primarily restrict attention to simple parametric structures, such as those considered in Chapters 2 and 3. Even though this practice limits flexibility in modeling error covariance structures, one reason why it is common is that nonparametric structures (e.g., one obtained by a functional principal components analysis) generally make the error covariance matrix singular. This would imply constraints on the model, which would raise challenges for fitting and interpreting the model.

The issue of singular error covariance matrix has been studied in the context of linear models. The book of[101]—called SJ here for brevity—provides a nice, systematic summary of the literature on this topic. However, to our knowledge, this issue has not been taken up in the literature on mixed models. This chapter represents the first step in the development of such models, which we call *singular mixed models*. Note that singularity here refers to only the error covariance matrix as we continue to assume that the random effect covariance matrix is nonsingular. Upon reviewing some basic concepts and results from matrix algebra and singular linear models in Section 4.2 that are relevant for our development, we define singular mixed models in Section 4.3. A simple example is presented in Section 4.4 to illustrate the constraints imposed by singularity. Some preliminary ideas on how such models may be fit to data, assuming known error covariance matrix up to a scaling parameter, are presented in Section 4.5. Section 4.6 lists the next steps in the development that are planned for the near future.

4.2 Preliminaries

In this section, we review some basic concepts and results from matrix algebra and singular linear models that are relevant for our development. We use the following notation. For a matrix $\mathbf{A}$, let $\mathbf{A}'$ be its transpose; $\mathcal{C}(\mathbf{A})$ be its column space; $\mathbf{P_A}$ be the projection matrix for $\mathcal{C}(\mathbf{A})$; and $\mathbf{I} - \mathbf{P_A}$ be the projection matrix for $\mathcal{C}(\mathbf{A})^{\perp}$. By projection matrix, we mean the orthogonal projection matrix. Further, $\mathbf{A}^{-}$ is a generalized inverse of $\mathbf{A}$ and $\mathbf{A}^{+}$ is its Moore-Penrose inverse. If $\mathbf{A}$ is square, $|\mathbf{A}|$ is its determinant. An identity matrix of order $n \times n$ is denoted as $\mathbf{I}_n$. It will be written as $\mathbf{I}$ if its dimension is clear from the context. Finally, $\mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a m-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

4.2.1 Some matrix algebra concepts and results

Definition 1. (SJ, p. 40) Any non-null $m \times n$ matrix of $\mathbf{A}$ of rank r can be written as $\mathbf{A} = \mathbf{BC}$, where $\mathbf{B}$ is a $m \times r$ matrix of full column rank and $\mathbf{C}$ is a $r \times n$ matrix of full row rank. This is called *rank factorization* of $\mathbf{A}$.

Proposition 4.1. [102] *p. 72-73, 120-121 Consider a linear system* $\mathbf{Ax} = \mathbf{b}$*, where* $\mathbf{A}$ *is a* $m \times n$ *coefficient matrix,* $\mathbf{x}$ *is a* $n \times 1$ *vector of unknowns, and* $\mathbf{b}$ *is a* $m \times 1$ *vector. Each of the following conditions is necessary and sufficient for the system to be consistent.*

(a) $\mathbf{b} \in \mathcal{C}(\mathbf{A})$.

(b) $\mathcal{C}(\mathbf{A} : \mathbf{b}) = \mathcal{C}(\mathbf{A})$.

(c) $rank(\mathbf{A} : \mathbf{b}) = rank(\mathbf{A})$.

(d) $\mathbf{AA}^{-}\mathbf{b} = \mathbf{b}$*, or equivalently,* $(\mathbf{I} - \mathbf{AA}^{-})\mathbf{b} = \mathbf{0}$.

Proposition 4.2. *(SJ, p. 38-43) Let* $\mathbf{V}$ *be a* $n \times n$ *symmetric, non-negative definite matrix of rank* $r \leq n$*. Suppose* $\mathbf{V} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}'$ *is a spectral decomposition of* $\mathbf{V}$*, where* $\mathbf{\Lambda}$ *is a diagonal matrix with the eigenvalues of* $\mathbf{V}$ *in nonincreasing order as the diagonal elements and* $\mathbf{P}$ *is a* $n \times n$ *orthogonal matrix with columns as the corresponding orthonormal eigenvectors of* $\mathbf{V}$*. Let* $\mathbf{P}_1$ *and* $\mathbf{P}_2$ *be* $n \times r$ *and* $n \times (n-r)$ *matrices respectively containing the first* r *and the last* $n-r$ *columns of* $\mathbf{P}$*. Also, let* $\mathbf{\Lambda}_1$ *be a diagonal matrix with the first* r *eigenvalues of* $\mathbf{V}$ *(which are positive) as the diagonal elements.*

(a) $\mathbf{V} = \mathbf{P}_1\mathbf{\Lambda}_1\mathbf{P}_1'$.

(b) $\mathbf{V}^+ = \mathbf{P}_1\mathbf{\Lambda}_1^{-1}\mathbf{P}_1'$.

(c) $\mathbf{V} = \mathbf{C}\mathbf{C}'$*, where the* $n \times r$ *matrix* $\mathbf{C} = \mathbf{P}_1\mathbf{\Lambda}_1^{1/2}$ *has* r *orthogonal columns and full column rank, provides a rank factorization of* $\mathbf{V}$*. Moreover,* $|\mathbf{C}'\mathbf{C}|$ *equals the product of* r *positive eigenvalues of* $\mathbf{V}$*.*

(d) $\mathbf{P}_\mathbf{V} = \mathbf{P}_1\mathbf{P}_1'$*, providing a rank factorization of* $\mathbf{P}_\mathbf{V}$.

(e) $\mathbf{I} - \mathbf{P}_\mathbf{V} = \mathbf{P}_2\mathbf{P}_2'$*, providing a rank factorization of* $\mathbf{I} - \mathbf{P}_\mathbf{V}$.

4.2.2 Some distribution theory

Proposition 4.3. *(SJ, ch. 3) Suppose* $\mathbf{Y}$ *is a* $n \times 1$ *random vector with finite mean* $\boldsymbol{\mu}$ *and covariance matrix* $\mathbf{V}$*, which may be singular or nonsingular.*

(a) $\mathbf{Y} - \boldsymbol{\mu} \in \mathcal{C}(\mathbf{V})$ *with probability 1.*

(b) $(\mathbf{I} - \mathbf{P}_\mathbf{V})(\mathbf{Y} - \boldsymbol{\mu}) = \mathbf{0}$ *with probability 1.*

(c) Suppose $\mathbf{V}$ *is singular of rank* $r \leq n$*. Let* $\mathbf{A}\mathbf{A}'$ *be a rank factorization of* $\mathbf{I} - \mathbf{P}_\mathbf{V}$*, i.e.,* $\mathbf{I} - \mathbf{P}_\mathbf{V} = \mathbf{A}\mathbf{A}'$*, where* $\mathbf{A}$ *is a* $n \times (n-r)$ *matrix with full column rank. Then,* $\mathbf{A}'(\mathbf{Y} - \boldsymbol{\mu}) = \mathbf{0}$ *with probability 1.*

When $\mathbf{V}$ is singular of rank r, the columns of $\mathbf{V}$ satisfy $n-r$ linearly independent constraints. This in turn implies a deterministic relationship that $n-r$ linear combinations of the random vector $\mathbf{Y}-\boldsymbol{\mu}$ are zero with probability 1. Part (c) above identifies such linear combinations and also gives the linear deterministic relationship as $\mathbf{A}'\mathbf{Y}=\mathbf{A}'\boldsymbol{\mu}$ with probability 1. Further, from Proposition 4.2, we may take $\mathbf{A}=\mathbf{P}_2$—a matrix whose columns are eigenvectors of $\mathbf{V}$ associated with its zero eigenvalues—in the above result.

Proposition 4.4. *(SJ, p. 56) If $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are two random vectors with finite mean and covariance matrices, $\mathcal{C}(cov(\mathbf{Y}_1,\mathbf{Y}_2))\subseteq\mathcal{C}(var(\mathbf{Y}_1))$.*

Next, we define a multivariate normal distribution that covers both singular and nonsingular covariance cases and present some characterizations and properties.

Definition 2. [103] p. 436 A $n\times 1$ random vector $\mathbf{Y}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{V}$ follows a $\mathcal{N}_n(\boldsymbol{\mu},\mathbf{V})$ distribution if $\mathbf{Y}$ is equal in distribution to $\mathbf{CZ}+\boldsymbol{\mu}$, where $\mathbf{C}$ is any $n\times r$ matrix so that $\mathbf{CC}'$ is a rank factorization of $\mathbf{V}$, and $\mathbf{Z}\sim\mathcal{N}_r(\mathbf{0},\mathbf{I})$.

The normal distribution is invariant to the choice of $\mathbf{C}$. An alternative equivalent definition is as follows:

Definition 3. [103] p. 436 The random vector $\mathbf{Y}$ is multivariate normal if and only if $\mathbf{a}'\mathbf{Y}$ is univariate normal for all $\mathbf{a}$. If $Y=\mu$ with probability one, we define Y to be $\mathcal{N}_1(\mu,0)$.

Proposition 4.5. *(SJ, p. 58) Suppose $\mathbf{Y}\sim\mathcal{N}_n(\boldsymbol{\mu},\mathbf{V})$, where $\mathbf{V}$ may be singular or nonsingular and has rank $r\le n$. Let $\mathbf{V}^-$ be a generalized inverse of $\mathbf{V}$ and $\mathbf{CC}'$ be a rank factorization of $\mathbf{V}$. The probability density function of $\mathbf{Y}$ is*

$$f(\mathbf{y})=\begin{cases}(2\pi)^{-n/2}|\mathbf{C}'\mathbf{C}|^{-1/2}\exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'\mathbf{V}^-(\mathbf{y}-\boldsymbol{\mu})\right), & \text{if } (\mathbf{I}-\mathbf{P}_{\mathbf{V}})(\mathbf{y}-\boldsymbol{\mu})=\mathbf{0},\\ 0, & \text{otherwise.}\end{cases}$$

This density is invariant to the choice of $\mathbf{V}^-$ and $\mathbf{C}$. Moreover, when $\mathbf{V}$ is nonsingular, $r = n$, $|\mathbf{C}'\mathbf{C}| = |\mathbf{V}|$, $\mathbf{V}^- = \mathbf{V}^{-1}$, $\mathbf{P}_\mathbf{V} = \mathbf{I}$, and $(\mathbf{I}-\mathbf{P}_\mathbf{V})(\mathbf{y}-\boldsymbol{\mu}) = \mathbf{0}$ holds for all $\mathbf{y} \in \mathbb{R}^n$. When $\mathbf{V}$ is singular with rank r, the probability density function exists only in a r-dimensional subspace of $\mathbb{R}^n$.

Proposition 4.6. [103] *p. 436-437 Let* $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \mathbf{V})$*, where the distribution may be singular or nonsingular.*

(a) The moment generating function of $\mathbf{Y}$ *is*

$$M_\mathbf{Y}(\mathbf{t}) = \exp\left(\mathbf{t}'\boldsymbol{\mu} + \mathbf{t}'\mathbf{V}\mathbf{t}/2\right), \quad \mathbf{t} \in \mathbb{R}^n.$$

(b) If $\mathbf{A}$ *is* $m \times n$*, then* $\mathbf{AY} \sim \mathcal{N}_m(\mathbf{A}\boldsymbol{\mu}, \mathbf{AVA}')$*. The distribution is nonsingular if* $\mathbf{V}$ *is nonsingular and* $\mathbf{A}$ *has rank* m*.*

(c) Any subset of a multivariate normal distribution is multivariate normal.

(d) If the covariance of any two vectors that contain disjoint subsets of $\mathbf{Y}$ *is zero, then the two vectors are statistically independent.*

(e) If $cov(\mathbf{AY}, \mathbf{BY}) = \mathbf{0}$*, then* $\mathbf{AY}$ *and* $\mathbf{BY}$ *are statistically independent.*

(f) Suppose $\mathbf{V}$ *is nonsingular, and let*

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}, \ \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \ \mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix},$$

where $\mathbf{Y}_i$ *and* $\boldsymbol{\mu}_i$ *are* $n_i \times 1$*,* $\mathbf{V}_{ii}$ *is* $n_i \times n_i$ *(*$i = 1, 2$*), and* $n_1 + n_2 = n$*. We then have the following conditional distribution:*

$$\mathbf{Y}_2|\mathbf{Y}_1 \sim \mathcal{N}_{n_2}(\boldsymbol{\mu}_{2.1}, \mathbf{V}_{22.1}),$$

where $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \mathbf{V}_{21}\mathbf{V}_{11}^{-1}(\mathbf{Y}_1 - \boldsymbol{\mu}_1)$ *and* $\mathbf{V}_{22.1} = \mathbf{V}_{22} - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12}$*.*

(g) The result (f) still holds if $\mathbf{V}$ *is singular upon replacing* $\mathbf{V}_{11}^{-1}$ *with* $\mathbf{V}_{11}^-$*, any generalized inverse of* $\mathbf{V}_{11}$*.*

4.2.3 Singular linear model

Our development in this section is primarily based on[101]. Consider a linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{Y}$ is a $n \times 1$ response vector; $\mathbf{X}$ is a $n \times p$ design matrix which may not have full column rank; $\boldsymbol{\beta}$ is a $p \times 1$ unknown coefficient vector; and $\boldsymbol{\epsilon}$ is a $n \times 1$ random error vector with $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\text{var}(\boldsymbol{\epsilon}) = \sigma^2\mathbf{R}$. Here $\mathbf{R}$ is assumed to be known. For $\mathbf{Y}$, it follows that $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$ and $\text{var}(\mathbf{Y}) = \sigma^2\mathbf{R}$. In a *standard linear model,* $\mathbf{R} = \mathbf{I}$, whereas in a *general linear model*, $\mathbf{R}$ is nonsingular (full rank). However, in a *singular linear model*, $\mathbf{R}$ is singular (rank-deficient).

Singularity of $\mathbf{R}$ implies that certain linear combinations of the error vector are zero with probability 1. This in turn implies that certain linear combinations of response $\mathbf{Y}$ are constant with probability 1. Thus, singularity of $\mathbf{R}$ implies a partly *deterministic* structure on $\mathbf{Y}$, which has to be taken into account while performing statistical inference. To be specific, the constraint on $\mathbf{Y}$ from Proposition 4.3 is: $(\mathbf{I} - \mathbf{P_R})\mathbf{Y} = (\mathbf{I} - \mathbf{P_R})\mathbf{X}\boldsymbol{\beta}$ with probability 1. When there exists a $\boldsymbol{\beta}$ such that this condition holds, the data are said to be *consistent* with the linear model. This consistency is related to consistency of a linear system considered in Proposition 4.1. Other equivalent forms of this consistency condition that do not explicitly require a $\boldsymbol{\beta}$ are given by the following result.

Proposition 4.7. *(SJ, p. 246) For a singular linear model, the following conditions are equivalent.*

(a) $\mathbf{Y} \in \mathcal{C}(\mathbf{X} : \mathbf{R})$ *with probability 1.*

(b) $(\mathbf{I} - \mathbf{P_R})\mathbf{Y} \in \mathcal{C}((\mathbf{I} - \mathbf{P_R})\mathbf{X})$ *with probability 1.*

(c) $(\mathbf{I} - \mathbf{P_X})\mathbf{Y} \in \mathcal{C}((\mathbf{I} - \mathbf{P_X})\mathbf{R})$ *with probability 1.*

If $\mathbf{R}$ is nonsingular, these conditions are automatically satisfied. However, this may not necessarily be the case if $\mathbf{R}$ is singular. If this happens, we say that the data are *inconsistent*

with the linear model. In practice, it is recommended to verify that one of these conditions is met before proceeding with inference.

Definition 4. (SJ, p. 250-251) A linear parametric function $\mathbf{p}'\boldsymbol{\beta}$ is called *estimable* if there is a linear estimator $\mathbf{l}'\mathbf{Y}$ of $\mathbf{p}'\boldsymbol{\beta}$ such that $E(\mathbf{l}'\mathbf{Y}) = \mathbf{p}'\boldsymbol{\beta}$ for *all* $\boldsymbol{\beta}$ (not just those that are possible under the model). A linear parametric function is *identifiable* if and only if it is estimable.

Proposition 4.8. *(SJ, p. 250) A necessary and sufficient condition for estimability of a linear parametric function* $\mathbf{p}'\boldsymbol{\beta}$ *is* $\mathbf{p} \in \mathcal{C}(\mathbf{X}')$.

Proposition 4.9. *(SJ, p. 252) Consider the vector* $\mathbf{L}\mathbf{X}\boldsymbol{\beta}$ *of linear parametric functions which are estimable under the model. Its best linear unbiased estimator (BLUE) is given by*

$$\widehat{\mathbf{L}\mathbf{X}\boldsymbol{\beta}} = \mathbf{L}\left[\mathbf{I} - \mathbf{R}(\mathbf{I} - \mathbf{P}_\mathbf{X})\{(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{R}(\mathbf{I} - \mathbf{P}_\mathbf{X})\}^-(\mathbf{I} - \mathbf{P}_\mathbf{X})\right]\mathbf{Y}.$$

This BLUE is unique.

The BLUE does not depend on the choice of the generalized inverse. Substituting $\mathbf{L} = \mathbf{I}$ in this result gives the vector of fitted values as (SJ, p. 253)

$$\hat{\mathbf{Y}} = \left[\mathbf{I} - \mathbf{R}(\mathbf{I} - \mathbf{P}_\mathbf{X})\{(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{R}(\mathbf{I} - \mathbf{P}_\mathbf{X})\}^-(\mathbf{I} - \mathbf{P}_\mathbf{X})\right]\mathbf{Y}. \tag{4.1}$$

Further, since any estimable linear parametric function $\mathbf{A}\boldsymbol{\beta}$ can be written as $\mathbf{A}\mathbf{X}^-\mathbf{X}\boldsymbol{\beta}$, it follows from Proposition 4.9 that its BLUE is $\widehat{\mathbf{A}\boldsymbol{\beta}} = \mathbf{A}\mathbf{X}^-\hat{\mathbf{Y}}$. This motivates the following as an estimator of $\boldsymbol{\beta}$ (SJ, p. 253):

$$\hat{\boldsymbol{\beta}} = \mathbf{X}^-\hat{\mathbf{Y}}. \tag{4.2}$$

This estimator is not unique as it depends on the choice of $\mathbf{X}^-$. However, it is unique if $\mathbf{X}$ has full column rank, in which case it is the BLUE of $\boldsymbol{\beta}$. The vector $\mathbf{e}$ of residuals is $\mathbf{Y} - \hat{\mathbf{Y}}$, which from (4.1) can be written as

$$\mathbf{e} = \mathbf{R}(\mathbf{I} - \mathbf{P}_\mathbf{X})\{(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{R}(\mathbf{I} - \mathbf{P}_\mathbf{X})\}^-(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{Y}. \tag{4.3}$$

Finally, an estimator of σ^2 (SJ, p. 260) is

$$\hat{\sigma}^2 = \frac{\mathbf{e}'\mathbf{R}^-\mathbf{e}}{\rho(\mathbf{R}:\mathbf{X}) - \rho(\mathbf{X})}, \tag{4.4}$$

where $\rho(\mathbf{A})$ is rank of a matrix $\mathbf{A}$. The estimator $\hat{\sigma}^2$ does not depend on the choice of $\mathbf{R}^-$.

4.3 Singular mixed model

Consider a linear mixed model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}$, where $\mathbf{Y}$ is a $m \times 1$ response vector; $\mathbf{X}$ is a $m \times p$ fixed effect design matrix; $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed coefficients; $\mathbf{Z}$ is a $m \times q$ random effect design matrix; $\mathbf{b}$ is a $q \times 1$ vector of random coefficients distributed as $\mathcal{N}_q(\mathbf{0}, \mathbf{G})$; and $\boldsymbol{\epsilon}$ is a $m \times 1$ random error vector with distributed as $\mathcal{N}_m(\mathbf{0}, \sigma^2\mathbf{R})$, independently of $\mathbf{b}$. Here we assume that $\mathbf{G}$ is nonsingular and $\mathbf{R}$ may be singular. We will refer to this model as the *singular mixed model*. If $\mathbf{R}$ is nonsingular, the model reduces to the classical mixed model which has been considered in the previous chapters.

Let $\mathbf{V} = \text{var}(\mathbf{Y}) = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \sigma^2\mathbf{R}$ denote the marginal covariance matrix of $\mathbf{Y}$. In a mixed model, $\mathbf{Z}$ generally does not have full row rank, making $\mathbf{Z}\mathbf{G}\mathbf{Z}'$ rank-deficient. If, in addition, $\mathbf{R}$ is singular, it may make $\mathbf{V}$ singular. The next result provides an analog of the distributional results for the nonsingular case.

Proposition 4.10. *For a singular mixed model, we have the following distributions:*

(a)

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{b} \end{pmatrix} \sim \mathcal{N}_{m+q}\left(\begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{V} & \mathbf{Z}\mathbf{G} \\ \mathbf{G}\mathbf{Z}' & \mathbf{G} \end{pmatrix} \right).$$

(b) $\mathbf{Y} \sim \mathcal{N}_m(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$. *Its probability density function* $f(\mathbf{y})$ *is given as*

$$\begin{cases} (2\pi)^{-r/2}|\mathbf{C}'\mathbf{C}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^-(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right), & \text{if } (\mathbf{I} - \mathbf{P}_\mathbf{V})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0}, \\ 0, & \text{otherwise}, \end{cases}$$

where r *is the rank of* $\mathbf{V}$ *and* $\mathbf{C}\mathbf{C}'$ *is a rank factorization of* $\mathbf{V}$.

(c) $\mathbf{b}|\mathbf{Y} \sim \mathcal{N}_q\left(E(\mathbf{b}|\mathbf{Y}), var(\mathbf{b}|\mathbf{Y})\right)$*, where*

$$E(\mathbf{b}|\mathbf{Y}) = \mathbf{G}\mathbf{Z}'\mathbf{V}^{-}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}),\ var(\mathbf{b}|\mathbf{Y}) = \mathbf{G} - \mathbf{G}\mathbf{Z}'\mathbf{V}^{-}\mathbf{Z}\mathbf{G}.$$

Proof. Let $\mathbf{t}_1 \in \mathbb{R}^m$, $\mathbf{t}_2 \in \mathbb{R}^q$ and $\mathbf{t}' = (\mathbf{t}_1', \mathbf{t}_2')$. The moment generating function of $(\mathbf{Y}, \mathbf{b})$ can be written as

$$M(\mathbf{t}) = E\{\exp(\mathbf{t}_1'\mathbf{Y} + \mathbf{t}_2'\mathbf{b})\} = \exp(\mathbf{t}_1'\mathbf{X}\boldsymbol{\beta})E\{\exp(\mathbf{t}_1'\boldsymbol{\epsilon})\}E\{\exp((\mathbf{Z}'\mathbf{t}_1 + \mathbf{t}_2)'\mathbf{b})\},$$

where the last equality follows from independence of $\boldsymbol{\epsilon}$ and $\mathbf{b}$. Upon substituting the moment generating functions of $\boldsymbol{\epsilon}$ and $\mathbf{b}$ from part (a) of Proposition 4.6 and simplifying, we get

$$M(\mathbf{t}) = \exp\left\{\mathbf{t}_1'\mathbf{X}\boldsymbol{\beta} + \frac{1}{2}\left(\mathbf{t}_1'\mathbf{V}\mathbf{t}_1 + \mathbf{t}_1'\mathbf{Z}\mathbf{G}\mathbf{t}_2 + \mathbf{t}_2'\mathbf{G}\mathbf{Z}'\mathbf{t}_1 + \mathbf{t}_2'\mathbf{G}\mathbf{t}_2\right)\right\}.$$

Now (a) follows by noting that $M(\mathbf{t})$ is the moment generating function of the multivariate distribution given in the result. The first part of (b) follows from (a) upon applying part (c) of Proposition 4.6. The density follows from Proposition 4.5. The result in (c) also follows from (a) upon applying part (g) of Proposition 4.6. □

As in Section 4.2.3, singularity of $\mathbf{R}$ implies a partly deterministic structure on the error $\boldsymbol{\epsilon}$, which in turn implies deterministic structures on $(\mathbf{Y}, \mathbf{b})$ within the context of their joint distribution and on $\mathbf{Y}$ within the context of its marginal distribution. The next result provides these constraints.

Proposition 4.11. *For a singular mixed model, the following conditions are equivalent.*

(a) $\mathbf{R}$ *is singular.*

(b) $(\mathbf{I} - \mathbf{P}_{\mathbf{R}})(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{b}) = \mathbf{0}$ *with probability 1 with respect to* $(\mathbf{Y}, \mathbf{b})$.

(c) $(\mathbf{I} - \mathbf{P}_{\mathbf{R}})(\mathbf{I} - \mathbf{V}\mathbf{V}^{-})(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0}$ *with probability 1 with respect to* $\mathbf{Y}$.

Proof. Suppose $\mathbf{R}$ is singular. From Proposition 4.3, this means $(\mathbf{I}-\mathbf{P_R})\boldsymbol{\epsilon}=\mathbf{0}$ with probability 1. The result in (b) follows since $\boldsymbol{\epsilon}=\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Zb}$. To get (c) from (b), fix any $\mathbf{Y}$. Then, given this $\mathbf{Y}$, the conditional probability

$$P_{\mathbf{b}|\mathbf{Y}}((\mathbf{I}-\mathbf{P_R})\mathbf{Zb}=(\mathbf{I}-\mathbf{P_R})(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})|\mathbf{Y})=1,$$

implying that $(\mathbf{I}-\mathbf{P_R})\mathbf{Zb}$ equals the constant $(\mathbf{I}-\mathbf{P_R})(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})$ with probability 1 with respect to $\mathbf{b}|\mathbf{Y}$. If a random quantity is constant with probability one, the constant must be its expectation. Thus, for the given $\mathbf{Y}$, we have

$$(\mathbf{I}-\mathbf{P_R})(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})=(\mathbf{I}-\mathbf{P_R})\mathbf{Z}E(\mathbf{b}|\mathbf{Y})=(\mathbf{I}-\mathbf{P_R})\mathbf{V}\mathbf{V}^{-}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}), \tag{4.5}$$

where the last equality follows from substituting the expression for $E(\mathbf{b}|\mathbf{Y})$ from Proposition 4.10 and using $(\mathbf{I}-\mathbf{P_R})\mathbf{P_R}=\mathbf{0}$. This gives (c) since the fixed $\mathbf{Y}$ is arbitrary.

Next, we need to get (a) from (c). Suppose (c) holds. Then, reversing the arguments, we have from (4.5) that for any fixed $\mathbf{Y}$, $E((\mathbf{I}-\mathbf{P_R})\mathbf{Zb}|\mathbf{Y})=(\mathbf{I}-\mathbf{P_R})(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})$. Further, using the expression for $\text{var}(\mathbf{b}|\mathbf{Y})$ from Proposition 4.10 and applying the facts that $\mathbf{V}\mathbf{V}^{-}\mathbf{V}=\mathbf{V}$ and $(\mathbf{I}-\mathbf{P_R})\mathbf{P_R}=\mathbf{0}$, it can be seen that $\text{var}((\mathbf{I}-\mathbf{P_R})\mathbf{Zb}|\mathbf{Y})=\mathbf{0}$. Thus, it follows that, for any fixed $\mathbf{Y}$, $(\mathbf{I}-\mathbf{P_R})\mathbf{Zb}$ equals the constant $(\mathbf{I}-\mathbf{P_R})(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})$ with probability 1 with respect to $\mathbf{b}|\mathbf{Y}$, i.e., (4.5) holds. Since $\mathbf{Y}$ is arbitrary, this implies (b), which in turn implies that $(\mathbf{I}-\mathbf{P_R})\boldsymbol{\epsilon}=\mathbf{0}$ with probability 1 with respect to $\boldsymbol{\epsilon}$, giving (a). □

When $\mathbf{R}$ is singular, it implies the constraint $(\mathbf{I}-\mathbf{P_R})\boldsymbol{\epsilon}=\mathbf{0}$ on the marginal distribution of $\boldsymbol{\epsilon}$. The results (b) and (c) above imply the corresponding constraints on the joint distribution of $(\mathbf{Y},\mathbf{b})$ and the marginal distribution of $\mathbf{Y}$. Thus, from this result, $\mathbf{R}$ is singular if and only if $(\mathbf{I}-\mathbf{P_R})(\mathbf{I}-\mathbf{V}\mathbf{V}^{-})(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})=\mathbf{0}$ with probability 1 with respect to $\mathbf{Y}$. If here we use the Moore-Penrose inverse $\mathbf{V}^{+}$ in place of a generalized inverse $\mathbf{V}^{-}$, then $\mathbf{P_V}=\mathbf{V}\mathbf{V}^{+}$. Upon applying Proposition 4.12 below, the constraint becomes

$$(\mathbf{I}-\mathbf{P_V})(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})=\mathbf{0} \text{ with probability 1 with respect to } \mathbf{Y}, \tag{4.6}$$

which is same as the condition needed in Proposition 4.10 for the probability density function of $\mathbf{Y}$ to be non-zero. This is same as saying $\mathbf{Y} \in \mathcal{C}(\mathbf{X} : \mathbf{V})$ with probability 1, which is the condition needed for consistency of data with the model (see also Proposition 4.7).

Proposition 4.12. *Under the singular mixed model,* $(\mathbf{I} - \mathbf{P_R})(\mathbf{I} - \mathbf{P_V}) = (\mathbf{I} - \mathbf{P_V})$.

Proof. Under the model, we have $\text{cov}(\mathbf{Y}, \boldsymbol{\epsilon}) = \sigma^2\mathbf{R}$. From Proposition 4.4, this implies $\mathcal{C}(\mathbf{R}) \subseteq \mathcal{C}(\mathbf{V})$. Next, take any any $\mathbf{x} \in \mathcal{C}(\mathbf{V})^\perp = \mathcal{C}(\mathbf{I} - \mathbf{P_V})$. Since this $\mathbf{x}$ is also orthogonal to $\mathcal{C}(\mathbf{R})$, we have $\mathbf{P_R x} = \mathbf{0}$. Thus, it follows that $\mathbf{P_R}(\mathbf{I} - \mathbf{P_V}) = \mathbf{0}$, which implies the result. □

For the example in Section 4.4, we also have $(\mathbf{I} - \mathbf{P_R})(\mathbf{I} - \mathbf{VV}^-) = \mathbf{I} - \mathbf{P_V}$, although it is not clear if this holds in general for any generalized inverse $\mathbf{V}^-$.

4.4 A simple illustration

For illustration of constraints imposed by singularity of $\mathbf{R}$, consider a simple example with $m = 4$. Take $\sigma^2 = 1$ and

$$\mathbf{R} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The rank of $\mathbf{R}$ is 2. The singularity of $\mathbf{R}$ implies that ϵ_1 and ϵ_2 have correlation 1 and $\text{var}(\epsilon_4) = 0$. Since the errors have mean zero, this constraint becomes $\epsilon_1 = \epsilon_2$ and $\epsilon_4 = 0$. It can be seen that

$$\mathbf{I} - \mathbf{P_R} = \frac{1}{2}\begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \ (\mathbf{I} - \mathbf{P_R})\boldsymbol{\epsilon} = \frac{1}{2}\begin{pmatrix} \epsilon_1 - \epsilon_2 \\ -\epsilon_2 + \epsilon_1 \\ 0 \\ \epsilon_4 \end{pmatrix}.$$

Thus, the constraint on errors from Proposition 4.3 that $(\mathbf{I}-\mathbf{P_R})\boldsymbol{\epsilon}=\mathbf{0}$ becomes $\epsilon_1=\epsilon_2$ and $\epsilon_4=0$. This is identical to what we get above from a direct examination of $\mathbf{R}$. Next, take

$$\mathbf{X}=\begin{pmatrix}1&0\\0&1\\1&0\\1&0\end{pmatrix},\ \mathbf{Z}=\begin{pmatrix}1&0\\1&0\\0&1\\0&1\end{pmatrix},$$

yielding

$$\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{b}=\begin{pmatrix}\beta_1+b_1\\\beta_2+b_1\\\beta_1+b_2\\\beta_1+b_2\end{pmatrix},\ (\mathbf{I}-\mathbf{P_R})\mathbf{Y}=\frac{1}{2}\begin{pmatrix}Y_1-Y_2\\-Y_2+Y_1\\0\\Y_4\end{pmatrix},\ (\mathbf{I}-\mathbf{P_R})(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{b})=\frac{1}{2}\begin{pmatrix}\beta_1-\beta_2\\-\beta_2+\beta_1\\0\\\beta_1+b_2\end{pmatrix}.$$

Therefore, the constraint on $(\mathbf{Y},\mathbf{b})$ from Proposition 4.11 that $(\mathbf{I}-\mathbf{P_R})(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{b})=\mathbf{0}$ becomes $Y_1-Y_2=\beta_1-\beta_2$ and $Y_4=\beta_1+b_2$. We may also get this constraint by directly substituting $\epsilon_1=\epsilon_2$ and $\epsilon_4=0$ in the model equation $\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{b}+\boldsymbol{\epsilon}$.

Next, to get the constraint on the marginal distribution of $\mathbf{Y}$, take $\mathbf{G}=\mathbf{I}$, leading to

$$\mathbf{V}=\begin{pmatrix}2&2&0&0\\2&2&0&0\\0&0&2&1\\0&0&1&1\end{pmatrix},\ \mathbf{I}-\mathbf{P_V}=\frac{1}{2}\begin{pmatrix}1&-1&0&0\\-1&1&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix}$$

and

$$(\mathbf{I}-\mathbf{P_V})(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})=\frac{1}{2}\begin{pmatrix}(Y_1-\beta_1)-(Y_2-\beta_2)\\-(Y_1-\beta_1)+(Y_2-\beta_2)\\0\\0\end{pmatrix}.$$

The rank of $\mathbf{V}$ is 3, one more than that of $\mathbf{R}$, and it is rank deficient by one. The rank deficiency corresponds to a correlation of one between Y_1 and Y_2. We get the same constraint from (4.6) as $(\mathbf{I}-\mathbf{P_V})(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})=\mathbf{0}$ becomes $Y_1 = Y_2 + (\beta_1 - \beta_2)$. Further, by taking

$$\mathbf{V}^- = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix},$$

as a generalized inverse of $\mathbf{V}$, we see that $(\mathbf{I}-\mathbf{P_R})(\mathbf{I}-\mathbf{V}\mathbf{V}^-) = (\mathbf{I}-\mathbf{P_V})$. Therefore, the constraint from Proposition 4.11 that $(\mathbf{I}-\mathbf{P_R})(\mathbf{I}-\mathbf{V}\mathbf{V}^-)(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})=\mathbf{0}$ is also $Y_1 = Y_2 + (\beta_1 - \beta_2)$. It may be noted that $\mathbf{P_R}(\mathbf{I}-\mathbf{P_V})=\mathbf{0}$ as given by Proposition 4.12 holds here but $\mathbf{P_R}(\mathbf{I}-\mathbf{V}\mathbf{V}^-)\neq\mathbf{0}$.

4.5 Fitting a singular mixed model

Suppose there are n subjects in the data and subject i provides a $m\times 1$ response vector $\mathbf{Y}_i$, $i=1,\ldots,n$. We assume that each of these response vectors follows the singular mixed model from Section 4.3 and the responses from different subjects are independent. Thus, the model for all the data together is

$$\mathbf{Y}_i = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}_i + \boldsymbol{\epsilon}_i,\ \mathbf{b}_i \sim \mathcal{N}_q(\mathbf{0},\mathbf{G}),\ \boldsymbol{\epsilon}_i \sim \mathcal{N}_m(\mathbf{0},\sigma^2\mathbf{R}), \tag{4.7}$$

where $\mathbf{b}_i$ and $\boldsymbol{\epsilon}_i$ are independent for $i=1,\ldots,n$ and the two vectors are also mutually independent. Although by not letting $\mathbf{X}$ and $\mathbf{Z}$ matrices depend on i, we are assuming a balanced design for data, but this assumption can be relaxed. Further, as in singular linear models, we assume that $\mathbf{R}$ is known. Thus, the unknowns in this model are $\boldsymbol{\beta}$, $\mathbf{G}$, and σ^2.

The singularity of $\mathbf{R}$ imposes the constraint (4.6) for each i, i.e.,

$$(\mathbf{I}-\mathbf{P_V})(\mathbf{Y}_i-\mathbf{X}\boldsymbol{\beta})=\mathbf{0} \text{ with probability 1 with respect to } \mathbf{Y}_i,\ i=1,\ldots,n.$$

Thus, altogether there are n constraints related to singularity. Assuming that all the response vectors are individually consistent with the model, i.e., $\mathbf{Y}_i \in \mathcal{C}(\mathbf{X} : \mathbf{V})$ with probability 1 for $i = 1, \ldots, n$, fitting this model involves finding a common $\boldsymbol{\beta}$ that satisfies all the constraints. But such a $\boldsymbol{\beta}$ usually does not exist, limiting the practical utility of the model. However, we can make progress by taking an approximate, two-step approach.

To describe the first step, let $\mathbf{C} = (\mathbf{X} : \mathbf{Z})$ be the combined $m \times (p + q)$ design matrix and $\boldsymbol{\beta}_i = (\boldsymbol{\beta} : \mathbf{b}_i)^T$. The model for subject i becomes $\mathbf{Y}_i = \mathbf{C}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i$, $\boldsymbol{\epsilon} \sim \mathcal{N}_m(\mathbf{0}, \sigma^2\mathbf{R})$, which is treated as a singular linear model from Section 4.2.3 with $\mathbf{X} = \mathbf{C}$. Assuming individual consistency between $\mathbf{Y}_i$ and the model, we fit this model *separately* for each subject. This yields estimates $\hat{\boldsymbol{\beta}}_i$ for $\boldsymbol{\beta}_i$ and $\hat{\sigma}_i^2$ for σ^2 and the fitted response vector as $\hat{\mathbf{Y}}_i$ using (4.2), (4.4), and (4.1), respectively. To get a common estimate $\hat{\sigma}^2$ of σ^2, we can simply pool the individual estimates in the usual manner.

In the second step, we separate the common fixed effect from the random subject effects. For this, we take $\hat{\mathbf{Y}}_i$, $i = 1, \ldots, n$ obtained in the first step as the observations and perform a functional principal components analysis[4,5]. This eventually amounts to fitting a mixed effects model $\hat{\mathbf{Y}}_i \approx \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}^*\mathbf{b}_i^* + \boldsymbol{\delta}_i$, where $\mathbf{Z}^*$ is a new $m \times K$ design matrix obtained via eigenanalysis, with K as the number of functional principal components selected; $\mathbf{b}_i^*$ is the associated $K \times 1$ random coefficient vector with mean zero and diagonal variance matrix $\mathbf{G}^*$; and $\boldsymbol{\delta}_i$ is the error term. It yields the estimates $\hat{\boldsymbol{\beta}}$, $\hat{\mathbf{G}}^*$, and $\hat{\mathbf{b}}_i^*$ of $\boldsymbol{\beta}$, $\mathbf{G}^*$, and $\mathbf{b}_i^*$, respectively, and the error $\boldsymbol{\delta}_i$ is ignored. Thus, the final fitted model on which all subsequent inference is based is

$$\mathbf{Y}_i = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}^*\mathbf{b}_i^* + \boldsymbol{\epsilon}_i, \quad \mathbf{b}_i^* \sim \mathcal{N}_K(\mathbf{0}, \hat{\mathbf{G}}^*), \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}_m(\mathbf{0}, \hat{\sigma}^2\mathbf{R}).$$

Under this model, $\hat{\boldsymbol{V}}^* = \mathbf{Z}^*\hat{\mathbf{G}}^*\mathbf{Z}^{*\prime} + \hat{\sigma}^2\mathbf{R}$. However, due to the approximation resulting from selecting a presumably small number of principal components and ignoring $\boldsymbol{\delta}_i$, $(\mathbf{I} - \mathbf{P}_{\hat{\mathbf{V}}^*})(\mathbf{Y}_i - \mathbf{X}\hat{\boldsymbol{\beta}})$ may not be equal to $\mathbf{0}$.

4.6 Summary and future work

As mentioned in Section 4.1, our development in this chapter for singular mixed models amounts to a first step. Further research is needed to make the methodology viable. We specifically plan to explore the following in the near future:

- Implement the fitting method described in Section 4.5 for known $\mathbf{R}$ and test it on real and simulated data.
- Extend the methodology to allow for dependent subjects from family data (as in Chapter 3) for known $\mathbf{R}$ and test it on real and simulated data.
- Having family data as in Chapter 3 would allow us to separate covariance structure due to random effects and error covariance structure without any parametric assumption for the latter. We can then allow $\mathbf{R}$ to be unknown and estimate it using, e.g., low-rank smoothing or functional principal components analysis, within an iterative framework. Implement this methodology, test it on real and simulated data, and compare it with alternative competing methods.
- Develop an `R` package to implement the methodology and make it publicly available.

REFERENCES

1. Ramsay JO, Silverman BW. *Functional Data Analysis*. 2nd edn. Springer, 2005.

2. Greven S, Scheipl F. A general framework for functional regression modelling. *Statistical Modeling* 2017; **17**:1–35.

3. Sorensen H, Goldsmith J, Sangalli LM. An introduction with medical applications to functional data analysis. *Stat Med* 2013; **32**:5222–5240.

4. Yao F, Müller HG, Wang JL. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* 2005; **100**:577–590.

5. Goldsmith J, Greven S, Crainiceanu C. Corrected confidence bands for functional data using principal components. *Biometrics* 2013; **69**:41–51.

6. Shi M, Weiss RE, Taylor JMG. An analysis of paediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves. *Applied Statistics* 1996; **2**:151–163.

7. James GM, Hastie TJ, Sugar CA. Principal component models for sparse functional data. *Biometrika* 2000; **87**:587–602.

8. Rice JA, Wu CO. Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* 2001; **57**:253–259.

9. Hall P, Müller HG, Fang Y. Modelling sparse generalized longitudinal observations with latent gaussian processes. *Journal of the Royal Statistical Society, Series B* 2008; **70**:703–723.

10. Gertheiss J, Goldsmith J, Staicu AM. A note on modeling sparse exponential-family functional response curves. *Computational Statistics and Data Analysis* 2017; **105**:45–52.

11. Pinheiro JC, Bates DM. *Mixed-Effects Models in S and S-PLUS*. Springer, 2000.

12. Falconer DS, Mackay TFC. *Introduction to Quantitative Genetics*. 4th edn. Prentice Hall, 1996.

13. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era–concepts and misconceptions. *Nature Reviews Genetics* 2008; **9**:255–266.

14. Fisher RA. The correlation between relatives on the supposition of mendelian inheritance. *Transactions of the Royal Society Edinburgh* 1918; **52**:399–433.

15. Wright S. Systems of mating. *Genetics* 1921; **6**:111–178.

16. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; **38**:963–974.

17. Ruppert D, Wand MP, Carroll RJ. *Semiparametric Regression*. Cambridge University Press, 2005.

18. Pinheiro JC, Bates DM. Unconstrained parametrizations for variance-covariance matrices. *Statistics and Computing* 1996; **6**:289–296.

19. Carroll RJ, Ruppert D. *Transformation and Weighting in Regression*. Chapman & Hall, New York, 1988.

20. Davidian M, Giltinan DM. *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall/CRC, Boca Raton, FL, 1995.

21. Pinheiro JC, Liu C, Wu YN. Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *Journal of Computational and Graphical Statistics* 2001; **10**:249–276.

22. Box GEP, Jenkins GM, Reinsel GC. *Time Series Analysis: Forecasting and Control*. 3rd edn. Holden–Day, 1994.

23. Jones RH. *Longitudinal Data with Spatial Correlation: A State-space Approach*. Chapman & Hall, London, 1993.

24. Cressie NAC. *Statistics for Spatial Data*. Wiley, 1993.

25. Diggle PJ, Liang KY, Zeger SL. *Analysis of Longitudinal Data*. Oxford University Press, New York, 1994.

26. de Silva GSR, Rathnayake LN, Choudhary PK. Modeling and analysis of functional method comparison data. *Under Review* 2020; **0**:00–00.

27. Choudhary PK, Nagaraja HN. *Measuring Agreement: Models, Methods, and Applications*. Wiley, 2017.

28. Hothorn T, Bretz F, Westfall P. Simultaneous inference in general parametric models. *Biometrical Journal* 2008; **50**:346–363.

29. Ramsay JO, Silverman BW. *Functional Data Analysis*. 2nd edn. Springer, New York, 2006.

30. Ramsay JO, Silverman BW. *Applied Functional Data Analysis*. 2nd edn. Springer, New York, 2002.

31. Tuddenham RD, Snyder MM. "physical growth of california boys and girls from birth to age 18". *University of California Publications in Child Development* 1954; **1**:183–364.

32. Wang Y. Flexible estimation of covariance function by penalized spline with application to longitudinal family data. *Statististics in Medicine* 2011; **30**:1883–1897.

33. Chen Z, Dunson DB. Random effects selection in linear mixed models. *Biometrics* 2003; **59**:762–769.

34. Diggle PJ, Heagerty P, Liang KY, Zeger SL. *Analysis of Longitudinal Data*. 2nd edn. Oxford University Press, 2013.

35. Jones RH. *Longitudinal Data with Spatial Correlation: A State-space Approach*. 2nd edn. Chapman & Hall / CRC, 2009.

36. Chen H, Wang Y. A penalized spline approach to functional mixed-effects model analysis. *Biometrics* 2011; **67**:861–870.

37. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models*, 2018. URL `http://CRAN.R-project.org/package=nlme`, r package version 3.1-137.

38. Ruppert D, Wand MP, Carroll RJ. *Semiparametric Regression*. Cambridge University Press, New York, 2003.

39. Kokoszka P, Reimherr M. *Inroduction to Functional Data Analysis*. CRC Press, 2017.

40. Lange K, Westlake J, Spence MA. Extensions to pedigree analysis. III. Variance components by the scoring method. *Ann. Hum. Genet.* 1976; **39**:485–491.

41. Hopper JL, Mathews JD. Extensions to multivariate normal models for pedigree analysis. *Ann. Hum. Genet.* 1982; **46**:373–383.

42. Amos CI. Robust variance-components approach for assessing genetic linkage in pedigrees. *Am. J. Hum. Genet.* 1994; **54**:535–543.

43. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedig. *American Journal of Human Genetics* 1998; **62**:1198–1211.

44. Lange K, Boehnke M. Extensions to pedigree analysis. IV. Covariance components models for multivariate traits. *Am. J. Med. Genet.* 1983; **14**:513–524.

45. de Andrade M, Thiel TJ, Yu L, Amos CI. Assessing linkage on chromosome 5 using components of variance approach: univariate versus multivariate. *Genetic Epidemiology* 1997; **14**:773–778.

46. de Andrade M, Gueguen R, Visvikis S, Sass C, Siest G, Amos C. Extension of variance components approach to incorporate temporal trends and longitudinal pedigree data analysis. *Genetic Epidemiology* 2002; **22**:221–232.

47. Kraft P, Bauman L, Yuan JY, Horvath S. Multivariate variance-components analysis of longitudinal blood pressure measurements from the Framingham Heart Study. *BMC Genet.* 2003; **4 Suppl 1**:S55.

48. Macgregor S, Knott SA, White I, Visscher PM. Quantitative trait locus analysis of longitudinal quantitative trait data in complex pedigrees. *Genetics* 2005; **171**:1365–1376.

49. Wang Y, Huang C, Fang Y, Yang Q, Li R. Flexible semiparametric analysis of longitudinal genetic studies by reduced rank smoothing. *Appl. Statist.* 2012; **61**:1–24.

50. Wang Y, Huang C. Semiparametric variance components models for genetic studies with longitudinal phenotypes. *Biostatistics* 2012; **13**:482–496.

51. Ritland K. A Marker-based method for inferences about quantitative inheritance in natural populations. *Evolution* 1996; **50**:1062–1073.

52. Ritland K. Marker-inferred relatedness as a tool for detecting heritability in nature. *Mol. Ecol.* 2000; **9**:1195–1204.

53. Thomas SC. The estimation of genetic relationships using molecular markers and their efficiency in estimating heritability in natural populations. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2005; **360**:1457–1467.

54. Visscher PM, Medland SE, Ferreira MA, Morley KI, Zhu G, Cornes BK, Montgomery GW, Martin NG. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet.* 2006; **2**:e41.

55. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 2010; **42**:565–569.

56. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 2011; **88**:76–82.

57. Zaitlen N, Kraft P. Heritability in the genome-wide association era. *Hum. Genet.* 2012; **131**:1655–1664.

58. Vattikuti S, Guo J, Chow CC. Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits. *PLoS Genet.* 2012; **8**:e1002637.

59. Zaitlen N, Pasaniuc B, Sankararaman S, Bhatia G, Zhang J, Gusev A, Young T, Tandon A, Pollack S, Vilhjalmsson BJ, Assimes TL, Berndt SI, Blot WJ, Chanock S, Franceschini N, Goodman PG, He J, Hennis AJ, Hsing A, Ingles SA, Isaacs W, Kittles RA, Klein EA, Lange LA, Nemesure B, Patterson N, Reich D, Rybicki BA, Stanford JL, Stevens VL, Strom SS, Whitsel EA, Witte JS, Xu J, Haiman C, Wilson JG, Kooperberg C, Stram D, Reiner AP, Tang H, Price AL. Leveraging population admixture to characterize the heritability of complex traits. *Nat. Genet.* 2014; **46**:1356–1362.

60. Orozco LD, Morselli M, Rubbi L, Guo W, Go J, Shi H, Lopez D, Furlotte NA, Bennett BJ, Farber CR, Ghazalpour A, Zhang MQ, Bahous R, Rozen R, Lusis AJ, Pellegrini M. Epigenome-wide association of liver methylation patterns and complex metabolic traits in mice. *Cell Metab.* 2015; **21**:905–917.

61. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 2015; **67**:1–48.

62. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL `https://www.R-project.org/`.

63. Ziyatdinov A, Vazquez-Santiago M, Brunel H, Martinez-Perez A, Aschard H, Soria J. lme4qtl: Linear mixed models with flexible covariance structure for genetic studies of related individuals. *BMC Bioinformatics* 2018; **19**:68.

64. Rabe-Hesketh S, Skrondal A, Gjessing HK. Biometrical modeling of twin and family data using standard mixed model software. *Biometrics* 2008; **64**:280–288.

65. Lange K. *Mathematical and Statistical Methods for Genetic Analysis*. 2nd edn. Springer, 2003.

66. Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 1992; **69**:915–926.

67. Reik W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* 2007; **447**:425–432.

68. Huang K, Fan G. DNA methylation in cell differentiation and reprogramming: An emerging systematic view. *Regen Med* 2010; **5**:531–544.

69. Smith ZD, Meissner A. DNA methylation: Roles in mammalian development. *Nat. Rev. Genet.* 2013; **14**:204–220.

70. Richardson B. Primer: Epigenetics of autoimmunity. *Nat Clin Pract Rheumatol* 2007; **3**:521–527.

71. Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat. Genet.* 2007; **39**:61–69.

72. Migeon B. *X Chromosome Inactivation in Human Cells*, 2010. In The Biomedical & Life Sciences Collection, Henry Stewart Talks. Retrieved January 12, 2019, from `https://hstalks.com/bs/1676/`.

73. Horvath S, Zhang Y, Langfelder P, Kahn RS, Boks MP, van Eijk K, van den Berg LH, Ophoff RA. Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol.* 2012; **13**:R97.

74. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol.* 2013; **14**:R115.

75. Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev.* 2002; **16**:6–21.

76. Grundberg E, Small KS, Hedman AK, Nica AC, Buil A, Keildson S, Bell JT, Yang TP, Meduri E, Barrett A, Nisbett J, Sekowska M, Wilk A, Shin SY, Glass D, Travers M, Min JL, Ring S, Ho K, Thorleifsson G, Kong A, Thorsteindottir U, Ainali C, Dimas AS, Hassanali N, Ingle C, Knowles D, Krestyaninova M, Lowe CE, Di Meglio P, Montgomery SB, Parts L, Potter S, Surdulescu G, Tsaprouni L, Tsoka S, Bataille V, Durbin R, Nestle FO, O'Rahilly S, Soranzo N, Lindgren CM, Zondervan KT, Ahmadi KR, Schadt EE, Stefansson K, Smith GD, McCarthy MI, Deloukas P, Dermitzakis ET, Spector TD. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat. Genet.* 2012; **44**:1084–1089.

77. McRae AF, Powell JE, Henders AK, Bowdler L, Hemani G, Shah S, Painter JN, Martin NG, Visscher PM, Montgomery GW. Contribution of genetic variation to transgenerational inheritance of DNA methylation. *Genome Biol.* 2014; **15**:R73.

78. Shah S, McRae AF, Marioni RE, Harris SE, Gibson J, Henders AK, Redmond P, Cox SR, Pattie A, Corley J, Murphy L, Martin NG, Montgomery GW, Starr JM, Wray NR, Deary IJ, Visscher PM. Genetic and environmental exposures constrain epigenetic drift over the human life course. *Genome Res.* 2014; **24**:1725–1733.

79. Gaunt TR, Shihab HA, Hemani G, Min JL, Woodward G, Lyttleton O, Zheng J, Duggirala A, McArdle WL, Ho K, Ring SM, Evans DM, Davey Smith G, Relton CL. Systematic identification of genetic influences on methylation across the human life course. *Genome Biol.* 2016; **17**:61.

80. Rowlatt A, Hernandez-Suarez G, Sanabria-Salas MC, Serrano-Lopez M, Rawlik K, Hernandez-Illan E, Alenda C, Castillejo A, Soto JL, Haley CS, Tenesa A. The heritability and patterns of DNA methylation in normal human colorectum. *Hum. Mol. Genet.* 2016; **25**:2600–2611.

81. Joo JE, Dowty JG, Milne RL, Wong EM, Dugue PA, English D, Hopper JL, Goldgar DE, Giles GG, Southey MC, kConFab. Heritable DNA methylation marks associated with susceptibility to breast cancer. *Nat Commun* 2018; **9**:867.

82. van Dongen J, Nivard MG, Willemsen G, Hottenga JJ, Helmer Q, Dolan CV, Ehli EA, Davies GE, van Iterson M, Breeze CE, Beck S, Suchiman HE, Jansen R, van Meurs JB, Heijmans BT, Slagboom PE, Boomsma DI. Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat Commun* 2016; **7**:11115.

83. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, Burton J, Cox TV, Davies R, Down TA, Haefliger C, Horton R, Howe K, Jackson DK, Kunde J, Koenig C, Liddle J, Niblett D, Otto T, Pettett R, Seemann S, Thompson C, West T, Rogers J, Olek A, Berlin K, Beck S. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat. Genet.* 2006; **38**:1378–1385.

84. Ferguson-Smith AC. Genomic imprinting: The emergence of an epigenetic paradigm. *Nat. Rev. Genet.* 2011; **12**:565–575.

85. Frankhouser DE, Murphy M, Blachly JS, Park J, Zoller MW, Ganbat JO, Curfman J, Byrd JC, Lin S, Marcucci G, Yan P, Bundschuh R. PrEMeR-CG: Inferring nucleotide level DNA methylation values from MethylCap-seq data. *Bioinformatics* 2014; **30**:3567–3574.

86. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 2014; **30**:1363–1369.

87. Ayyala DN, Frankhouser DE, Ganbat JO, Marcucci G, Bundschuh R, Yan P, Lin S. Statistical methods for detecting differentially methylated regions based on MethylCap-seq data. *Brief. Bioinformatics* 2016; **17**:926–937.

88. Qin Z, Li B, Conneely KN, Wu H, Hu M, Ayyala D, Park Y, Jin VX, Zhang F, Zhang H, Li L, Lin S. Statistical challenges in analyzing methylation and long-range chromosomal interaction data. *Stat Biosci* 2016; **8**:284–309.

89. Park J, Lin S. Detection of differentially methylated regions using Bayesian curve credible bands. *Stat Biosci* 2018; **10**:20–40.

90. Han C, Tang H, Lou S, Gao Y, Cho MH, Lin S. Evaluation of recent statistical methods for detecting differential methylation using BS-seq data. *OBM Genetics* 2018; **2**:041.

91. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* 2012; **13**:R87.

92. Kapourani CA, Sanguinetti G. Higher order methylation features for clustering and prediction in epigenomic studies. *Bioinformatics* 2016; **32**:i405–i412.

93. Kapourani CA, Sanguinetti G. BPRMeth: a flexible Bioconductor package for modelling methylation profiles. *Bioinformatics* 2018; **34**:2485–2486.

94. Lawson JT, Tomazou EM, Bock C, Sheffield NC. MIRA: an R package for DNA methylation-based inference of regulatory activity. *Bioinformatics* 2018; **34**:2649–2650.

95. Lee W, Morris JS. Identification of differentially methylated loci using wavelet-based functional mixed models. *Bioinformatics* 2016; **32**:664–672.

96. Shafi A, Mitrea C, Nguyen T, Draghici S. A survey of the approaches for identifying differential methylation using bisulfite sequencing data. *Brief. Bioinformatics* 2018; **19**:737–753.

97. Quon G, Lippert C, Heckerman D, Listgarten J. Patterns of methylation heritability in a genome-wide analysis of four brain regions. *Nucleic Acids Res.* 2013; **41**:2095–2104.

98. Lea AJ, Tung J, Zhou X. A flexible, efficient binomial mixed model for identifying differential DNA methylation in bisulfite sequencing data. *PLoS Genet.* 2015; **11**:e1005650.

99. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, Geman D, Baggerly K, Irizarry RA. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet.* 2010; **11**:733–739.

100. Meyer KN, Lacey MR. Modeling methylation patterns with long read sequencing data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2018; **15**:1379–1389.

101. Sengupta D, Jammalamadaka SR. *Linear Models: An Integrated Approach.* World Scientific, 2003.

102. Harville DA. *Matrix Algebra from A Statistician's Perspective.* Springer, 2008.

103. Seber GAF. *A Matrix Handbook for Statisticians.* Wiley, 2008.

BIOGRAPHICAL SKETCH

Mohammad Shaha Alam Patwary grew up in Feni, Bangladesh. He has received a Bachelor of Science degree and a Master of Science degree in Statistics with distinctions from the University of Chittagong, Bangladesh in 2008 and 2009, respectively. Before joining The University of Texas at Dallas to pursue his doctoral degree in Statistics in Fall 2016, he received a Master of Science degree in Applied Statistics from Minnesota State University, Mankato in 2016.

CURRICULUM VITAE

Mohammad Shaha Alam Patwary

Contact Information:

Email: mohammad.patwary@utdallas.edu
LinkedIn: https://www.linkedin.com/in/mohammad-shaha-patwary-a2b6b7b7/
ResearchGate: https://www.researchgate.net/profile/Mohammad_Patwary4

Educational History:

- Doctor of Philosophy (PhD) in Statistics, August 2020
  The University of Texas at Dallas, Richardson, TX
- Master of Science (MS) in Applied Statistics, May 2016
  Minnesota State University, Mankato, MN
- Master of Science (MS) in Statistics, October 2009
  University of Chittagong, Chittagong, Bangladesh
- Bachelor of Science (BS) in Statistics , April 2008
  University of Chittagong, Chittagong, Bangladesh.

Teaching and Employment Experience:

- Teaching Associate, The University of Texas at Dallas, August 2018–January, 2019 and August 2019–January 2020.
- Teaching Assistant, The University of Texas at Dallas, August 2016–Present.
- Teaching Assistant, Department of Mathematics and Statistics, Minnesota State University, Mankato, MN, August 2014–May 2016.
- Assistant Professor, University of Chittagong, Chittagong, Bangladesh, April 2013–August 2014 (On Leave).
- Lecturer, University of Chittagong, Chittagong, Bangladesh, April 2011–April 2013.
- Lecturer, Bangladesh University of Professionals, Dhaka, Bangladesh, January 2011–April 2011.
- Lecturer, Bangladesh University of Business and Technology, Dhaka, Bangladesh, Jun 2010–January 2011.

Honors and Awards:

- Outstanding Teaching Assistant Award, Mathematical Sciences, UT Dallas, 2019.
- President's Teaching Excellence Award (Honorable Mention in Teaching Assistant Category), CTL, UT Dallas, 2020.
- Dr. A. R. Mollick Memorial Award, University of Chittagong, Chittagong, Bangladesh, 2011.
- Vice-Chancellor Award, University of Chittagong, Chittagong, Bangladesh, 2010.

Teaching and Professional Certifications:

- Graduate Teaching Certification, CTL, UT Dallas, 2019.
- Advanced Graduate Teaching Certification, CTL, UT Dallas, 2019.
- Online Teaching Certification, eLearning Management, UT Dallas, 2019.
- University Teaching Certification, The University of Hong Kong, Hong Kong, 2019.
- SAS Base Certified Programmer, SAS, 2015.

Courses Taught:

Introductory Statistics,Elementary Statistics, Statistical Decision Making, Introduction to Business Statistics, Business Statistics, Managerial Statistics, Fundamentals of Population Studies, Introduction to Design of Experiments, Statistical Inference, Mathematical Programming, Pre-Calculus, Applied Calculus, Calculus I, and Calculus.

Selected Journal Publications:

Chowdhury, M., VanBrackle, L., and **Patwary, M. S. A.** (2017). Two-step Smoothing Estimation of the Time-variant Parameter with Application to Temperature Data, *Journal of the Iranian Statistical Society*, 16(2), 33-50.

Wu, H., Rahman, M., and **Patwary, M. S. A.** (2016). A Study of a Generalized Modified Weibull Model: An Exponentiated Weibull Case, *Advances and Applications in Statistics*, 49(5), 357-367.

Patwary, M. S. A., Gokhale, D. V., and Rahman, M. (2016). On Testing Equality of Two Independent Proportions: To Pool or Not To Pool, *Far East Journal of Theoretical Statistics*, 52(3), 193-213.

Patwary, M. S. A. and Parveen, R. (2010). Breastfeeding Correlates: A Bangladesh Scenario, *Bangladesh Univ. of Business & Tech. Journal*, 3, 94-105.

Leadership Experience:

- Assistant Provost, Suhrawardy Residential Hall, University of Chittagong, Bangladesh, 2011–2014.
- Assistant Provost, Bangladesh University of Professionals Male Residence Hall, Dhaka, Bangladesh, 2010–2011.