

# Linear Mixed Effects Models in Functional Data Analysis

by

Wei Wang

B.Sc., University of Science and Technology of China, 2001  
M.Sc., National University of Singapore, 2003

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

The Faculty of Graduate Studies

(Statistics)

The University Of British Columbia

January, 2008

© Wei Wang 2008

# Abstract

Regression models with a scalar response and a functional predictor have been extensively studied. One approach is to approximate the functional predictor using basis function or eigenfunction expansions. In the expansion, the coefficient vector can either be fixed or random. The random coefficient vector is also known as random effects and thus the regression models are in a mixed effects framework.

The random effects provide a model for the within individual covariance of the observations. But it also introduces an additional parameter into the model, the covariance matrix of the random effects. This additional parameter complicates the covariance matrix of the observations. Possibly, the covariance parameters of the model are not identifiable.

We study identifiability in normal linear mixed effects models. We derive necessary and sufficient conditions of identifiability, particularly, conditions of identifiability for the regression models with a scalar response and a functional predictor using random effects.

We study the regression model using the eigenfunction expansion approach with random effects. We assume the random effects have a general covariance matrix and the observed values of the predictor are contaminated with measurement error. We propose methods of inference for the regression model's functional coefficient.

As an application of the model, we analyze a biological data set to investigate the dependence of a mouse's wheel running distance on its body mass trajectory.

# Table of Contents

<b>Abstract</b> . . . . .	ii
<b>Table of Contents</b> . . . . .	iii
<b>List of Tables</b> . . . . .	vi
<b>List of Figures</b> . . . . .	vii
<b>Acknowledgements</b> . . . . .	x
<b>Statement of Co-Authorship</b> . . . . .	xi
<b>1 Introduction</b> . . . . .	1
1.1 Summary of the biology experiment . . . . .	1
1.2 Review of smoothing methods . . . . .	2
1.3 Functional regression models . . . . .	5
1.4 Introduction to the thesis work . . . . .	9
<b>Bibliography</b> . . . . .	11
<b>2 Identifiability discussion of linear mixed effects models</b> . . . . .	13
2.1 Introduction . . . . .	13
2.2 Motivation . . . . .	13
2.3 Simple sufficient conditions of identifiability . . . . .	16
2.4 Sufficient conditions of identifiability for a structured $\Sigma_\epsilon$ . . . . .	17
2.5 Extensions . . . . .	22
<b>Bibliography</b> . . . . .	25

<b>3</b>	<b>Linear mixed models for measurement error in functional regression</b>	26
3.1	Introduction	26
3.2	Notation and Preliminaries	29
3.3	Parameter estimation	31
3.3.1	Initial estimates of parameters other than $\boldsymbol{\mu}$ and $\beta_0$	32
3.3.2	Updating $\boldsymbol{\Sigma}_x^{(t)}$	32
3.3.3	Updating $\sigma_\epsilon^{2(t)}$	34
3.3.4	Updating $\boldsymbol{\beta}^{(t)}$ and $\sigma^{2(t)}$	36
3.4	Inference for $\beta$	38
3.4.1	Hypothesis testing for $\beta$	38
3.5	Model assumption checking	43
3.6	Model application	44
3.6.1	Choosing the basis functions	44
3.6.2	Data description	45
3.6.3	Choice of basis functions	45
3.6.4	Estimation and residual analysis	46
3.6.5	Inference for $\beta(\cdot)$	47
3.6.6	Inference for $\beta^s - \beta^c$	48
3.7	Simulation study	48
3.7.1	Two stage estimate	49
3.7.2	One sample comparison	50
3.7.3	Two sample comparison	51
3.7.4	Edge effect discussion in one-sample MSE comparison	52
	<b>Bibliography</b>	73
<b>4</b>	<b>Dependence of average lifetime wheel-running on body mass ontogeny</b>	75
4.1	Introduction	75
4.2	Methods	78
4.3	Results	82
4.4	Discussion	85

<b>Bibliography</b> . . . . .	95
<b>5 Conclusions and future plan</b> . . . . .	98
5.1 Models with correlated individuals . . . . .	98
5.2 Proposing models with a functional response . . . . .	101
5.2.1 Models of a functional response . . . . .	101
5.2.2 Outline of the model study . . . . .	103
5.2.3 Literature review . . . . .	104
<b>Bibliography</b> . . . . .	106
 <b>Appendices</b>	
<b>A Appendix to Chapter 2</b> . . . . .	107
A.1 Proof of Corollary 2.4.1 . . . . .	107
A.2 Proof of Corollary 2.4.2 . . . . .	108
<b>B Appendix to Chapter 3</b> . . . . .	110
B.1 Definition of the first differential . . . . .	110
B.2 Definition of the second differential . . . . .	113
B.3 Matrix algebraic and differential rules . . . . .	115
B.4 Calculations in Section 3.3.2 . . . . .	116
B.5 Calculations in Section 3.3.4 . . . . .	118

# List of Tables

3.1	P-values of the test $H_o : \beta(t) = 0$ , for all $t \in [-1, 60]$ in each group. . . . .	54
3.2	P-values of the test $H_o : \beta^s(t) = \beta^c(t)$ , for all $t \in [-1, 60]$ , within each gender. . . . .	54
4.1	P-values of the test $H_o : \beta(t) = 0$ , for all $t \in [-1, 60]$ , using test statistic $U_f$ . . . . .	87
4.2	P-values of the test $H_o : \beta^s(t) = \beta^c(t)$ , for all $t \in [-1, 60]$ , using test statistic $U_f$ . . . . .	87

# List of Figures

3.1	Plots of log body mass versus week for the four groups of laboratory mice. . . . .	55
3.2	Histogram of the response: averaged wheel running from weeks 5 to 60 for the four groups of laboratory mice. . . . .	56
3.3	Plots of the proportion of cumulative variance of the centered log body mass explained by the first ten principal components of the four groups of mice, after the individual average log body mass has been removed. . . . .	57
3.4	The constant function and the first three smoothed eigenfunctions of the covariance of centered log body mass of the four groups of mice, calculated as described in Section 3.6.3. . . .	58
3.5	Residual plots of the fit of the $Y_i =$ averaged wheel running of the four groups of mice. . . . .	59
3.6	Plots of residuals of $Y_i =$ averaged wheel running in the male control group of mice before and after removing the outlier. . . . .	60
3.7	Plots of $\hat{\beta}$ and its standard errors computed from the Hessian matrix (solid) and from the bootstrap (dash-dot) for the four groups of mice, as described in Section 3.4. . . . .	61
3.8	Comparing the permuted values of the generalized likelihood ratio statistic $U_l$ with the Wald statistic $U_w$ for four groups of mice. The equivalence of $U_l$ and $U_w$ were observed in Table 3.1. . . . .	62
3.9	Plots of $\hat{\beta}^s - \hat{\beta}^c$ and the standard errors of the difference computed from the Hessian matrix (solid) and from the bootstrap (dash-dot) for both genders. . . . .	63

3.10 MSE of the estimate of $\beta$ for each $\gamma$ value in one sample simulation as described in Section 3.7.2. Compare the ECME method with the two stage method. . . . .	64
3.11 Proportion of times $H_o$ is rejected using level $\alpha = 0.01$ , where $H_o : \beta(t) = 0$ , for all $t \in [-1, 60]$ in one sample simulation as described in Section 3.7.2. Two test statistics are considered, $U_w$ and $U_f$ . . . . .	65
3.12 Proportion of times $H_o$ is rejected using level $\alpha = 0.05$ , where $H_o : \beta(t) = 0$ , for all $t \in [-1, 60]$ in one sample simulation as described in Section 3.7.2. Two test statistics are considered, $U_w$ and $U_f$ . . . . .	66
3.13 MSE of the estimate of $\beta^s$ for each $\gamma$ value in two sample simulation as described in Section 3.7.3. Compare the ECME method with the two stage method. . . . .	67
3.14 MSE of the estimate of $\beta^c$ for each $\gamma$ value in two sample simulation as described in Section 3.7.3. Compare the ECME method with the two stage method. . . . .	68
3.15 MSE of the estimate of $\beta^s - \beta^c$ for each $\gamma$ value in two sample simulation as described in Section 3.7.3. Compare the ECME method with the two stage method. . . . .	69
3.16 Proportion of times $H_o$ is rejected using level $\alpha = 0.01$ , where $H_o : \beta^s = \beta^c$ , for all $t \in [-1, 60]$ in two sample simulation as described in Section 3.7.3. Four test statistics are considered $U_l, U_w, U_e$ and $U_f$ . . . . .	70
3.17 Proportion of times $H_o$ is rejected using level $\alpha = 0.05$ , where $H_o : \beta^s = \beta^c$ , for all $t \in [-1, 60]$ in two sample simulation as described in Section 3.7.3. Four test statistics are considered $U_l, U_w, U_e$ and $U_f$ . . . . .	71
3.18 MSE of the estimate of $\beta$ for each $\gamma$ value using truncated log body mass as described in Section 3.7.4. Compare the ECME method with the two stage method. . . . .	72



4.1	Plots of log body mass versus week for the four groups of laboratory mice. . . . .	88
4.2	Histogram of the response: averaged wheel running from weeks 5 to 60 for the four groups of laboratory mice. . . . .	89
4.3	Plots of the proportion of cumulative variance of the centered log body mass explained by the first ten principal components for four groups of mice, after the individual average log body mass has been removed. . . . .	90
4.4	The constant function and the first three smoothed eigenfunctions of the covariance of centered log body mass of four group. The four groups' functions are together in one panel. The eigenfunctions are calculated as described in Section 4.3.	91
4.5	Plots of $\hat{\beta}$ and its standard errors computed from the Hessian matrix (solid) and from the bootstrap (dash-dot) for four groups of mice. Standard errors are calculated as described in Section 4.2. . . . .	92
4.6	Plots of the covariances between the average wheel running and the log body mass trajectory at week $j$ , $j = -1, \dots, 60$ , for four groups of mice. Solid lines are the covariance from the model and dash-dot lines are the pointwise sample covariances between average wheel running and log body mass. For the calculation, refer to Heckmand and Wang (2007). . . . .	93
4.7	Plots of $\hat{\beta}^s - \hat{\beta}^c$ , the difference between the selection $\hat{\beta}$ and the control $\hat{\beta}$ , and the standard errors of the difference computed from the Hessian matrix (solid) and from the bootstrap (dash-dot) for both genders. . . . .	94

# Acknowledgements

For the completion of this thesis, I would like very much to express my heartfelt gratitude to my supervisor Professor Nancy Heckman for all her invaluable advice and guidance, endless patience, kindness and encouragement during the mentor period in the Department of Statistics of University of British Columbia. I have learned many things from her, particularly regarding academic research and analytical writing. I truly appreciate all the time and effort she has spent in helping me to solve the problems encountered even when she is in the midst of her work.

I would also like to express my sincere gratitude and appreciation to my committee members, namely Professors Gustafson, Paul and Wu, Lang (alphabetically) for their precious advice and help in my thesis.

Special thanks to Professor Patrick Carter for providing us the biological data and the biological background of the thesis work.

Thank you to all faculty, staff and graduate students in the Department of Statistics, University of British Columbia, for making my stay here such an enriching experience.

# Statement of Co-Authorship

The thesis is finished under the supervision of my supervisor, Professor Nancy Heckman.

Chapter 2 is completed with the help of Professor Heckman.

Chapter 3 is co-authored with Professor Nancy Heckman. My main contributions are the derivation of the model parameter estimation in Section 3.3 and the model diagnostics in Section 3.5. I conduct the programming work of the data analysis and the simulation studies.

Chapter 4 is co-authored with Professor Patrick Carter, School of Biological Sciences, Washington State University. Professor Carter conducted the biological experiment and provided us the data. Professor Carter also gives a detailed description of the experiment and the biological background. We describes the statistical method and carry out the data analysis.

# Chapter 1

## Introduction

This thesis work is motivated by a biology experiment involving data on individuals, data that can be considered as scalar responses depending on function-valued predictors. In the thesis, we provide a functional regression approach to study the biology problem of interest. In this introductory section, we first give a brief summary of the experiment and then a review of smoothing methods in functional data analysis. We give a literature review of the functional regression model followed by a description of our approach.

### 1.1 Summary of the biology experiment

Professor Patrick Carter of the School of Biological Sciences of Washington State University conducted an experiment on house mice (*Mus domesticus*) selectively bred for increased voluntary wheel-running exercise. Details of the experiment are described in Morgan, Garland and Carter, 2003. In summary, mice were divided into four groups according to gender and the two treatments “selection” and “control”. The selection group mice were from lines bred with selection being on high wheel-running activity at age eight weeks. Control mice were from lines bred at random. Body mass and wheel running activity were recorded for each mouse for sixty two consecutive weeks, indexed from  $-1$  to  $60$ , except for weeks 34, 38, 39, 50. The research interest is to know how body mass and wheel running are related and if the relationship depends on the treatment.

Unfortunately, the wheel running distance data have many missing values and are very noisy. In addition, the wheels were cleaned every four weeks, which resulted in spikes in wheel-running activity every fourth week.

One way to eliminate the problem is to average the wheel running distance over weeks and use the averaged value in the analysis. After this averaging, a scalar is obtained. We treat the averaged wheel running as the response and the body mass trajectory as the predictor. We can build a regression model to study the dependence of the averaged wheel running on the body mass. To make use of the functional form of the body mass, we use a functional regression model approach. More details of this model are in Section 1.3. In the next section, we give a review of smoothing methods which can be used to study the body mass trajectories. Some of these methods will be used in Section 1.3.

## 1.2 Review of smoothing methods

One can view a body mass trajectory as a process which can be described as a continuous function of time. Though the measurements of the trajectory are made at a finite discrete set of points, the trajectory lies in an infinite dimensional parameter space. This functional type of data is the study object of a statistics branch, namely *functional data analysis*. For a reference, please see Ramsay and Silverman (2005). In this section, we give a brief summary of some popular smoothing methods to study functional data. Ruppert, Wand and Carroll (2003) give a detailed summary of these methods. These methods can be applied to study the body mass trajectories of the mice. To further investigate the relationship between the body mass and the averaged wheel running, an additional method needs to be used.

First, we consider analyzing data on one individual. Fix the  $i$ th individual and let  $Z_i(\cdot)$  denote its trajectory. Let  $t_{ij}$  denote  $Z_i$ 's  $j$ th observational point and let  $z_{ij}$  denote its observed value at that time. Data collected on this individual are the  $t_{ij}$ 's and the  $z_{ij}$ 's,  $j = 1, \dots, n_i$ . The dependence of  $z_{ij}$  on  $t_{ij}$  through the function  $Z_i$  is written as

$$z_{ij} = Z_i(t_{ij}) + \epsilon_{ij}, \quad j = 1, \dots, n_i. \quad (1.1)$$

where  $\epsilon_{ij}$  represents an error, either measurement error or modelling error,

with mean zero. Estimating  $Z_i$  from the  $z_{ij}$ 's is often referred to as *nonparametric regression*, since the function  $Z_i$  is only assumed to be “smooth”. One may view  $z_{ij}$  as the perturbed  $Z_i$  at time  $t_{ij}$ .

One of the popular methods to estimate  $Z_i$  is local polynomial smoothing. See Fan and Gijbels (1996) for details. Let  $K$  be a symmetric positive function with  $K(t)$  decreasing as  $|t|$  increases. Let  $h > 0$  be the smoothing parameter which is usually referred to as the “bandwidth”. To estimate  $Z_i(t)$  by a local polynomial of degree  $p$ , we minimize

$$\sum_{j=1}^n [z_j - \beta_0 - \beta_1(t_j - t) - \dots - \beta_p(t_j - t)^p]^2 K\left(\frac{t_j - t}{h}\right)$$

to obtain  $(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)$ . The estimate of  $Z_i(t)$  is then  $\hat{Z}_i(t) = \hat{\beta}_0$ . Compared with other smoothing methods, one advantage of local polynomial smoothing is its simpler theoretical analysis which has allowed greater insight into its asymptotic properties.

Another popular smoothing method uses spline basis functions to model the function  $Z_i$ . A spline function with degree  $q$  is defined on an interval. One first divides the interval into subintervals by breakpoints, called knots. On each subinterval, the spline function is a polynomial of degree  $q$ . On the entire interval, the function has  $q - 1$  continuous derivatives. For a discussion of spline functions, please see Ramsay and Silverman (2005) p. 47. A popular choice of spline basis functions is the B-spline basis. In general, let  $\phi_1, \dots, \phi_K$  be the spline basis functions and assume that  $Z_i$  can be modelled as

$$Z_i(t) \sim \sum_{k=1}^K x_{ik} \phi_k(t) \equiv \boldsymbol{\phi}(t)' \mathbf{x}_i, \quad \mathbf{x}_i \equiv (x_{i1}, \dots, x_{iK})'. \quad (1.2)$$

It thus follows from (1.1) that

$$z_{ij} = \sum_{k=1}^K x_{ik} \phi_k(t_{ij}) + \epsilon_{ij} \equiv \boldsymbol{\phi}(t_{ij})' \mathbf{x}_i + \epsilon_{ij}, \quad j = 1, \dots, n_i. \quad (1.3)$$

One way to estimate the unknown coefficient vector  $\mathbf{x}_i$  is via minimizing

$$\sum_{j=1}^{n_i} [z_{ij} - \phi(t_{ij})' \mathbf{x}_i]^2. \quad (1.4)$$

To use the spline basis functions, one must specify the number and the location of “knots”. This is not a simple task. Improper choice of knots may result in overfitting or underfitting the data. One strategy to overcome this problem is to use many knots, but to penalize to prevent overfitting. In this approach, called penalized splines, we minimize a penalized version of (1.4) to estimate  $\mathbf{x}_i$ . That is, we minimize

$$\sum_{j=1}^{n_i} [z_{ij} - \phi(t_{ij})' \mathbf{x}_i]^2 + \lambda \mathbf{x}_i' \mathbf{P} \mathbf{x}_i, \quad (1.5)$$

where  $\lambda$  is a smoothing parameter and  $\mathbf{P}$  is a penalty matrix. For examples of  $\mathbf{P}$  and how to choose  $\lambda$  using the method of cross validation, please see Ramsay and Silverman (2005) p. 94.

Interestingly, there is a close connection between the penalized smoothing estimate of  $Z_i(t)$  using basis functions and the best linear unbiased predictor (BLUP) of  $Z_i(t)$  in a random effects model framework. Intuitively, consider the vector coefficient  $\mathbf{x}_i$  in (1.3) being random and having a covariance matrix  $\Sigma_x$ . Under a normality assumption of the  $\mathbf{x}_i$  and the  $\epsilon_{ij}$ 's, the log likelihood function of the  $z_{ij}$ 's and  $\mathbf{x}_i$  can be written in a form similar to (1.5), up to constant. The smoothing parameter  $\lambda$  is related to the variance of  $\epsilon_{ij}$  and the penalty matrix  $\mathbf{P}$  is in the same position as the inverse of  $\Sigma_x$ . For more discussion on the BLUP, see Robinson (1991).

The random effects view of penalized splines is particularly useful when analyzing data from  $N$  individuals: the  $z_{ij}$ 's,  $j = 1, \dots, n_i$ ,  $i = 1, \dots, N$ . We model each  $z_{ij}$  as in (1.3), with the random effects  $\mathbf{x}_i$ 's assumed independently and identically distributed with covariance matrix  $\Sigma_x$ . The random effects induce a covariance structure on the  $z_{ij}$ 's that provides a reasonable, yet flexible model for within subject dependence. The parameter  $\Sigma_x$  is to be estimated using all the  $z_{ij}$ 's. Standard methods to fit random effects or

mixed models (containing both random and fixed effects) can thus be used. For a reference, see McCulloch and Searle (2001). Rice and Wu (2001) use a B-spline basis in (1.3) and use the EM algorithm to estimate the model parameters. The authors also calculate the best linear unbiased predictor of each individual trajectory.

### 1.3 Functional regression models

Some ideas and methods in functional data analysis can be viewed as extensions of those in classical statistics. One of the extensions is the functional regression model with a scalar response and a functional predictor. This model can be used to study the dependence of the averaged wheel running on the body mass trajectory. In this section, we introduce the functional regression model and sketch the methods to study this model. Among the rich literature on this topic, we select those papers most relevant to our approach.

Let  $Y_i$  be the averaged wheel running of the  $i$ th individual. Referring to (1.1), we view  $Z_i$  as the predictor while  $z_{ij}$  is the perturbed value of  $Z_i$  at time  $t_{ij}$ . Data collected on  $N$  individuals are  $Y_i$  and  $z_{ij}, j = 1, \dots, n_i, i = 1, \dots, N$ . The functional regression model is

$$Y_i = \beta_0 + \int \beta(t)Z_i(t)dt + e_i. \quad (1.6)$$

The unknown functional coefficient  $\beta(\cdot)$  measures the dependence of the scalar response  $Y_i$  on the predictor  $Z_i$  and is of primary interest. Usually it is assumed that the function  $\beta$  is smooth.

The functional regression model (1.6) is loosely related to ordinary multiple regression in that the integral replaces the summation. The ordinary multiple linear regression model for a fixed individual uses “covariates”  $Z_i(t_{i1}), \dots, Z_i(t_{in_i})$ . and models  $Y_i$  as

$$Y_i = \beta_0 + \sum_j \beta_j Z_i(t_{ij}) + e_i.$$



In our case, where the  $Z_i(t_{ij})$ 's are unobserved, we are essentially carrying out multiple regression with covariates measured with error. Note that the  $\beta_j$ 's do not depend on  $i$ . Thus, to use all individuals in this model, the  $Z_i$ 's have to be measured at the same time points to obtain the  $z_{ij}$ 's, i.e.  $t_{ij} = t_j$  and  $n_i = n$ . When individuals are measured at different sets of time points or have different number of observations, it is not clear how to apply the multiple regression model. Another problem with multiple regression is that for large values of  $n$ , the high-dimensional vector of parameters,  $\boldsymbol{\beta}$ , may be hard to estimate.

The functional model (1.6) can easily accommodate the case when different time points are observed for different individuals. And exploiting the continuity of  $\beta$  and the  $Z_i$ 's allows us to reduce the number of parameters used. In the literature, there are several approaches to calculate the integral  $\int \beta(t)Z_i(t)dt$  and to estimate  $\beta$ . In the remainder of this section, we give a summary of some of these approaches and focus on those closest to ours.

One approach uses a discretized version of the integral and estimates  $\beta$  by minimizing a penalized version of the residual sum of squares. Cardot, Crambes, Kneip and Sarda (2007) assumed that data were observed at the same  $n$  equally spaced time points, i.e.  $n_i \equiv n$ ,  $t_{ij} = t_j$  and  $t_j - t_{j-1} = 1/n$  for all  $j = 2, \dots, n$ . Cardot et al. considered model (1.6) but without the intercept  $\beta_0$ , i.e. the model

$$Y_i = \int \beta(t)Z_i(t)dt + e_i.$$

The authors discretized the integral  $\int \beta(t)Z_i(t)dt$  using a grid of  $t$  values equal to the  $t_j$ 's and wrote

$$Y_i = \frac{1}{n} \sum_{j=1}^n \beta(t_j)Z_i(t_j) + e_i.$$

Let  $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{in})'$ ,  $\mathbf{z}_i = (z_{i1}, \dots, z_{in})'$  and  $\boldsymbol{\beta} = (\beta(t_1), \dots, \beta(t_n))'$ . The authors used a *Total Least Squares* method to estimate  $\boldsymbol{\beta}$  and predict  $Z_i$  as

the solutions of the minimization problem

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^n, \mathbf{Z}_i \in \mathbb{R}^n} \left\{ \frac{1}{N} \sum_{i=1}^N \left[ \left( Y_i - \frac{1}{n} \mathbf{Z}'_i \boldsymbol{\beta} \right)^2 + \frac{1}{n} \|\mathbf{z}_i - \mathbf{Z}_i\|^2 \right] + \frac{\lambda}{n} \boldsymbol{\beta}' \mathbf{P} \boldsymbol{\beta} \right\},$$

where  $\lambda$  is a smoothing parameter and  $\mathbf{P}$  is a penalty matrix. Unfortunately, it is not clear how to apply this approach directly when individuals are measured at different sets of time points.

Another approach uses basis functions to approximate the  $Z_i$ 's and  $\beta$ . In this basis expansion approach, individuals can be measured at different time points and they don't need to have the same number of observations. This approach approximates the  $Z_i$ 's as in (1.2) and similarly approximates  $\beta$  using basis functions,  $\psi_1, \dots, \psi_J$ , as

$$\beta(t) = \sum_{j=1}^J \beta_j \psi_j(t) \equiv \boldsymbol{\beta}' \boldsymbol{\psi}(t), \quad \boldsymbol{\beta} = (\beta_1, \dots, \beta_J)'. \quad (1.7)$$

An advantage of expansions (1.2) and (1.7) is that they reduce the infinite dimensional functional uncertainty into finite dimensional unknowns which are conveniently handled by vectors. Estimating  $\beta(\cdot)$  thus reduces to estimating the coefficient vector  $\boldsymbol{\beta}$ . We also notice that using (1.2) and (1.7), the integral  $\int \beta(t) Z_i(t) dt$  in (1.6) can be written as

$$\int \beta(t) Z_i(t) dt = \boldsymbol{\beta}' \mathbf{T} \mathbf{x}_i, \quad \text{where } \mathbf{T}[j, k] = \int \psi_j(t) \phi_k(t) dt. \quad (1.8)$$

In this approach, the  $\mathbf{x}_i$ 's in (1.2) can be assumed to be fixed or random. When the  $\mathbf{x}_i$ 's are assumed random, they are usually modelled as described at the end of Section 1.2. That is, they are assumed independently and identically distributed with a normal distribution with covariance matrix  $\boldsymbol{\Sigma}_x$ . This assumption contributes to the construction of the likelihood function of the data, the  $(Y_i, z_{ij})$ 's. The vector  $\boldsymbol{\beta}$  is estimated by maximizing the log likelihood or a penalized log likelihood.

In the following, we give brief summaries of three statistical papers and a book chapter that use a basis expansion approach. In expansions (1.2) and

(1.7), the  $\phi_k$ 's and  $\psi_j$ 's are usually chosen to be splines, although this is not necessary. For instance, the  $\phi_k$ 's can be equal to the principal component functions of the  $Z_i$ 's.

James (2002), James and Silverman (2005) and Müller (2005) considered more general models. In the following, we describe their approaches with an emphasis on studying model (1.6).

James (2002) used  $\phi_k$ 's in (1.3) equal to a basis for natural cubic splines and also considered  $\psi = \phi$  in (1.7). The  $Z_i$ 's were subject to error, and so the data were the perturbed  $z_{ij}$ 's, given in (1.3). The  $\mathbf{x}_i$ 's were multivariate normal and  $\beta$  was estimated by maximizing the log likelihood. With  $\mathbf{x}_i$ 's as the missing data and the  $\{(z_{ij}, j = 1, \dots, n_i), Y_i\}$ 's as the observed data, the author used the EM algorithm to estimate unknown parameters.

James and Silverman (2005) used  $\phi_k$ 's in (1.3) equal to the estimated principal component functions of the  $Z_{ij}$ 's and used  $\psi_j$ 's equal to cubic spline basis functions. The coefficient vector's, the  $\mathbf{x}_i$ 's, were taken to be fixed. When the  $Z_i$ 's were observed without error, the authors estimated  $\beta$  by maximizing the log likelihood of the  $(Y_i, Z_{ij})$ 's subject to a penalty term which penalized  $\beta$ 's that have  $\int \beta(t)f(t)dt$  big when  $\text{Var}(\int Z(t)f(t)dt)$  is small. When the  $z_{ij}$ 's were observed, the log likelihood was modified to include the measurement error and a penalty term was added to ensure the smooth fits of the  $Z_i$ 's.

Müller (2005) used the idea of Karhunen-Loève expansion of the  $Z_i$  process to estimate the  $\phi_k$ 's and the corresponding  $\mathbf{x}_i$ 's. In this approach, the  $\phi_k$ 's in (1.3) are equal to the first  $K$  estimated eigenfunctions of the  $Z_i$  process. The  $\mathbf{x}_i$ 's are random and have diagonal covariance matrix, and Müller assumes that the estimated  $\mathbf{x}_i$ 's have diagonal covariance matrix too. Müller's method of estimating can be separated into two parts where the  $\mathbf{x}_i$ 's are estimated in the first part and  $\beta$  is estimated in the second part. A detailed description of Müller's method is provided in Chapter 3.

Ramsay and Silverman (2005, Chapter 15) first calculated  $\hat{Z}_i(\cdot)$  by fitting the  $z_{ij}$ 's using least squares and a fairly large number of basis functions. They then estimated  $\beta$  by minimizing a penalized residual sum of squares

which was defined as

$$\sum_{i=1}^N \left[ Y_i - \beta_0 - \int \beta(t) Z_i(t) dt \right]^2 + \lambda \boldsymbol{\beta}' \mathbf{P} \boldsymbol{\beta}.$$

This method is implemented in the R library *fda* (Ramsay, Wickham and Graves, 2007).

## 1.4 Introduction to the thesis work

In (1.3), when the  $\mathbf{x}_i$ 's are random, the distribution of the observed vector  $(z_{i1}, \dots, z_{in_i})$  will not, in general, be identifiable. This concern motivates our work in Chapter 2. Assuming the  $\mathbf{x}_i$ 's random introduces an additional parameter  $\boldsymbol{\Sigma}_x$  into the model. This additional parameter complicates the covariance matrix of the  $z_{ij}$ 's and the variance of the  $Y_i$ 's. Possibly, the covariance/variance parameters of the model are not identifiable.

In Chapter 2, we study identifiability in normal linear mixed effects models. We derive necessary and sufficient conditions of identifiability, including conditions of identifiability for the scalar response and functional predictor mixed effects model.

In Chapter 3 of the thesis, we study model (1.6) assuming the observed  $Z_i$ 's are contaminated with measurement error. We use principal component functions to expand  $Z_i$  in (1.2) and use the same functions to expand  $\beta$  in (1.7). This choice of basis functions can fit the  $Z_i$ 's well with  $K$  fairly small. An added benefit of this eigenfunction approach is that we focus on “estimable directions” of  $\beta$ . We assume that  $\mathbf{x}_i$  in (1.3) is a normal random vector with a general covariance matrix  $\boldsymbol{\Sigma}_x$ .

In Chapter 3, we implement the ECME (Expectation/Conditional Maximization Either) algorithm of Liu and Rubin (1994) to estimate the model parameters. We consider Hessian matrix based and boot-strapped standard errors of the estimate of  $\boldsymbol{\beta}$ . We propose a new integrated t-statistic for hypothesis testing involving  $\beta(\cdot)$ . Expressions of the residuals of the model fit are derived and we discuss model diagnostics based on the analysis of residuals.

In Chapter 4, we apply our model to analyze a biological data set, providing a detailed biological background.

Chapter 5 summarizes the thesis work and discusses possible future research directions.

# Bibliography

- [1] Cardot, H., Crambes, C., Kneip, A. and Sarda, P. (2007) Smoothing splines estimators in functional linear regression with errors-in-variables. *Computational Statistics & Data Analysis* 51, 10, 4832-4848.
- [2] James, G. (2002) Generalized linear models with functional predictors. *Journal of the Royal Statistical Society B* 64, 3, 411-432.
- [3] James, G. M. and Silverman, B. W. (2005) Functional adaptive model estimation. *Journal of the American Statistical Association* 100, 470, 565-576.
- [4] Fan, J., and Gijbels, I. (1996) *Local Polynomial Modelling and Its Applications*. London: Chapman & Hall.
- [5] Liu, C. H. and Rubin, D. B. (1994) The ECME Algorithm - a simple extension of EM and ECM with faster monotone convergence. *Biometrika* 81 4, 633-648.
- [6] McCulloch, C. E. and Searle, S. R. (2001) *Generalized, Linear and Mixed Models*. New York: Wiley.
- [7] Morgan, T.J., T. Garland, Jr., and Carter, P. A. (2003) Ontogenetic trajectories in mice selected for high wheel- running activity. I. Mean ontogenetic trajectories. *Evolution* 57, 646-657.
- [8] Müller, H. G. (2005) Functional modeling and classification of longitudinal data. *Scandinavian Journal Statistics* 32, 223-240.
- [9] Ramsay, J. O. and Silverman, B. W. (2005) *Functional Data Analysis*. 2nd edition, Springer Series in Statistics.

- [10] Ramsay, J. O., Wickham, H. and Graves, S. (2007) R *fda* Library.
- [11] Rice, J. A. and Wu, C. O. (2001) Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* 57, 253-259.
- [12] Robinson, G.K. (1991) That BLUP is a good thing: the estimation of random effects. *Statistical Science* 6, 15-32.
- [13] Ruppert, D., Wand, M. P. and Carroll, R. J. (2003) *Semiparametric Regression*. Cambridge University Press.

## Chapter 2

# Identifiability discussion of linear mixed effects models

### 2.1 Introduction

Mixed effects models have moved beyond the original simple models, becoming more complicated, containing more parameters. However, two different sets of parameters producing the same covariance matrix for the observations may cause problems for parameter estimation and for inference. In principle, parameter identifiability is the first thing that should be verified when building a model. As Demidenko (2004, p.118) states, “identifiability may be viewed as a necessary property for the adequacy of a statistical model”.

We study identifiability in normal linear mixed effects models. In the next section, we define the classical mixed effects model and show that in an unrestricted form and a specific restricted form, the model is not identifiable. Sections 2.3 and 2.4 give sufficient conditions to identify parameters in various models. In Section 2.5 we discuss identifiability of extended models.

### 2.2 Motivation

In this section, we give the definition of nonidentifiability. Then we introduce the general unrestricted classical linear mixed effects model and give two models that are nonidentifiable.

**Definition 2.2.1** *Let  $\mathbf{y}$  be the vector of observable random variables with*

---

<sup>0</sup>A version of this chapter will be submitted for publication. Author: Wang, W.



distribution function  $\mathcal{P}_\theta$  where  $\theta$  is in the parameter space  $\Theta$ . This probability model for  $\mathbf{y}$  is not identifiable if and only if there exist  $\theta, \theta^* \in \Theta$ , with  $\theta \neq \theta^*$  and  $\mathcal{P}_\theta = \mathcal{P}_{\theta^*}$ .

Throughout the paper, we assume all random variables follow normal distributions. As the normal distribution is uniquely characterized by its first two moments, identifiability of a normal distribution function then reduces to the identifiability of its mean vector and covariance matrix (Demidenko, 2004, *Proposition 10*, p.118).

In the standard linear mixed effects model,  $\mathbf{y}$  is the observable random vector of length  $n$  and  $\mathbf{X}$  and  $\mathbf{Z}$  are known, non-random design matrices with dimensions  $n \times p$ ,  $n > p$  and  $n \times q$ ,  $n > q$  respectively. We assume throughout that both  $\mathbf{X}$  and  $\mathbf{Z}$  have full column rank. Then

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon},$$

$$\mathbf{u} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_u), \boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon), \mathbf{u} \text{ independent of } \boldsymbol{\epsilon}. \quad (2.1)$$

The random effects vector  $\mathbf{u}$  and the error vector  $\boldsymbol{\epsilon}$  are unobservable. This model has been studied and applied by, for instance, McCulloch and Searle (2001).

Unknown parameters in the model are  $(\boldsymbol{\beta}, \boldsymbol{\theta})$ , where  $\boldsymbol{\beta} \in \mathbf{B} \subseteq \mathbb{R}^p$ ,  $\boldsymbol{\theta} = (\boldsymbol{\Sigma}_\epsilon, \boldsymbol{\Sigma}_u) \in \tilde{\Theta} \subseteq \Theta =$  the set of all  $(\boldsymbol{\Sigma}_\epsilon, \boldsymbol{\Sigma}_u)$  with  $\boldsymbol{\Sigma}_\epsilon$ ,  $n \times n$ , and  $\boldsymbol{\Sigma}_u$ ,  $q \times q$ , both symmetric and positive definite. Throughout the paper, we assume that  $\boldsymbol{\beta}$  and  $\boldsymbol{\theta}$  do not have common elements, i.e. we assume that  $(\boldsymbol{\beta}, \boldsymbol{\theta}) \in \mathbf{B} \otimes \tilde{\Theta}$ . We also assume that  $\boldsymbol{\Sigma}_\epsilon$  and  $\boldsymbol{\Sigma}_u$  do not have common elements. That is, we sometimes assume that  $\tilde{\Theta} = \tilde{\Theta}_\epsilon \otimes \tilde{\Theta}_u$  where  $\tilde{\Theta}_\epsilon \subseteq \Theta_\epsilon$  and  $\tilde{\Theta}_u \subseteq \Theta_u$ , and  $\Theta_\epsilon$  contains all  $n \times n$  positive definite symmetric matrices and  $\Theta_u$  contains all  $q \times q$  positive definite symmetric matrices. We take as our operating definition of identifiability the ability to identify the parameters  $\boldsymbol{\beta}$ ,  $\boldsymbol{\Sigma}_\epsilon$  and  $\boldsymbol{\Sigma}_u$ .

The linear mixed effects model has become popular in the analysis of longitudinal and functional data. For instance, the response of individual  $i$  at time  $t$  can be modelled as  $\sum \alpha_j \phi_j(t) + \sum u_{ik} \psi_k(t)$  plus error, where

the  $\alpha_j$ 's are fixed population effects and the  $u_{ik}$ 's are individual-specific random effects. The inclusion of random effects allows us to realistically model covariance within an individual. Taking  $\phi_j$ 's and  $\psi_k$ 's equal to B-splines allows us to flexibly model a wide variety of responses. In addition, if we allow some of the  $\alpha_j$ 's to be random but with a very specific covariance structure, then the resulting best linear unbiased predictors of the  $\alpha_j$ 's are computationally equivalent to the parameter estimates in a smoothing spline fit or a penalized B-spline fit. See Verbyla (1999) or Ruppert, Wand and Carroll (2003).

The parameter vector  $\beta$  is always identifiable since  $\mathbf{E}\mathbf{y} = \mathbf{X}\beta$  and  $\mathbf{X}$  is of full column rank. So to study identifiability in the normal model, it suffices to study the covariance matrix of  $\mathbf{y}$ ,  $\Sigma_y = \mathbf{Z}\Sigma_u\mathbf{Z}' + \Sigma_\epsilon$ , to see when  $\Sigma_\epsilon$  and  $\Sigma_u$  are identifiable. Thus, nonidentifiability of model (2.1) with parameter space  $\mathbf{B} \otimes \tilde{\Theta}$  is equivalent to the existence of  $\theta, \theta^* \in \tilde{\Theta}$ , with  $\theta \neq \theta^*$  giving the same  $\Sigma_y$ , i.e.

$$\mathbf{Z}\Sigma_u\mathbf{Z}' + \Sigma_\epsilon = \mathbf{Z}\Sigma_u^*\mathbf{Z}' + \Sigma_\epsilon^*. \quad (2.2)$$

In Example 2.2.1 below, we show that the unrestricted model where  $\tilde{\Theta} = \Theta$  is nonidentifiable. A similar argument shows that the restricted model in Example 2.2.2 below is not identifiable. Example 2.2.2 uses the covariance structure assumed for the random effects in the penalized Bspline method. Thus, when using penalized Bsplines, one cannot assume a general form for  $\Sigma_\epsilon$ .

### Example 2.2.1

Suppose we place no further restrictions on  $\Theta$ . Then the model is not identifiable. To see this, let  $(\Sigma_\epsilon, \Sigma_u) \in \Theta$  and  $0 < a < 1$ . Then  $\Sigma_u^* = (1 - a)\Sigma_u$  and  $\Sigma_\epsilon^* = \Sigma_\epsilon + a\mathbf{Z}\Sigma_u\mathbf{Z}'$ . So  $\Sigma_\epsilon^*$  and  $\Sigma_u^*$  satisfy (2.2) and  $(\Sigma_\epsilon^*, \Sigma_u^*) \in \Theta$ .

### Example 2.2.2

Suppose that  $\Theta = \Theta_\epsilon \otimes \tilde{\Theta}_u$  where  $\tilde{\Theta}_u$  contains all matrices of the form  $\sigma^2\mathbf{R}$  where  $\mathbf{R}$  is positive definite and known, and  $\sigma^2 > 0$ . To see that this

model is not identifiable, use the same argument as in Example 2.2.1 and note that the constructed  $\Sigma_u^*$  is in  $\tilde{\Theta}_u$  if  $\Sigma_u$  is in  $\tilde{\Theta}_u$ .

In practice, one usually assumes a more specific structure for  $\Sigma_\epsilon$ , such as  $\Sigma_\epsilon = \sigma_\epsilon^2 \mathbf{I}$ . Restrictions may lead to identifiability, and such restrictions and their effects on identifiability will be discussed in the next two sections.

### 2.3 Simple sufficient conditions of identifiability

In this section, we find sufficient conditions of identifiability of model (2.1) assuming  $\tilde{\Theta} = \Theta$ .

A further examination of (2.2) gives us the following sufficient conditions. Clearly, if  $\Sigma_u$  is known, then  $\mathbf{Z}\Sigma_u\mathbf{Z}'$  is known, and so  $\Sigma_\epsilon$  is completely determined.

If  $\Sigma_\epsilon$  is known, then the model is identifiable. To see this, consider (2.2) with  $\Sigma_\epsilon = \Sigma_\epsilon^*$ . It follows that  $\mathbf{Z}\Sigma_u\mathbf{Z}' = \mathbf{Z}\Sigma_u^*\mathbf{Z}'$  and so  $\Sigma_u = \Sigma_u^*$  since  $\mathbf{Z}$  is of full column rank.

If  $\mathbf{Z}\Sigma_u\mathbf{Z}'\Sigma_\epsilon^{-1} = \mathbf{K}$ , where  $\mathbf{K}$  is known and  $\mathbf{K} + \mathbf{I}$  is of full column rank, then the model is identifiable. Suppose by way of contradiction the model is not identifiable. Then (2.2) holds for  $(\Sigma_u, \Sigma_\epsilon) \neq (\Sigma_u^*, \Sigma_\epsilon^*)$  both in  $\tilde{\Theta}$  with, by assumption,  $\mathbf{Z}\Sigma_u\mathbf{Z}' = \mathbf{K}\Sigma_\epsilon$  and  $\mathbf{Z}\Sigma_u^*\mathbf{Z}' = \mathbf{K}\Sigma_\epsilon^*$ . Substituting these expressions into (2.2) yields

$$\mathbf{K}\Sigma_\epsilon + \Sigma_\epsilon = \mathbf{K}\Sigma_\epsilon^* + \Sigma_\epsilon^*,$$

that is,

$$(\mathbf{K} + \mathbf{I})(\Sigma_\epsilon - \Sigma_\epsilon^*) = \mathbf{0}.$$

Since  $\mathbf{K} + \mathbf{I}$  is of full rank, we must have  $\Sigma_\epsilon = \Sigma_\epsilon^*$ . But, as shown in the previous paragraph, this implies that  $\Sigma_u = \Sigma_u^*$ .

The last condition is similar to a common condition for identifiability in simple linear regression models with measurement errors. The model assumes

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \epsilon_i \sim (0, \sigma_\epsilon^2),$$

where  $x_i$  is observed with error having variance  $\sigma_u^2$ . The response  $y_i$  then has variance  $\sigma_u^2 + \sigma_\epsilon^2$ . One of the common conditions of model identifiability is to assume the ratio  $\sigma_u^2/\sigma_\epsilon^2$  is known. The inverse  $\Sigma_\epsilon^{-1}$  appearing in our last condition could be viewed as multivariate version of “denominator”.

If there are any supplementary data, we may then be able to find an estimate of  $\Sigma_u$ ,  $\Sigma_\epsilon$  or  $\mathbf{K}$  and we can treat this estimate as the true value. The sufficient conditions for identifiability can then be satisfied.

## 2.4 Sufficient conditions of identifiability for a structured $\Sigma_\epsilon$

As we observed from Examples 2.2.1 and 2.2.2, the model is not identifiable even if we restrict  $\Sigma_u$  to be a scalar multiple of a known matrix. In this section, we study the effect of putting restrictions on  $\Sigma_\epsilon$ . In Theorem 2.4.1 below, we give a necessary and sufficient condition of nonidentifiability, a condition that relies mainly on the design matrix  $\mathbf{Z}$  via  $\mathbf{H}_\mathbf{Z} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ . The theorem leads to four corollaries: Corollaries 2.4.1 and 2.4.2 give necessary and sufficient conditions for identifiability when  $\Sigma_\epsilon$  arises from an exchangeable covariance structure or is diagonal. Corollary 2.4.3 states an easily checked condition on  $\Sigma_\epsilon$  that guarantees identifiability of the model. That corollary is then applied to two commonly used error structures. Using Corollary 2.4.4, we can generalize a known identifiability result, giving a shorter proof under weaker conditions.

**Theorem 2.4.1** *Let  $\tilde{\Theta} \subseteq \Theta$  and define  $\mathbf{H}_\mathbf{Z} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ . Then model (2.1) with parameter space  $\mathbf{B} \otimes \tilde{\Theta}$  is nonidentifiable if and only if there exist  $(\Sigma_\epsilon, \Sigma_u) \in \tilde{\Theta}$  and  $(\Sigma_\epsilon^*, \Sigma_u^*) \in \tilde{\Theta}$ , with  $\Sigma_\epsilon^* \neq \Sigma_\epsilon$  such that*

$$\mathbf{H}_\mathbf{Z} [\Sigma_\epsilon - \Sigma_\epsilon^*] = \Sigma_\epsilon - \Sigma_\epsilon^*, \quad (2.3)$$

and

$$\Sigma_u^* = \Sigma_u + (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' [\Sigma_\epsilon - \Sigma_\epsilon^*] \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}. \quad (2.4)$$

**Proof:**

Nonidentifiability of the model is equivalent to the existence of  $(\boldsymbol{\Sigma}_\epsilon, \boldsymbol{\Sigma}_u)$  and  $(\boldsymbol{\Sigma}_\epsilon^*, \boldsymbol{\Sigma}_u^*)$  in  $\tilde{\Theta}$ , not equal, satisfying (2.2). Note that this is equivalent to having  $(\boldsymbol{\Sigma}_\epsilon, \boldsymbol{\Sigma}_u)$  and  $(\boldsymbol{\Sigma}_\epsilon^*, \boldsymbol{\Sigma}_u^*)$  in  $\tilde{\Theta}$  with  $\boldsymbol{\Sigma}_\epsilon^* \neq \boldsymbol{\Sigma}_\epsilon$  satisfying

$$\mathbf{Z}(\boldsymbol{\Sigma}_u - \boldsymbol{\Sigma}_u^*)\mathbf{Z}' = \boldsymbol{\Sigma}_\epsilon^* - \boldsymbol{\Sigma}_\epsilon. \quad (2.5)$$

Suppose the model is nonidentifiable. We premultiply (2.5) by  $\mathbf{Z}'$ , postmultiply it by  $\mathbf{Z}$  and then pre- and postmultiply by  $(\mathbf{Z}'\mathbf{Z})^{-1}$  to get

$$\boldsymbol{\Sigma}_u - \boldsymbol{\Sigma}_u^* = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'[\boldsymbol{\Sigma}_\epsilon^* - \boldsymbol{\Sigma}_\epsilon]\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}. \quad (2.6)$$

This gives (2.4). To derive (2.3), premultiply (2.6) by  $\mathbf{Z}$ , postmultiply (2.6) by  $\mathbf{Z}'$  to get

$$\mathbf{Z}(\boldsymbol{\Sigma}_u - \boldsymbol{\Sigma}_u^*)\mathbf{Z}' = \mathbf{H}_Z[\boldsymbol{\Sigma}_\epsilon^* - \boldsymbol{\Sigma}_\epsilon]\mathbf{H}_Z \quad (2.7)$$

which, by (2.5), is the same as

$$\boldsymbol{\Sigma}_\epsilon - \boldsymbol{\Sigma}_\epsilon^* = \mathbf{H}_Z[\boldsymbol{\Sigma}_\epsilon - \boldsymbol{\Sigma}_\epsilon^*]\mathbf{H}_Z. \quad (2.8)$$

Premultiplying (2.8) by the idempotent matrix  $\mathbf{H}_Z$  gives

$$\mathbf{H}_Z[\boldsymbol{\Sigma}_\epsilon - \boldsymbol{\Sigma}_\epsilon^*] = \mathbf{H}_Z[\boldsymbol{\Sigma}_\epsilon - \boldsymbol{\Sigma}_\epsilon^*]\mathbf{H}_Z.$$

Substituting (2.8) into the right side of the above yields (2.3).

To prove the converse, we want to show that (2.3) and (2.4) lead to (2.5). It is clear from (2.4) that (2.7) holds. If we can show that (2.8) holds then we are done since substituting (2.8) into the right side of (2.7) yields (2.5). To show (2.8), from (2.3) and the symmetry of  $\boldsymbol{\Sigma}_\epsilon - \boldsymbol{\Sigma}_\epsilon^*$ , we see that

$$\mathbf{H}_Z[\boldsymbol{\Sigma}_\epsilon - \boldsymbol{\Sigma}_\epsilon^*] = [\boldsymbol{\Sigma}_\epsilon - \boldsymbol{\Sigma}_\epsilon^*]\mathbf{H}_Z.$$

Premultiplying the above identity by the idempotent matrix  $\mathbf{H}_Z$  gives  $\mathbf{H}_Z[\boldsymbol{\Sigma}_\epsilon - \boldsymbol{\Sigma}_\epsilon^*] = \mathbf{H}_Z[\boldsymbol{\Sigma}_\epsilon - \boldsymbol{\Sigma}_\epsilon^*]\mathbf{H}_Z$ . Substituting (2.3) for the left side of the equation, we see that (2.8) holds.  $\square$

The proofs of the next two corollaries are in Appendix A.

**Corollary 2.4.1** *Let  $\mathbf{1}$  be an  $n$ -vector with each element being one. Suppose that the distribution of  $\epsilon_1, \dots, \epsilon_n$  is exchangeable, that is, the covariance matrix of  $\epsilon$  is of the form  $\sigma^2 [(1 - \rho)\mathbf{I} + \rho\mathbf{J}]$  where  $\mathbf{J} = \mathbf{1}\mathbf{1}'$ . Let  $\tilde{\Theta}_\epsilon = \{\Sigma_\epsilon = \sigma^2 [(1 - \rho)\mathbf{I} + \rho\mathbf{J}], \sigma^2 > 0, -1/(n - 1) < \rho < 1\}$ . Suppose the matrix  $\mathbf{Z}$  satisfies  $\mathbf{1}'\mathbf{Z} \neq \mathbf{0}$  and  $\text{rank}(\mathbf{Z}) = q$  with  $1 \leq q < n - 1$ . Suppose the parameter space is  $\mathbf{B} \otimes \tilde{\Theta}_\epsilon \otimes \Theta_{\mathbf{u}}$ . Then model (2.1) is identifiable if and only if  $\mathbf{H}_Z\mathbf{J} \neq \mathbf{J}$ .*

**Comments.** The condition  $\mathbf{H}_Z\mathbf{J} = \mathbf{J}$  means the sum of the elements of each row of  $\mathbf{H}_Z$  is equal to one, and this is an easy condition to check. For the case that  $q = 1$ , i.e.  $\mathbf{Z}$  is a column vector  $(z_1, \dots, z_n)'$  where  $\sum z_i \neq 0$ ,

$$\mathbf{H}_Z = \begin{pmatrix} \frac{z_1^2}{s_z^2} & \frac{z_1 z_2}{s_z^2} & \dots & \frac{z_1 z_n}{s_z^2} \\ \vdots & & & \vdots \\ \frac{z_n z_1}{s_z^2} & \frac{z_n z_2}{s_z^2} & \dots & \frac{z_n^2}{s_z^2} \end{pmatrix}, \text{ where } s_z^2 = \sum z_i^2.$$

The model is identifiable if and only if  $\mathbf{Z}$  is not a constant vector.

When  $q = 2$ , suppose we have the usual simple linear regression model with centered covariates:

$$\mathbf{Z} = \begin{pmatrix} 1 & z_1 \\ \vdots & \vdots \\ 1 & z_n \end{pmatrix}, \text{ with } \sum z_i = 0. \quad (2.9)$$

Then

$$\mathbf{H}_Z = \begin{pmatrix} \frac{1}{n} + \frac{z_1^2}{s_z^2} & \frac{1}{n} + \frac{z_1 z_2}{s_z^2} & \dots & \frac{1}{n} + \frac{z_1 z_n}{s_z^2} \\ \vdots & & & \vdots \\ \frac{1}{n} + \frac{z_n z_1}{s_z^2} & \frac{1}{n} + \frac{z_n z_2}{s_z^2} & \dots & \frac{1}{n} + \frac{z_n^2}{s_z^2} \end{pmatrix}.$$

and each row of  $\mathbf{H}_Z$  sums to one. Thus unfortunately, the model is not identifiable under this  $\mathbf{Z}$  combined with the exchangeable covariance structure.

**Corollary 2.4.2** *Suppose that  $\tilde{\Theta}_\epsilon$  equals the collection of all diagonal positive definite  $n \times n$  matrices. Then model (2.1) with parameter space  $\mathbf{B} \otimes$*

$\tilde{\Theta}_\epsilon \otimes \Theta_u$  is identifiable if and only none of the diagonal elements of  $\mathbf{H}_Z$  is equal to one.

**Comments.** Again, the condition on  $\mathbf{H}_Z$  is easy to check. Consider the case  $q = 1$ . As we have seen, diagonal elements of  $\mathbf{H}_Z$  equal  $z_i^2 / \sum z_j^2$ ,  $i = 1, \dots, n$ . Therefore, the model is identifiable if and only if  $\mathbf{Z}$  does not have  $n - 1$  zero elements. Consider  $q = 2$  with  $\mathbf{Z}$  as in (2.9). The model is identifiable provided, for all  $i$ ,  $(1/n) + z_i^2 / \sum z_j^2$  doesn't equal 1. So typically, the model is identifiable.

The following corollary provides a sufficient condition for identifiability, a condition that can sometimes be easily checked. Consider (2.3). Note that the rank of  $\mathbf{H}_Z(\Sigma_\epsilon - \Sigma_\epsilon^*)$  is at most  $q$ , since the rank of  $\mathbf{H}_Z$  is  $q$ . Thus, for (2.3) to hold, we must be able to find some  $\Sigma_\epsilon$  and  $\Sigma_\epsilon^*$  with the rank of  $\Sigma_\epsilon - \Sigma_\epsilon^*$  less than or equal to  $q$ . This proves the following.

**Corollary 2.4.3** *Suppose that  $\tilde{\Theta} \subseteq \Theta$ . Then model (2.1) with parameter space  $\mathbf{B} \times \tilde{\Theta}$  is identifiable if  $\text{rank}(\Sigma_\epsilon - \Sigma_\epsilon^*) > q$  for all  $\Sigma_\epsilon, \Sigma_\epsilon^*$  in the parameter space.*

Now we apply Corollary 2.4.3 to show model identifiability under the “multiple of a known positive definite matrix” and the “MA(1)” covariance structures respectively in next two examples.

**Example 2.4.1** *Multiple of a known positive definite matrix*

Fix  $\mathbf{R}$ , symmetric and positive definite, and suppose  $(\Sigma_\epsilon, \Sigma_u) \in \tilde{\Theta}$  implies that  $\Sigma_\epsilon = \sigma^2 \mathbf{R}$ , some  $\sigma^2 > 0$ . Consider  $\Sigma_\epsilon = \sigma^2 \mathbf{R}$  and  $\Sigma_\epsilon^* = \sigma^{*2} \mathbf{R}$ ,  $\sigma^{*2} \neq \sigma^2$ . Clearly  $\Sigma_\epsilon - \Sigma_\epsilon^* = (\sigma^2 - \sigma^{*2}) \mathbf{R}$  is invertible, and so is of rank  $n$ , which we have assumed is greater than  $q$ . Thus, the model is identifiable.

To show the model in Example 2.4.2 below is identifiable, we need the following lemma which is a result in (Graybill, 1983, p.285)

**Lemma 2.4.1** *Let  $\mathbf{T}$  be the  $n \times n$  Toeplitz matrix with ones on the two parallel subdiagonals and zeroes elsewhere. Given two scalars  $a_0$  and  $a_1$ , the*

eigenvalues of the  $n \times n$  matrix  $\mathbf{C} = a_0\mathbf{I} + a_1\mathbf{T}$  are

$$\mu_i = a_0 + 2|a_1| \cos \frac{i\pi}{n+1}, \quad i = 1, \dots, n.$$

**Example 2.4.2** *MA(1)*

Suppose that  $n - 1 > q$ . Let the components of  $\boldsymbol{\epsilon}$  have the MA(1) covariance structure, i.e. of the form  $\sigma^2(\mathbf{I} + \rho\mathbf{T})$ . Let  $\tilde{\Theta}_{\boldsymbol{\epsilon}} = \{\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = \sigma^2(\mathbf{I} + \rho\mathbf{T}), \sigma^2 > 0, |\rho| < 1/2\}$  and suppose  $(\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}, \boldsymbol{\Sigma}_u) \in \tilde{\Theta}$  implies that  $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} \in \tilde{\Theta}_{\boldsymbol{\epsilon}}$ .

Let  $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$  and  $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^* \in \tilde{\Theta}_{\boldsymbol{\epsilon}}$ . By Lemma 2.4.1, the eigenvalues of the difference matrix  $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} - \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^* = (\sigma^2 - \sigma^{*2})\mathbf{I} + (\sigma^2\rho - \sigma^{*2}\rho^*)\mathbf{T}$  are

$$\lambda_i = (\sigma^2 - \sigma^{*2}) + 2 \left| \sigma^2\rho - \sigma^{*2}\rho^* \right| \cos \frac{i\pi}{n+1}, \quad i = 1, \dots, n.$$

Given any  $(\sigma^2, \rho)$  and  $(\sigma^{*2}, \rho^*)$ , with  $(\sigma^2, \rho) \neq (\sigma^{*2}, \rho^*)$ , the number of zero  $\lambda_i$ 's is at most one. Hence, the rank of the difference matrix is greater than or equal to  $n - 1$ . Therefore, model (2.1) is identifiable under this MA(1) covariance structure.

In longitudinal or functional data analysis, usually there are  $N$  individuals with the  $i$ th individual modelled as in (2.1):

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\epsilon}_i,$$

$$\mathbf{u}_i \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}_u), \quad \boldsymbol{\epsilon}_i \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_i}),$$

$$(\boldsymbol{\Sigma}_u, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_i}) \in \Theta_{\mathbf{u}} \otimes \tilde{\Theta}_{\boldsymbol{\epsilon}}^i, \quad \mathbf{u}_i \text{ and } \boldsymbol{\epsilon}_i \text{ independent.} \quad (2.10)$$

Statistical inference is normally based on the joint model, the model of these  $N$  individuals. The following corollary gives sufficient conditions for identifiability of the joint model. The intuition behind the result is that, if we can identify  $\boldsymbol{\Sigma}_u$  from one individual, then we can identify all of the  $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}_i}$ 's.

**Corollary 2.4.4** *If an individual model (2.10) is identifiable, then the joint model is identifiable.*



**Proof:**

We notice each individual model (2.10) shares a common parameter, the covariance matrix  $\Sigma_u$ . If one individual model uniquely determines  $\Sigma_u$  and its  $\Sigma_{\epsilon_i}$ , the identified  $\Sigma_u$  will then yield identifiability of all the individual  $\Sigma_{\epsilon_i}$ 's since, if  $\mathbf{Z}_i \Sigma_u \mathbf{Z}_i' + \epsilon_i = \mathbf{Z}_i \Sigma_u \mathbf{Z}_i' + \epsilon_i^*$ , clearly  $\epsilon_i = \epsilon_i^*$ . Therefore, the joint model is identifiable.  $\square$

Corollary 2.4.4 reduces the verification of a joint model's identifiability to the individuals'. For instance, if the  $i$ -th individual model has  $\mathbf{Z}_i$  of full column rank and  $\Sigma_{\epsilon_i} = \sigma_\epsilon^2 \mathbf{I}_{n_i}$ , where  $n_i$  is the length of  $\mathbf{y}_i$ , then this individual model is identifiable by Corollary 2.4.1 and thus so is the joint model. Note that the other individual models can still have their  $\mathbf{Z}_j$ 's not of full column rank.

Demidenko (2004, Chapters 2 & 3) studies the joint model but assumes the covariance matrix of  $\epsilon_i$  is  $\sigma_\epsilon^2 \mathbf{I}_{n_i}$ . Suppose each  $\mathbf{Z}_i$  is of dimension  $n_i \times q$ . Demidenko shows that the joint model is identifiable if at least one matrix  $\mathbf{Z}_i$  is of full column rank and  $\sum_{i=1}^N (n_i - q) > 0$ . Using our argument in the previous paragraph, the condition  $\sum_{i=1}^N (n_i - q) > 0$  can be dropped. Furthermore, our result can be applied to more general  $\Sigma_\epsilon$ 's.

## 2.5 Extensions

In this section, we discuss identifiability of a model in functional regression for a functional predictor  $y(\cdot)$  and a scalar response  $w$ . We derive a necessary and sufficient condition of nonidentifiability for this model.

We model  $y(t)$  as  $\sum_j \alpha_j \phi_j(t) + \sum_k u_k \psi_k(t)$  plus error, with the  $\phi_j$ 's and  $\psi_k$ 's known, the  $\alpha_j$ 's unknown and the  $u_k$ 's unknown and random. The dependence of the response  $w$  on  $y$  is modelled through an unknown functional coefficient,  $\beta$ :  $w = \beta_0 + \int \beta(t) [y(t) - E(y(t))] dt + \eta$  where  $\eta$  is mean 0 normal noise. Thus, for appropriately defined  $\boldsymbol{\rho}$  and with  $\mathbf{u} = (u_1, \dots, u_q)'$ , we can write

$$w = \beta_0 + \boldsymbol{\rho}' \mathbf{u} + \eta,$$

$$\beta_0 \in \mathbb{R}, \boldsymbol{\rho} \in \mathbb{R}^q \text{ unknown, } \eta \sim (0, \sigma_\eta^2) \text{ independent of } \mathbf{u}. \quad (2.11)$$

The predictor  $y$  is observed at a sequence of discretized points and the observed values are contained in the vector  $\mathbf{y}$ , which then follows model (2.1). James (2002), Müller (2005) and Heckman and Wang (2007) consider this modelling of  $\mathbf{y}$  and  $w$  and propose different approaches to estimate the functional coefficient,  $\beta$ .

We assume model (2.1) and (2.11). For our purpose of identifiability discussion here, we consider the unknown parameters to be  $(\beta_0, \beta)$  and  $\theta = (\Sigma_\epsilon, \Sigma_u, \sigma_\eta^2, \rho)$ . We suppose that  $(\beta_0, \beta) \in \mathbf{B} \subseteq \mathfrak{R} \otimes \mathfrak{R}^p$  and that  $\theta \in \tilde{\Theta} \subseteq \Theta = \Theta_\epsilon \otimes \Theta_u \otimes \mathfrak{R}^+ \otimes \mathfrak{R}^q$  where  $\Theta_\epsilon$  and  $\Theta_u$  are as before and  $\mathfrak{R}^+$  is the set of positive real numbers.

To study identifiability, we must study the distribution of the random vector  $(\mathbf{y}', w)$ . We see that  $E(\mathbf{y}) = \mathbf{X}\beta$ ,  $E(w) = \beta_0$  and  $(\mathbf{y}', w)'$  has covariance matrix

$$\Sigma = \begin{pmatrix} \mathbf{Z}\Sigma_u\mathbf{Z}' + \Sigma_\epsilon & \mathbf{Z}\Sigma_u\rho \\ \rho'\Sigma_u\mathbf{Z}' & \rho'\Sigma_u\rho + \sigma_\eta^2 \end{pmatrix}.$$

We know the parameter  $\beta$  is identifiable if the matrix  $\mathbf{X}$  is of full column rank. The identifiability of  $\beta_0$  is also clear. So we focus on identifying the covariance parameters  $\theta$ .

Our discussion in Section 2.2 suggests the unrestricted model won't be identifiable. In fact, we can construct an example following Example 2.2.1 to show the existence of nonidentical  $\theta$  and  $\theta^*$  both in  $\Theta$  such that (2.2) holds and

$$\mathbf{Z}\Sigma_u\rho = \mathbf{Z}\Sigma_u^*\rho^*, \quad (2.12)$$

$$\rho'\Sigma_u\rho + \sigma_\eta^2 = \rho^{*'}\Sigma_u^*\rho^* + \sigma_\eta^{*2}. \quad (2.13)$$

**Example 2.5.1** *Example 2.2.1 continued.*

Let  $0 < a < 1$ ,  $\Sigma_u^*$  and  $\Sigma_\epsilon^*$  be as in Example 2.2.1, and let  $\rho^* = \rho/(1-a)$  and  $\sigma_\eta^{*2} = \sigma_\eta^2 - a\rho'\Sigma_u\rho/(1-a)$ . It is not hard to see (2.12) and (2.13) are satisfied. If, in addition, we restrict  $a < \sigma_\eta^2/(\sigma_\eta^2 + \rho'\Sigma_u\rho)$ , we see that  $\sigma_\eta^{*2}$  is positive.

The following theorem gives a necessary and sufficient condition of non-

identifiability.

**Theorem 2.5.1** *Let  $\tilde{\Theta}_\epsilon \subseteq \Theta_\epsilon$ , let  $\tilde{\Theta} = \tilde{\Theta}_\epsilon \otimes \Theta_u \otimes \mathbb{R}^+ \otimes \mathbb{R}^q$ . and define  $\mathbf{H}_Z = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ . Then model (2.1) and (2.11) with parameter space  $\mathbf{B} \otimes \tilde{\Theta}$  is nonidentifiable if and only if there exist  $(\Sigma_\epsilon, \Sigma_u, \sigma_\eta^2, \rho)$  and  $(\Sigma_\epsilon^*, \Sigma_u^*, \sigma_\eta^{*2}, \rho^*)$  both in  $\tilde{\Theta}$ , with  $\Sigma_\epsilon^* \neq \Sigma_\epsilon$  such that the following hold*

- (a)  $\mathbf{H}_Z [\Sigma_\epsilon - \Sigma_\epsilon^*] = \Sigma_\epsilon - \Sigma_\epsilon^*$ ,
- (b)  $\Sigma_u^* = \Sigma_u + (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' [\Sigma_\epsilon - \Sigma_\epsilon^*] \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}$ ,
- (c)  $\rho^* = \Sigma_u^{*-1}\Sigma_u\rho$ ,
- (d)  $\sigma_\eta^{*2} = \sigma_\eta^2 + \rho'\Sigma_u(\Sigma_u^{-1} - \Sigma_u^{*-1})\Sigma_u\rho$ .

**Proof:**

Nonidentifiability of the model is equivalent to the existence of nonidentical  $\theta$  and  $\theta^*$  in  $\tilde{\Theta}$ , satisfying (2.2), (2.12) and (2.13).

First suppose that (2.2), (2.12) and (2.13) hold for some  $\theta \neq \theta^*$ . By the argument given in Theorem 2.4.1, (2.2) implies that (a) and (b) hold. Since  $\mathbf{Z}$  is of full column rank, (2.12) yields (c). Substituting (c) into (2.13) yields (d). We easily see that, if  $\theta \neq \theta^*$  and if (a) through (d) hold, then  $\Sigma_\epsilon \neq \Sigma_\epsilon^*$ .

Now suppose that conditions (a) through (d) hold for some  $\Sigma_\epsilon \neq \Sigma_\epsilon^*$ . Again, by the argument given in Theorem 2.4.1, (2.2) holds. We easily see that (2.12) and (2.13) also hold.  $\square$

# Bibliography

- [1] Demidenko, E. (2004) *Mixed Models: Theory and Applications*, John Wiley & Sons.
- [2] Graybill, F. A. (1983) *Matrices with Applications in Statistics*, Second Edition, Wadsworth, Inc.
- [3] Heckman, N. and Wang, W. (2007) Linear mixed models for measurement error in functional regression. manuscript, Department of Statistics, University of British Columbia.
- [4] James, G. (2002) Generalized linear models with functional predictors. *Journal of the Royal Statistical Society B* 64, 3, 411-432.
- [5] McCulloch, C. & Searle, S. R. (2001) *Generalized, Linear and Mixed Models*, John Wiley & Sons.
- [6] Müller, H.G. (2005) Functional modeling and classification of longitudinal data. *Scandinavian Journal of Statistics* 32, 223-240.
- [7] Ruppert, D., Wand, M.P. & Carroll, R.J. (2003). *Semiparametric Regression*, Cambridge University Press.
- [8] Verbyla, A. P., Cullis, B. R., Kenward, M. G., & Welham, S. J. (1999). The analysis of designed experiments and longitudinal data by using smoothing splines. *Applied Statistics* 48, 269-311.

# Chapter 3

## Linear mixed models for measurement error in functional regression

### 3.1 Introduction

Regression models with a functional predictor  $Z(\cdot)$  and a scalar response  $Y$  have been extensively studied (see, eg, Ramsay and Silverman, 2005, and references therein). For individual  $i$ , the dependence of  $Y_i$  on  $Z_i$  is modelled as

$$Y_i = \beta_0 + \int_a^b \beta(t) \left[ Z_i(t) - \mathbb{E}(Z_i(t)) \right] dt + e_i. \quad (3.1)$$

The goal is to estimate  $\beta$ .

Data are collected from  $N$  independent individuals, with data on individual  $i$ ,  $i = 1, \dots, N$ , being  $Y_i, Z_{ij} \equiv Z_i(t_{ij})$ ,  $j = 1, \dots, n_i$ . If the  $Z_i$  processes are observed with error, then our data on individual  $i$  are  $Y_i$  and  $z_{ij}, j = 1, \dots, n_i$ , with

$$z_{ij} = Z_i(t_{ij}) + \epsilon_{ij}, \quad \text{Cov}(\epsilon_{i1}, \dots, \epsilon_{in_i}) = \Sigma_{\epsilon_i}. \quad (3.2)$$

Here, we consider data where the  $Z_{ij}$ 's are observed with error which could be measurement error or modelling error, and we model the function  $Z_i$  using random effects with a set of basis functions  $\phi_1, \dots, \phi_K$ . In our

---

<sup>0</sup>A version of this chapter will be submitted for publication. Authors: Heckman, N. and Wang, W.

estimation process, we approximate  $Z_i(\cdot)$  as

$$Z_i(t) = \mu(t) + \sum_{k=1}^K x_{ik} \phi_k(t), \quad (3.3)$$

where  $\mu$  is an arbitrary function and  $\mathbf{x}_i \equiv (x_{i1}, \dots, x_{iK})'$  are independent and normally distributed with mean vector  $\mathbf{0}$  and covariance matrix  $\Sigma_x$ .

This approach was taken by James (2002), Müller (2005) and James and Silverman (2005) for data with and without measurement error. James used  $\phi_k$ 's equal to a basis for natural cubic splines and also used this basis for modelling  $\mu$ . James's approach is similar to that described in Section 3.3. However, we implement a faster algorithm, the ECME (Expectation/Conditional Maximization Either) algorithm in Liu and Rubin (1994), and consider Hessian matrix based and boot-strapped standard errors.

Müller used  $\phi_k$ 's equal to the first  $K$  estimated eigenfunctions of the  $Z_i$  process. Müller's approach can be separated into two parts: the first part uses only the  $z_{ij}$ 's to determine  $\mu$ , the  $\phi_k$ 's and the  $x_{ij}$ 's. The second part incorporates the  $Y_i$ 's to estimate  $\beta_0$  and  $\beta$ . The first part uses the PACE method (principal analysis through conditional expectation) of Yao, Müller and Wang (2005). In PACE, Yao et al. smooth the observed  $z_{ij}$ 's to obtain an estimate  $\hat{\mu}$  of  $\mu$  and then centre the data by subtracting  $\hat{\mu}$ . Next the authors smooth the centred data to estimate the covariance function of the  $Z_i$ 's and then estimate the first  $K$  eigenfunctions and eigenvalues of the estimated covariance function. Denote these estimates by  $\phi_1, \dots, \phi_K$  and  $\hat{\lambda}_1, \dots, \hat{\lambda}_K$ . Let  $\hat{x}_{ik}$  be the best linear unbiased estimate of individual  $i$ 's  $k$ th PC score,  $\hat{x}_{ik} = E(x_{ik} | z_{i1}, \dots, z_{in_i})$ , calculated assuming normality, (3.2) with  $\Sigma_{\epsilon_i} = \sigma^2 \mathbf{I}$  and (3.3) with  $\Sigma_x = \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_K)$ . The second part of Müller's methodology involves regressing  $Y_i$  on  $\hat{x}_{i1}, \dots, \hat{x}_{iK}$  to estimate  $\beta_k = \int \beta(t) \phi_k(t) dt$  and then setting  $\hat{\beta}(t) = \sum \hat{\beta}_k \phi_k(t)$ . Müller justifies the use of this regression by showing that  $E(Y_i | z_{ij}, j = 1, \dots, n_i) = \beta_0 + \sum_1^\infty \beta_k x_{ik}$ , for the true PC scores  $x_{ik}$  calculated with the true eigenfunctions.

Müller's approach has a big computational advantage over James's (2002), in that it can fit the  $Z_i$ 's well with  $K$  fairly small. An added benefit of

this eigenfunction approach is that we focus on “estimable directions” of  $\beta$ . For instance, consider the extreme case where  $Z_i$  can be written exactly as  $Z_i(t) = \mu(t) + \sum_1^K x_{ik}\phi_k(t)$ . So  $Z_i$  has no variability in directions orthogonal to the  $\phi_k$ 's. Since  $\int \beta(t)[Z_i(t) - E(Z_i(t))] dt = \sum_1^K x_{ik} \int \beta(t) \phi_k(t) dt$ , we can only hope to estimate  $\int \beta(t) \phi_k(t) dt$ ,  $k = 1, \dots, K$ . We cannot estimate  $\int \beta(t) f(t) dt$  for  $f$  orthogonal to the  $\phi_k$ 's. This issue was noted by James and Silverman (2005) who handled it by adding a penalty term which penalizes  $\beta$ 's that have  $\int \beta(t)f(t)dt$  big when  $\text{Var}(\int Z(t)f(t)dt)$  is small.

Müller's approach has the disadvantage that it does not fully use the  $Y_i$ 's: the  $x_{ik}$ 's in (3.3) are estimated using only the  $z_{ij}$ 's. Clearly, the  $Y_i$ 's also provide information about the  $x_{ik}$ 's if there is a relationship between  $Y_i$  and  $Z_i$ , that is, if  $\beta \neq 0$ . As James and Silverman (2005) note “It is an interesting feature of this problem that the responses provide additional information in the estimation of the  $Z_i$ 's”. Also, Müller's calculation, that  $E(Y_i|z_{ij}, j = 1, \dots, n_i) = \beta_0 + \sum_1^\infty \beta_k x_{ik}$  does not hold if the eigenvalues or eigenfunctions are incorrect. In particular, the calculation relies on  $\text{Cov}(x_{ij}, x_{ik}) = 0$  for  $j \neq k$ .

We consider a hybrid approach. Like Müller, we use  $\phi_1, \dots, \phi_K$  equal to the first  $K$  estimated eigenfunctions of the  $Z_i$  process. Thus we not only improve on James's choice of  $\phi_k$ 's but also focus on “estimable directions” of  $\beta$ . We then treat these  $\phi_k$ 's as fixed and known in our analysis. We use all of the data, the  $Y_i$ 's and the  $z_{ij}$ 's, to estimate the  $x_{ik}$ 's and we do not place any restrictions on  $\Sigma_x$ . Thus we improve on Müller's procedure, where the  $x_{ik}$ 's are estimated using only the  $z_{ij}$ 's, and  $\Sigma_x$  is assumed diagonal and is estimated completely by the eigenanalysis of the  $z_{ij}$ 's.

Our detailed parameter estimation procedure using the ECME algorithm is in Section 3.3. In this work, we also propose test statistics for hypothesis testing of  $\beta(\cdot)$ . In Section 3.4.1, we consider testing the nullity of the function  $\beta$ , i.e. testing  $H_o : \beta(t) = 0$ , for all  $t \in [a, b]$ . In the two sample situation, we consider testing the equality of a selection  $\beta^s$  and a control  $\beta^c$ , i.e. testing  $H_o : \beta^s(t) = \beta^c(t)$ , for all  $t \in [a, b]$ . We propose a new integrated t-statistic and three test statistics that are more standard. In Section 3.5, we derive expressions of the residuals of the model fit and discuss model diagnostics

based on the analysis of residuals. In Section 3.6.2, we apply the model to analyze a biological data set. In Section 3.7, via a simulation study, we compare our ECME estimate of  $\beta$  to a modification of Müller's (2005) two-stage estimate and we also study the performance of the different test statistics. Our detailed calculations to derive the ECME estimates are in Appendix B.

### 3.2 Notation and Preliminaries

Before we fit the model, we introduce some notation and carry out preliminary calculations. In this section and the next, we suppose that the  $\phi_k$ 's are fixed and known. In practice, however, we will estimate them from an eigenanalysis of the  $z_{ij}$ 's.

For ease, assume  $n_i \equiv n$  and  $t_{ij} = t_j$ . Suppose that (3.1), (3.2) and (3.3) above hold. Let  $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{in})'$ ,  $\boldsymbol{\mu} = (\mu(t_1), \dots, \mu(t_n))'$ ,  $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{in})'$  represent errors and  $\mathbf{z}_i = (z_{i1}, \dots, z_{in})'$  be the observed values. Write

$$\mathbf{z}_i = \mathbf{Z}_i + \boldsymbol{\epsilon}_i \equiv \boldsymbol{\mu} + \mathbf{A}\mathbf{x}_i + \boldsymbol{\epsilon}_i, \quad (3.4)$$

where  $\mathbf{x}_i \stackrel{i.i.d.}{\sim} \text{N}(0, \boldsymbol{\Sigma}_x)$ ,  $\mathbf{A}_{jk} = \phi_k(t_j)$ ,  $j = 1, \dots, n$ ,  $k = 1, \dots, K$ ,

$\boldsymbol{\epsilon}_i \stackrel{i.i.d.}{\sim} \text{N}(0, \sigma_\epsilon^2 \mathbf{R})$ ,  $\mathbf{R}$  known, symmetric and positive definite,

$\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  independent of  $\{\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n\}$ .

If  $\mathbf{R}$  is known, then the model is identifiable (Wang, 2007). However, if  $\sigma_\epsilon^2 \mathbf{R}$  is unknown, then the model is not identifiable (Wang, 2007). Wang also discusses model identifiability under other assumptions on the covariance matrix of  $\boldsymbol{\epsilon}_i$ .

We choose a basis for  $\beta$  and write

$$\beta(t) = \sum_{j=1}^J \beta_j \psi_j(t) \equiv \boldsymbol{\beta}' \boldsymbol{\psi}(t).$$



Typically, we will take  $J = K$  and  $\psi_j = \phi_j$ , but will write for the general case. Thus we can write

$$Y_i = \beta_0 + \boldsymbol{\beta}' \mathbf{T} \mathbf{x}_i + e_i, \quad \text{where } \mathbf{T}_{jk} = \int_a^b \psi_j(t) \phi_k(t) dt \quad \text{and } e_i \sim N(0, \sigma^2). \quad (3.5)$$

We easily see that  $\mathbf{z}_i, \mathbf{x}_i$  and  $Y_i$  are jointly multivariate normal with

$$E(\mathbf{z}_i) = \boldsymbol{\mu}, \quad E(Y_i) = \beta_0, \quad E(\mathbf{x}_i) = \mathbf{0},$$

and

$$\text{Var}(Y_i) \equiv \sigma_Y^2 = \boldsymbol{\beta}' \mathbf{T} \boldsymbol{\Sigma}_x \mathbf{T}' \boldsymbol{\beta} + \sigma^2, \quad \text{Cov}(\mathbf{z}_i) \equiv \boldsymbol{\Sigma}_z = \mathbf{A} \boldsymbol{\Sigma}_x \mathbf{A}' + \sigma_\epsilon^2 \mathbf{R},$$

$$\text{Cov}(\mathbf{z}_i, Y_i) \equiv \boldsymbol{\Sigma}_{z,Y} = \mathbf{A} \boldsymbol{\Sigma}_x \mathbf{T}' \boldsymbol{\beta}, \quad \text{Cov}(\mathbf{z}_i, \mathbf{x}_i') \equiv \boldsymbol{\Sigma}_{z,x} = \mathbf{A} \boldsymbol{\Sigma}_x$$

$$\text{and } \text{Cov}(Y_i, \mathbf{x}_i') \equiv \boldsymbol{\Sigma}_{Y,x} = \boldsymbol{\beta}' \mathbf{T} \boldsymbol{\Sigma}_x. \quad (3.6)$$

Let  $\mathbf{W}_i = (\mathbf{z}_i', Y_i)'$  be the  $i$ th subject's observation and let  $\boldsymbol{\mu}_W$  denote  $E(\mathbf{W}_i)$ .

Let

$$\mathbf{C} = \begin{pmatrix} \mathbf{A} \\ \boldsymbol{\beta}' \mathbf{T} \end{pmatrix} \quad (3.7)$$

and  $\boldsymbol{\Sigma}_d$  be a block diagonal matrix as

$$\boldsymbol{\Sigma}_d = \text{diag}(\sigma_\epsilon^2 \mathbf{R}, \sigma^2). \quad (3.8)$$

Then

$$\text{Cov}(\mathbf{W}_i) \equiv \boldsymbol{\Sigma}_W = \mathbf{C} \boldsymbol{\Sigma}_x \mathbf{C}' + \boldsymbol{\Sigma}_d \quad (3.9)$$

and the log-likelihood of the observed data is

$$\Lambda_N = -\frac{N}{2} \ln \det (\mathbf{C} \boldsymbol{\Sigma}_x \mathbf{C}' + \boldsymbol{\Sigma}_d) - \frac{1}{2} \sum_{i=1}^N (\mathbf{W}_i - \boldsymbol{\mu}_W)' (\mathbf{C} \boldsymbol{\Sigma}_x \mathbf{C}' + \boldsymbol{\Sigma}_d)^{-1} (\mathbf{W}_i - \boldsymbol{\mu}_W)$$

up to an additive constant term. Unknown model parameters are  $\boldsymbol{\theta} = (\boldsymbol{\mu}, \beta_0, \boldsymbol{\Sigma}_x, \sigma_\epsilon^2, \boldsymbol{\beta}, \sigma^2)$ . Directly maximizing  $\Lambda_N$  over  $\boldsymbol{\theta}$  does not give us closed forms of the parameter estimates except for  $\boldsymbol{\mu}$  and  $\beta_0$ . So we need to rely

on iterative numerical methods to find the estimates. We will elaborate on these methods in the next section.

### 3.3 Parameter estimation

In this section, we use the ECME (Expectation/Conditional Maximization Either) algorithm (Liu and Rubin, 1994) to estimate the model parameters in  $\boldsymbol{\theta}$ .

In this balanced design, the MLEs of  $\boldsymbol{\mu}$  and  $\beta_0$  are  $\bar{\mathbf{z}}$  and  $\bar{Y}$  respectively. So throughout we take

$$\boldsymbol{\mu}^{(t)} = \hat{\boldsymbol{\mu}} = \bar{\mathbf{z}} \quad \text{and} \quad \beta_0^{(t)} = \hat{\beta}_0 = \bar{Y} \quad \text{and so} \quad \boldsymbol{\mu}_W^{(t)} = \bar{\mathbf{W}}, \quad t = 0, 1, \dots$$

We estimate the other parameters iteratively and sequentially. Given  $\boldsymbol{\theta}^{(t)}$ , the parameter estimates at iteration  $t$ , we update one component of  $\boldsymbol{\theta}^{(t)}$  at a time, holding the other components fixed. We treat  $(\mathbf{z}_i, Y_i, \mathbf{x}_i)$ ,  $i = 1, \dots, N$ , as the complete data. We update  $\sigma_\epsilon^2$  by finding its EM estimate. That is, we find its estimate by maximizing the conditional expected complete data log-likelihood function, where we condition on the observed data. The other components of  $\boldsymbol{\theta}^{(t)}$  are updated by maximizing  $\Lambda_N$  directly.

Throughout our ECME procedure in Sections 3.3.2-3.3.4, we make the following assumptions.

- (a)  $\mathbf{A}$  is of full column rank;
- (b)  $\mathbf{T}$  is of full row rank;
- (c) there exists no  $\mathbf{u}$  and  $\mathbf{v}$  such that, for all  $i = 1, \dots, n$ ,  $\mathbf{z}_i = \mathbf{u} + \mathbf{v}'\mathbf{x}_i$ ;
- (d) there exists no  $\mathbf{v}$  such, for all  $i = 1, \dots, n$ ,  $Y_i = \bar{Y} + \mathbf{v}'(\mathbf{z}_i - \bar{\mathbf{z}})$ .

The restrictions on  $\mathbf{A}$  and  $\mathbf{T}$  are easily satisfied. Assumption (b) requires  $J$ , the number of the  $\psi_j$  basis functions to be no larger than  $K$ , the number of the  $\phi_k$  basis functions. Typically, we will take  $J = K$  and  $\psi_j = \phi_j$ . Assumptions (c) and (d) are common for data where there is noise. We

notice assumption (a) implies that the matrix  $\mathbf{C}$  defined in (3.7) is also of full column rank.

### 3.3.1 Initial estimates of parameters other than $\boldsymbol{\mu}$ and $\beta_0$

We choose the initial estimates as follows.

We take  $\boldsymbol{\beta}^{(0)}$  to be a vector of zeroes. If this were the true value of  $\boldsymbol{\beta}$ , then we could simply estimate  $\sigma^2$  by the sample variance of the  $Y_i$ 's. To account for the fact that  $\boldsymbol{\beta}$  may not be zero and thus the sample variance of the  $Y_i$ 's would overestimate  $\sigma^2$ , we take  $\sigma^{2(0)}$  equal to 0.6 times the sample variance of the  $Y_i$ 's.

The values of  $\boldsymbol{\Sigma}_x^{(0)}$  and  $\sigma_\epsilon^{2(0)}$  are based on the penalized eigenanalysis of the  $\mathbf{z}_i$ 's sample covariance matrix described in Section 3.6.1. These initial estimates are sensible if  $\mathbf{R}$  is the identity matrix, but can still be used if  $\mathbf{R}$  is not the identity. Roughly, the eigenanalysis in Section 3.6.1 partitions the variability of the  $\mathbf{z}_i$ 's into two parts: variability from the  $Z_i$  process and variability from the noise. Let  $\lambda_1, \dots, \lambda_n$  denote the calculated eigenvalues and suppose the largest  $K$  sufficiently describe the variability in the  $\mathbf{z}_i$ 's. So we will use  $K$  eigenfunctions. A reasonable first estimate of  $\boldsymbol{\Sigma}_x$  is  $\boldsymbol{\Sigma}_x^{(0)}$  diagonal with entries  $\lambda_1, \dots, \lambda_K$ . We take  $\sigma_\epsilon^{2(0)}$  equal to  $\sum_{K+1}^n \lambda_k / (n - K)$ , explaining the remaining variability in the  $z_{ij}$ 's.

Clearly, under assumptions (a)-(d), we can force  $\sigma^{2(0)}$  and  $\sigma_\epsilon^{2(0)}$  to be positive and  $\boldsymbol{\Sigma}_x^{(0)} > \mathbf{0}$ . Given  $\boldsymbol{\theta}^{(t)}$ , we update  $\boldsymbol{\Sigma}_x$ ,  $\sigma_\epsilon^2$ ,  $\boldsymbol{\beta}$ , and  $\sigma^2$ , as described below.

### 3.3.2 Updating $\boldsymbol{\Sigma}_x^{(t)}$

We update  $\boldsymbol{\Sigma}_x^{(t)}$  by maximizing  $\Lambda_N$  over  $\boldsymbol{\Sigma}_x$  while keeping the other parameters fixed. Let  $\mathbf{S}_W = \sum_{i=1}^N (\mathbf{W}_i - \bar{\mathbf{W}})(\mathbf{W}_i - \bar{\mathbf{W}})' / N$ . We show that if  $\sigma^{2(t)}$  and  $\sigma_\epsilon^{2(t)}$  are positive and if  $\mathbf{S}_W - \boldsymbol{\Sigma}_d^{(t)} > \mathbf{0}$ , then our update  $\boldsymbol{\Sigma}_x^{(t+1)}$  is positive definite and using it in the log likelihood instead of  $\boldsymbol{\Sigma}_x^{(t)}$  increases the log likelihood.

With detailed derivation in Section B.4, differentiating  $\Lambda_N$  with respect

to  $\Sigma_x$  and equating to zero yields the first order condition

$$\mathbf{C}'\Sigma_W^{-1}\mathbf{C} = \mathbf{C}'\Sigma_W^{-1}\mathbf{S}_W\Sigma_W^{-1}\mathbf{C}. \quad (3.10)$$

Here,  $\mathbf{C}$  depends on  $\beta^{(t)}$  and  $\Sigma_W = \mathbf{C}\Sigma_x\mathbf{C}' + \Sigma_d^{(t)}$  from (3.9). Equation (3.10) holds provided  $\Sigma_W$  is invertible at the critical value of  $\Sigma_x$ . Since we assume that  $\sigma^{2(t)}$  and  $\sigma_\epsilon^{2(t)}$  are positive,  $\Sigma_d^{(t)}$  is positive definite. So  $\Sigma_W$  will be invertible provided  $\Sigma_x$  is non-negative definite.

We now solve (3.10) for  $\Sigma_x$ , first deriving two useful identities, (3.11) and (3.12). For ease, we drop the hats and superscript  $t$ 's on the parameter estimates that are being held fixed, that is, on  $\hat{\boldsymbol{\mu}}_W, \sigma_\epsilon^{2(t)}, \beta^{(t)}$ , and  $\sigma^{2(t)}$ . Direct multiplication and some manipulation of the left hand side of the following shows that

$$\left(\mathbf{C}'\Sigma_d^{-1}\mathbf{C}\right) \times \left[\left(\mathbf{C}'\Sigma_d^{-1}\mathbf{C}\right)^{-1} + \Sigma_x\right] \mathbf{C}'\Sigma_W^{-1} = \mathbf{C}'\Sigma_d^{-1}.$$

Solving this for  $\mathbf{C}\Sigma_W^{-1}$  yields

$$\mathbf{C}'\Sigma_W^{-1} = \left[\left(\mathbf{C}'\Sigma_d^{-1}\mathbf{C}\right)^{-1} + \Sigma_x\right]^{-1} \left(\mathbf{C}'\Sigma_d^{-1}\mathbf{C}\right)^{-1} \mathbf{C}'\Sigma_d^{-1}. \quad (3.11)$$

Postmultiplying both sides of identity (3.11) by  $\mathbf{C}$  yields

$$\mathbf{C}'\Sigma_W^{-1}\mathbf{C} = \left(\left(\mathbf{C}'\Sigma_d^{-1}\mathbf{C}\right)^{-1} + \Sigma_x\right)^{-1}. \quad (3.12)$$

Substituting (3.11) into the right side of (3.10) and (3.12) into the left side of (3.10) yields

$$\left(\mathbf{C}'\Sigma_d^{-1}\mathbf{C}\right)^{-1} + \Sigma_x = \mathbf{F}\mathbf{S}_W\mathbf{F}',$$

where  $\mathbf{F} = \left(\mathbf{C}'\Sigma_d^{-1}\mathbf{C}\right)^{-1} \mathbf{C}'\Sigma_d^{-1}$ . Note that  $\mathbf{F}$  is of full row rank. Thus, the critical point is

$$\hat{\Sigma}_x = \mathbf{F}\mathbf{S}_W\mathbf{F}' - \left(\mathbf{C}'\Sigma_d^{-1}\mathbf{C}\right)^{-1} = \mathbf{F}(\mathbf{S}_W - \Sigma_d)\mathbf{F}', \quad (3.13)$$

which is strictly positive definite. And so, clearly we have  $\Sigma_W$  invertible at

the critical point.

To see that the updated  $\hat{\Sigma}_x$  leads to an increase in  $\Lambda_N$ , we show that the Hessian matrix,  $\mathbf{H}(\Sigma_x)$  evaluated at  $\hat{\Sigma}_x$  is negative definite. The  $ij$ th element of  $\mathbf{H}(\Sigma_x)$  is the second order partial derivative of  $\Lambda_N$  with respect to the  $i$ th and  $j$ th elements of the vectorized  $\Sigma_x$ . From calculations in Section B.4, we have

$$\mathbf{H}(\hat{\Sigma}_x) = -(N/2) (\hat{\mathbf{D}} \otimes \hat{\mathbf{D}}), \quad \text{where } \hat{\mathbf{D}} = \mathbf{C}' \hat{\Sigma}_W^{-1} \mathbf{C}, \quad (3.14)$$

which is clearly negative definite.

### 3.3.3 Updating $\sigma_\epsilon^{2(t)}$

We update  $\sigma_\epsilon^{2(t)}$ , holding all other parameter estimates fixed, using one E-step and one M-step of the EM algorithm. We show that if  $\sigma_\epsilon^{2(t)}$  and  $\sigma_\epsilon^{2(t)}$  are positive and if  $\Sigma_x^{(t)} > \mathbf{0}$ , then our update  $\sigma_\epsilon^{2(t+1)}$  is positive. Increase of the log likelihood after updating  $\sigma_\epsilon^{2(t)}$  by  $\sigma_\epsilon^{2(t+1)}$  is guaranteed by the property of the EM algorithm.

Recall  $(\mathbf{z}_i, Y_i, \mathbf{x}_i)$ ,  $i = 1, \dots, N$ , are our complete data and  $\mathbf{W}_i \equiv (\mathbf{z}'_i, Y_i)'$ ,  $i = 1, \dots, N$ , are the observed data. In conditional expectations, we let “ $\cdot$ ” stand for the observed data. Abusing notation slightly, we let  $f$  denote a generic density function with the exact meaning clear from the arguments. The E-Step of the EM algorithm calculates  $E_{\boldsymbol{\theta}^{(t)}} \left( \sum_{i=1}^N \ln f(\mathbf{z}_i, Y_i, \mathbf{x}_i) \mid \cdot \right)$  and the M-step maximizes this conditional expectation over  $\sigma_\epsilon^2$  to obtain  $\sigma_\epsilon^{2(t+1)}$ . By the conditional independence of  $\mathbf{z}_i$  and  $Y_i$  given  $\mathbf{x}_i$ ,

$$\ln f(\mathbf{z}_i, Y_i, \mathbf{x}_i) \equiv \ln f(\mathbf{z}_i | \mathbf{x}_i) + \ln f(y_i | \mathbf{x}_i) + \ln f(\mathbf{x}_i).$$

Since only  $\ln f(\mathbf{z}_i | \mathbf{x}_i)$  contains  $\sigma_\epsilon^2$ , we can ignore the last two terms and obtain  $\sigma_\epsilon^{2(t+1)}$  via maximizing  $E_{\boldsymbol{\theta}^{(t)}} \left( \sum_{i=1}^N \ln f(\mathbf{z}_i | \mathbf{x}_i) \mid \cdot \right)$  over  $\sigma_\epsilon^2$ .

From (3.4), we first get

$$\sum_{i=1}^N \ln f(\mathbf{z}_i | \mathbf{x}_i) = -\frac{N}{2} \ln(\det \sigma_\epsilon^2 \mathbf{R}) - \frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^N (\mathbf{z}_i - \boldsymbol{\mu} - \mathbf{A}\mathbf{x}_i)' \mathbf{R}^{-1} (\mathbf{z}_i - \boldsymbol{\mu} - \mathbf{A}\mathbf{x}_i).$$

Following (3.6), we have

$$\text{Cov}(\mathbf{W}_i, \mathbf{x}_i) = \mathbf{C}\boldsymbol{\Sigma}_x \quad (3.15)$$

which then leads to the conditional mean and covariance of  $\mathbf{x}_i$  given  $\mathbf{W}_i$  as

$$\mathbb{E}[\mathbf{x}_i|\mathbf{W}_i] \equiv \boldsymbol{\mu}_{x_i|W_i} = \boldsymbol{\Sigma}_x \mathbf{C}' \boldsymbol{\Sigma}_W^{-1} (\mathbf{W}_i - \boldsymbol{\mu}_W), \quad (3.16)$$

$$\text{Cov}[\mathbf{x}_i|\mathbf{W}_i] \equiv \boldsymbol{\Sigma}_{x|W} = \boldsymbol{\Sigma}_x - \boldsymbol{\Sigma}_x \mathbf{C}' \boldsymbol{\Sigma}_W^{-1} \mathbf{C} \boldsymbol{\Sigma}_x. \quad (3.17)$$

Let

$$\tilde{s} = \sum_{i=1}^N (\mathbf{z}_i - \hat{\boldsymbol{\mu}} - \mathbf{A} \boldsymbol{\mu}_{x_i|W_i}^{(t)})' \mathbf{R}^{-1} (\mathbf{z}_i - \hat{\boldsymbol{\mu}} - \mathbf{A} \boldsymbol{\mu}_{x_i|W_i}^{(t)}).$$

Routine calculations yield

$$\mathbb{E}_{\boldsymbol{\theta}^{(t)}} \left( \sum_{i=1}^N \ln f(\mathbf{z}_i|\mathbf{x}_i) \right) = -\frac{N}{2} \ln(\det \mathbf{R}) - \frac{nN}{2} \ln \sigma_\epsilon^2 - \frac{1}{2\sigma_\epsilon^2} \left[ \tilde{s} + N \text{tr}(\mathbf{R}^{-1} \mathbf{A} \boldsymbol{\Sigma}_{x|W}^{(t)} \mathbf{A}') \right].$$

Differentiating this conditional mean with respect to  $\sigma_\epsilon^2$  and equating the derivative to zero yields

$$\sigma_\epsilon^{2(t+1)} = \frac{1}{nN} \tilde{s} + \frac{1}{n} \text{tr}[\mathbf{R}^{-1} \mathbf{A} \boldsymbol{\Sigma}_{x|W}^{(t)} \mathbf{A}']. \quad (3.18)$$

We show the update  $\sigma_\epsilon^{2(t+1)}$  is positive in the following. The first term in  $\sigma_\epsilon^{2(t+1)}$  is positive, by assumption (c) and the fact that  $\mathbf{R}$  is positive definite. The second term is nonnegative by the following argument.

Using the famous matrix identity

$$(\mathbf{V}\boldsymbol{\Sigma}\mathbf{V}' + \boldsymbol{\Sigma}_0)^{-1} = \boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\Sigma}_0^{-1} \mathbf{V} \left( \boldsymbol{\Sigma}^{-1} + \mathbf{V}' \boldsymbol{\Sigma}_0^{-1} \mathbf{V} \right)^{-1} \mathbf{V}' \boldsymbol{\Sigma}_0^{-1}$$

provided the matrix orders properly defined, we see that

$$\boldsymbol{\Sigma}_{x|W}^{(t)} = \left( \boldsymbol{\Sigma}_x^{(t)-1} + \mathbf{C}^{(t)'} \boldsymbol{\Sigma}_d^{(t)-1} \mathbf{C}^{(t)} \right)^{-1}$$

which is positive definite. Given  $\boldsymbol{\Sigma}_{x|W}^{(t)} > \mathbf{0}$ , assumption (a) then implies

that  $\mathbf{A}\Sigma_x^{(t)}\mathbf{W}\mathbf{A}' \geq \mathbf{0}$ . Together with the fact that  $\mathbf{R} > \mathbf{0}$ , the second term in (3.18) is thus nonnegative.

### 3.3.4 Updating $\boldsymbol{\beta}^{(t)}$ and $\sigma^{2(t)}$

The updates of  $\boldsymbol{\beta}^{(t)}$  and  $\sigma^{2(t)}$  maximize  $\Lambda_N$  over  $\boldsymbol{\beta}$  and  $\sigma^2$ , holding the other parameters fixed. Suppose that  $\sigma_\epsilon^{2(t)} > 0$  and  $\Sigma_x^{(t)} > \mathbf{0}$ . We find unique critical points,  $\hat{\boldsymbol{\beta}}$  and  $\hat{\sigma}^2$ , and show that they increase the log likelihood provided  $\hat{\sigma}^2 > 0$ .

Note that  $\log f(y_i, \mathbf{z}_i) = \log f(y_i|\mathbf{z}_i) + \log f(\mathbf{z}_i)$ , that  $\log f(\mathbf{z}_i)$  doesn't depend on  $\boldsymbol{\beta}$  or  $\sigma^2$ . We also note given  $\mathbf{z}_i$ ,  $y_i$  is normal with mean

$$\mathbb{E}(Y_i|\mathbf{z}_i) \equiv \beta_0 + \boldsymbol{\beta}'\mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu}),$$

and variance

$$\sigma_{Y|z}^2 \equiv \text{Var}(Y_i|\mathbf{z}_i) = \boldsymbol{\beta}'\mathbf{K}\boldsymbol{\beta} + \sigma^2 \quad (3.19)$$

where

$$\mathbf{G} = \mathbf{T}\Sigma_x\mathbf{A}'\Sigma_z^{-1} \quad (3.20)$$

and

$$\mathbf{K} = \mathbf{T}\Sigma_x\mathbf{T}' - \mathbf{T}\Sigma_x\mathbf{A}'\Sigma_z^{-1}\mathbf{A}\Sigma_x\mathbf{T}'.$$

Therefore, to maximize  $\Lambda_N$  with respect to  $\boldsymbol{\beta}$  and  $\sigma^2$ , we maximize

$$\tilde{\Lambda}_N = -\frac{N}{2} \ln(\boldsymbol{\beta}'\mathbf{K}\boldsymbol{\beta} + \sigma^2) - \frac{1}{2(\boldsymbol{\beta}'\mathbf{K}\boldsymbol{\beta} + \sigma^2)} \sum_{i=1}^N (Y_i - \beta_0 - \boldsymbol{\beta}'\mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu}))^2 \quad (3.21)$$

With detailed derivation in Section B.5, equating  $\partial\tilde{\Lambda}_N/\partial\boldsymbol{\beta}$  and  $\partial\tilde{\Lambda}_N/\partial\sigma^2$  to zero yields respectively

$$\frac{1}{\boldsymbol{\beta}'\mathbf{K}\boldsymbol{\beta} + \sigma^2} \sum_{i=1}^N (Y_i - \beta_0 - \boldsymbol{\beta}'\mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu})) \mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu}) = 0 \quad (3.22)$$

$$\frac{1}{(\boldsymbol{\beta}'\mathbf{K}\boldsymbol{\beta} + \sigma^2)^2} \left[ \sum_{i=1}^N (Y_i - \beta_0 - \boldsymbol{\beta}'\mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu}))^2 - N(\boldsymbol{\beta}'\mathbf{K}\boldsymbol{\beta} + \sigma^2) \right] = 0. \quad (3.23)$$

Note that  $\mathbf{G}$  is of full row rank because of the following two observations. First,  $\mathbf{T}$  is of full row rank by assumption (b). Second, the matrix  $\boldsymbol{\Sigma}_z = \boldsymbol{\Sigma}_x^{(t)} + \sigma_\epsilon^{2(t)}\mathbf{R}$  is invertible since it is positive definite.

Let

$$\mathbf{M} = \mathbf{G} \sum_{i=1}^N (\mathbf{z}_i - \boldsymbol{\mu})(\mathbf{z}_i - \boldsymbol{\mu})'\mathbf{G}'.$$

Then, by assumption (c),  $\mathbf{M}$  is positive definite.

Solving (3.22) for  $\boldsymbol{\beta}$  and (3.23) for  $\sigma^2$  gives

$$\begin{aligned} \boldsymbol{\beta}^{(t+1)} &= \hat{\boldsymbol{\beta}} = \mathbf{M}^{-1}\mathbf{G} \sum_{i=1}^N (\mathbf{z}_i - \boldsymbol{\mu})(Y_i - \beta_0) \\ \sigma^{2(t+1)} &= \hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (Y_i - \beta_0 - \boldsymbol{\beta}'\mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu}))^2 - \boldsymbol{\beta}'\mathbf{K}\boldsymbol{\beta}. \end{aligned} \quad (3.24)$$

Unfortunately, we are not guaranteed that  $\hat{\sigma}^2$  is positive. However, in all of our data analyses and simulation studies, the final estimate of  $\sigma^2$  was always positive.

Again, to check if the update increases  $\tilde{\Lambda}_N$ , we show that the Hessian matrix is negative definite. We notice that (3.24) implies

$$\hat{\sigma}_{Y|z}^2 \equiv \hat{\boldsymbol{\beta}}'\mathbf{K}\hat{\boldsymbol{\beta}} + \hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N \left( Y_i - \beta_0 - \hat{\boldsymbol{\beta}}'\mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu}) \right)^2, \quad (3.25)$$

which is positive by assumption (d). With detailed calculation in Section B.5, the Hessian matrix  $\mathbf{H}\tilde{\Lambda}(\boldsymbol{\beta}, \sigma^2)$  when evaluated at  $\hat{\boldsymbol{\beta}}$  and  $\hat{\sigma}^2$  equals

$$\mathbf{H}\tilde{\Lambda}(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) = -\frac{N}{(\hat{\sigma}_{Y|z}^2)^2} \begin{pmatrix} 2\mathbf{K}\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}'\mathbf{K} + \frac{\hat{\sigma}_{Y|z}^2}{N}\mathbf{M} & \mathbf{K}\hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\beta}}'\mathbf{K} & \frac{1}{2} \end{pmatrix}. \quad (3.26)$$

It follows that  $\mathbf{H}\tilde{\Lambda}(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) < \mathbf{0}$  by the following argument. Let  $\mathbf{x}_1 \in \Re^J$  and



$x_2 \in \Re$  with at least one of  $\mathbf{x}_1$  or  $x_2$  non-zero. Direct calculation yields

$$\begin{aligned} & (\mathbf{x}'_1, x_2) \mathbf{H}\Lambda_N(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) \begin{pmatrix} \mathbf{x}_1 \\ x_2 \end{pmatrix} \\ &= -\frac{N}{(\hat{\sigma}_{Y|z}^2)^2} \left[ \left( \frac{x_2}{\sqrt{2}} + \sqrt{2}\mathbf{x}'_1\mathbf{K}\hat{\boldsymbol{\beta}} \right)^2 + \frac{\hat{\sigma}_{Y|z}^2}{N} \mathbf{x}'_1\mathbf{M}\mathbf{x}_1 \right] < 0. \end{aligned}$$

### 3.4 Inference for $\beta$

Given the estimate  $\hat{\boldsymbol{\beta}}$  of  $\boldsymbol{\beta}$ , we estimate the function  $\beta$  by  $\hat{\beta} = \hat{\boldsymbol{\beta}}'\boldsymbol{\psi}$ . If the covariance matrix of  $\hat{\boldsymbol{\beta}}$  is  $\boldsymbol{\Sigma}_\beta$ , then the covariance function of  $\hat{\beta}$ , denoted  $V_\beta$ , is

$$V_\beta(s, t) = \text{Cov}(\hat{\boldsymbol{\beta}}'\boldsymbol{\psi}(s), \hat{\boldsymbol{\beta}}'\boldsymbol{\psi}(t)) = \boldsymbol{\psi}(s)'\boldsymbol{\Sigma}_\beta\boldsymbol{\psi}(t).$$

We base inference for  $\beta$  on  $\hat{\beta}$  and an estimate of  $\boldsymbol{\Sigma}_\beta$ .

We estimate  $\boldsymbol{\Sigma}_\beta$  in two ways, using bootstrap by resampling the observed  $\mathbf{W}_i$ 's with replacement or using the observed Hessian matrix  $\mathbf{H}\tilde{\Lambda}_N(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2)$  defined in (3.26). We take  $\hat{\boldsymbol{\Sigma}}_\beta$  as the  $K \times K$  upper corner of the inverse of  $-\mathbf{H}\tilde{\Lambda}_N(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2)$ . In doing this, we are treating the other parameters as known, ignoring the variability introduced by estimating them. Thus, we expect that we may underestimate  $V_\beta(t, t)$ , while we don't anticipate underestimation with the bootstrap estimate. However, the Hessian-based estimate is very fast to compute.

#### 3.4.1 Hypothesis testing for $\beta$

**Testing that  $\beta \equiv 0$**

To determine if  $Y_i$  depends on  $Z_i(\cdot)$  we test

$$H_o : \beta(t) = 0, \text{ for all } t \in [a, b].$$

We consider three test statistics.

The first test statistic is the generalized likelihood ratio statistic

$$U_l = \sup_{\beta = \mathbf{0}} \Lambda_N - \sup \Lambda_N.$$

The unrestricted supremum of  $\Lambda_N$  is achieved at the ECME estimates described in Section 3.3. We can also use the ECME procedure to calculate the first supremum. To obtain the first supremum, we observe under the restriction  $\beta = \mathbf{0}$ ,  $\mathbf{z}_i$  and  $Y_i$  are independent, with  $\mathbf{z}_i \sim N(\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}_x\mathbf{A}' + \sigma_\epsilon^2\mathbf{R})$  and  $Y_i \sim N(\beta_0, \sigma^2)$ . Thus  $\hat{\boldsymbol{\mu}} = \bar{\mathbf{z}}$ ,  $\hat{\beta}_0 = \bar{Y}$  and  $\hat{\sigma}^2 = \sum(Y_i - \bar{Y})^2/N$ . We then calculate  $\boldsymbol{\Sigma}_x^{(t)}$  and  $\sigma_\epsilon^{2(t)}$  by an ECME method treating these estimates of  $\boldsymbol{\mu}$ ,  $\hat{\beta}$  and  $\sigma^2$  as known. We update  $\boldsymbol{\Sigma}_x$  by maximizing  $\Lambda_N$  directly while holding  $\sigma_\epsilon^2$  fixed. We update  $\sigma_\epsilon^2$  by finding its EM estimate  $\sigma_\epsilon^{2(t+1)}$  while holding  $\boldsymbol{\Sigma}_x$  fixed. We iterate until convergence occurs.

The second statistic considered is Wald's test statistic using  $\hat{\beta}$ , the vector of estimated basis coefficients:

$$U_w = \hat{\beta}' \hat{\boldsymbol{\Sigma}}_\beta^{-1} \hat{\beta}.$$

It is interesting to note that this test statistic can be re-written in terms of a vector of function evaluations of  $\hat{\beta}$ . To see this, let  $t_i^*, i = 1, \dots, n^*$ , be a sequence of time points, let  $\tilde{\beta}$  be the vector containing the values of  $\hat{\beta}$  at the  $t_i^*$ 's, and let  $\tilde{\boldsymbol{\Sigma}}_\beta$  be  $\tilde{\beta}$ 's covariance matrix. The Wald test statistic based on  $\tilde{\beta}$  is  $\tilde{\beta}' \hat{\boldsymbol{\Sigma}}_{\tilde{\beta}}^+ \tilde{\beta}$ , where  $\hat{\boldsymbol{\Sigma}}_{\tilde{\beta}}^+$  is the Moore-Penrose inverse of an estimate of  $\boldsymbol{\Sigma}_{\tilde{\beta}}$ . We now argue that, under mild conditions on the  $t_i^*$ 's,  $\tilde{\beta}' \hat{\boldsymbol{\Sigma}}_{\tilde{\beta}}^+ \tilde{\beta} = U_w$ . Define the  $n^* \times J$  matrix  $\boldsymbol{\Psi}$  as  $\boldsymbol{\Psi}_{ij} = \psi_j(t_i^*)$  and suppose that  $\boldsymbol{\Psi}$  is of full column rank. Since  $\tilde{\beta} = \boldsymbol{\Psi}\hat{\beta}$ ,  $\boldsymbol{\Sigma}_{\tilde{\beta}} = \boldsymbol{\Psi}\boldsymbol{\Sigma}_\beta\boldsymbol{\Psi}'$  and thus it is natural to take  $\hat{\boldsymbol{\Sigma}}_{\tilde{\beta}} = \boldsymbol{\Psi}\hat{\boldsymbol{\Sigma}}_\beta\boldsymbol{\Psi}'$ . Since  $\hat{\boldsymbol{\Sigma}}_{\tilde{\beta}}^+ = \boldsymbol{\Psi}^+\hat{\boldsymbol{\Sigma}}_\beta^{-1}\boldsymbol{\Psi}^+$  and  $\boldsymbol{\Psi}^+\boldsymbol{\Psi} = \mathbf{I}$ ,  $\tilde{\beta}' \hat{\boldsymbol{\Sigma}}_{\tilde{\beta}}^+ \tilde{\beta} = U_w$ .

The third statistic is the integrated t-statistic

$$U_f = \int_a^b \frac{\hat{\beta}^2(t)}{\hat{V}_\beta(t, t)} dt.$$

To calculate the null distribution of a test statistic, and thus calculate a p-value, we use a permutation type method, one that does not rely

on distributional assumptions. Under the null hypothesis that  $\beta(t) \equiv 0$  for all  $t \in [a, b]$ ,  $Y_i$  and  $\mathbf{z}_i$  are independent. Therefore the joint distribution of  $(\mathbf{z}_i, Y_i)$  is the same as that of  $(\mathbf{z}_i, Y_{i'})$  where  $i'$  is chosen at random from  $\{1, \dots, N\}$ . To simulate the null distribution of a test statistic, we make  $Q$  new data sets via permuting the  $Y_i$ 's: the  $q$ th data set is simply  $\{\mathbf{z}_1, Y_{i_1}, \dots, \mathbf{z}_N, Y_{i_N}\}$  where  $i_1, \dots, i_N$  is a random permutation of  $1, \dots, N$ . The p-value is then calculated as the proportion of the resulting  $Q$  statistic values larger than the original observed value.

### Testing equality of two $\beta$ 's

Often we want to know if the  $\beta$ 's governing  $Y$ 's and  $Z$ 's in two different groups are equal. To denote group membership, we use the superscript “ $s$ ” or “ $c$ ” to indicate the “selection” or “control” group respectively. We have data collected independently from the two groups and we want to test

$$H_o : \beta^s(t) = \beta^c(t), \text{ for all } t \in [a, b].$$

We consider four test statistics. We assume that the selection and control log-likelihoods,  $\Lambda_{N^s}^s$  and  $\Lambda_{N^c}^c$ , each have the same expressions as in (3.21) but with possibly different parameter values, superscripted by  $s$  or  $c$ .

The first test statistic is from a likelihood ratio test

$$\begin{aligned} U_l &= \sup_{\beta^s = \beta^c} \{\Lambda_{N^s}^s + \Lambda_{N^c}^c\} - \sup\{\Lambda_{N^s}^s + \Lambda_{N^c}^c\} \\ &= \sup_{\beta^s = \beta^c} \{\Lambda_{N^s}^s + \Lambda_{N^c}^c\} - \sup \Lambda_{N^s}^s - \sup \Lambda_{N^c}^c. \end{aligned}$$

Each of the last two suprema is calculated separately, using the ECME estimates from Section 3.3.

We can also apply this ECME procedure to calculate the first supremum. Under the restriction  $\beta^s = \beta^c$ ,  $\Lambda_{N^s}^s$  and  $\Lambda_{N^c}^c$  have a common parameter  $\beta \equiv \beta^s = \beta^c$ . Following the same argument as in Section 3.3, we have

$$\boldsymbol{\mu}^{c(t)} = \bar{\mathbf{z}}^c, \quad \boldsymbol{\mu}^{s(t)} = \bar{\mathbf{z}}^s, \quad \beta_0^{c(t)} = \bar{Y}^c \quad \text{and} \quad \beta_0^{s(t)} = \bar{Y}^s.$$

To update each of  $\Sigma_x^{c(t)}$  and  $\Sigma_x^{s(t)}$ , we follow the steps outlined in Section 3.3.2. To update each of  $\sigma_\epsilon^{c2(t)}$  and  $\sigma_\epsilon^{s2(t)}$ , we follow the steps outlined in Section 3.3.3.

To update  $\boldsymbol{\beta}^{(t)}$ ,  $\sigma^{s2(t)}$  and  $\sigma^{c2(t)}$ , we proceed as in Section 3.3.4, but with some modification for our updating of  $\boldsymbol{\beta}^{(t)}$ .

To describe the procedure to update  $\boldsymbol{\beta}^{(t)}$ ,  $\sigma^{s2(t)}$  and  $\sigma^{c2(t)}$ , define  $\tilde{\Lambda}_{N^s}^s$  and  $\tilde{\Lambda}_{N^c}^c$  in a manner analagous to  $\tilde{\Lambda}_N$  in (3.21). By an argument similar to that in Section 3.3.4, we must find  $\boldsymbol{\beta}$ ,  $\sigma^{s2}$  and  $\sigma^{c2}$  to maximize  $\tilde{\Lambda}_{N^s}^s + \tilde{\Lambda}_{N^c}^c$ . Differentiating  $\tilde{\Lambda}_{N^s}^s + \tilde{\Lambda}_{N^c}^c$  with respect to  $\sigma^{s2}$  and  $\sigma^{c2}$  and setting equal to zero yields equations analagous to the equation for  $\sigma^{2(t+1)}$  in (3.24).

Unfortunately, differentiating  $\tilde{\Lambda}_{N^s}^s + \tilde{\Lambda}_{N^c}^c$  with respect to  $\boldsymbol{\beta}$  and setting equal to zero yields an intractable equation. So we modify our calculation of  $\boldsymbol{\beta}$ , but retain the above-described updates for  $\sigma^{s2}$  and  $\sigma^{c2}$ . Instead of maximizing  $\tilde{\Lambda}_{N^s}^s + \tilde{\Lambda}_{N^c}^c$  with respect to  $\boldsymbol{\beta}$  we maximize  $\tilde{\tilde{\Lambda}}_{N^s}^s + \tilde{\tilde{\Lambda}}_{N^c}^c$  with respect to  $\boldsymbol{\beta}$ , where the  $\tilde{\tilde{\Lambda}}$ 's are defined as follows. First consider the term  $\boldsymbol{\beta}'\mathbf{K}^s\boldsymbol{\beta} + \sigma^{s2}$ . Calculate the matrix  $\mathbf{K}^s$  using  $\Sigma_x^{s(t+1)}$  and  $\sigma_\epsilon^{s2(t+1)}$ . Take  $\boldsymbol{\beta} = \boldsymbol{\beta}^{(t)}$  and  $\sigma^{s2} = \sigma^{s2(t)}$ . Then the resulting expression for  $\boldsymbol{\beta}'\mathbf{K}^s\boldsymbol{\beta} + \sigma^{s2}$ , which we denote  $\sigma_{Y|z}^{s2(t)}$ , no longer contains the parameters  $\boldsymbol{\beta}$  and  $\sigma^{s2}$ . Let

$$\tilde{\tilde{\Lambda}}_{N^s}^s = -\frac{N^s}{2} \ln(\sigma_{Y|z}^{s2(t)}) - \frac{1}{2\sigma_{Y|z}^{s2(t)}} \sum (Y_i^s - \beta_0^s - \boldsymbol{\beta}'\mathbf{G}^s(\mathbf{z}_i^s - \boldsymbol{\mu}^s))^2.$$

Define  $\sigma_{Y|z}^{c2(t)}$  and  $\tilde{\tilde{\Lambda}}_{N^c}^c$  similarly.

Differentiating  $\tilde{\tilde{\Lambda}}_{N^s}^s + \tilde{\tilde{\Lambda}}_{N^c}^c$  with respect to  $\boldsymbol{\beta}$  and setting equal to zero yields the update

$$\boldsymbol{\beta}^{(t+1)} = \left( \frac{1}{\sigma_{Y|z}^{s2(t)}} \mathbf{M}^s + \frac{1}{\sigma_{Y|z}^{c2(t)}} \mathbf{M}^c \right)^{-1} \left( \frac{1}{\sigma_{Y|z}^{s2(t)}} \mathbf{G}^s \sum (\mathbf{z}_i^s - \boldsymbol{\mu}^s)(Y_i^s - \beta_0^s) + \frac{1}{\sigma_{Y|z}^{c2(t)}} \mathbf{G}^c \sum (\mathbf{z}_i^c - \boldsymbol{\mu}^c)(Y_i^c - \beta_0^c) \right).$$

where

$$\mathbf{M}^s = \mathbf{G}^s \sum (\mathbf{z}_i^s - \boldsymbol{\mu}^s)(\mathbf{z}_i^s - \boldsymbol{\mu}^s)' \mathbf{G}^{s'}$$

and  $\mathbf{M}^c$  is defined similarly.

To define the two sample Wald's statistic  $U_w$ , let  $\hat{\boldsymbol{\beta}}^s$  be the estimate of  $\boldsymbol{\beta}^s$  with estimated covariance matrix  $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}^s$  and define  $\hat{\boldsymbol{\beta}}^c$  and  $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}^c$  similarly. The two sample statistic is

$$U_w = (\hat{\boldsymbol{\beta}}^s - \hat{\boldsymbol{\beta}}^c)' \left( \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}^s} + \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}^c} \right)^{-1} (\hat{\boldsymbol{\beta}}^s - \hat{\boldsymbol{\beta}}^c).$$

We consider a two sample statistic based on function evaluations, rather than on basis coefficients, recalling the notation in Section 3.4.1. Let  $\tilde{\boldsymbol{\beta}}^s = \boldsymbol{\Psi}^s \hat{\boldsymbol{\beta}}^s$  be the vector of function evaluations of  $\hat{\boldsymbol{\beta}}^s$  at a sequence of time points. Let  $\hat{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{\beta}}^s} = \boldsymbol{\Psi}^s \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}^s \boldsymbol{\Psi}^{s'}$  be the estimate of  $\tilde{\boldsymbol{\beta}}^s$ 's covariance matrix. Similarly define  $\tilde{\boldsymbol{\beta}}^c$  and  $\hat{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{\beta}}^c}$ . The two sample Wald test statistic based on  $\tilde{\boldsymbol{\beta}}^s$  and  $\tilde{\boldsymbol{\beta}}^c$  is

$$U_e = (\tilde{\boldsymbol{\beta}}^s - \tilde{\boldsymbol{\beta}}^c)' \left( \hat{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{\beta}}^s} + \hat{\boldsymbol{\Sigma}}_{\tilde{\boldsymbol{\beta}}^c} \right)^+ (\tilde{\boldsymbol{\beta}}^s - \tilde{\boldsymbol{\beta}}^c).$$

In Section 3.4.1, the one sample situation, we argued that we needn't use the function evaluation statistic  $U_e$  as it is equivalent to the Wald test statistic  $U_w$  under mild conditions. In the two sample situation here, however, the two sample  $U_w$  and  $U_e$  may not agree unless  $\boldsymbol{\Psi}^s = \boldsymbol{\Psi}^c$  and both of them are of full column rank.

To define the two sample integrated t-statistic  $U_f$ , let  $\hat{V}_{\boldsymbol{\beta}^s}(s, t) = \boldsymbol{\psi}^s(s)' \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}^s \boldsymbol{\psi}^s(t)$  be the estimate of the covariance function of  $\hat{\boldsymbol{\beta}}^s$  and define  $\hat{V}_{\boldsymbol{\beta}^c}(s, t)$  similarly. The two sample  $U_f$  is

$$U_f = \int \frac{[\hat{\boldsymbol{\beta}}^s(t) - \hat{\boldsymbol{\beta}}^c(t)]^2}{\hat{V}_{\boldsymbol{\beta}^s}(t, t) + \hat{V}_{\boldsymbol{\beta}^c}(t, t)} dt.$$

Again, we use the permutation method to calculate the null distribution of the test statistics and thus the p-values. Under the null hypothesis that  $\boldsymbol{\beta}^s(t) = \boldsymbol{\beta}^c(t)$  for all  $t \in [a, b]$ , the dependence of  $Y_i$  on  $\mathbf{z}_i$  is identical in both groups. We generate  $Q$  "data sets" from the original data set, data sets that follow the null hypothesis. To construct the  $q$ th "data set", we

randomly split the  $N^s + N^c$  individuals into two groups of size  $N^s$  and  $N^c$  and calculate the resulting test statistic. We use the empirical distribution of the obtained  $Q$  test statistic values to approximate the null distribution of our statistic. The p-value is then calculated as the proportion of the  $Q$  statistic values larger than the original observed value.

### 3.5 Model assumption checking

After fitting the model, we need to check if the model assumptions are satisfied. Our model diagnostics rely on the analysis of residuals. In this section, we derive expressions of the fitted values for  $\mathbf{W}_i$  and for the residuals. Fitted values and residuals for  $\mathbf{z}_i$  and  $Y_i$  are then obtained as components. We can then plot the residuals to check model assumptions and to look for outliers and influential points.

To simplify notation, unknown parameters below stand for their estimates. Using model (3.4) and (3.5), we base our fitted values  $\hat{\mathbf{W}}_i$  on the BLUP of the random effects  $\mathbf{x}_i$ ,  $\boldsymbol{\mu}_{x_i|W_i}$  in (3.16),

$$\begin{aligned}\hat{\mathbf{W}}_i &= \boldsymbol{\mu}_W + \mathbf{C}\boldsymbol{\mu}_{x_i|W_i} \\ &= \boldsymbol{\mu}_W + \mathbf{C}\boldsymbol{\Sigma}_x\mathbf{C}'\boldsymbol{\Sigma}_W^{-1}(\mathbf{W}_i - \boldsymbol{\mu}_W)\end{aligned}$$

Recall the expression of  $\boldsymbol{\Sigma}_W$  in (3.9). We get

$$\mathbf{C}\boldsymbol{\Sigma}_x\mathbf{C}'\boldsymbol{\Sigma}_W^{-1} = (\mathbf{C}\boldsymbol{\Sigma}_x\mathbf{C}' + \boldsymbol{\Sigma}_d - \boldsymbol{\Sigma}_d)\boldsymbol{\Sigma}_W^{-1} = \mathbf{I} - \boldsymbol{\Sigma}_d\boldsymbol{\Sigma}_W^{-1}$$

and thus

$$\hat{\mathbf{W}}_i = (\mathbf{I} - \boldsymbol{\Sigma}_d\boldsymbol{\Sigma}_W^{-1})\mathbf{W}_i + \boldsymbol{\Sigma}_d\boldsymbol{\Sigma}_W^{-1}\boldsymbol{\mu}_W.$$

It then follows

$$\mathbf{r}_i = \mathbf{W}_i - \hat{\mathbf{W}}_i = \boldsymbol{\Sigma}_d\boldsymbol{\Sigma}_W^{-1}(\mathbf{W}_i - \boldsymbol{\mu}_W).$$

The last element of  $\hat{\mathbf{W}}_i$  gives the fitted value of  $Y_i$  and the last element of  $\mathbf{r}_i$  gives the  $Y_i$  residual. As we focus on modelling the dependence of  $Y_i$  on

$Z_i$ , plotting residuals against fitted values of  $Y_i$  is useful to detect outliers or influential points of the model fit.

## 3.6 Model application

In this section, we analyze a biological data set using our model. First we give a general description of how to choose the basis functions  $\phi_1, \dots, \phi_K$ . Then we conduct the data analysis.

### 3.6.1 Choosing the basis functions

We choose the basis functions, the  $\phi_k$ 's, as functions that give good approximations to  $Z_i(\cdot)$ : we choose them as the first few estimated eigenfunctions of the  $Z_i$  process. To do so, we apply a smoothed principle component analysis to the observed  $\mathbf{z}_i$ 's, penalizing an approximation of the second derivative of the eigenfunction. Let  $\hat{\Sigma}$  be the sample covariance matrix of the  $\mathbf{z}_i$ 's. We find a sequence of orthogonal vectors  $\tilde{\mathbf{v}}_j \in \mathbf{R}^n$  to maximize

$$\frac{\mathbf{v}'_j \hat{\Sigma} \mathbf{v}_j}{\mathbf{v}'_j \mathbf{v}_j + \lambda \mathbf{v}'_j \mathbf{D}' \mathbf{D} \mathbf{v}_j},$$

where  $\lambda$  is the smoothing parameter and  $\mathbf{D} \mathbf{v}_j$  calculates second divided differences of the vector  $\mathbf{v}_j$ . The  $(n-2) \times n$  matrix  $\mathbf{D}$  depends on  $t_1, \dots, t_n$  and is defined to differentiate quadratic functions exactly. That is, if  $\mathbf{v}_j[i] = a + bt_i + ct_i^2$ , then  $(\mathbf{D} \mathbf{v}_j)[i] = 2c$ . Given  $\lambda$ , the vectors  $\tilde{\mathbf{v}}_j$  are eigenvectors of the matrix  $\mathbf{G}^{-1/2} \hat{\Sigma} \mathbf{G}^{-1/2}$ , where  $\mathbf{G} = \mathbf{I} + \lambda \mathbf{D}' \mathbf{D}$ . The approach is similar to Ramsay and Silverman's (2005) but we don't use basis function expansions of the  $Z_i$ .

Choice of  $\lambda$  can be done by cross-validation, but for simplicity here, we select  $\lambda$  by examining the smoothness of the resulting  $\tilde{\mathbf{v}}_j$ 's. In the data analysis, we chose  $\lambda = 100$ .

### 3.6.2 Data description

Data were provided by Patrick Carter, School of Biological Sciences, Washington State University, and are described in Morgan et al, 2003. Data are from mice divided into four groups according to gender and the two treatments “selection” and “control”. The selection group mice were bred over 16 generations, with selection being on high wheel-running activity at age eight weeks. Control mice were bred at random. In the final generation, body mass and wheel running activity were recorded for each mouse for sixty two consecutive weeks, indexed from  $-1$  to  $60$ , except for weeks  $34$ ,  $38$ ,  $39$ ,  $50$ . The research interest is to know how body mass and wheel running are related and if the relationship depends on the treatment.

The wheel running distance data have many missing values and are very noisy. In addition, the wheels were cleaned every four weeks, and so we see spikes in wheel-running activity every fourth week. So in our analysis, we average the nonmissing wheel running distances over weeks  $5$  to  $60$  and use this average as the response  $Y$ . The predictor  $Z(\cdot)$  is the log transformed body mass. We want to know if any of the groups have  $\beta$  non-zero and if there is any difference between the selection  $\beta^s$  and the control  $\beta^c$  within each gender.

Plots of the observed  $\mathbf{z}_i$ 's and histograms of the  $Y_i$ 's in each group are in Figures 3.1 and 3.2 respectively. We see that log body mass is roughly monotone with a high rate of increase in weeks  $-1$  to  $4$ . The log body mass in the males is more variable than that in the females.

### 3.6.3 Choice of basis functions

A smoothed eigenanalysis in each of the four groups yielded a first eigenfunction that was close to constant, indicating that the biggest source of variability in log body mass in each group was overall size of the mouse. Since a constant eigenfunction is biologically meaningful, we forced our first basis function to be constant and, within each group, calculated the remaining functions via a smoothed eigenanalysis on the centered log body mass



as follows. We let

$$\bar{z}_i = \frac{1}{58} \sum_{\substack{k=-1 \\ k \neq 34, 38, 39, 50}}^{60} z_{ik}$$

be the  $i$ th mouse's average log body mass and

$$z_{ij} - \bar{z}_i \quad j = -1, \dots, 60, j \neq 34, 38, 39, 50$$

the  $i$ th mouse's centered log body masses. Within each group, we calculated the sample covariance matrix of the centered log body mass vectors. We then applied the smoothed eigenanalysis to this covariance matrix. Figure 3.3 shows the proportion of cumulative variance explained by the first ten principal components of this analysis in each group. The variance explained by the constant function was 85% in the male selection group, 72% in the male control group, 70% in the female selection group and 63% in the female control group. Figure 3.4 shows the constant function and the first three eigenfunctions for each group. They are displayed one by one with the four groups' functions together in one panel. There we see the three smoothed eigenfunctions are very similar in the four groups.

Figure 3.3 suggests that, if we choose the first three smoothed eigenfunctions in each group, we will capture about 90% of the variability of the log body mass trajectories, beyond the variability captured by the constant function. We will use a constant function plus these three eigenfunctions in our analysis.

### 3.6.4 Estimation and residual analysis

Before proceeding with inference, we study residuals of our fits to see if there are any outliers.

We fit the model (3.4) and (3.5) within each group, using  $R = \mathbf{I}$ , and choose the same basis functions for  $\beta$  as for the log body mass, calculated in Section 3.6.3. Within each group, we plot the  $Y_i$  residuals against the fitted  $Y_i$ 's to detect outliers as outlined in Section 3.5. Residual plots of each group are in Figure 3.5. In the control males, we see an outlier at the left

side. This outlier may be influential in the estimation of  $\beta$ . So we remove this outlier and refit the model to the male control group. The first panel of Figure 3.6 is the same as the plot in the second panel of Figure 3.5, showing the outlier in the control males. The second panel of Figure 3.6 shows the residual plot of the fit after removing the outlier. There we see no new outliers. We remove the one outlier in the control males in our subsequent analyses.

### 3.6.5 Inference for $\beta(\cdot)$

In each of the four groups, we estimated  $\beta$  and calculated standard errors from both the Hessian matrix and the bootstrap, as described in Section 3.4. These are shown in Figure 3.7.

To determine if  $\beta(t) = 0$  for all  $t \in [-1, 60]$ , we can study the plots in Figure 3.7 and we can conduct a formal hypothesis test. From the plots in Figure 3.7, we see that, except for the female control group, all groups show a region where the zero line is not within one standard error of  $\hat{\beta}$ . This suggests that, within these groups, perhaps there may be a dependence of averaged wheel running on log body mass.

We conduct a hypothesis test of  $H_o : \beta(t) = 0$  for all  $t \in [-1, 60]$  in each group using the test statistics in Section 3.4.1. We compute the standard errors using the Hessian matrix (3.26). The results are in Table 3.1 below. The last three columns contain the permutation p-values of  $U_l$ ,  $U_w$  and  $U_f$ , computed by using 500 permutations. The second column gives the observed value of the test statistic  $U_l$  and the next column gives the p-value of  $U_l$  based on the fact that negative twice the log-likelihood ratio statistic is asymptotically chi-squared distributed with 4 degrees of freedom.

From the Table, we see that, in selected and control males, average wheel-running depends on the log body mass trajectories. However, there is insufficient evidence to make this claim in the other three groups.

We observe that the p-values of  $U_l$  and  $U_w$  are identical. We plot the permuted values of  $U_l$  and  $U_w$  in Figure 3.8 and see that  $U_l$  is an increasing function of  $U_w$ . This may be due to the normality assumption in the model.

### 3.6.6 Inference for $\beta^s - \beta^c$

Within each gender, we want to compare the selection  $\beta(\cdot)$  with the control  $\beta(\cdot)$ . Figure 3.9 shows the pointwise difference between the selection  $\hat{\beta}$  and the control  $\hat{\beta}^c$  together with the standard errors computed from the Hessian matrix and from the bootstrap. For both males and females, the region within one standard error of  $\hat{\beta}^s - \hat{\beta}^c$  contains much of the zero line, which suggests we can't distinguish between the selection  $\beta$  and the control  $\beta$ .

Table 3.2 gives results of the test  $H_o : \beta^s(t) = \beta^c(t)$ , for all  $t \in [-1, 60]$  in each gender. The last three columns contain the permutation p-values of  $U_l$ ,  $U_w$  and  $U_f$ , computed by using 500 permutations. The second column gives the observed value of the test statistic  $U_l$  and the next column gives the p-value of  $U_l$  based on the fact that negative twice the log-likelihood ratio statistic is asymptotically chi-squared distributed with 4 degrees of freedom.

Given the results, we can not reject  $H_o$  in either gender. That is, within each gender, there is no evidence of a difference between the selected group and the control group in terms of average wheel running's dependence on log body mass.

In conclusion, we find a strong dependence of average wheel running on the log body mass in the selected males and control males. We don't have enough evidence to distinguish the difference between the selected group and the control group in terms of average wheel running's dependence on log body mass in either gender.

## 3.7 Simulation study

In the simulation study, we compare the pointwise mean squared errors of our ECME estimate of  $\beta$  with those of a modified version of the two stage estimate proposed by Müller (2005). We also compare the power of the test statistics proposed in Section 3.4.1 for one-sample and two-sample comparisons.

We will see that, in terms of mean squared error, in the one-sample case, both the ECME estimate and the two stage estimate suffer from an edge

effect. In Section 3.7.4, we look into this edge effect further. The ECME estimate typically has slightly smaller MSE than the two stage estimate, with the improvement in MSE being more noticeable as the dependence of  $Y$  on  $Z$  increases. In the two-sample case, the two methods are comparable.

For testing, in the one-sample case, there is little difference in power between the two statistics considered. In the two sample case, the integrated t-statistic has more power to distinguish the difference between  $\beta^s$  and  $\beta^c$ .

In the one-sample comparison in Section 3.7.2, we simulate data using parameter values estimated from the male selection group data analyzed in Section 3.6.2. In the two-sample comparison in Section 3.7.3, we simulate selection group data using estimates from the male selection group data and the control group data using estimates from the male control group data without the outlier.

### 3.7.1 Two stage estimate

Recall from Section 3.1 that the calculation of Müller's (2005) estimate of  $\beta$  had two parts. The first part ignored the  $y_i$ 's and only used the  $\mathbf{z}_i$ 's to predict the underlying  $\mathbf{x}_i$ 's. The second part of the analysis was essentially a linear regression of the  $y_i$ 's on the predicted  $\mathbf{x}_i$ 's. We would like to study this procedure's ability to estimate  $\beta$ . We suspect that ignoring the  $y_i$ 's in the first part of the procedure will lead to poorer estimation of  $\beta$ .

To study this in a way that is comparable to our over-all methodology, we slightly modify the Müller's method. We use linear mixed effects methodology to predict the  $\mathbf{x}_i$ 's from the  $\mathbf{z}_i$ 's. Specifically, we first use our smoothed principal components analysis of the  $\mathbf{z}_i$ 's to determine basis functions, the  $\phi_k$ 's, see (3.3). We then construct the matrix  $\mathbf{T}$  and fit model (3.4) using the EM algorithm. Laird (1982) gave an elaboration on the EM computation in linear mixed models. We use the resulting estimated covariances of the  $\mathbf{x}_i$ 's and  $\epsilon_i$ 's to calculate our predictors, the BLUPs of  $\mathbf{x}_i$  given  $\mathbf{z}_i$ , that is, to calculate  $E(\mathbf{x}_i|\mathbf{z}_i)$ . For the second part of the analysis, we find the least squares estimate of  $\beta$  according to the regression model

$$Y_i = \beta_0 + E(\mathbf{x}_i'|\mathbf{z}_i)\mathbf{T}'\beta + \epsilon_i.$$

In our fit of (3.4), we force  $\Sigma_x$  to be diagonal. This is in keeping with Müller, and stems from the fact that the components of the  $\mathbf{x}_i$ 's are principal component scores.

The main difference between our approach and Müller's is in the way we estimate the required variances and covariances for calculating the BLUPs. We use a linear mixed effects model based on eigenfunctions. Müller used a smoothing method to directly estimate the components of the covariance structure.

### 3.7.2 One sample comparison

We simulate data based on parameter and eigenfunction estimates from the data analysis of the male selection group in Section 3.6.2. Let  $\boldsymbol{\mu}^s, \Sigma_x^s, \sigma_\epsilon^{s2}, \sigma^{s2}, \beta_0^s, \boldsymbol{\beta}^s$  and  $\beta^s(t) = \boldsymbol{\psi}(t)' \boldsymbol{\beta}^s$  denote these estimates and let  $\mathbf{A}^s$  be the matrix constructed from the basis functions, evaluated at the same  $t_j$ 's as in the data analysis.

We simulate the unperturbed predictor  $\mathbf{Z}_i$  according to a multivariate normal  $N(\boldsymbol{\mu}^s, \mathbf{A}^s \Sigma_x^s \mathbf{A}^{s'})$  and let the observed  $\mathbf{z}_i$  be  $\mathbf{Z}_i$  plus  $N(\mathbf{0}, \sigma_\epsilon^{s2} \mathbf{I})$  noise. We consider four possible  $\beta$  functions to describe the relationship between  $Y_i$  and  $Z_i$ :

$$Y_i = \beta_0^s + \int_{-1}^{60} \beta(t) [Z_i(t) - \bar{Z}_i(t)] dt + e_i.$$

The integral is calculated using the R function "sintegral" and we simulate  $e_i$  from  $N(0, \sigma^{s2})$ . We take  $\beta = \gamma \beta^s$  with  $\gamma = 0, 2/3, 4/3$  or  $2$ . When  $\gamma = 0$ ,  $Y_i$  does not depend on  $Z_i$ . When  $\gamma = 2$ , the dependence is large. Recall in our data analysis in Section 3.6.2, we found that  $\beta^s$  was significantly different from zero. Thus, for each value of  $\gamma$ , we can generate an observed data set  $(\mathbf{z}_i, Y_i^\gamma)$ ,  $i = 1, \dots, 39$ .

We first compare our estimate of  $\beta$  with the two stage estimate. We run 100 simulations and the MSE of the estimate  $\hat{\beta}$  is calculated as

$$\sum_1^{100} (\hat{\beta}(t) - \beta(t))^2 / 100$$

at each observation point  $t$ . The results are in Figure 3.10 where the first panel shows the pattern of the true  $\beta$ 's, the next two panels show the pointwise MSE's of our two estimates of  $\beta$  and the last panel shows the differences of these pointwise MSE's. In the plots, the MSE's increase as  $\gamma$  increases and both estimates of  $\beta$  have high MSE's for  $t$ 's at the edges. The MSE's of both methods are much worse at the left hand edge, probably due to the sharp decrease of the true  $\beta$  and the sharp increase in log body mass before week 10. The ECME method seems to be more affected by this edge. We will give a further study of this edge effect in Section 3.7.4. However, in the last panel we see that overall, the ECME estimate has a smaller MSE, with the superiority of ECME increasing as  $\gamma$  increases.

Therefore, if there is a significant dependence of  $Y_i$  on  $\mathbf{z}_i$  through  $\beta$ , the ECME method of estimate is preferred.

To test  $H_o : \beta(t) = 0$ , for all  $t \in [-1, 60]$ , we use the test statistics  $U_w$  and  $U_f$  with standard errors calculated using the Hessian matrix (3.26) as described in Section 3.4. For each data set, we run 300 permutations to calculate the p-values. We simulate 100 data sets for each value of  $\gamma$  and choose levels  $\alpha = 0.01$  and  $\alpha = 0.05$ . Figure 3.11 and Figure 3.12 summarize the proportion of times  $H_o$  was rejected using  $U_w$  and  $U_f$  but with different levels:  $\alpha = 0.01$  and  $\alpha = 0.05$  respectively. As expected, the powers increase as  $\gamma$  increases. There is little difference in power between the two statistics.

### 3.7.3 Two sample comparison

To simulate data from two independent samples, we choose model parameters using the male selection group data and the male control group data analysed in Section 3.6.2. We simulate data for the selection group and, separately, for the control group using the same methodology as in Section 3.7.2 but with different true  $\beta$ 's.

Let  $\beta^s$  and  $\beta^c$  be the estimates of  $\beta$  from the original male selection group and male control group data. Let  $\bar{\beta} = (\beta^s + \beta^c)/2$  and  $\Delta\beta = (\beta^s - \beta^c)/2$ . In the simulation study we set the  $\beta$  of the selection group to  $\bar{\beta} + \gamma\Delta\beta$  and that of the control group to  $\bar{\beta} - \gamma\Delta\beta$  with  $\gamma = (0, 2/3, 4/3, 2)$ . Thus, the

difference between the selection and control  $\beta$  is  $2\gamma\Delta\beta = \gamma(\beta^s - \beta^c)$ .

To estimate  $\beta^s$  and  $\beta^c$ , we fit the selection data and control data separately. For each value of  $\gamma$  we simulate 100 data sets and calculate the MSE as in Section 3.7.2 in each group. Figures 3.13 and 3.14 show the MSE's of the two estimates in each group respectively. The pattern of the two MSE's are similar to their counterparts in Figure 3.10.

We calculate the MSE of the estimate  $\beta^s - \beta^c$  as

$$\sum_1^{100} \left( \hat{\beta}^s(t) - \hat{\beta}^c(t) - \beta^s(t) + \beta^c(t) \right)^2 / 100$$

at each observation point  $t$ . Figure 3.15 shows the results. The patterns of the two MSE's are similar to their counterparts in Figure 3.10. The MSE's of the two estimates are comparable but in general the MSE of the ECME estimate is smaller.

To test  $H_o : \beta^s(t) = \beta^c(t)$ , for all  $t \in [-1, 60]$ , we compare the four test statistics,  $U_l$ ,  $U_w$ ,  $U_e$  and  $U_f$ , with standard errors calculated using the Hessian matrix (3.26). For each data set, we run 300 permutations to calculate the p-values. We simulate 100 data sets for each value of  $\gamma$  and choose levels  $\alpha = 0.01$  and  $\alpha = 0.05$ . Figures 3.16 and 3.17 show the power of the four test statistics with levels 0.01 and 0.05 respectively. The statistic  $U_f$  is the most powerful, especially when  $\gamma$  is large, but  $U_f$  and  $U_w$  are comparable.

### 3.7.4 Edge effect discussion in one-sample MSE comparison

In Figure 3.10, we see that the MSE's of the ECME estimate of  $\beta$  are much worse at the left hand edge than the MSE's of the two stage estimate. We suspect this is probably due to the sharp decrease of the true  $\beta$  before week 10 as shown at the first panel of Figure 3.10, and the large increase in log body mass in that same period. In this section, we exclude the early weeks' data from our analysis.

To determine the various parameter values to use in a new simulation, we re-analyze the male selection group data but with  $\mathbf{z}_i$  containing log body

mass values a week 5 rather than at week  $-1$ . After obtaining the new parameter estimates, we simulate data in the same way as in Section 3.7.2. The simulation analysis result is in Figure 3.18. The edge effect still exists but this time the MSE's of ECME and the two stage method are comparable at the edges.



Group	Observed $U_l$	Asymptotic p- $U_l$	p- $U_l$	p- $U_w$	p- $U_f$
Male selected	13.723	0.008	0.008	0.008	0.008
Male control	10.940	0.027	0.044	0.044	0.034
Female selected	9.095	0.059	0.082	0.082	0.264
Female control	3.280	0.512	0.568	0.568	0.656

Table 3.1: P-values of the test  $H_o : \beta(t) = 0$ , for all  $t \in [-1, 60]$  in each group.

Gender	Observed $U_l$	Asymptotic p- $U_l$	p- $U_l$	p- $U_w$	p- $U_e$	p- $U_f$
Male	11.919	0.018	0.364	0.356	0.532	0.110
Female	5.852	0.210	0.564	0.568	0.884	0.444

Table 3.2: P-values of the test  $H_o : \beta^s(t) = \beta^c(t)$ , for all  $t \in [-1, 60]$ , within each gender.

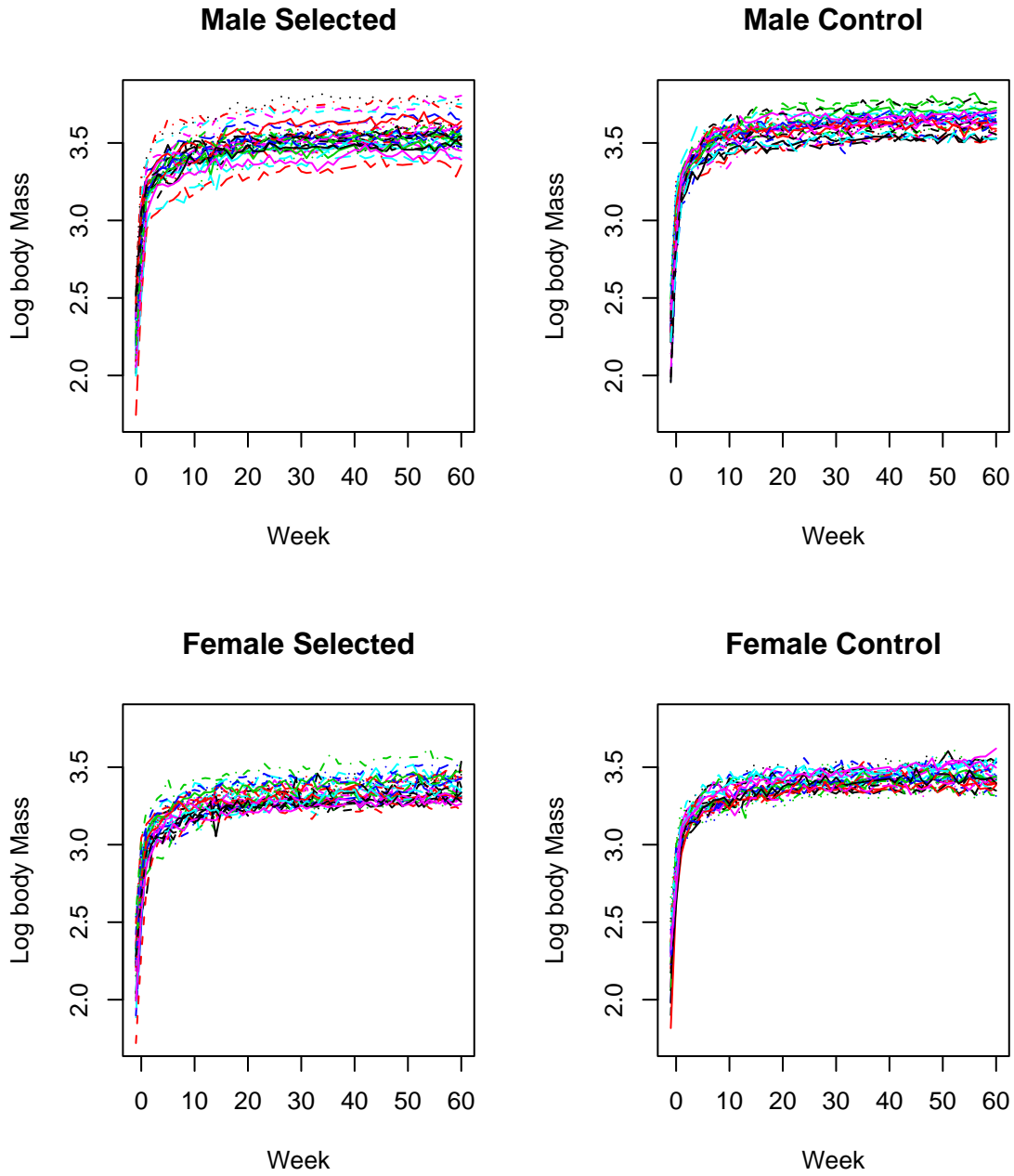


Figure 3.1: Plots of log body mass versus week for the four groups of laboratory mice.

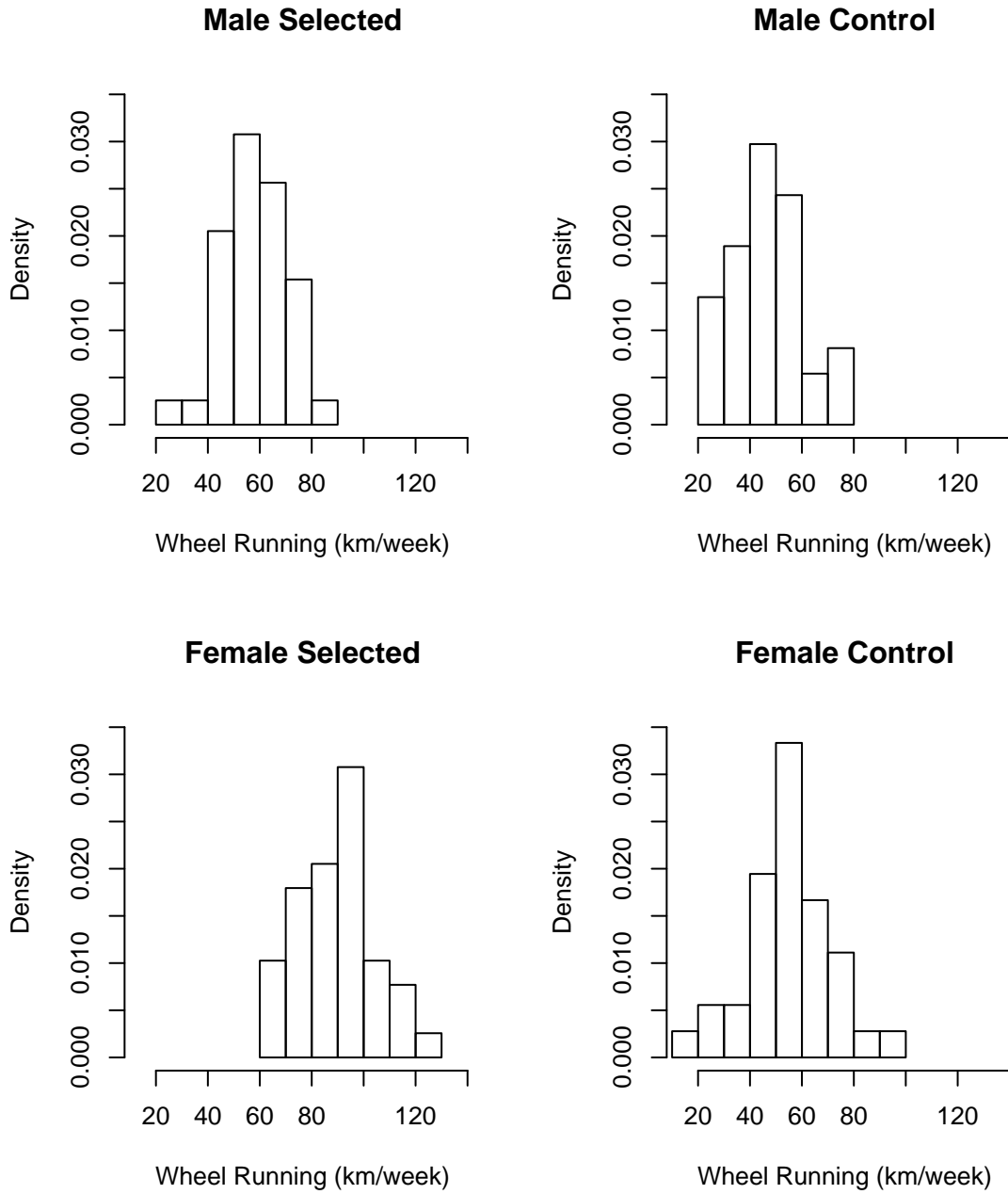


Figure 3.2: Histogram of the response: averaged wheel running from weeks 5 to 60 for the four groups of laboratory mice.

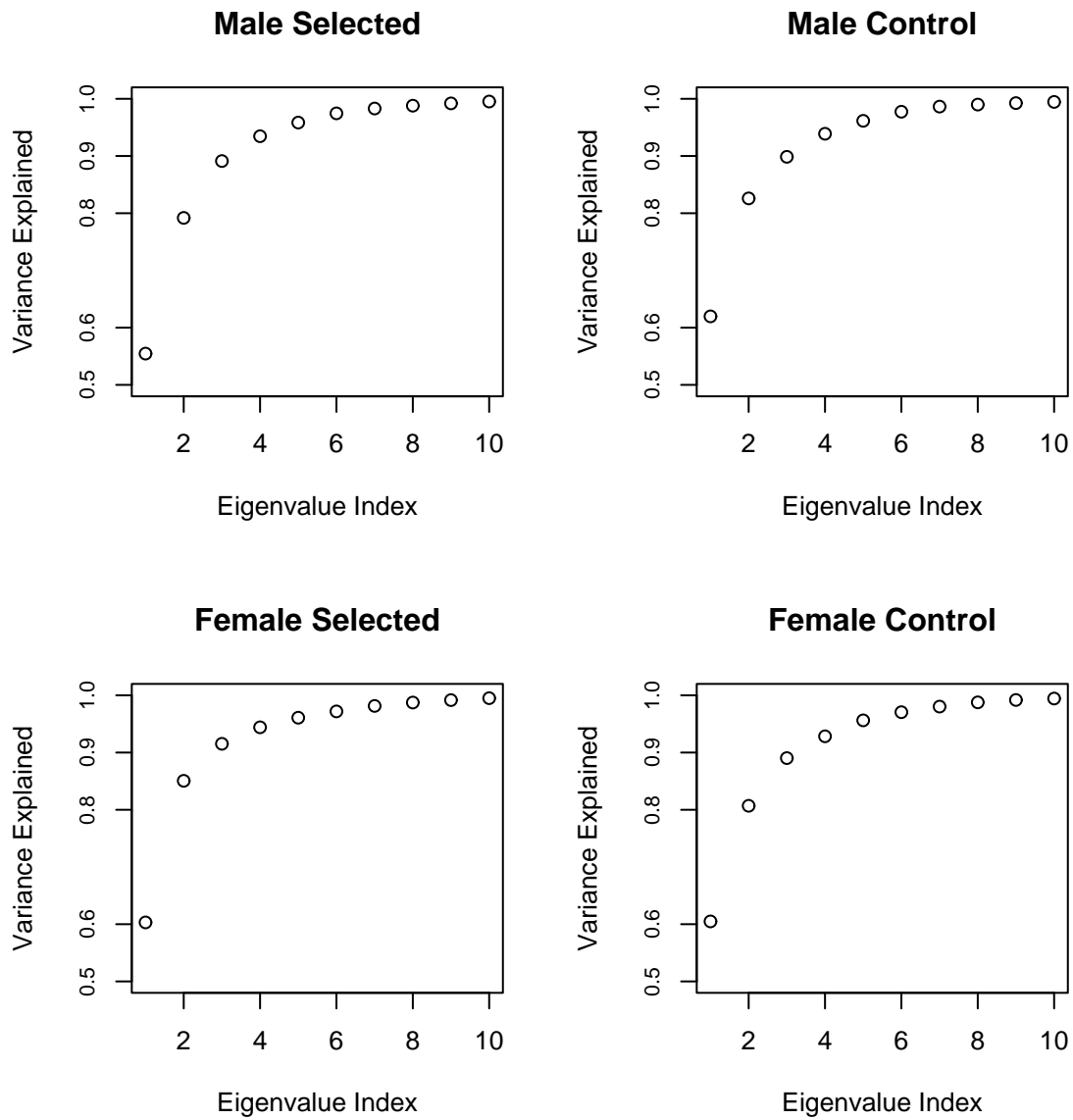


Figure 3.3: Plots of the proportion of cumulative variance of the centered log body mass explained by the first ten principal components of the four groups of mice, after the individual average log body mass has been removed.

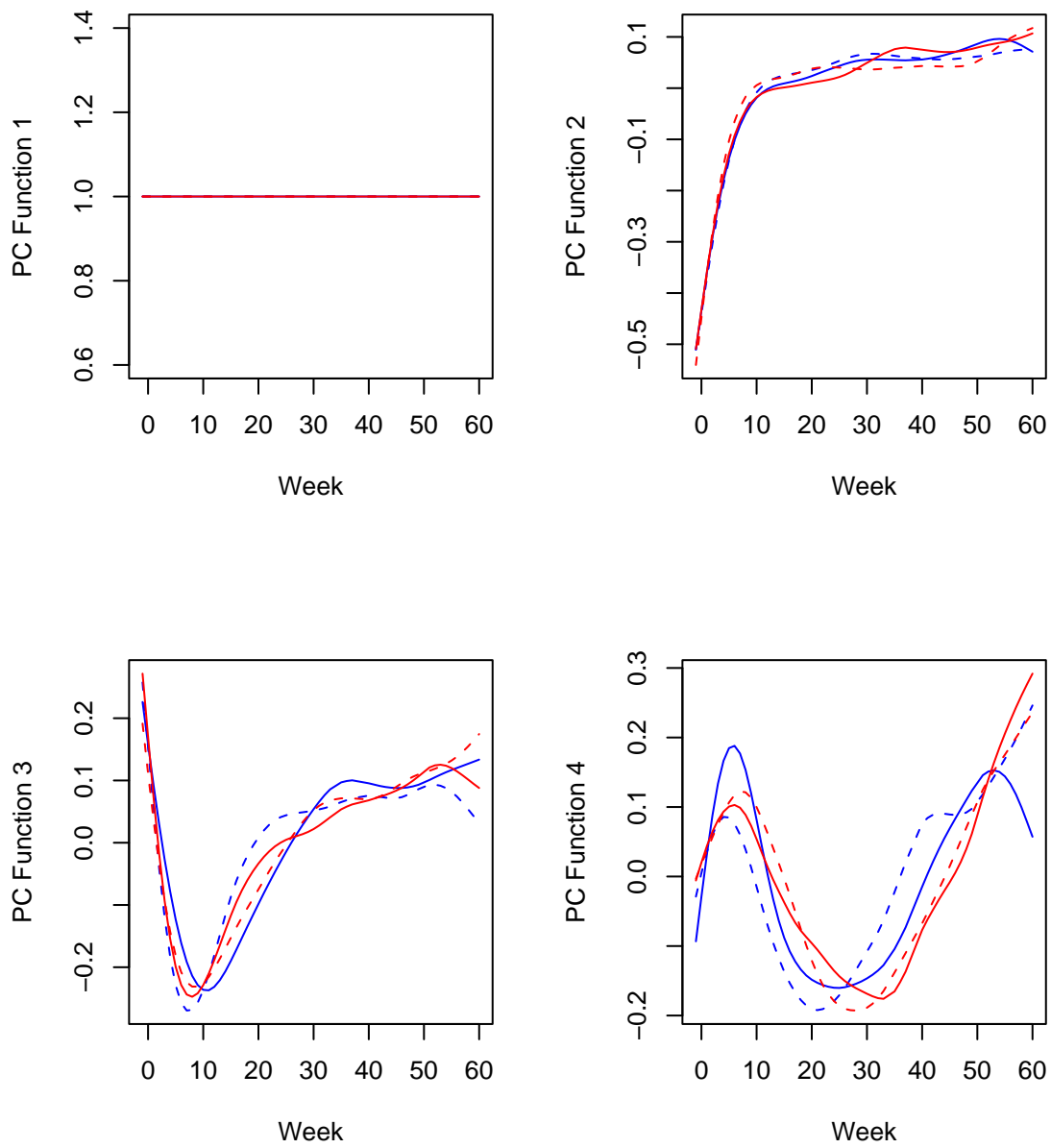


Figure 3.4: The constant function and the first three smoothed eigenfunctions of the covariance of centered log body mass of the four groups of mice, calculated as described in Section 3.6.3.

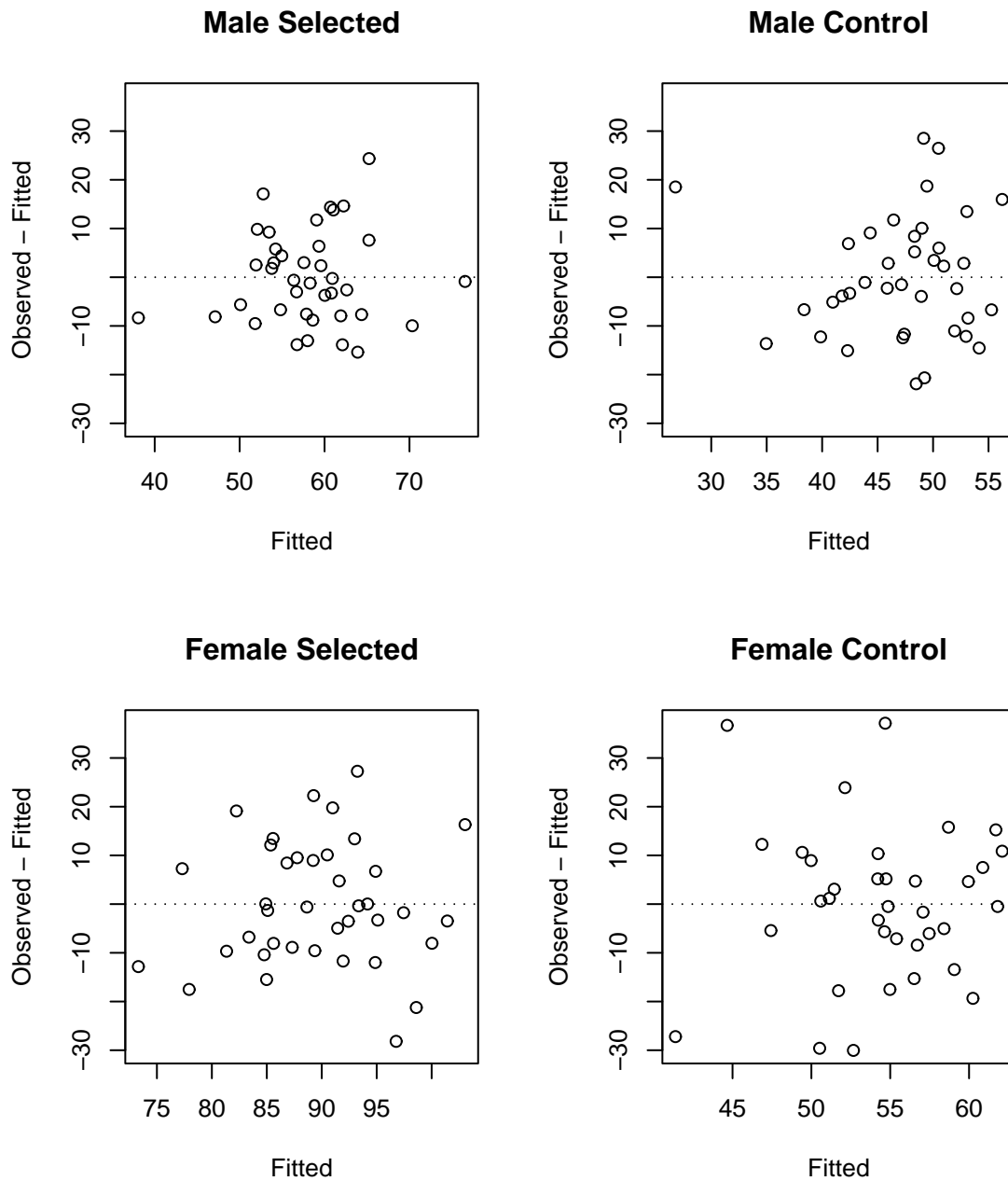


Figure 3.5: Residual plots of the fit of the  $Y_i =$  averaged wheel running of the four groups of mice.

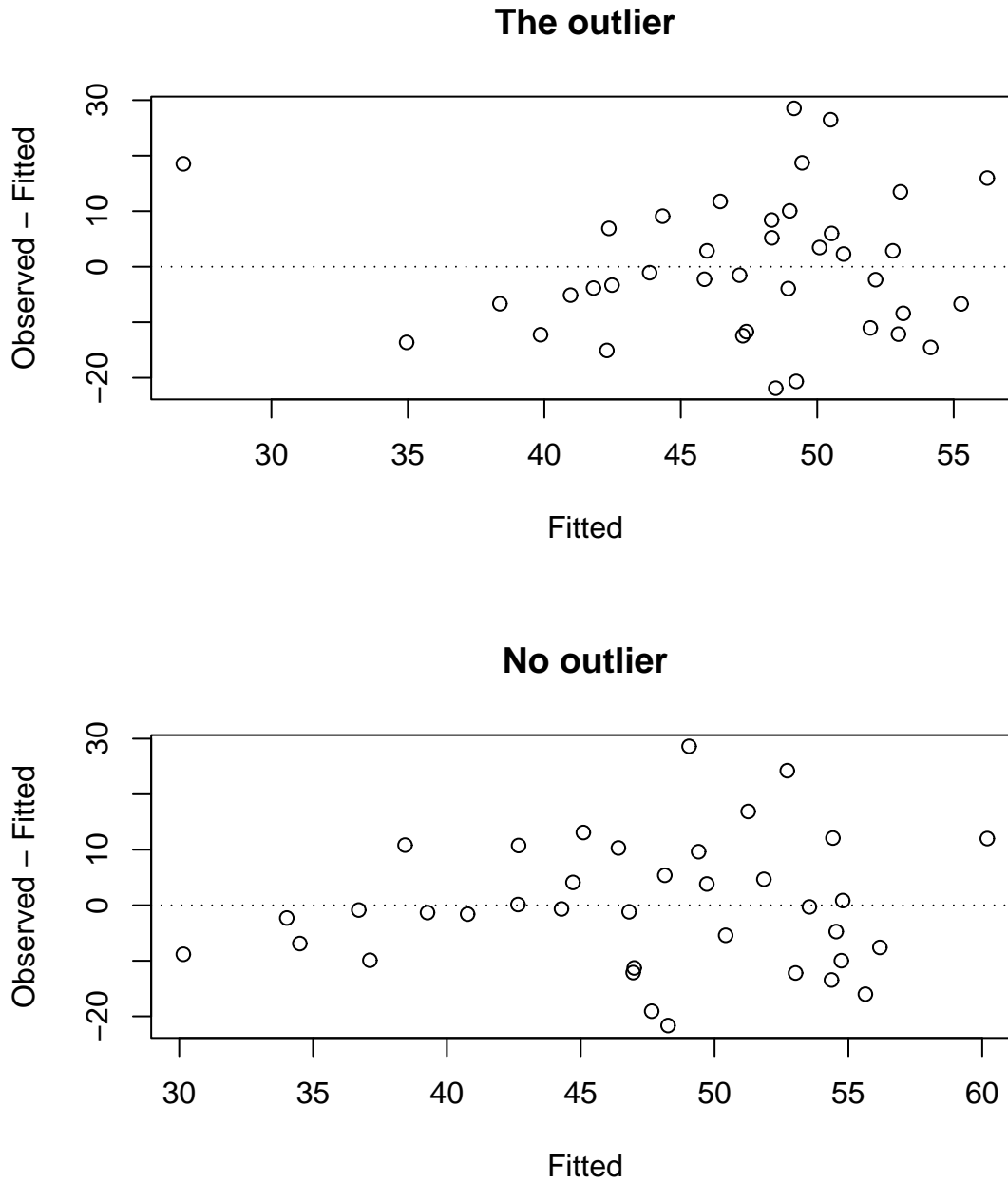


Figure 3.6: Plots of residuals of  $Y_i =$  averaged wheel running in the male control group of mice before and after removing the outlier.

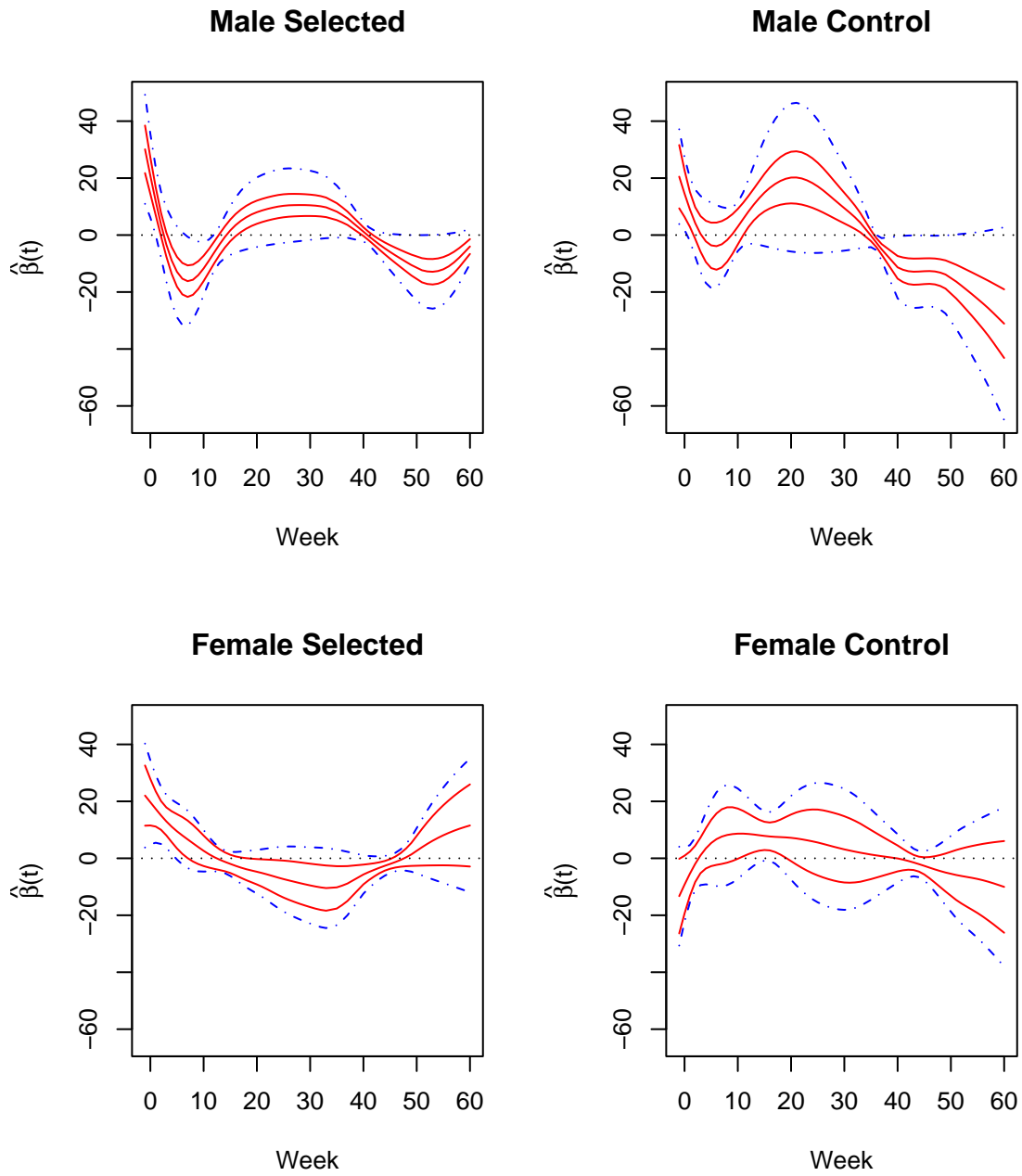


Figure 3.7: Plots of  $\hat{\beta}$  and its standard errors computed from the Hessian matrix (solid) and from the bootstrap (dash-dot) for the four groups of mice, as described in Section 3.4.



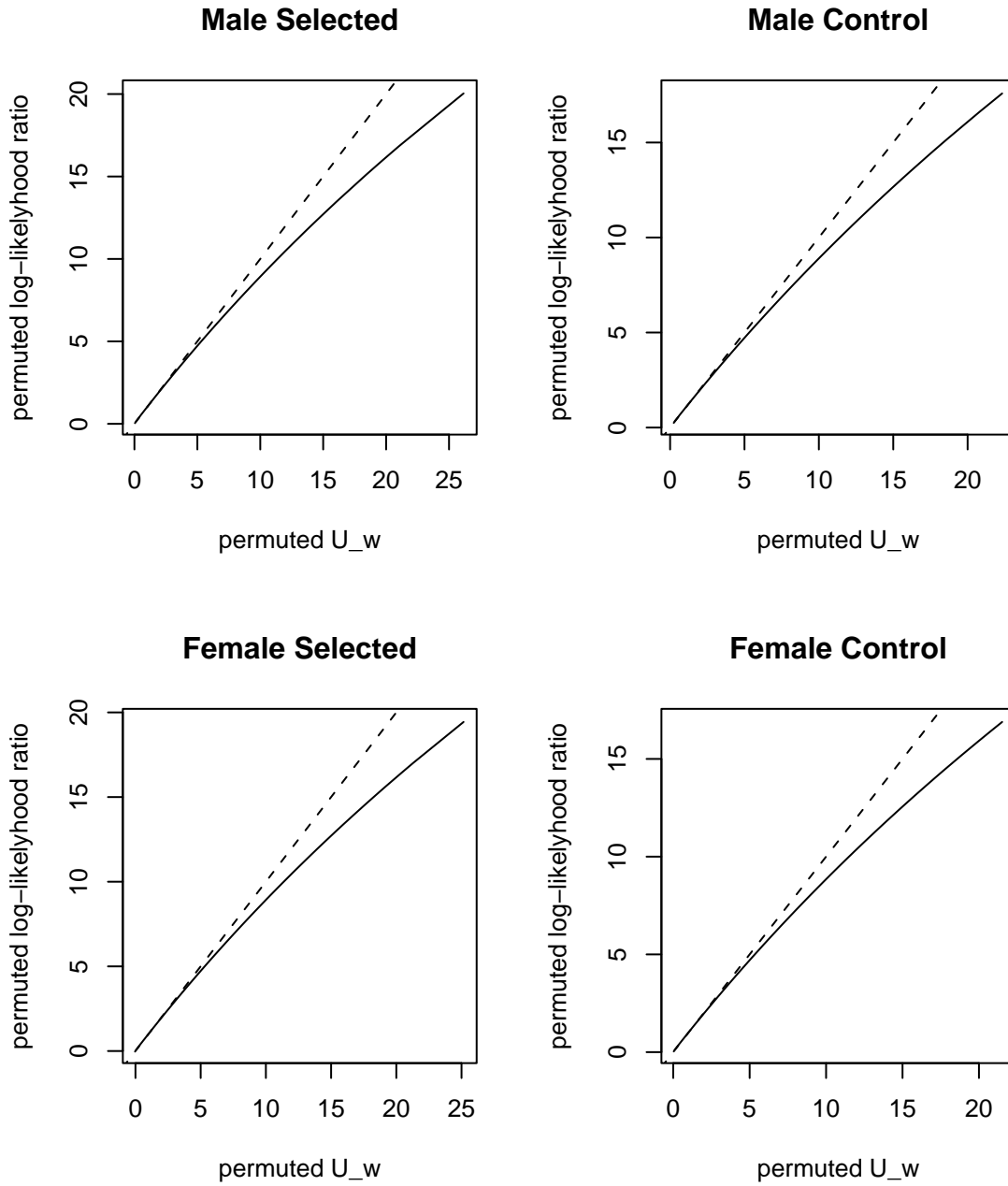


Figure 3.8: Comparing the permuted values of the generalized likelihood ratio statistic  $U_l$  with the Wald statistic  $U_w$  for four groups of mice. The equivalence of  $U_l$  and  $U_w$  were observed in Table 3.1.

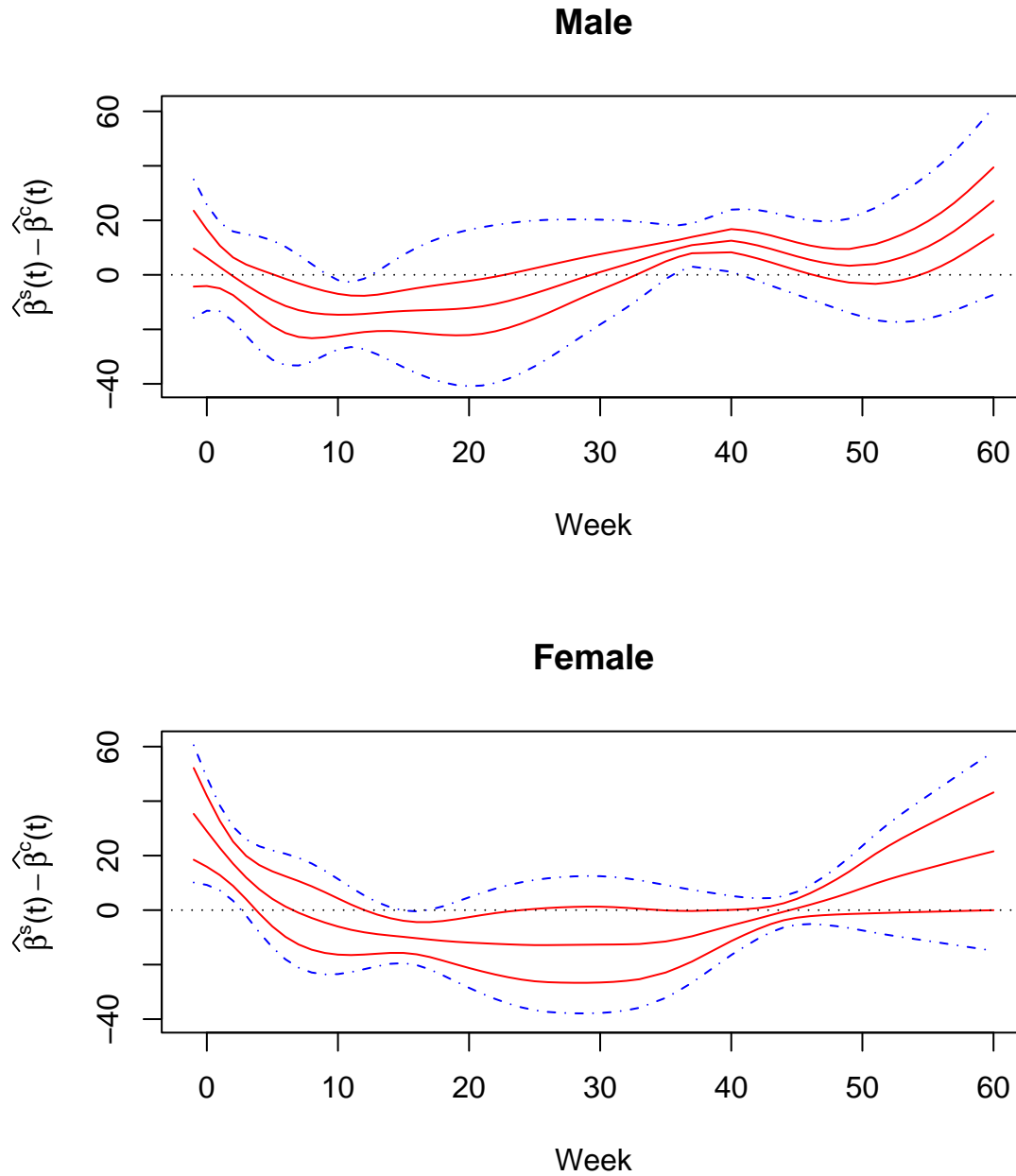


Figure 3.9: Plots of  $\hat{\beta}^s - \hat{\beta}^c$  and the standard errors of the difference computed from the Hessian matrix (solid) and from the bootstrap (dash-dot) for both genders.

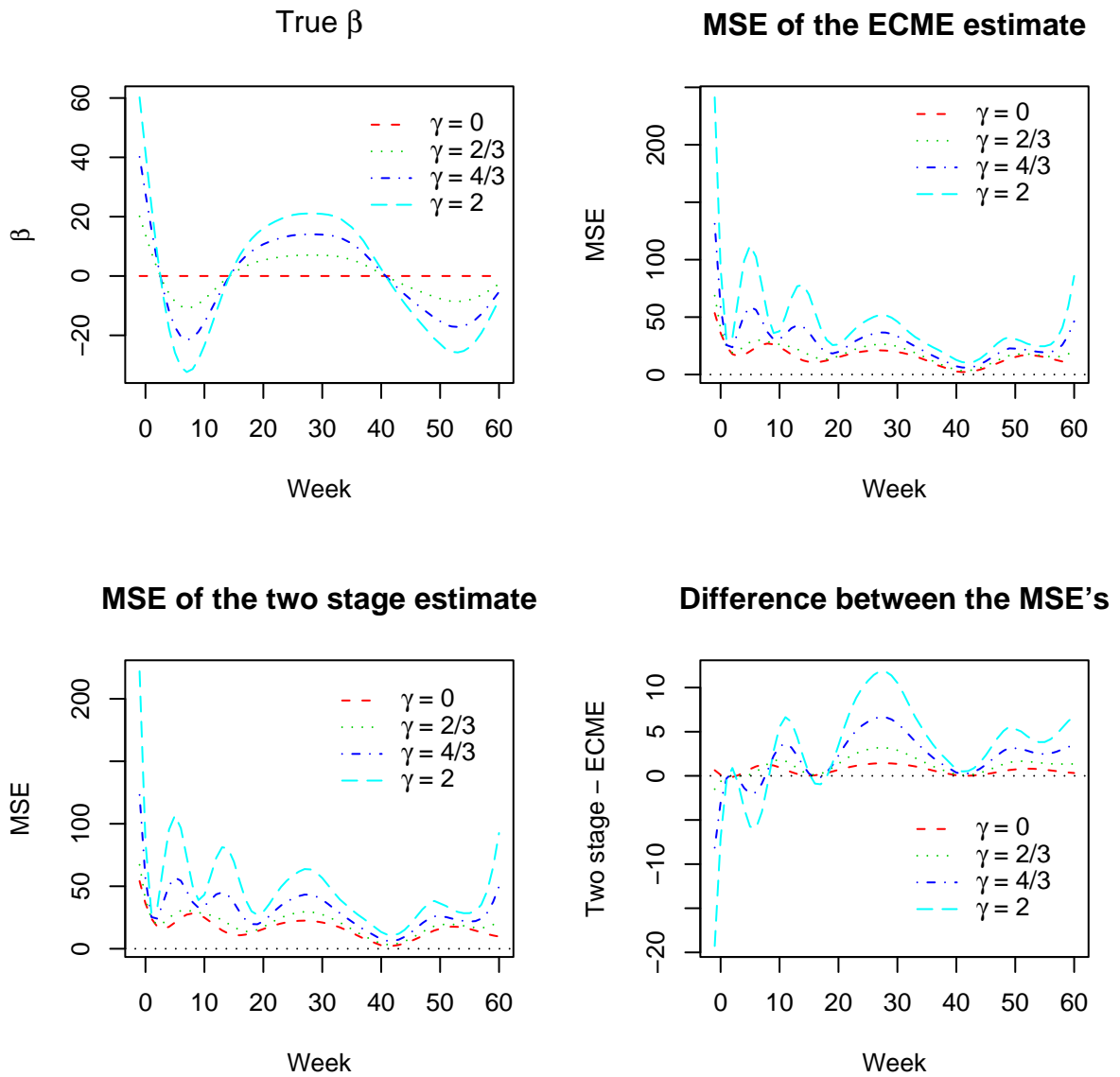


Figure 3.10: MSE of the estimate of  $\beta$  for each  $\gamma$  value in one sample simulation as described in Section 3.7.2. Compare the ECME method with the two stage method.

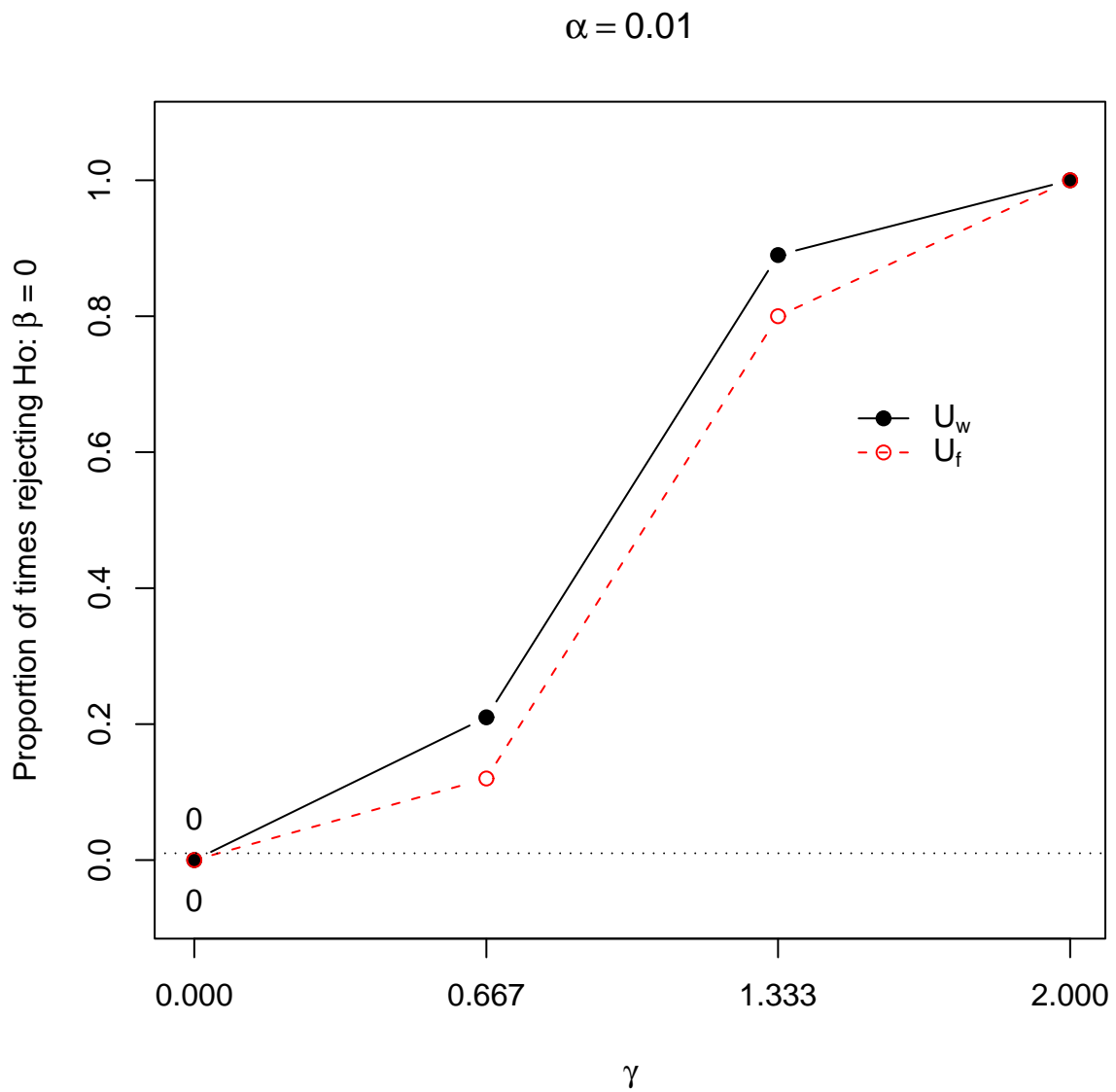


Figure 3.11: Proportion of times  $H_0$  is rejected using level  $\alpha = 0.01$ , where  $H_0 : \beta(t) = 0$ , for all  $t \in [-1, 60]$  in one sample simulation as described in Section 3.7.2. Two test statistics are considered,  $U_w$  and  $U_f$ .

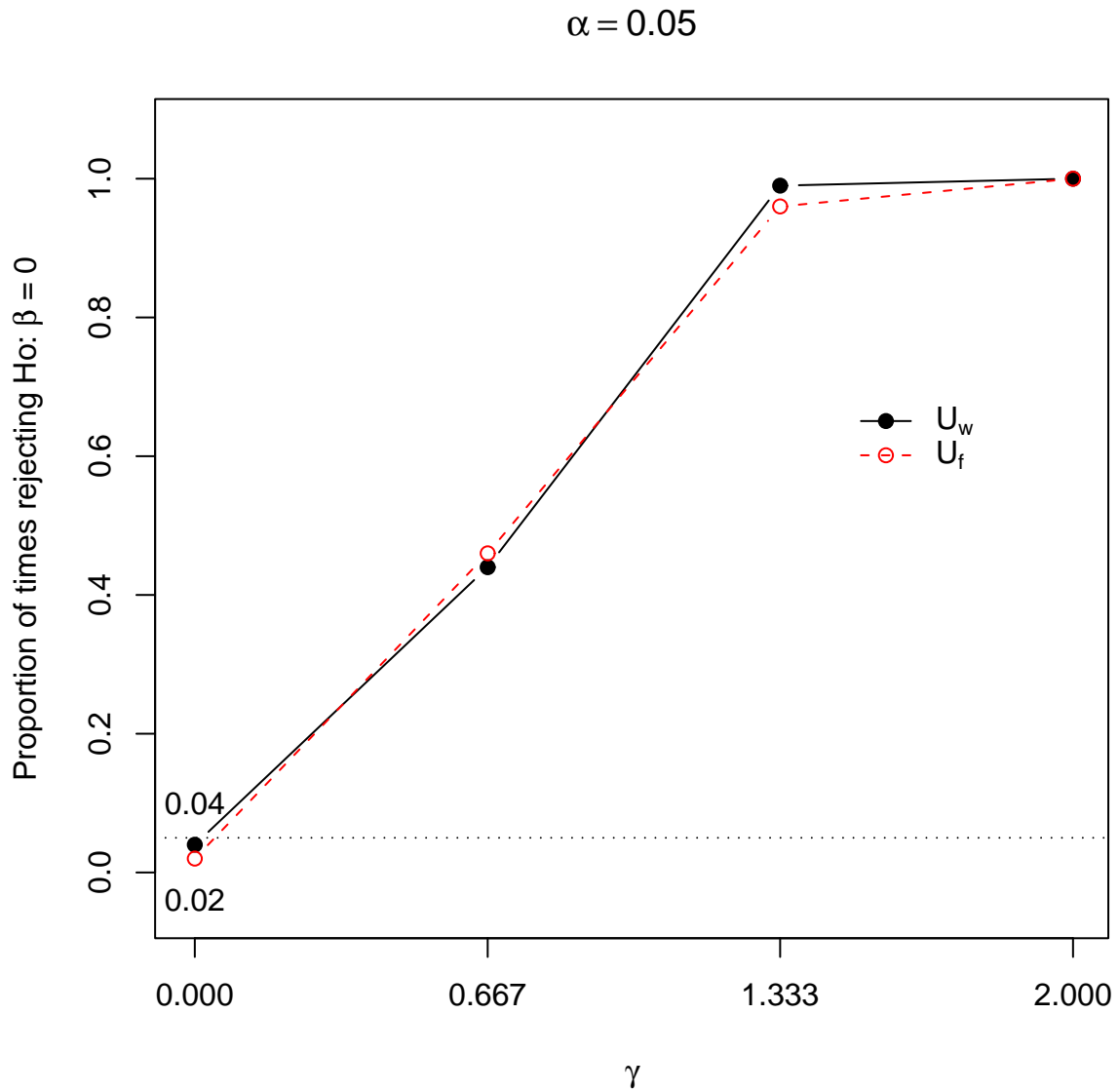


Figure 3.12: Proportion of times  $H_o$  is rejected using level  $\alpha = 0.05$ , where  $H_o : \beta(t) = 0$ , for all  $t \in [-1, 60]$  in one sample simulation as described in Section 3.7.2. Two test statistics are considered,  $U_w$  and  $U_f$ .

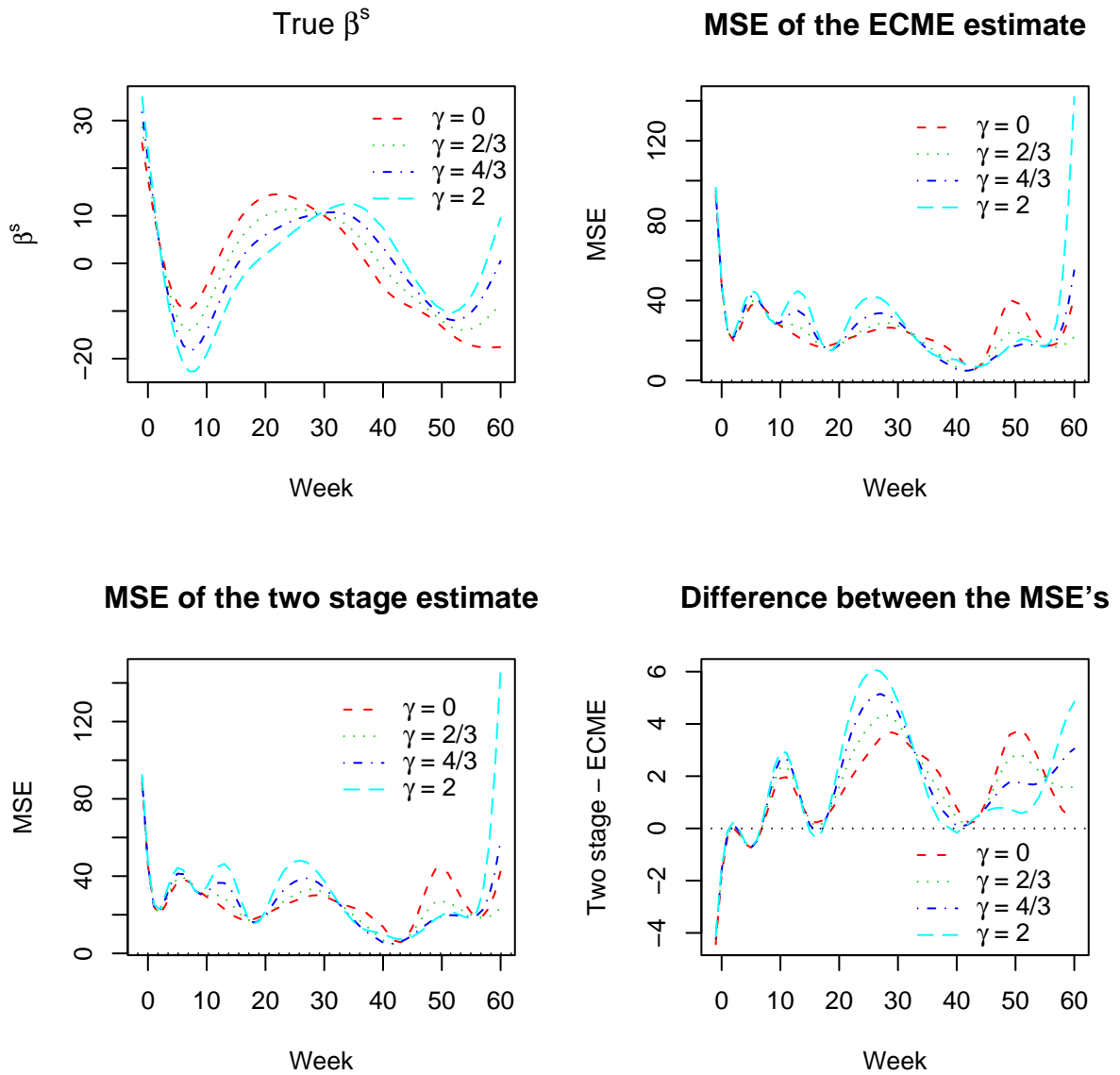


Figure 3.13: MSE of the estimate of  $\beta^s$  for each  $\gamma$  value in two sample simulation as described in Section 3.7.3. Compare the ECME method with the two stage method.

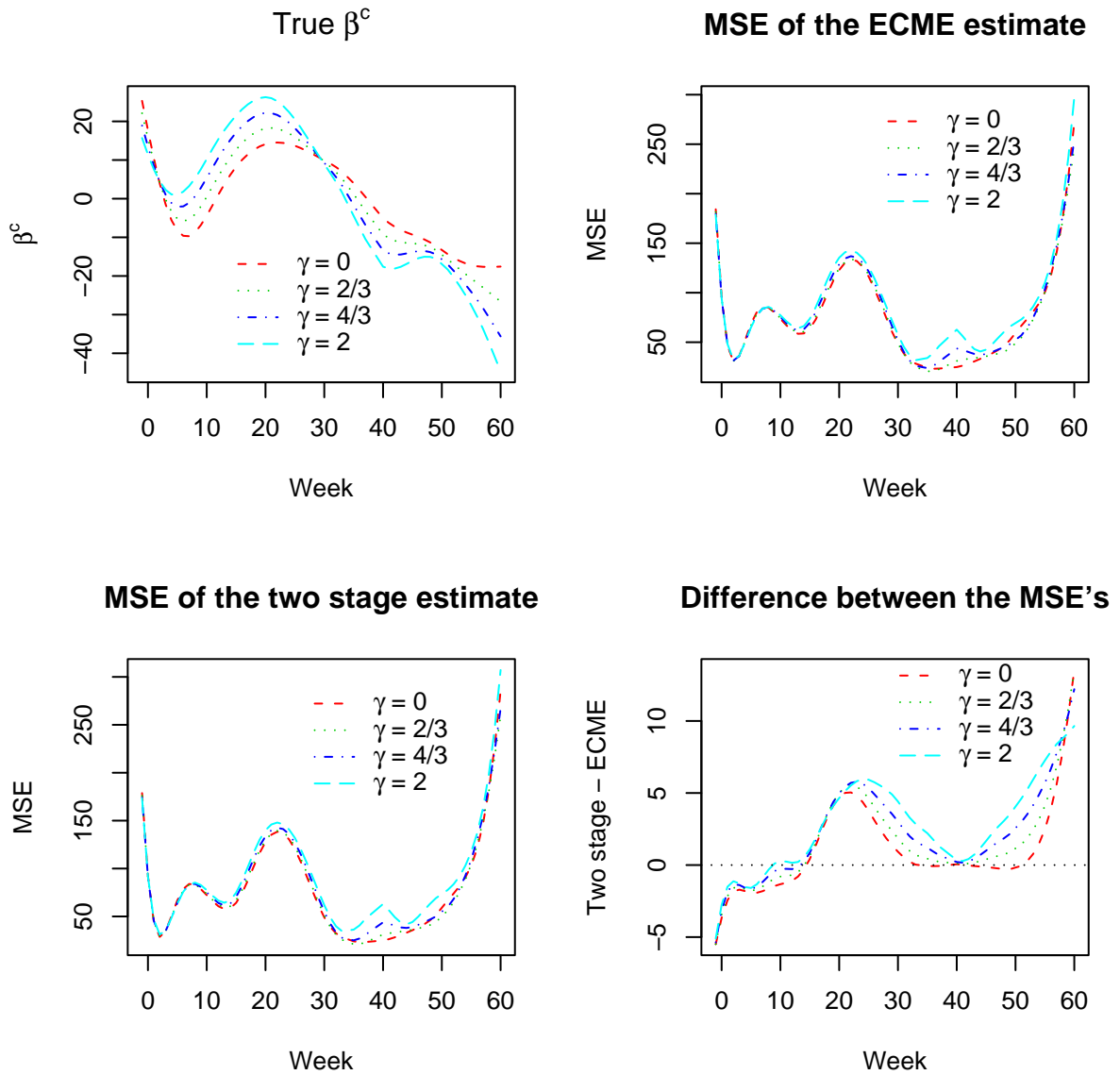


Figure 3.14: MSE of the estimate of  $\beta^c$  for each  $\gamma$  value in two sample simulation as described in Section 3.7.3. Compare the ECME method with the two stage method.

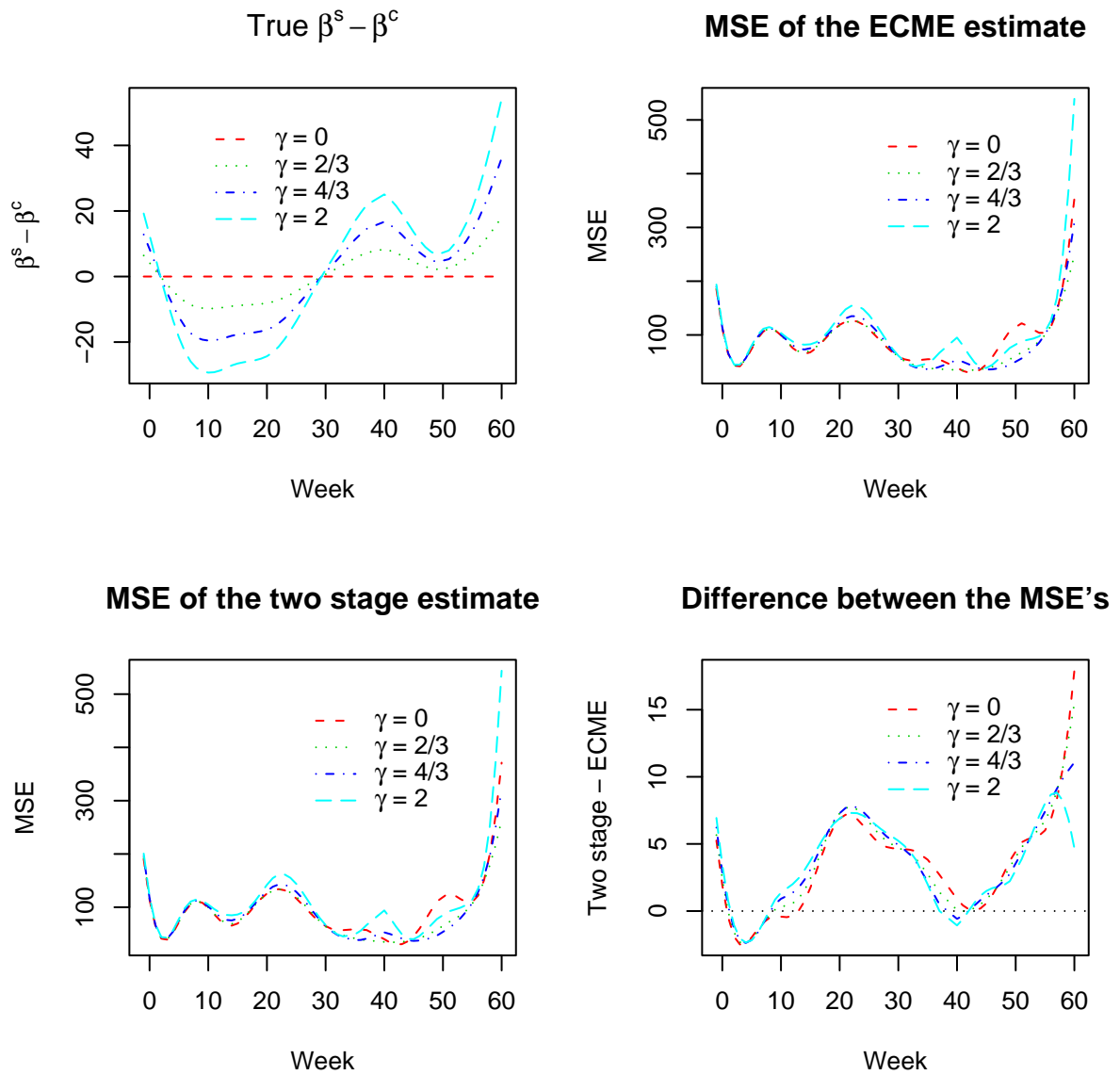


Figure 3.15: MSE of the estimate of  $\beta^s - \beta^c$  for each  $\gamma$  value in two sample simulation as described in Section 3.7.3. Compare the ECME method with the two stage method.



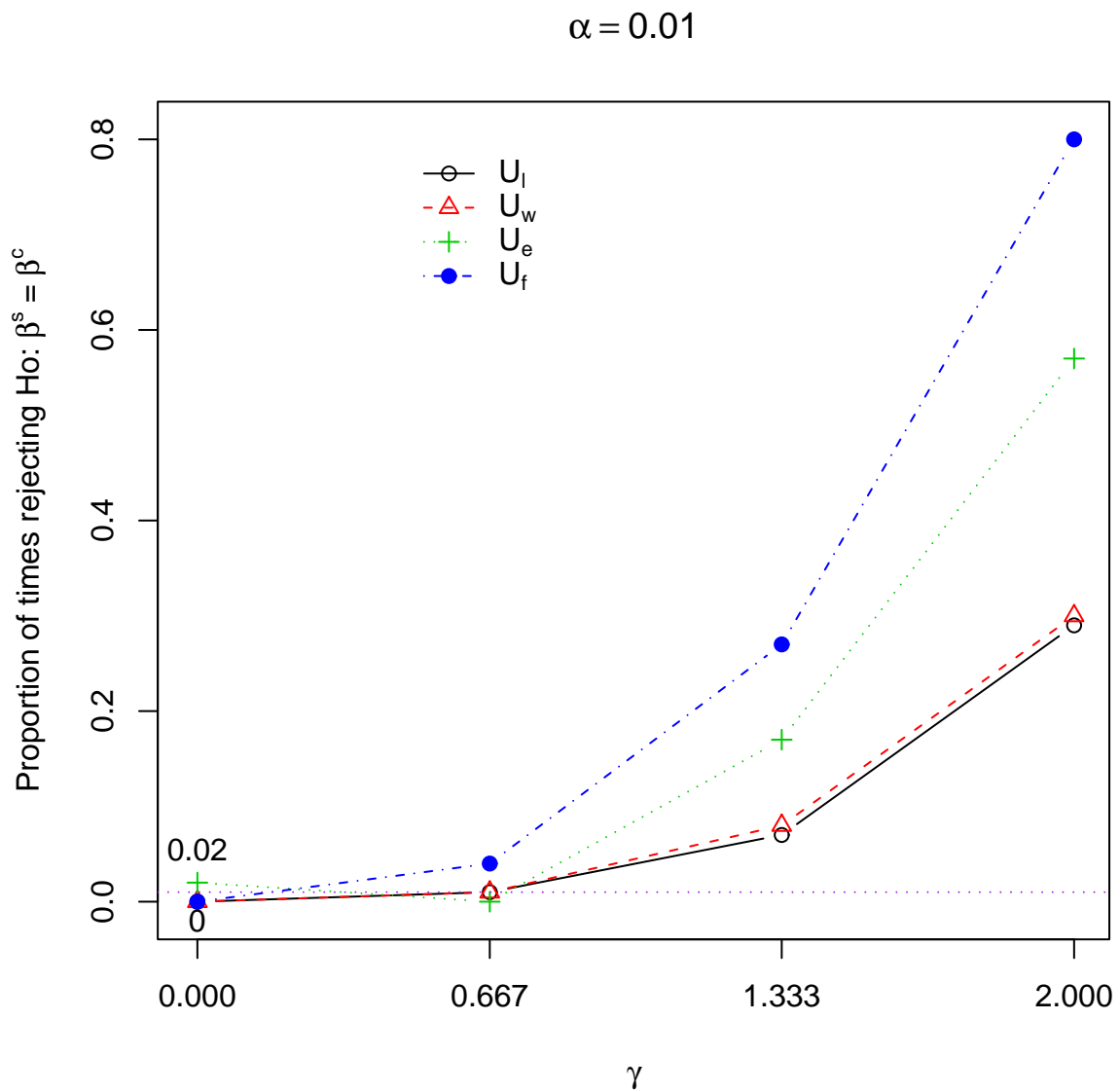


Figure 3.16: Proportion of times  $H_o$  is rejected using level  $\alpha = 0.01$ , where  $H_o : \beta^s = \beta^c$ , for all  $t \in [-1, 60]$  in two sample simulation as described in Section 3.7.3. Four test statistics are considered  $U_l$ ,  $U_w$ ,  $U_e$  and  $U_f$ .

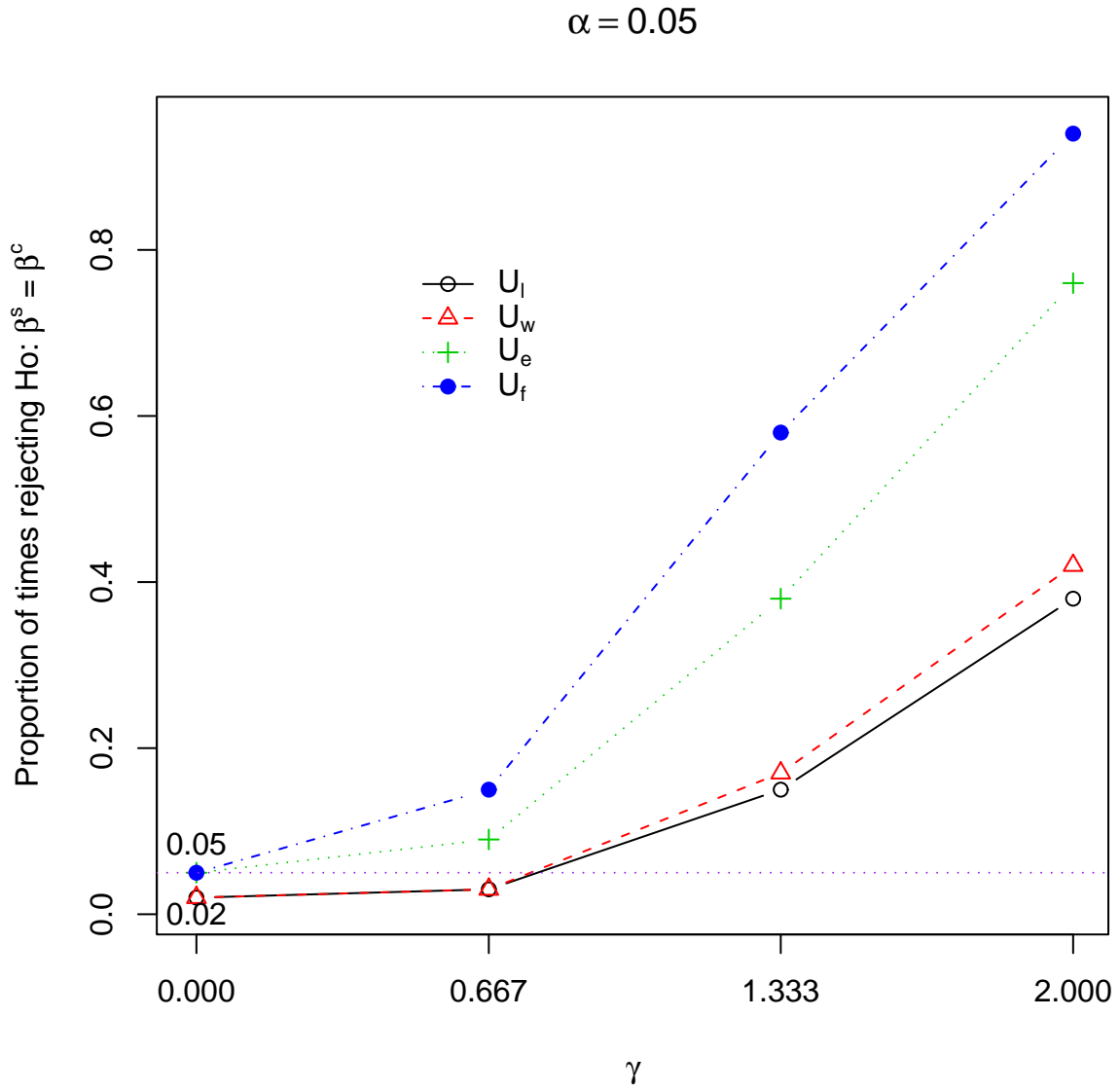


Figure 3.17: Proportion of times  $H_o$  is rejected using level  $\alpha = 0.05$ , where  $H_o : \beta^s = \beta^c$ , for all  $t \in [-1, 60]$  in two sample simulation as described in Section 3.7.3. Four test statistics are considered  $U_l, U_w, U_e$  and  $U_f$ .

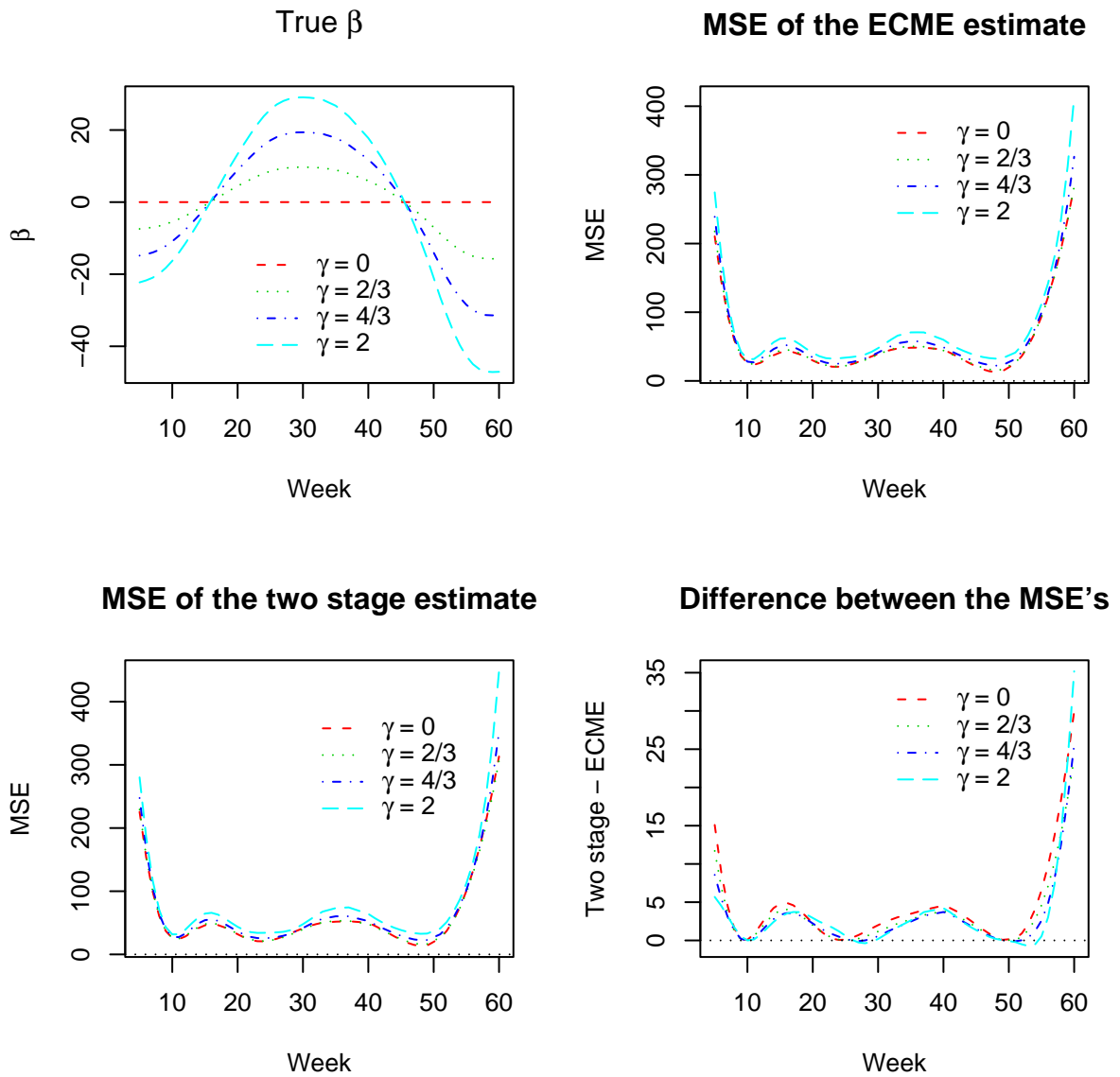


Figure 3.18: MSE of the estimate of  $\beta$  for each  $\gamma$  value using truncated log body mass as described in Section 3.7.4. Compare the ECME method with the two stage method.

# Bibliography

- [1] James, G. (2002) Generalized linear models with functional predictors. *Journal of the Royal Statistical Society B* 64, 3, 411-432.
- [2] James, G. M. and Silverman, B. W. (2005) Functional adaptive model estimation. *Journal of the American Statistical Association* 100, 470, 565-576.
- [3] Laird, N. M. (1982) Computation of variance components using the E-M Algorithm. *Journal of Statistical Computation and Simulation* 14, 295-303.
- [4] Liu, C. H. and Rubin, D. B. (1994) The ECME Algorithm - a simple extension of EM and ECM with faster monotone convergence. *Biometrika* 81 4, 633-648.
- [5] Magnus, J. R. and Neudecker, H. (1988) *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley and Sons.
- [6] Morgan, T.J., T. Garland, Jr., and Carter, P. A. (2003) Ontogenetic trajectories in mice selected for high wheel- running activity. I. Mean ontogenetic trajectories. *Evolution* 57, 646-657.
- [7] Müller, H.G. (2005) Functional modeling and classification of longitudinal data. *Scandinavian Journal of Statistics* 32, 223-240.
- [8] Ramsay, J. O. and Silverman, B. W. (2005) *Functional Data Analysis*. 2nd edition, Springer Series in Statistics.
- [9] Wang, W. (2007) Identifiability in linear mixed effects models. *Manuscript*, Department of Statistics, University of British Columbia.

- [10] Yao, F., Müller, H.G., Wang, J.L. (2005) Functional data analysis for sparse longitudinal data. *Journal of American Statistical Association* 100, 577-590.

## Chapter 4

# Dependence of average lifetime wheel-running on body mass ontogeny

### 4.1 Introduction

In natural populations of animals, strong correlations between large body size and increased Darwinian fitness are frequently measured: large body size is related to increased fecundity, reduced predation and, in general, an increased ability to survive in a given environment (Sogard 1997, Hone and Benton 2005). Variation in body mass is also associated with variation in a number of physiological traits, including metabolic rate and heart mass, and much effort has been expended understanding allometric relationships in animals (Garland and Carter, 1994, and references therein). Body mass has even been shown to impact mutation rates in nuclear and mitochondrial DNA (Gillooly et al., 2005; Estabrook et al., 2007). Body mass is also associated with locomotor performance; however, these relationships are complex. Body mass is positively correlated with home range area and daily movement distances in wild populations of mammals (Garland 1983), but negatively correlated with laboratory measures of voluntary activity (Dewsbery 1980).

Replicate lines of mice which have been selected for high levels of voluntary wheel running activity are a powerful model system with which to

---

<sup>0</sup>A version of this chapter will be submitted for publication. Authors: Carter, P. A., Heckman, N. and Wang, W.

investigate the relationships between body mass and voluntary locomotor activity (Garland 2003). After 10 generations of selection, mice selected for high voluntary wheel running activity at 8 weeks of age ran significantly more than their unselected controls; however, body mass at the age at which wheel running was measured and selected was not a significant predictor of the amount of wheel running nor was it significantly different between selected and control mice (Swallow et al. 1998). A subsequent study, 14 generations after selection, examined wheel running and body mass in this system of mice after 8 weeks of access to running wheels (Swallow et al., 1999). Three major results emerged from this study. First, body mass and wheel running share a negative genetic correlation of  $-.50$ ; selected mice were significantly lighter at maturity than control mice. Second, this result is not mediated solely through wheel running activity; the difference in body mass persisted even in mice without wheel access, and in mice with wheel access even when total wheel running was statistically removed. Third, access to running wheels did result in a decrease in body mass, but this decrease was only about a third of the decrease caused by selection.

A detailed study on body mass and wheel running ontogenies in the 16th generation of this system of mice was conducted by Morgan et al (2003). This experiment utilized 320 mice, half from the selected lines and half from the control lines; half of each of these groups were given lifetime access to running wheels while the other half did not have wheel access. Morgan et al (2003) compared the ontogenetic trajectories of the different experimental groups using repeated measures analysis and identified several interesting results. First, both the position and shape of the wheel running trajectory had responded to selection on wheel running at 8 weeks of age; selected mice ran more across the trajectory and showed a steeper decline in activity at older ages than control mice. This correlated response of the wheel running ontogeny to early-age selection on wheel running indicates a positive genetic covariance between wheel running at younger ages and wheel running at older ages, with the covariance declining as age increases. Second, the position but not the shape of the body mass trajectory had responded to selection, regardless of access to running wheels; selected mice with and

without wheel access were significantly lighter than their counterpart control mice across all ages. This result indicates a negative response of the body mass ontogeny to selection on early-age wheel running and hence a negative genetic covariance between wheel running at 8 weeks of age and body mass across ontogeny; in addition, it indicates strong positive genetic covariances among body masses at different ages. Finally, the analysis of the wheel running trajectory revealed that age-specific body mass was a significant covariate for age-specific wheel running; smaller individuals ran more at each age than did larger individuals.

The fact that Swallow et al. (1999) measured a negative genetic covariance between age-specific wheel running and body mass and that Morgan et al. (2003) showed that body mass is a significant covariate for wheel running at each age, with smaller mice running more than larger mice, raises the question of whether wheel running through ontogeny explicitly depends on the body mass trajectory. Given that body mass at each age was a significant covariate for wheel running at that age, we hypothesized that wheel running is significantly dependent on the body mass trajectory. In addition, given the age-specific negative genetic covariance between body mass and wheel running, we can ask whether this dependence has evolved in response to early-age selection on wheel running. We hypothesize that this dependence has evolved such that we will measure quantitative differences between the selected and control mice in the dependence of wheel running on the body mass trajectory.

To test these hypotheses, we will generally take a function-valued approach. In essence, we regress average wheel running on the log body mass trajectory. We then test if the “regression coefficients” are zero within each group and if they are equal between two groups. To model the log body mass trajectory, we use basis functions with expansion coefficients modeled as random effects.



## 4.2 Methods

Experiments were performed on house mice (*Mus domesticus*) from the 16th generation of a selective-breeding experiment for increased voluntary wheel-running exercise (Swallow et al. 1998; Garland 2003). The selection experiment comprises four selection and four control lines; each line is propagated with 10 breeding pairs that contribute one male and one female to the next generation, with the condition that siblings are not mated. Breeders are chosen either randomly with respect to wheel running (control lines) or selectively, within family, as the mice that run the most revolutions on days five and six of six days of wheel exposure administered at eight weeks of age (selection lines). The individuals in this report are the offspring of the generation 15 breeders for generation 16 of the selection experiment. Five breeding pairs from each of the eight lines were remated to produce second litters at Washington State University. Details of this design are presented in Morgan et al (2003). Briefly, pups were weaned at 21 days of age and placed in treatment groups at  $28 \pm 3$  days of age in the animal care facility at Washington State University. Four males and four females from each family were used, with half designated for the active treatment and half designated for the sedentary treatment. Each activity group thus contained two females and two males from each of the five families within each of the eight lines, for a total of 160 individuals per activity group, i.e., 320 individuals in the aging colony. Mice in the active treatment group were placed individually in cages with a 11.75 cm radius running wheel and electronic wheel-revolution counter built into the cage-top. The mouse thus had the option of voluntarily getting into the wheel and running, or of remaining in the cage and not running. On the same day, sedentary mice were placed individually in standard rodent cages. Photoperiod was 12 : 12 h, and water was available ad libitum. Mice were provided excess food weekly (Harland Teklad rodent diet W8604) and apparent food consumption was determined by weighing food hoppers. This measure does not account for possible variation in food wasting, as when mice shred food pellets and allow fragments to drop in the litter. A study of food wasting at generation 10

found no significant differences between selection and control lines, but significant variation among the replicate lines within selection group (Koteja et al. 2003). Additionally, once a week each animal was weighed and its weekly wheel revolutions downloaded from the counter device. Cages were cleaned weekly and running wheels were cleaned monthly. Extra sibs from all 40 families were placed in similar housing and were used as sentinels to monitor the colony monthly for the presence of specific pathogen exposure. At 24 months post start date, a sentinel tested positive for MHV exposure, presumably from a barrier breakdown. No other sentinels tested positive. No treatment was initiated for the virus. Subsequent necropsied mice were monitored specifically for the presence of diseased liver tissue. None were observed to have contracted hepatitis (necropsies performed by LAR veterinarians at WSU).

Before carrying out any analysis, we eliminate line effects in log body mass and wheel-running by subtracting out line means within each group (selection and control) and gender. Suppose individual  $i$  is in line  $k$  of the selection group and is female. We replace individual  $i$ 's log body mass with original log body mass minus line  $k$ 's average plus the selection group average. The line  $k$  average and the selection group average only involve females. Similarly, we do this for the wheel-running.

Let  $Z_i(\cdot)$  be the  $i$ th mouse's log body mass trajectory. We study how  $Z_i(\cdot)$  affects  $Y_i$ , the averaged wheel running response over weeks 5 to 60. We use average wheel-running since the wheel running distance data have many missing values and are very noisy. In addition, since the running wheels were cleaned every four weeks, we see spikes in wheel-running activity every fourth week.

In the following, we describe our model for the dependence of wheel-running on log body mass. This dependence is parameterized through an unknown function  $\beta$ . We estimate  $\beta$  and provide standard errors. We also test the null hypothesis that  $\beta = 0$  (no dependence of wheel-running on log body mass) and test that two groups have the same  $\beta$ 's. We use permutation type methods to calculate the p-values. For more details of the statistical methodology, see Heckman and Wang (2007).

We model  $Z_i$  using random effects with a set of basis functions  $\phi_1, \dots, \phi_K$  as

$$Z_i(t) = \mu(t) + \sum_{k=1}^K x_{ik} \phi_k(t),$$

where  $\mu$  is a smooth unknown function. The random effects  $\mathbf{x}_i \equiv (x_{i1}, \dots, x_{iK})'$  are assumed normally distributed as  $N(\mathbf{0}, \Sigma_x)$ . The log body mass trajectory  $Z_i$  is observed with error at a sequence of weeks, with the observed value of  $Z_i$  at week  $j$  equal to  $z_{ij} \equiv Z_i(j) + \epsilon_{ij}$ . The  $\epsilon_{ij}$ 's are independent and normally distributed with mean 0 and variance  $\sigma_\epsilon^2$ . Let  $\mathbf{z}_i$  contain the observed values of the  $z_{ij}$ 's. We can then write

$$\mathbf{z}_i = \boldsymbol{\mu} + \mathbf{A}\mathbf{x}_i + \boldsymbol{\epsilon}_i,$$

$$\boldsymbol{\mu} = (\mu(-1), \dots, \mu(60))', \quad \mathbf{A}_{jk} = \phi_k(j), \quad \boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{in})'.$$

The dependence of  $Y_i$  on  $Z_i(\cdot)$  is reflected through the unknown function  $\beta(\cdot)$  as

$$Y_i = \beta_0 + \int_{-1}^{60} \beta(t) \left[ Z_i(t) - E(Z_i(t)) \right] dt + e_i,$$

where  $e_i$  is an error term distributed as  $N(0, \sigma^2)$ . The functional parameter  $\beta$  is of primary interest and can be viewed as a generalization of the conventional multiple regression coefficient vector. Choosing a basis for  $\beta$  (for ease, the same as for the  $Z_i(\cdot)$ ), we approximate  $\beta$  as

$$\beta(t) = \sum_{k=1}^K \beta_k \phi_k(t).$$

Together with our modelling of  $Z_i(\cdot)$ , we can write the model of  $Y_i$  as

$$Y_i = \beta_0 + \boldsymbol{\beta}' \mathbf{T}\mathbf{x}_i + e_i,$$

$$\boldsymbol{\beta} = (\beta_1, \dots, \beta_K)', \quad \mathbf{T}_{jk} = \int_{-1}^{60} \phi_j(t) \phi_k(t) dt.$$

The goal is to use information of the observed data, the  $\mathbf{z}_i$ 's and  $Y_i$ 's, to estimate the coefficient vector  $\boldsymbol{\beta}$ .

Before estimating  $\beta$  along with the other unknown parameters in the model, we must first estimate the basis functions, the  $\phi_k$ 's. We apply a smoothed eigenanalysis to the  $z_{ij}$ 's, penalizing an approximation of the second derivative of the eigenfunction. We choose the amount of smoothing by eye. A detailed description of choosing the eigenfunctions is provided in Section 4.3. Treating the estimated  $\phi_k$ 's as fixed and known, we then estimate the model parameters via maximum likelihood. We use an iterative numerical method, an implementation of the ECME (Expectation/Conditional Maximization Either) algorithm of Liu and Rubin (1994). Obtaining the estimate  $\hat{\beta}$ , our estimate of  $\beta(t)$  is  $\hat{\beta}(t) = \sum_{k=1}^K \hat{\beta}_k \phi_k(t)$ . We estimate the covariance matrix of  $\hat{\beta}$  by both a Hessian matrix based method and a bootstrapped method. Using the estimated covariance matrix, we can estimate the variance of  $\hat{\beta}(t)$ , and thus make inference for  $\beta$ .

To determine if  $Y_i$  depends on  $Z_i(\cdot)$ , we test

$$H_o : \beta(t) = 0, \text{ for all } t \in [-1, 60]$$

via an integrated t-statistic

$$U_f = \int_{-1}^{60} \frac{\hat{\beta}^2(t)}{\text{var}(\hat{\beta}(t))} dt.$$

To calculate the null distribution of  $U_f$ , and thus calculate a p-value, we use a permutation type method, one that does not rely on distributional assumptions. Under the null hypothesis that  $\beta(t) \equiv 0$  for all  $t \in [-1, 60]$ ,  $Y_i$  and  $\mathbf{z}_i$  are independent. Therefore the joint distribution of  $(\mathbf{z}_i, Y_i)$  is the same as that of  $(\mathbf{z}_i, Y_{i'})$  where  $i'$  is chosen at random from  $\{1, \dots, N\}$  and  $N$  is the number of individuals. To simulate the null distribution of  $U_f$ , we make  $Q = 500$  new data sets via permuting the  $Y_i$ 's: the  $q$ th data set is simply  $\{\mathbf{z}_1, Y_{i_1}, \dots, \mathbf{z}_N, Y_{i_N}\}$  where  $i_1, \dots, i_N$  is a random permutation of  $1, \dots, N$ . The p-value is then calculated as the proportion of the resulting  $Q$  statistic values larger than the original observed value.

Consider the two-sample problem where there are observations independently from a selection group and from a control group. To see if the  $\beta$ 's

governing the  $Y$ 's and  $Z$ 's in the selection group and the control are equal, we test

$$H_o : \beta^s(t) = \beta^c(t), \text{ for all } t \in [-1, 60],$$

where the superscript “ $s$ ” or “ $c$ ” indicates the group membership, “selection” or “control” respectively. To test  $H_o$ , we use the two sample test statistic

$$U_f = \int \frac{[\hat{\beta}^s(t) - \hat{\beta}^c(t)]^2}{\widehat{\text{var}}(\hat{\beta}^s(t)) + \widehat{\text{var}}(\hat{\beta}^c(t))} dt.$$

Again, we use a permutation method to calculate the null distribution of the two sample  $U_f$  and thus the p-values. Under the null hypothesis that  $\beta^s(t) = \beta^c(t)$  for all  $t \in [-1, 60]$ , the dependence of  $Y_i$  on  $\mathbf{z}_i$  is identical in both groups. We generate  $Q = 500$  “data sets” from the original data set, data sets that follow the null hypothesis. To construct the  $q$ th “data set”, we randomly split the  $N^s + N^c$  individuals into two groups of size  $N^s$  and  $N^c$  and calculate the resulting  $U_f$ . We use the empirical distribution of the obtained  $Q$  test statistic values to approximate the null distribution of our statistic. The p-value is then calculated as the proportion of the  $Q$  statistic values larger than the original observed value.

The use of function-valued traits in evolutionary biology was first proposed by Kirkpatrick and Heckman (1989). The methodology has found applications in animal breeding, via basis function expansions with random coefficients. See Meyer (2005) and references therein.

### 4.3 Results

The body mass trajectories for each experimental group are presented in Figure 4.1. For all four groups, body mass is a roughly monotone increasing function of age, with the highest rate of increase from ages  $-1$  to 4 weeks. The body mass trajectories of males, both selected and control, are more variable than those of the females. Histograms of average wheel running from ages 5 to 60 are presented in Figure 4.2; within each sex, selected mice ran more than control mice and females ran more than males (Morgan et

al., 2003). In order to determine a few functions which adequately describe the body mass trajectories, a smoothed eigenanalysis of the trajectories was conducted. This smoothed eigenanalysis yielded a first eigenfunction that was close constant, indicating that the biggest source of variability in body mass was overall size of the mouse. Since a constant eigenfunction is biologically meaningful, the first basis function was forced to be constant and the remaining functions calculated via a smoothed eigenanalysis on the trajectories, centered as follows. Let

$$\bar{z}_i = \frac{1}{58} \sum_{\substack{k=-1 \\ k \neq 34, 38, 39, 50}}^{60} z_{ik}$$

be the  $i$ th mouse's average body mass. We apply a smoothed eigenanalysis to the sample covariance matrix of the centered body mass

$$z_{ij} - \bar{z}_i \quad j = -1, \dots, 60, j \neq 34, 38, 39, 50.$$

Using the constant as the first eigenfunction, we found that the corresponding first principal component, overall size, accounts for 85% of the variability in the male selection group, 72% in the male control group, 70% in the female selection group and 63% in the female control group. Figure 4.3 shows the proportion of cumulative variance explained by the next ten principal components in each group after removing the variability of the first principal component. The first three smoothed eigenfunctions in each group capture about 90% of the remaining variability of the body mass trajectories so these three eigenfunctions plus the constant function will be used in the subsequent analyses. Figure 4.4 shows each of these four leading eigenfunctions; the functions of the four experimental groups are presented together in each panel. Clearly the smoothed eigenfunctions are very similar among the four experimental groups.

We model the mean of the averaged wheel running as

$$E(Y) = \beta_0 + \int \beta(t)[Z(t) - EZ(t)]dt,$$

where  $Z(t)$  is the logarithm of body mass at age  $t$ . Since the observed body mass curves are somewhat rougher than one would expect, we model the observed log body mass at age  $t$  as  $Z(t)$  plus error. To estimate  $\beta(\cdot)$ , we approximate both  $\beta(\cdot)$  and  $Z(\cdot)$  using basis function expansions. The basis functions used are the estimated eigenfunctions. To model within individual variability in body mass, we use a random effects model, that is, we assume that the coefficients in each individual's  $Z(\cdot)$  expansion are random. We then estimate all parameters by maximum likelihood.

The  $\hat{\beta}$  is estimated for each group from the smoothed eigenfunctions. These  $\hat{\beta}$ 's are presented in Figure 4.5. Both the male selection and control groups show regions of  $\hat{\beta}$  that are non-overlapping with the 0 line and formal testing shows that  $\hat{\beta}$  is significantly different from 0 for both selected and control males (Figure 4.5; Table 4.1;  $p = 0.008$  and  $0.034$ , respectively). Hence in male selection and control mice average wheel running significantly depends on body mass trajectory. Interestingly,  $\hat{\beta}$  for the selection males is positive at very young ages, then negative just prior to 10 weeks, positive again at middle age and then negative or not different from 0 at older ages. For females,  $\hat{\beta}$  overlaps 0 and the results of the formal testing are not significant for either selection or control groups (Figure 4.5;  $p = 0.264$  and  $0.656$ , respectively). Hence for all females, there is no evidence that average wheel running depends on the body mass trajectory.

Figure 4.6 shows both the pointwise estimates and the model estimates of the covariance between average wheel running and body mass trajectory for all four experimental groups. The general shapes of both estimates are similar to each other and they are similar to the shapes of the curves seen for the  $\hat{\beta}$  estimates in Figure 4.5. For instance, in each group except for the female control, at the first few ages the covariance between average wheel-running and log body mass is large. Correspondingly, at these ages each  $\hat{\beta}$  is also large, where we believe that average wheel running is positively dependent on log body mass.

The  $\hat{\beta}$  for selection and control mice were compared within each gender to test for the evolution of the relationship of body mass trajectory and average wheel running (Figure 4.7). No significant differences were detected

between selected and control mice in either gender (Table 4.2); no evolution of the dependence of wheel running on body mass trajectory in response to selection on age-specific wheel running can be detected.

## **4.4 Discussion**

As hypothesized, both male selected and control mice showed a significant dependence of wheel running on the body mass ontogeny. This result generally fits those of Swallow et al. (1999), who measured a negative genetic correlation of  $-0.5$  between wheel running and body mass, and Morgan et al. (2003) who measured a negative correlated response of the body mass ontogeny to selection on early-age wheel running. However, the analyses used herein have revealed complexity that was not identified in the two previous studies: both previous studies showed clear negative relationships between body mass and wheel running, whereas we have shown that the relationship is sometimes negative and sometimes positive, depending on age. As can be seen in Figures 4.5 and 4.6, body mass has a positive effect on wheel running at young and some middle ages, with a negative effect at some younger ages and at older ages. Interestingly Swallow et al. (1999) conducted their study at the younger ages ( $\sim$  weeks 8 to 15) when the relationship is negative, and although Morgan et al. (2003) clearly showed a negative genetic covariance between the body mass ontogeny and early-age wheel running, that does not preclude the phenotypic covariance from being at times positive. Hence our results do not disagree with those of the earlier studies but rather elucidate complexity that the earlier studies were not able to identify.

A significant dependence of wheel running on the body mass trajectory was not identified in females. Different results in males and females is not surprising in the context of this model system of mice, in which males and females have differed in traits as diverse as wheel running itself (Swallow et al., 1998) and longevity (Bronikowski et al., 2006). On the other hand, the lack of a measurable difference in females may be caused by a lack of power.

Given that Morgan et al. (2003) showed a negative correlated response of the wheel running trajectory to selection on early age body mass, we



had hypothesized that the dependence of body mass on the wheel running trajectory would have evolved and hence be different between selected and control mice. However, no such differences were identified. Hence we can not demonstrate any evolution of the dependence of wheel running on body mass trajectory. Several reasons may explain this. First, it is possible that the dependence is evolving, just slowly, and that a similar study conducted on a later generation would reveal differences between selected and control mice. Second, the dependence of wheel running on body mass may be constrained by genetic covariances between wheel running and body mass, effectively making it difficult to alter the dependence. Such a constraint could be identified by detailed quantitative genetic analyses of the wheel running and body mass trajectories (Kingsolver et al., 2001). Third, the inability to demonstrate any evolution dependence of wheel running on body mass trajectory could simply be because of a lack of power.

Developing powerful yet flexible statistical techniques to analyze data sets with a large number of observations on each individual is a challenge. We often do not have enough degrees of freedom to carry out a “usual” regression analysis, modelling  $Y_i = \beta_0 + \sum \beta_j Z(t_j) + e_{ij}$ : for our data, within one group we have around 40 individuals and 58 values of  $\beta_j$ . The function-valued approach effectively reduces the number of parameters in the model. Using eigenfunctions as basis functions has allowed us to fit our data using only  $\beta_0$  and 4  $\beta_j$ 's. Griswold, Gomulkiewicz and Heckman (2007) have found via simulation that basis-function type techniques are more powerful than traditional ones. Although they have considered a slightly different problem, applicable to a comparison of log body mass trajectories of two groups, one would expect their results to hold qualitatively for the problem considered here. The statistics methodology was developed by Heckman and Wang (2007), with this its first application. Here we see its benefits: the method has exposed patterns in dependence that were previously unknown. The method allows full use of information from all individuals, through the use of random effects modelling. And by using a function-valued approach, we can handle large number of observations per individual.

Group	p-value
Male selected	0.008
Male control	0.034
Female selected	0.264
Female control	0.656

Table 4.1: P-values of the test  $H_o : \beta(t) = 0$ , for all  $t \in [-1, 60]$ , using test statistic  $U_f$ .

Gender	p-value
Male	0.110
Female	0.444

Table 4.2: P-values of the test  $H_o : \beta^s(t) = \beta^c(t)$ , for all  $t \in [-1, 60]$ , using test statistic  $U_f$ .

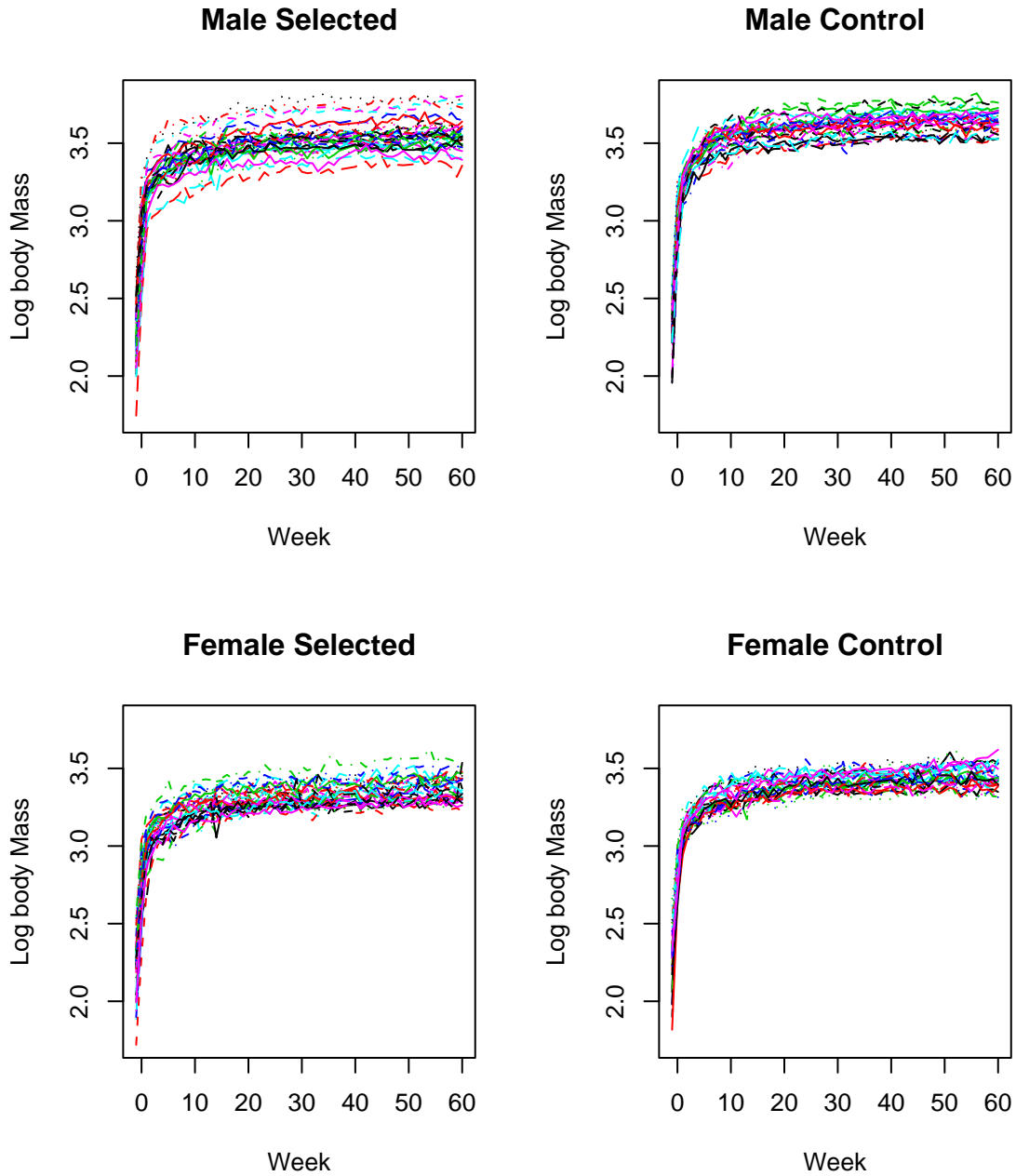


Figure 4.1: Plots of log body mass versus week for the four groups of laboratory mice.

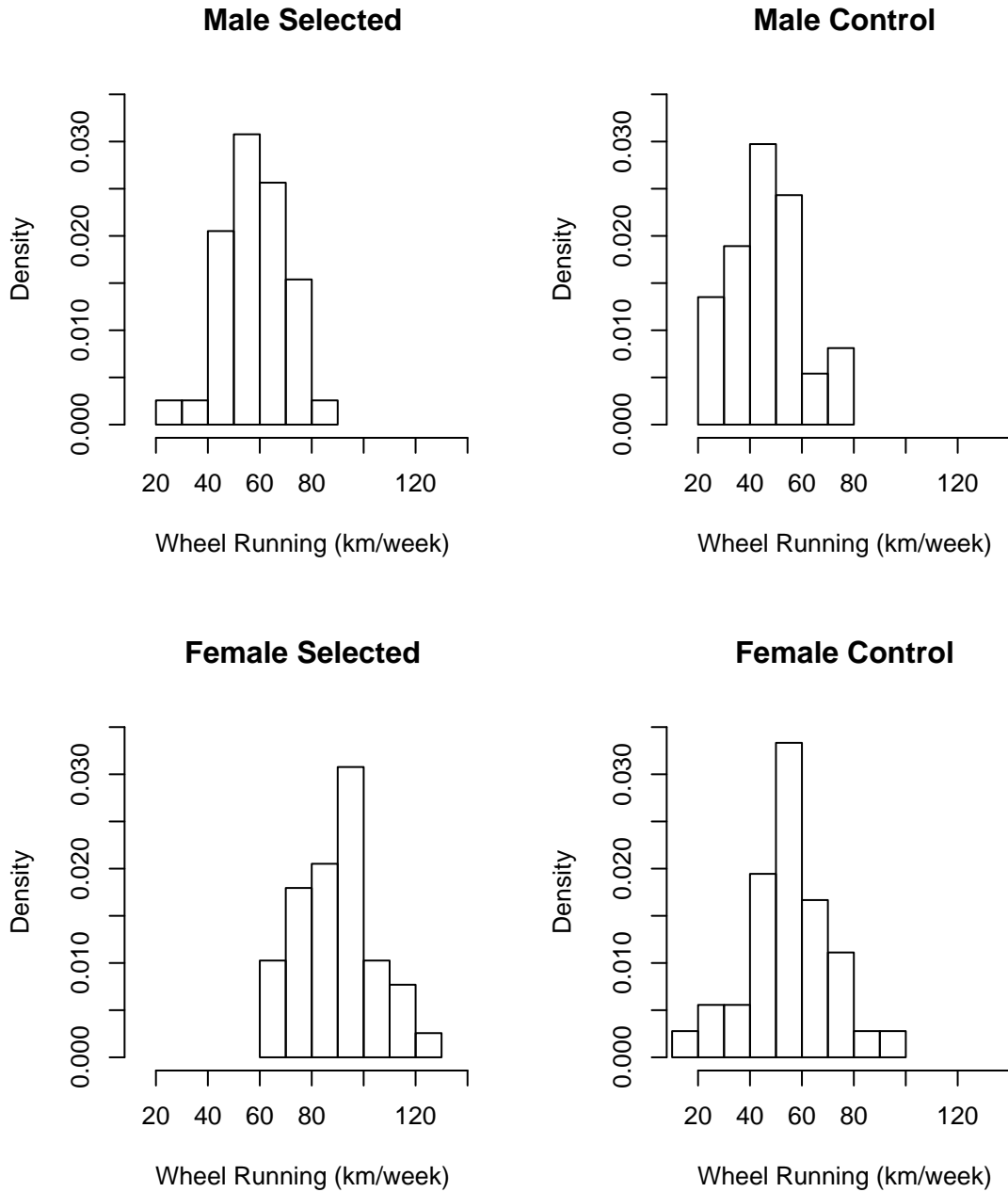


Figure 4.2: Histogram of the response: averaged wheel running from weeks 5 to 60 for the four groups of laboratory mice.

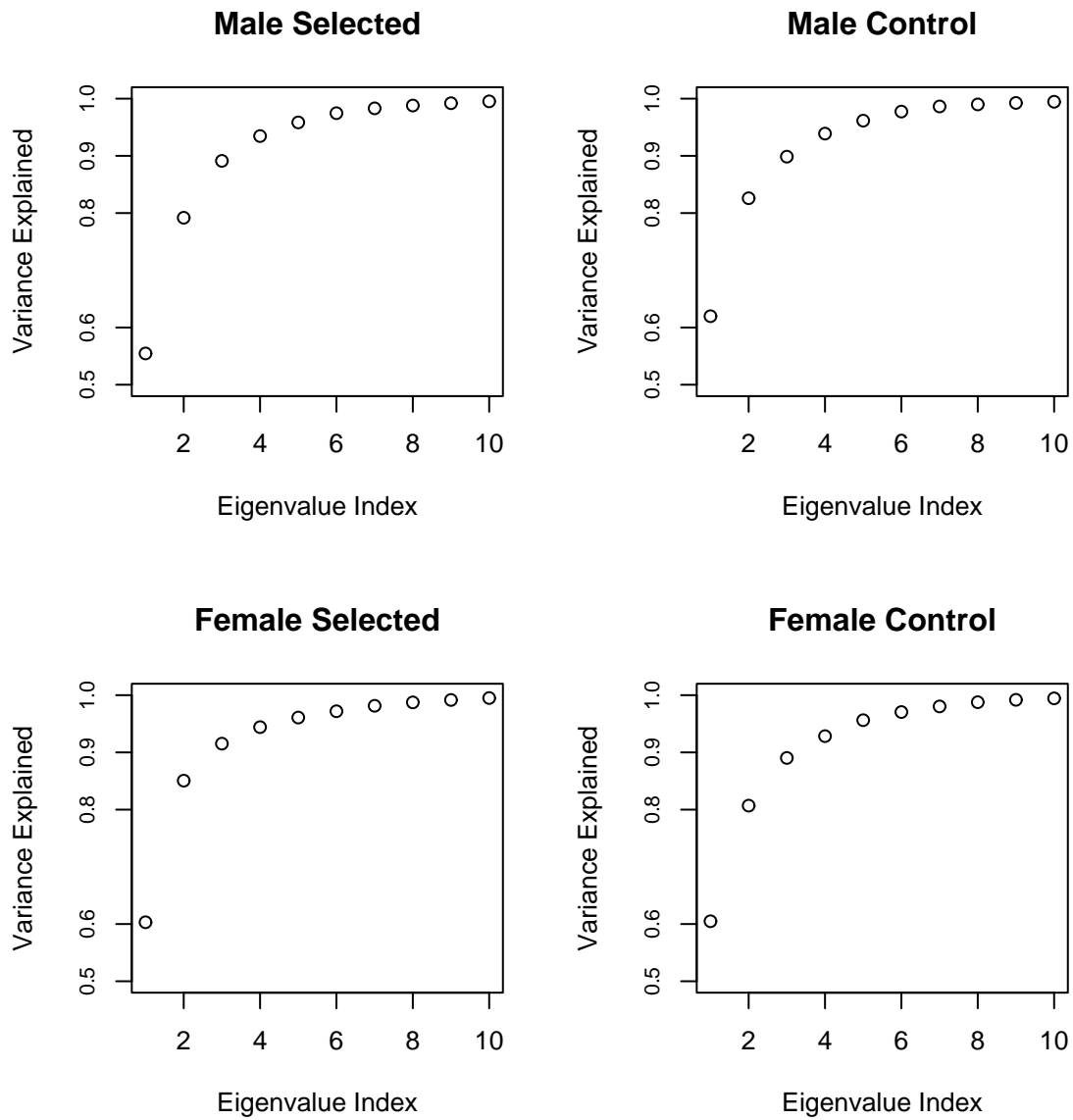


Figure 4.3: Plots of the proportion of cumulative variance of the centered log body mass explained by the first ten principal components for four groups of mice, after the individual average log body mass has been removed.

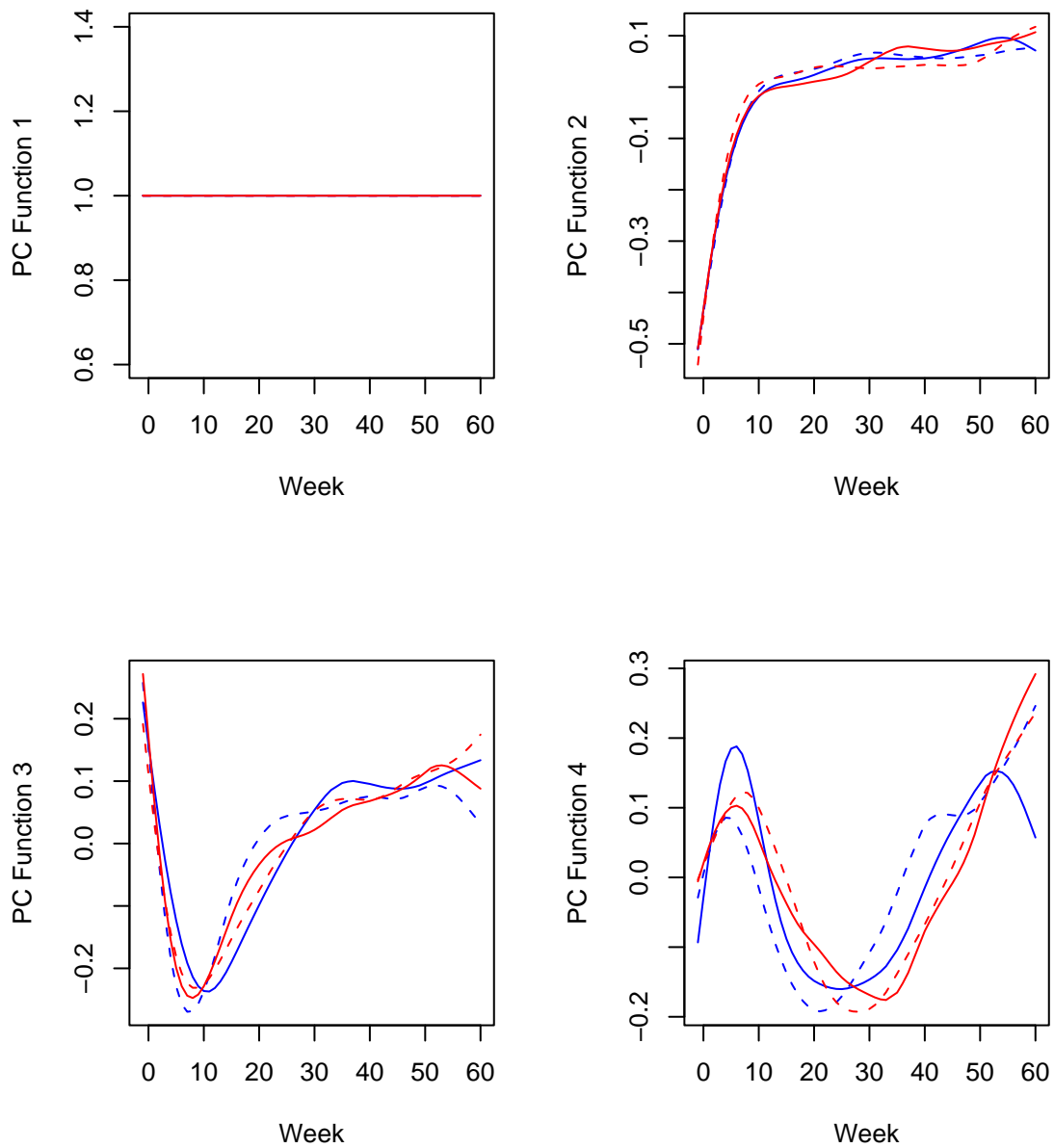


Figure 4.4: The constant function and the first three smoothed eigenfunctions of the covariance of centered log body mass of four group. The four groups' functions are together in one panel. The eigenfunctions are calculated as described in Section 4.3.

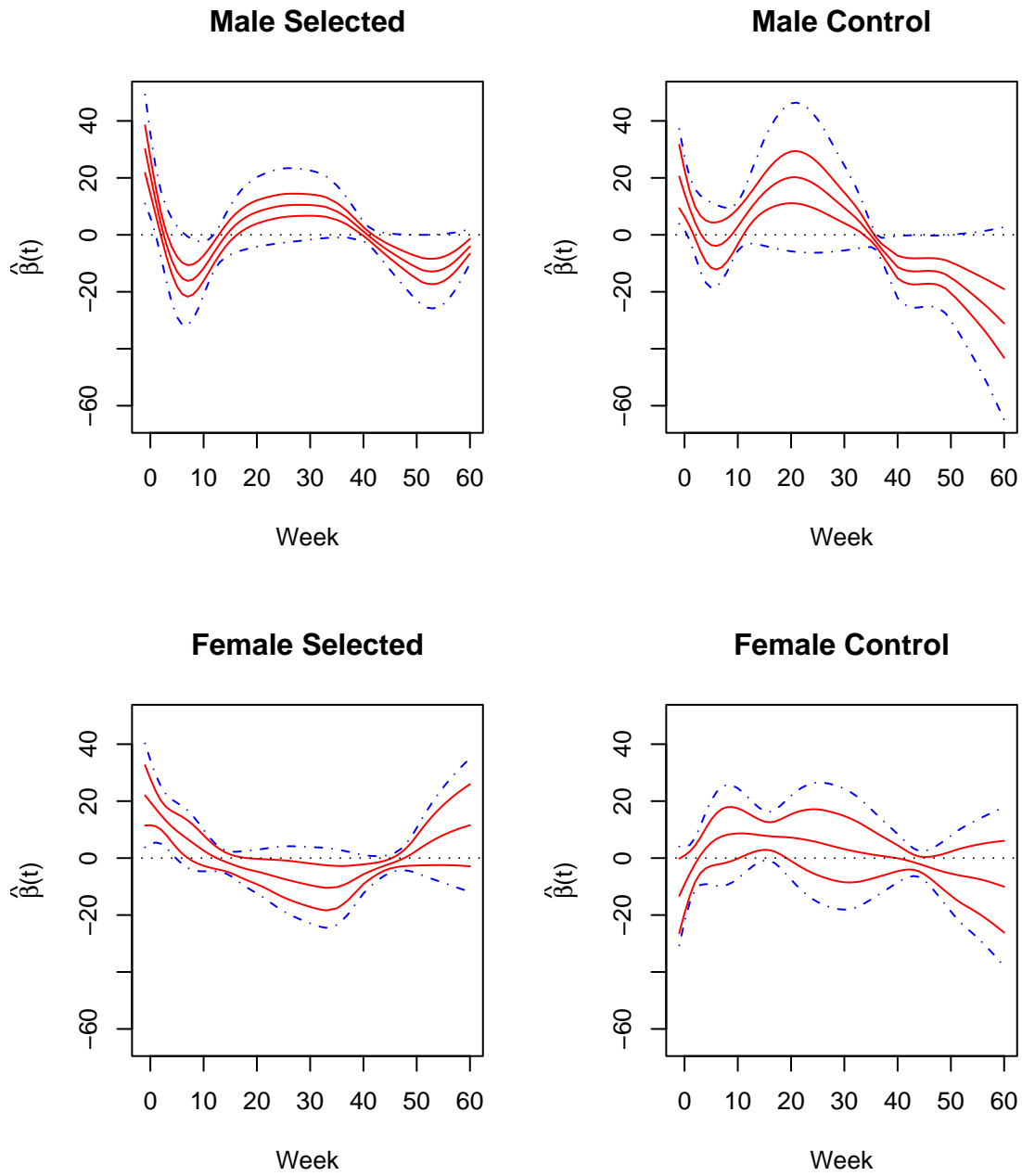


Figure 4.5: Plots of  $\hat{\beta}$  and its standard errors computed from the Hessian matrix (solid) and from the bootstrap (dash-dot) for four groups of mice. Standard errors are calculated as described in Section 4.2.

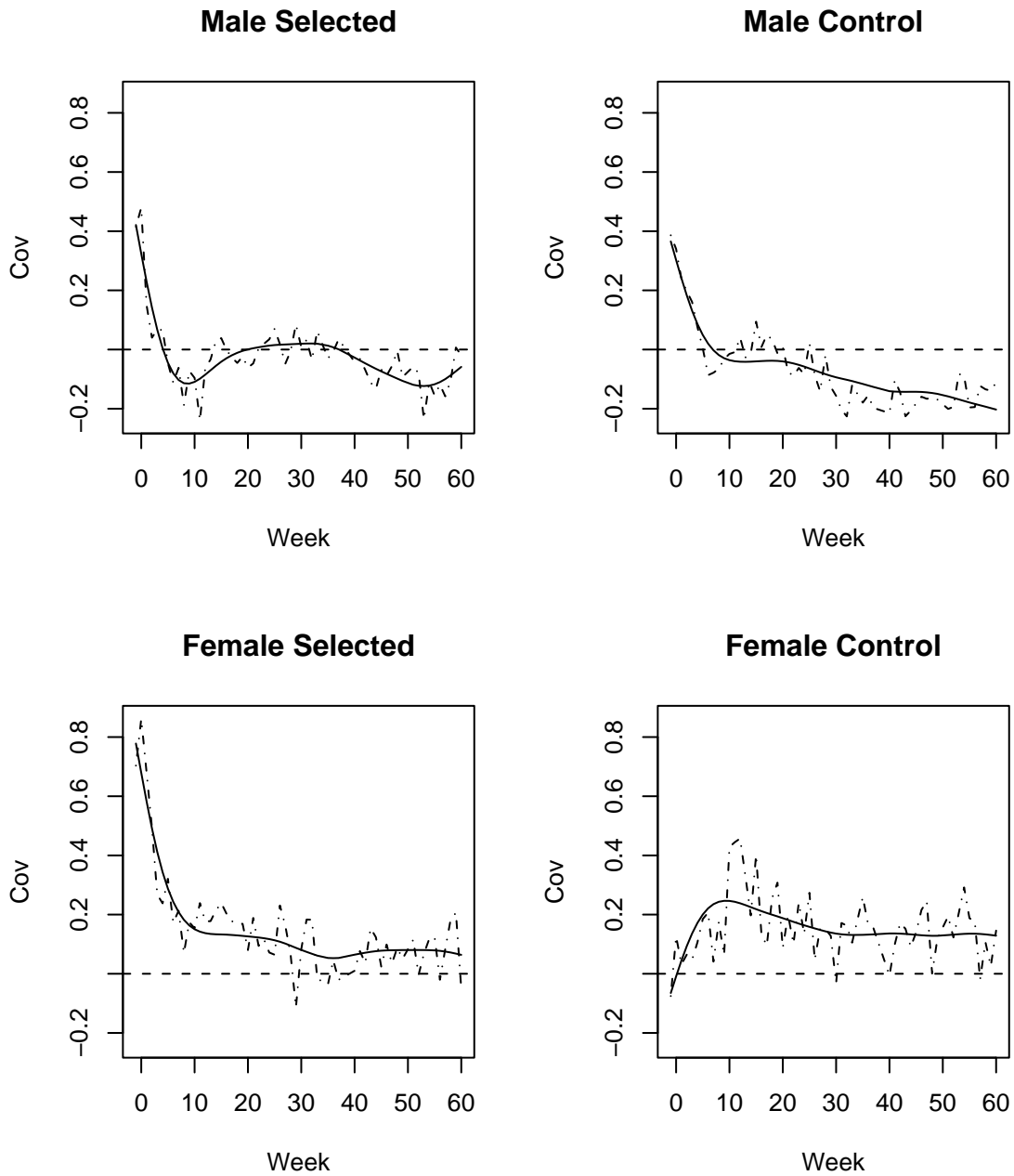


Figure 4.6: Plots of the covariances between the average wheel running and the log body mass trajectory at week  $j$ ,  $j = -1, \dots, 60$ , for four groups of mice. Solid lines are the covariance from the model and dash-dot lines are the pointwise sample covariances between average wheel running and log body mass. For the calculation, refer to Heckmand and Wang (2007).



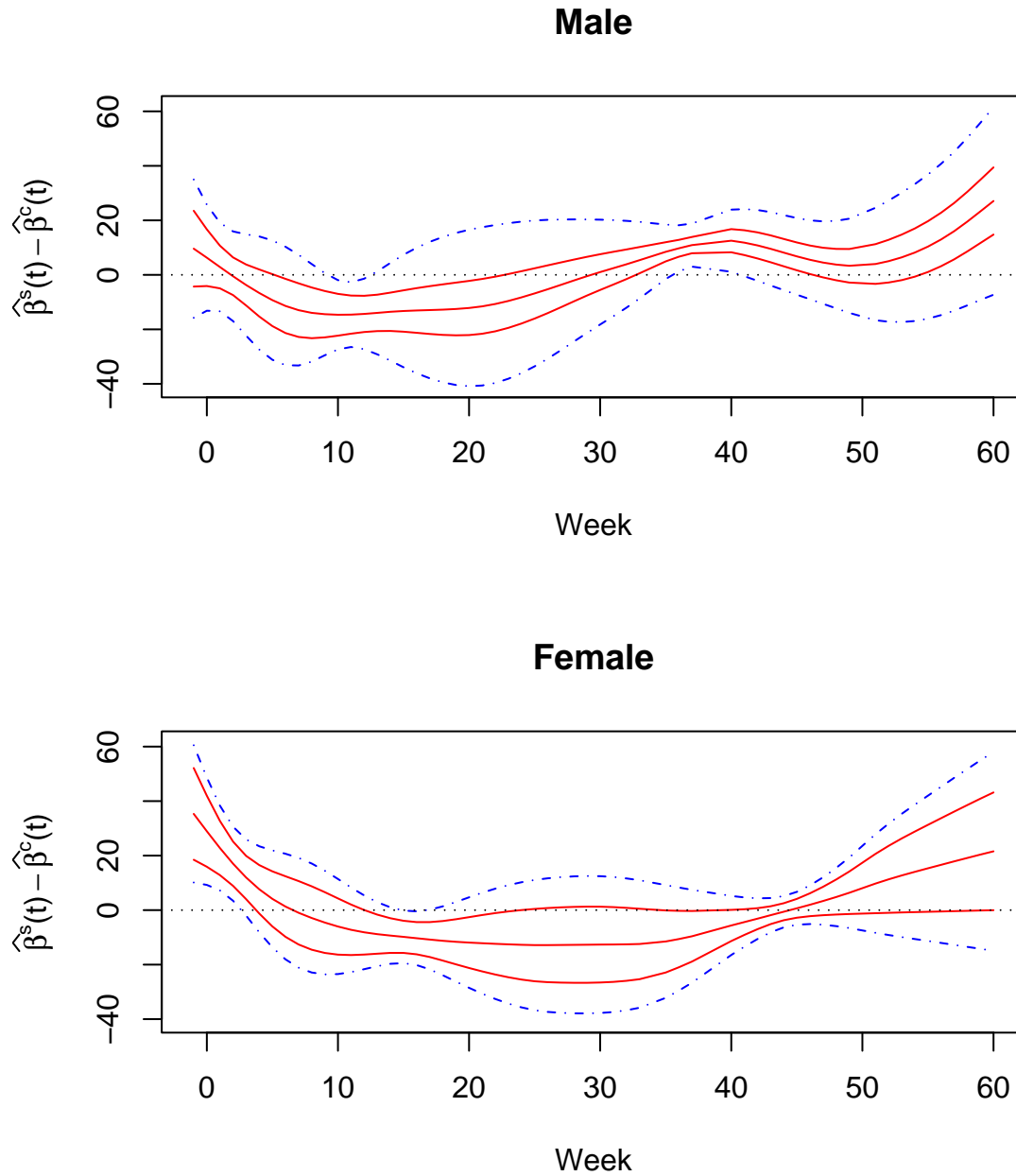


Figure 4.7: Plots of  $\hat{\beta}^s - \hat{\beta}^c$ , the difference between the selection  $\hat{\beta}^s$  and the control  $\hat{\beta}^c$ , and the standard errors of the difference computed from the Hessian matrix (solid) and from the bootstrap (dash-dot) for both genders.

# Bibliography

- [1] Bronikowski, A. M., Morgan, T. J., Garland, T., Jr. and Carter, P. A. (2006) The evolution of aging and age-related physical decline in mice selectively bred for high voluntary exercise. *Evolution* 60, 1494-1508.
- [2] Dewsbury, D. A. (1980) Wheel-running behavior in 12 species of murid rodents. *Behav. Process* 6, 271-280.
- [3] Estabrook, G. F., Smith, G. R. and Dowling, T. E. (2007) Body mass and temperature influence rates of mitochondrial DNA evolution in North American Cyprinid fish. *Evolution* 61, 1176-1187.
- [4] Garland, T., Jr. (1983) Scaling the ecological cost of transport to body mass in terrestrial mammals. *Am. Nat* 121, 571-587.
- [5] Garland, T., Jr. (2003) Selection experiments: an under-utilized tool in biomechanics and organismal biology. Pages 23-56 in Bels, V. L., Gasc, J. P., Casinos, A. (2003) *Vertebrate biomechanics and evolution*. BIOS Scientific Publishers, Oxford, U.K.
- [6] Garland, T., Jr and Carter, P. A. (1994) Evolutionary physiology. *Annu. Rev. Physiol* 56, 579-621.
- [7] Gillooly, J. F., Allen, A. P., West, G. B. and Brown, J.H. (2005) The rate of DNA evolution: effects of body size and temperature on the molecular clock. *PNAS*, 102, 140-145.
- [8] Gomulkiewicz, R. and Kirkpatrick, M. (1992) Quantitative genetics and the evolution of reaction norms. *Evolution* 46, 390-411.

- [9] Gomulkiewicz, R. and Beder, J. H. (1996) The selection gradient of an infinite-dimensional trait. *SIAM Journal of Applied Mathematics* 56, 509-523.
- [10] Griswold, C. K., Gomulkiewicz, R. and Heckman, N. (2007) Statistical power and hypothesis testing with functional data. manuscript.
- [11] Heckman, N. and Wang, W. (2007) Linear mixed models for measurement error in functional regression. manuscript, Department of Statistics, University of British Columbia.
- [12] Hone, D. W. E. and Benton, B. J. (2005) The evolution of large size: how does Cope's Rule work? *Trends in Ecology and Evolution* 20, 1, 4-6.
- [13] Kingsolver, J.G., R.S. Gomulkiewicz, and Carter, P. A. (2001) Variation, selection, and evolution of function-valued traits. *Genetica* 112, 87-104.
- [14] Kirkpatrick, M. and Heckman, N. (1989) A quantitative-genetic model for growth, shape, reaction norms, and other infinite-dimensional characters. *Journal of Mathematical Biology* 27, 429-450.
- [15] Koteja, P., Carter, P. A. and Swallow, J. G. and Garland, T., Jr. (2003) Food wasting in house mice: variation among individuals, families, and genetic lines. *Physiology & Behavior* 80, 375-383.
- [16] Meyer, K. (2005) Advances in methodology for random regression analyses. *Australian Journal of Experimental Agriculture* 45, 847-859.
- [17] Morgan, T. J., Garland, T., Jr. and Carter, P. A. (2003) Ontogenies in mice selected for high voluntary wheel-running activity. I. Mean ontogenies. *Evolution* 57, 646-657.
- [18] Sogard, S. M. (1997) Size-selective mortality in the juvenile stage of teleost fishes: A review. *Bulletin of Marine Science* 60, 1129-1157

- [19] Swallow, J. G., Carter, P. A. and Garland, T., Jr. (1998) Artificial selection for increased wheel-running behavior in house mice. *Behavior Genetics* 28, 227-237.
- [20] Swallow, J. G., Koteja, P. and Carter, P. A. and Garland, T., Jr. (1999) Artificial selection for increased wheel-running activity in house mice results in decreased body mass at maturity. *Journal of Experimental Biology* 202, 2513-2520.

## Chapter 5

# Conclusions and future plan

In the thesis, we study identifiability of linear mixed effects models in Chapter 2. Particularly in Section 2.5, we give necessary and sufficient conditions for identifiability of a model used in functional data analysis. The model is our study object in Chapter 3. There we show how to use the ECME algorithm to estimate the model parameters, develop statistics for hypothesis testing of the functional coefficient  $\beta$  and derive expressions of the residuals to check for outliers or influential points. We also apply the model to analyze a biological data set in Section 3.6. Chapter 4 provides a detailed background of the biological experiment, a description of the data, and biological interpretations of our analysis.

In this chapter, we give possible directions for future research. In Section 5.1, we adjust our Chapter 3 to account for correlated individuals. In Section 5.2, we consider generalizing our Chapter 3 model from a scalar response to a functional response.

### 5.1 Models with correlated individuals

In Chapter 4, we apply our model (3.4) and (3.5) to analyze a biological data set. Specifically, the observed  $i$ th mouse's log body mass,  $\mathbf{z}_i$ , is modelled as

$$\mathbf{z}_i = \boldsymbol{\mu} + \mathbf{A}\mathbf{x}_i + \boldsymbol{\epsilon}_i \quad (5.1)$$

and the  $i$ th mouse's average wheel running,  $Y_i$ , is modelled as

$$Y_i = \beta_0 + \boldsymbol{\beta}'\mathbf{T}\mathbf{x}_i + e_i. \quad (5.2)$$

We assume the random effects, the  $\mathbf{x}_i$ 's, are independent across individuals. However, in the biological experiment, mice may come from the same family and thus traits may be genetically correlated. This fact would lead to correlations among the random effects, the  $\mathbf{x}_i$ 's. Clearly, our independence model assumption is then violated and we should take the dependence of individuals into consideration. In this section, we show how to handle this family correlation more directly, by adjusting our model to account for this dependence.

In the experiment, two mice came from the same family. Let the subscript  $i$  index the family and the subscript  $k = 1$  or  $2$  differentiate the two family members. Models of the  $ik$ th mouse's log body mass and average wheel running are respectively

$$\mathbf{z}_{ik} = \boldsymbol{\mu} + \mathbf{A}\mathbf{x}_{ik} + \boldsymbol{\epsilon}_{ik}, \quad (5.3)$$

$$Y_{ik} = \beta_0 + \boldsymbol{\beta}'\mathbf{T}\mathbf{x}_{ik} + e_{ik}, \quad k = 1, 2. \quad (5.4)$$

We still assume that  $\text{Cov}(\mathbf{x}_{i1}) = \boldsymbol{\Sigma}_x = \text{Cov}(\mathbf{x}_{i2})$  but, in the  $i$ th family,  $\mathbf{x}_{i1}$  and  $\mathbf{x}_{i2}$  are correlated as

$$\text{Cov}(\mathbf{x}_{i1}, \mathbf{x}_{i2}) = \alpha\boldsymbol{\Sigma}_x, \quad (5.5)$$

where  $\alpha$  is a known constant. The constant,  $\alpha$ , comes from genetic theory (see, for instance, Heckman, 2003). We still assume the measurement errors are independent of the random effects and are independent across individuals as

$$\begin{aligned} \boldsymbol{\epsilon}_{ik} &\stackrel{i.i.d.}{\sim} \text{N}(\mathbf{0}, \sigma_{\epsilon}^2 \mathbf{R}), \quad e_{ik} \stackrel{i.i.d.}{\sim} \text{N}(0, \sigma^2), \\ &(\mathbf{x}_{11}, \mathbf{x}_{12}), (\mathbf{x}_{21}, \mathbf{x}_{22}), \dots, (\mathbf{x}_{N1}, \mathbf{x}_{N2}), \end{aligned}$$

$\boldsymbol{\epsilon}_{11}, \boldsymbol{\epsilon}_{12}, \dots, \boldsymbol{\epsilon}_{N1}, \boldsymbol{\epsilon}_{N2}, e_{11}, e_{12}, \dots, e_{N1}, e_{N2}$  are mutually independent.

The model (5.3) and (5.4) now includes the dependence between the two family members. In the following, we give another expression of the model, dropping the subscript  $k$ .

We first group each family's information together. Let

$$\mathbf{z}_i = \begin{pmatrix} \mathbf{z}_{i1} \\ \mathbf{z}_{i2} \end{pmatrix}, \mathbf{y}_i = \begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix}, \mathbf{x}_i = \begin{pmatrix} \mathbf{x}_{i1} \\ \mathbf{x}_{i2} \end{pmatrix}, \boldsymbol{\epsilon}_i = \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \end{pmatrix}, \text{ and } \mathbf{e}_i = \begin{pmatrix} e_{i1} \\ e_{i2} \end{pmatrix}.$$

To write a model for the  $i$ th family, let

$$\tilde{\boldsymbol{\mu}} = \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \end{pmatrix}, \tilde{\boldsymbol{\beta}}_0 = \begin{pmatrix} \beta_0 \\ \beta_0 \end{pmatrix}, \tilde{\mathbf{A}} = \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \end{pmatrix}, \text{ and } \widetilde{\boldsymbol{\beta}'\mathbf{T}} = \begin{pmatrix} \boldsymbol{\beta}'\mathbf{T} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\beta}'\mathbf{T} \end{pmatrix}.$$

From (5.3) and (5.4), our model for the  $i$ th family is

$$\mathbf{z}_i = \tilde{\boldsymbol{\mu}} + \tilde{\mathbf{A}}\mathbf{x}_i + \boldsymbol{\epsilon}_i, \quad (5.6)$$

$$\mathbf{y}_i = \tilde{\boldsymbol{\beta}}_0 + \widetilde{\boldsymbol{\beta}'\mathbf{T}}\mathbf{x}_i + \mathbf{e}_i. \quad (5.7)$$

To derive the covariances of the  $(\mathbf{z}'_i, \mathbf{y}'_i)$ , we let  $\boldsymbol{\Sigma}_{Ax} = \mathbf{A}\boldsymbol{\Sigma}_x\mathbf{A}'$ ,  $\boldsymbol{\Sigma}_{A,\beta} = \mathbf{A}\boldsymbol{\Sigma}_x\mathbf{T}'\boldsymbol{\beta}$  and  $\boldsymbol{\Sigma}_{\beta,x} = \boldsymbol{\beta}'\mathbf{T}\boldsymbol{\Sigma}_x\mathbf{T}'\boldsymbol{\beta}$ . From (5.5), it is not hard to see the following holds,

$$\begin{aligned} \text{Cov}(\mathbf{z}_i) &= \begin{pmatrix} \boldsymbol{\Sigma}_{Ax} + \boldsymbol{\Sigma}_\epsilon & \alpha\boldsymbol{\Sigma}_{Ax} \\ \alpha\boldsymbol{\Sigma}_{Ax} & \boldsymbol{\Sigma}_{Ax} + \boldsymbol{\Sigma}_\epsilon \end{pmatrix}, \\ \text{Cov}(\mathbf{z}_i, \mathbf{y}_i) &= \begin{pmatrix} \boldsymbol{\Sigma}_{A,\beta} & \alpha\boldsymbol{\Sigma}_{A,\beta} \\ \alpha\boldsymbol{\Sigma}_{A,\beta} & \boldsymbol{\Sigma}_{A,\beta} \end{pmatrix}, \\ \text{Cov}(\mathbf{y}_i) &= \begin{pmatrix} \boldsymbol{\Sigma}_{\beta,x} + \sigma^2 & \alpha\boldsymbol{\Sigma}_{\beta,x} \\ \alpha\boldsymbol{\Sigma}_{\beta,x} & \boldsymbol{\Sigma}_{\beta,x} + \sigma^2 \end{pmatrix}. \end{aligned}$$

Model (5.6) and (5.7) uses individuals' family information together and the correlation between family members is reflected in the model's covariance structure. To estimate the model parameters, we can use the ECME algorithm. But we may need to modify the algorithm as the model covariance structure is not in a simple form.

## 5.2 Proposing models with a functional response

In Chapter 3, we applied our model to analyze the mouse wheel running data set. In the experiment, the wheel running distances were observed every week for 60 consecutive weeks. We took the averaged wheel running distance as a scalar response, averaged to reduce nuisance effects in the wheel running. However, we would like to consider the entire wheel-running trajectory as a response and to study its dependence on the log body mass trajectory. In this section, we introduce a model which generalizes a scalar response to a functional response. We give an outline of the plan to study the model. Finally, we give a literature review of current work on this functional response and predictor model.

### 5.2.1 Models of a functional response

Recall our modelling of the scalar response  $Y_i$  is

$$Y_i = \beta_0 + \int_a^b \beta(s) [Z_i(s) - \mathbf{E}(Z_i(s))] ds + e_i.$$

Now we have a functional response  $Y_i$  indexed by  $t$ . Replacing the univariate  $\beta(\cdot)$  function with a bivariate  $\beta(\cdot, \cdot)$ , we consider generalizing the scalar response model to accomodate the functional response as

$$Y_i(t) = \alpha(t) + \int_a^b \beta(t, s) \left[ Z_i(s) - \mathbf{E}(Z_i(s)) \right] ds + e_i(t). \quad (5.8)$$

We consider a tensor-product expansion of  $\beta(\cdot, \cdot)$  using basis functions,  $\varphi_1, \dots, \varphi_J$  and  $\psi_1, \dots, \psi_L$ , as

$$\beta(t, s) = \sum_{j=1}^J \sum_{l=1}^L b_{jl} \varphi_j(t) \psi_l(s) \equiv \boldsymbol{\varphi}(t)' \mathbf{B} \boldsymbol{\psi}(s), \quad \mathbf{B}[j, l] = b_{jl}. \quad (5.9)$$

The modelling of the predictor  $Z_i$  remains the same as

$$Z_i(s) = \mu(s) + \sum_{k=1}^K x_{ik} \phi_k(s) \equiv \mu(s) + \boldsymbol{\phi}(s)' \mathbf{x}_i,$$



$\mu$  smooth,  $\mathbf{x}_i \equiv (x_{i1}, \dots, x_{iK})' \stackrel{i.i.d.}{\sim} N(\mathbf{0}, \Sigma_x)$ .

Thus we can write

$$Y_i(t) = \alpha(t) + \boldsymbol{\varphi}(t)' \mathbf{B} \mathbf{T} \mathbf{x}_i + e_i(t),$$

$$\mathbf{T}_{lk} = \int_a^b \psi_l(s) \phi_k(s) ds \equiv \int \boldsymbol{\psi}(s) \boldsymbol{\phi}(s)' ds. \quad (5.10)$$

Both the  $Z_i$  and  $Y_i$  are observed at a sequence of discrete time points. Typically, we assume the observed  $Z_i$  are perturbed with measurement errors. Let  $\epsilon_{ij}$  represent the measurement error of  $Z_i$  at time  $s_{ij}$ ,  $j = 1, \dots, n_i$ . Let  $z_{ij}$  be the contaminated observed value as

$$z_{ij} = Z_i(s_{ij}) + \epsilon_{ij} \equiv \mu(s_{ij}) + \boldsymbol{\phi}(s_{ij})' \mathbf{x}_i + \epsilon_{ij}. \quad (5.11)$$

Let  $\boldsymbol{\mu} = (\mu(s_{i1}), \dots, \mu(s_{in_i}))'$ ,  $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{in_i})'$  represent errors and  $\mathbf{z}_i = (z_{i1}, \dots, z_{in_i})'$  be the observed values. Write

$$\mathbf{z}_i = \boldsymbol{\mu} + \mathbf{A} \mathbf{x}_i + \boldsymbol{\epsilon}_i,$$

$$\mathbf{A}_{jk} = \phi_k(s_j), \quad j = 1, \dots, n_i, \quad k = 1, \dots, K,$$

$$\boldsymbol{\epsilon}_i \sim N(0, \Sigma_{\epsilon_i}),$$

$$\mathbf{x}_1, \dots, \mathbf{x}_n, \boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n \text{ are mutually independent.} \quad (5.12)$$

Suppose the process  $Y_i$  is observed at a sequence of discrete time points  $(t_1, \dots, t_{m_i})$ . Let  $\boldsymbol{\alpha} = (\alpha(t_1), \dots, \alpha(t_{m_i}))'$ ,  $\mathbf{e}_i = (e_i(t_1), \dots, e_i(t_{m_i}))'$  represent errors and  $\mathbf{y}_i = (y_i(t_1), \dots, y_i(t_{m_i}))' \equiv (y_{i1}, \dots, y_{im_i})'$  be the observed values. We then write

$$\mathbf{y}_i = \boldsymbol{\alpha} + \boldsymbol{\Phi} \mathbf{B} \mathbf{T} \mathbf{x}_i + \mathbf{e}_i, \quad \boldsymbol{\Phi}_{ij} = \varphi_j(t_i),$$

$$\mathbf{e}_i \text{ independent of both } \mathbf{x}_i \text{ and } \boldsymbol{\epsilon}_i,$$

$$\mathbf{e}_i \sim N(\mathbf{0}, \Sigma_{\mathbf{e}_i}),$$

$$\mathbf{x}_1, \dots, \mathbf{x}_n, \boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n, \mathbf{e}_1, \dots, \mathbf{e}_n \text{ are mutually independent.} \quad (5.13)$$

The data available are the observed  $\mathbf{y}_i$ 's and the  $\mathbf{z}_i$ 's and the unknown parameter of most interest is  $\mathbf{B}$ . We then use (5.12) and (5.13) as our model of the functional response and the functional predictor for data analysis.

### 5.2.2 Outline of the model study

We initiate our plan to study the model (5.12) and (5.13) in this section. The steps basically follow those in Chapter 3: model identifiability, parameter estimation, hypothesis testing and model diagnostics. Throughout, the basis functions are assumed known and fixed. We give a discussion of the choice of basis functions at the end.

In Section 2.5 we gave conditions of model identifiability for the scalar response model. We can use similar methods to study identifiability of this functional response model with covariance parameters now  $\{\Sigma_x, \Sigma_{\epsilon_i}, \Sigma_{e_i}, i = 1, \dots, N\}$ . New results will be needed to guarantee the model identifiability with restrictions on the  $\Sigma_{\epsilon_i}$ 's and the  $\Sigma_{e_i}$ 's.

After identifying the model parameters, we can proceed to estimate the unknown parameters using an ECME algorithm assuming the basis functions known and fixed. The parameter of most interest is the matrix  $\mathbf{B}$  in (5.9). After obtaining the estimate, similar approaches using the Hessian matrix or bootstrap can be applied to get the standard errors of  $\hat{\mathbf{B}}$ . For hypothesis testing of  $\beta$ , it may be of interest to know if, at certain fixed  $t$  point,  $\beta(t, \cdot)$  is zero or not. So the statistics developed in Section 4.4 can also be used to test the hypothesis  $H_o : \beta(t, s) = 0$  for all  $s \in [a, b]$  when  $t$  is fixed. After fitting the model, residuals of the fit of  $\mathbf{y}_i$  can be obtained based on the BLUP of the random effects  $\mathbf{x}_i$ . However, for each individual, the fitted value and the residual of  $\mathbf{y}_i$  now are vectors, not scalars. Appropriate methods for residual analysis need to be proposed.

The  $\varphi_j$ 's and  $\psi_l$ 's basis functions for  $\beta$  and the  $\phi_k$ 's for  $Z_i$  can be chosen using an eigenanalysis approach. As in the scalar response case, we apply an eigenanalysis to the observed  $z_{ij}$ 's, and take  $\phi_1, \dots, \phi_K$  as the first  $K$  estimated eigenfunctions. We take  $L = K$  and  $\psi_k = \phi_k$  to determine one set of the basis functions of  $\beta(\cdot, \cdot)$  in (5.9), the set of basis functions that

appear in the integral  $\mathbf{T}$  in (5.10). The other set of basis functions, the  $\varphi_j$ 's, can be chosen as the estimated eigenfunctions of the  $Y_i$  process.

An alternative to the use of a tensor-product expansion of  $\beta$  is to consider using bivariate basis function expansions. Let the  $\zeta_l$ 's be bivariate basis functions. We can approximate  $\beta$  as

$$\beta(s, t) = \sum_{l=1}^L \beta_l \zeta_l(s, t) \equiv \boldsymbol{\zeta}(s, t)' \boldsymbol{\beta}.$$

But we also need to consider what basis functions to choose.

### 5.2.3 Literature review

There are two research results closest to our work on the functional response and functional predictor model.

Ramsay and Silverman (2005, chapter 16) study a slight modification of model (5.8)

$$y_i(t) = \alpha(t) + \int \beta(t, s) Z_i(s) ds + e_i(t).$$

Besides using the tensor-product expansion of  $\beta$  as in (5.9), the authors expand  $\alpha$  using the  $\varphi_j$ 's as

$$\alpha(t) = \sum_{j=1}^{J_1} \alpha_j \varphi_j(t) \equiv \boldsymbol{\varphi}(t)' \boldsymbol{\alpha}.$$

Assuming the  $Z_i$ 's are observed without error, the authors estimate  $\boldsymbol{\alpha}$  and  $\mathbf{B}$  by minimizing the integrated residual sum of squares, which is

$$\sum_{i=1}^N \int \left[ y_i(t) - \boldsymbol{\varphi}(t)' \boldsymbol{\alpha} - \int Z_i(s) \boldsymbol{\psi}(s)' \mathbf{B} \boldsymbol{\varphi}(t) ds \right]^2 dt.$$

Yao, Müller and Wang (2005) consider (5.8) with the  $Z_i$  process observed with error. They use  $\phi_k$ 's equal to the eigenfunctions of the  $Z_i$  covariance function in (5.11). The random effects  $\mathbf{x}_i$  has a diagonal covariance matrix with elements  $\sigma_k^2$ 's being the eigenvalues of the  $Z_i$  covariance function. The errors  $\epsilon_{ij}$  are assumed to be i.i.d. with mean zero and variance  $\sigma_\epsilon^2$ .

Yao et al. carry out an eigenanalysis of the  $Y_i$ 's covariance function, with resulting eigenfunctions equal to the  $\psi_l$ 's and corresponding eigenvalues equal to the  $\sigma_l^2$ 's. They expand the observed  $y_{ik}$ , contaminated with measurement errors, as

$$y_{ik} = \mu_Y(t_{ik}) + \sum_{l=1}^L u_{il} \psi_l(t_{ik}) + e_{ik},$$

$$\mu_Y \text{ smooth, } Eu_{il} = 0, Eu_{il}^2 = \sigma_l^2, Eu_{il}u_{il'} = 0, l \neq l'.$$

$$e_{ik} \stackrel{i.i.d.}{\sim} (0, \sigma_e^2), \text{ independent of the } u'_{il}s. \quad (5.14)$$

Let  $C(s, t)$  denote the two-dimensional scatterplot smoothed cross-covariance surface  $\text{Cov}(Z(s), Y(t))$  of the  $z_{ij}$ 's and  $y_{ik}$ 's. The authors estimate  $\beta$  as

$$\hat{\beta}(s, t) = \sum_{k=1}^K \sum_{l=1}^L \frac{\sigma_{kl}}{\sigma_k^2} \phi_k(s) \psi_l(t),$$

where  $\sigma_{kl}$  is calculated as

$$\sigma_{kl} = \int \int \phi_k(s) C(s, t) \psi_l(t) ds dt.$$

Similarities between our modelling and Yao et al's are that the coefficient vector  $\mathbf{x}_i$  in the expansion of the  $Z_i$  is random and the  $Z_i$ 's are observed with measurement error. However, Yao et al's  $\mathbf{x}_i$  has a diagonal covariance structure while our covariance matrix of the  $\mathbf{x}_i$ ,  $\Sigma_x$ , can be more general. We use both the  $\mathbf{z}_i$ 's and the  $\mathbf{y}_i$ 's information to estimate  $\Sigma_x$  and other model parameters. In Yao et al., variance elements  $\sigma_k^2$ 's of the  $\mathbf{x}_i$  and variance of the error  $\epsilon_{ij}$ ,  $\sigma_e^2$ , are estimated completely by the  $\mathbf{z}_i$ 's. Similarly, the  $\sigma_l^2$ 's of the  $\mathbf{u}_i$  and  $\sigma_e^2$  of the  $e_{ik}$  are estimated completely using the  $\mathbf{y}_i$ 's.

# Bibliography

- [1] Heckman, N. (2003) Functional data analysis in evolutionary biology. *Recent Advances and Trends in Nonparametric Statistics*, 49-60, Akritas, M. G. and Politis, D. N. (editors).
- [2] Ramsay, J. O. and Silverman, B. W. (2005) *Functional Data Analysis*. 2nd edition, Springer Series in Statistics.
- [3] Yao, F., Müller, H.G., Wang, J.L. (2005) Functional linear regression analysis for longitudinal data. *Annals of Statistics* 33, 6, 2873-2903.

# Appendix A

## Appendix to Chapter 2

### A.1 Proof of Corollary 2.4.1

To prove it, we use the following result from Graybill, 1983, p.206.

**Lemma A.1.1** *Given two scalars  $a$  and  $b$ , the characteristic equation of the matrix  $\mathbf{C} = (a - b)\mathbf{I} + b\mathbf{J}$  in  $\lambda$  is*

$$(a + (n - 1)b - \lambda)(a - b - \lambda)^{n-1},$$

*and hence  $n - 1$  characteristic roots are equal to  $a - b$  and one root is equal to  $a + (n - 1)b$ .*

**Proof:**

To prove the corollary, we use Theorem 2.4.1 and a proof by contradiction.

First suppose that the model is identifiable and suppose, by way of contradiction,  $\mathbf{H}\mathbf{Z}\mathbf{J} = \mathbf{J}$ . Fix  $\boldsymbol{\Sigma}_\epsilon \in \tilde{\Theta}_\epsilon$ . Let  $s > 1$ ,  $\sigma^{*2} = s\sigma^2$  and  $\rho^* = (\rho - 1)/s + 1$ . It is not hard to check  $-1/(n - 1) < \rho^* < 1$ . Define  $\boldsymbol{\Sigma}_\epsilon^* = \sigma^{*2}[(1 - \rho^*)\mathbf{I} + \rho^*\mathbf{J}]$ . Then  $\boldsymbol{\Sigma}_\epsilon - \boldsymbol{\Sigma}_\epsilon^* = (\sigma^2 - \sigma^{*2})\mathbf{J}$  and, since  $\mathbf{H}\mathbf{Z}\mathbf{J} = \mathbf{J}$ , it is clear that (2.3) is satisfied. We now show that, for any  $\boldsymbol{\Sigma}_u \in \Theta_u$ , there exists  $s^* > 1$  so that  $\boldsymbol{\Sigma}_u^*$  defined as in (2.4) is positive definite whenever  $1 < s < s^*$ . This will show that the model is not identifiable, which contradicts our assumption. Plugging  $\boldsymbol{\Sigma}_\epsilon - \boldsymbol{\Sigma}_\epsilon^* = (\sigma^2 - \sigma^{*2})\mathbf{J}$  into (2.4) yields

$$\boldsymbol{\Sigma}_u^* = \boldsymbol{\Sigma}_u + \sigma^2(1 - s)(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{J}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}. \quad (\text{A.1})$$

By assumption  $\mathbf{1}'\mathbf{Z} \neq \mathbf{0}$  and  $\mathbf{Z}$  is of full column rank, the matrix  $(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{J}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}$  is non-negative definite and of rank one since  $\mathbf{J} = \mathbf{1}\mathbf{1}'$ . Let  $\lambda$  be its non-zero and thus the largest eigenvalue of  $(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{J}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}$ . Let  $\lambda_m$  be the

smallest eigenvalue of the matrix  $\Sigma_u$ , and let  $s^* = \lambda_m/(\lambda\sigma^2) + 1$ . For any  $\mathbf{x} \in \mathfrak{R}^q$ ,  $\mathbf{x} \neq \mathbf{0}$ ,  $\mathbf{x}'\Sigma_u^*\mathbf{x} > 0$  by the following argument.

$$\begin{aligned} \mathbf{x}'\Sigma_u^*\mathbf{x} &= \mathbf{x}'\Sigma_u\mathbf{x} + \sigma^2(1-s)\mathbf{x}'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{J}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{x} \\ &\geq \lambda_m\mathbf{x}'\mathbf{x} + \sigma^2(1-s)\lambda\mathbf{x}'\mathbf{x} \\ &> 0 \end{aligned}$$

Now suppose that  $\mathbf{H}_Z\mathbf{J} \neq \mathbf{J}$  and suppose, by contradiction, that the model is not identifiable. Then, by Theorem 2.4.1, there exist nonidentical  $\Sigma_\epsilon$  and  $\Sigma_\epsilon^*$  satisfying (2.3) and, since the rank of  $\mathbf{H}_Z$  is  $q$ , the rank of  $\Sigma_\epsilon - \Sigma_\epsilon^*$  is at most  $q$ . We have

$$\Sigma_\epsilon - \Sigma_\epsilon^* = \left[ (\sigma^2 - \sigma^{*2}) - (\sigma^2\rho - \sigma^{*2}\rho^*) \right] \mathbf{I} + (\sigma^2\rho - \sigma^{*2}\rho^*)\mathbf{J}.$$

By Lemma A.1.1, the eigenvalues of  $\Sigma_\epsilon - \Sigma_\epsilon^*$  are  $(\sigma^2 - \sigma^{*2}) - (\sigma^2\rho - \sigma^{*2}\rho^*)$ , which is of multiplicity  $n - 1$  and  $(\sigma^2 - \sigma^{*2}) + (n - 1)(\sigma^2\rho - \sigma^{*2}\rho^*)$ , of multiplicity 1. Since  $\Sigma_\epsilon - \Sigma_\epsilon^*$  is not the zero matrix, all of the eigenvalues cannot be equal to 0: we must either have no eigenvalues equal to 0, one eigenvalue equal to 0, or  $n - 1$  eigenvalues equal to 0. In order to have  $\text{rank}(\Sigma_\epsilon - \Sigma_\epsilon^*) \leq q$ , the  $n - 1$  multiple eigenvalues have to be zero since  $1 \leq q < n - 1$  by assumption. That is,  $\sigma^2 - \sigma^{*2} = \sigma^2\rho - \sigma^{*2}\rho^*$  and so  $\Sigma_\epsilon - \Sigma_\epsilon^* = (\sigma^2 - \sigma^{*2})\mathbf{J}$ . But plugging this into (2.3) yields  $\mathbf{H}_Z\mathbf{J} = \mathbf{J}$ , contradicting our assumption. □

## A.2 Proof of Corollary 2.4.2

**Proof:**

We first note a fact about the matrix  $\mathbf{H}_Z$ . Since  $\mathbf{H}_Z$  is symmetric and idempotent,

$$\mathbf{H}_Z[k, k] = \sum_l (\mathbf{H}_Z[k, l])^2 = (\mathbf{H}_Z[k, k])^2 + \sum_{l \neq k} (\mathbf{H}_Z[k, l])^2.$$

Thus, if  $\mathbf{H}_Z[k, k] = 1$ , then  $\mathbf{H}_Z[k, i] = \mathbf{H}_Z[i, k] = 0$  for all  $i \neq k$ .

To prove the corollary, we use Theorem 2.4.1 and a proof by contradiction.

First suppose that the model is identifiable and suppose, by way of contradiction, a diagonal element of  $\mathbf{H}_Z$  is equal to 1. Without loss of generality, we assume  $\mathbf{H}_Z[1, 1] = 1$ . Then by our observation,  $\mathbf{H}_Z[1, i] = \mathbf{H}_Z[i, 1] = 0$  for all  $i \neq 1$ . Fix  $\Sigma_\epsilon = \text{diag}\{\sigma_1^2, \dots, \sigma_n^2\} \in \tilde{\Theta}_\epsilon$ . Let  $\sigma_1^{*2}$  satisfy  $0 < \sigma_1^{*2} < \sigma_1^2$  and define  $\Sigma_\epsilon^* = \text{diag}\{\sigma_1^{*2}, \sigma_2^2, \dots, \sigma_n^2\}$ . Then  $\Sigma_\epsilon - \Sigma_\epsilon^* = \text{diag}\{\sigma_1^2 - \sigma_1^{*2}, 0, \dots, 0\}$ . It is not hard to check that (2.3) is satisfied. Clearly, for any  $\Sigma_u \in \Theta_u$ ,  $\Sigma_u^*$  defined as in (2.4) is also in  $\Theta_u$ . Thus, the model is not identifiable, which contradicts our assumption.

Now suppose that no diagonal element of  $\mathbf{H}_Z$  is equal to one and suppose, by contradiction, that the model is not identifiable. Then there exists nonidentical diagonal matrices,  $\Sigma_\epsilon$  and  $\Sigma_\epsilon^*$ , satisfying (2.3). As  $\Sigma_\epsilon \neq \Sigma_\epsilon^*$ , at least one of the diagonal elements of  $\Sigma_\epsilon - \Sigma_\epsilon^*$  is not zero. Suppose the  $k$ -th diagonal element is not zero. By (2.3), the  $k$ -th diagonal element of  $\mathbf{H}_Z$  must be one, contradicting our assumption.  $\square$



# Appendix B

## Appendix to Chapter 3

In this appendix, we provide the calculations in Sections 3.3.2 and 3.3.4 where we find the updates of  $\Sigma_x$  and  $\{\beta, \sigma^2\}$  in the ECME procedure.

In Section B.4, we derive the first order condition (3.10) and the Hessian matrix (3.14) of Section 3.3.2 where we maximize the log-likelihood  $\Lambda_N$  over  $\Sigma_x$  while holding the other parameters fixed. In Section B.5, we derive the first order conditions (3.22) and (3.23), and the Hessian matrix (3.26) of Section 3.3.4 where we maximize  $\tilde{\Lambda}_N$  over  $\{\beta, \sigma^2\}$  holding the other parameters fixed.

We use the tool of matrix differential calculus, calculating first differentials to obtain the first order conditions and second differentials to obtain the Hessian matrices. The book by Magnus and Neudecker (1988) gives an elegant description on this subject. In Sections B.1-B.3, we follow the book to introduce some definitions and provide some background, mainly from *Part Two* of the book. We keep the same notation as in the book. Throughout this section, chapters and page numbers all refer to (Magnus and Neudecker, 1988).

### B.1 Definition of the first differential

We first give the definition of the first differential for a vector function (a vector valued function with a vector argument). We show that the function's first differential is connected with its Jacobian matrix. We then give an extension of the definition to a matrix function (a matrix valued function with a matrix argument) and show how to identify the Jacobian matrix from the first differential.

**Definition B.1.1** Let  $\mathbf{f} : \mathbf{S} \rightarrow \mathbb{R}^m$  be a function defined on a set  $\mathbf{S}$  in  $\mathbb{R}^n$ . Let  $\mathbf{c}$  be an interior point of  $\mathbf{S}$ , and let  $\mathbf{B}(\mathbf{c}; r)$  be an  $n$ -ball lying in  $\mathbf{S}$  centred at  $\mathbf{c}$  of radius  $r$ . If there exists a real  $m \times n$  matrix  $\mathbf{A}$ , depending on  $\mathbf{c}$  but not on  $\mathbf{u}$ , such that

$$\mathbf{f}(\mathbf{c} + \mathbf{u}) = \mathbf{f}(\mathbf{c}) + \mathbf{A}(\mathbf{c})\mathbf{u} + \mathbf{r}_{\mathbf{c}}(\mathbf{u})$$

for all  $\mathbf{u} \in \mathbb{R}^n$  with  $\|\mathbf{u}\| < r$  and

$$\lim_{\mathbf{u} \rightarrow 0} \frac{\mathbf{r}_{\mathbf{c}}(\mathbf{u})}{\|\mathbf{u}\|} = 0,$$

then the function  $\mathbf{f}$  is said to be differentiable at  $\mathbf{c}$ ; the  $m \times n$  matrix  $\mathbf{A}(\mathbf{c})$  is then called the first derivative of  $\mathbf{f}$  at  $\mathbf{c}$ , and the  $m \times 1$  vector

$$d\mathbf{f}(\mathbf{c}; \mathbf{u}) = \mathbf{A}(\mathbf{c})\mathbf{u},$$

which is a linear function of  $\mathbf{u}$ , is called the first differential of  $\mathbf{f}$  at  $\mathbf{c}$  (with increment  $\mathbf{u}$ ). If  $\mathbf{f}$  is differentiable at every point of an open subset  $\mathbf{E}$  of  $\mathbf{S}$ , we say  $\mathbf{f}$  is differentiable on  $\mathbf{E}$ .

After calculating the first differential, we identify the *Jacobian matrix* as follows. Let  $D\mathbf{f}$  be the  $m \times n$  *Jacobian matrix* of  $\mathbf{f}$  whose  $ij$ th element is  $D_j f_i$ : the partial derivative of the  $i$ th component function  $f_i$  of  $\mathbf{f}$  with respect to the  $j$ th coordinate. The *First Identification Theorem* (p.87) states that the first derivative  $\mathbf{A}(\mathbf{c})$  is  $D\mathbf{f}(\mathbf{c})$  when  $\mathbf{f}$  is differentiable at  $\mathbf{c}$ .

To extend the definition of a vector function to a matrix function with a matrix argument is straightforward using the *vec* operator. The *vec* operator transforms a matrix into a vector by stacking the columns of the matrix one underneath the other.

Recall the norm of a real matrix  $\mathbf{X}$  is defined by

$$\|\mathbf{X}\| = (\text{tr} \mathbf{X}'\mathbf{X})^{1/2}.$$

Let  $\mathfrak{R}^{n \times q}$  contains all the real  $n \times q$  matrices. Define a ball  $\mathbf{B}(\mathbf{C}; r)$  with center  $\mathbf{C}$  and radius  $r$  in  $\mathfrak{R}^{n \times q}$  by

$$\mathbf{B}(\mathbf{C}; r) = \{\mathbf{X} : \mathbf{X} \in \mathfrak{R}^{n \times q}, \|\mathbf{X} - \mathbf{C}\| < r\}.$$

Let  $\mathbf{F} : \mathbf{S} \rightarrow \mathfrak{R}^{m \times p}$  be a matrix function defined on a set  $\mathbf{S}$  in  $\mathfrak{R}^{n \times q}$ . That is,  $F$  maps an  $n \times q$  matrix into an  $m \times p$  matrix  $\mathbf{F}(\mathbf{X})$ . We consider the the vector function  $\mathbf{f} : \text{vec } \mathbf{S} \rightarrow \mathfrak{R}^{m \times p}$  defined by

$$\mathbf{f}(\text{vec } \mathbf{X}) = \text{vec } \mathbf{F}(\mathbf{X})$$

and the following gives the definition of the first differential of  $\mathbf{F}$ .

**Definition B.1.2** *Let  $\mathbf{F} : \mathbf{S} \rightarrow \mathfrak{R}^{m \times p}$  be a matrix function defined on a set  $\mathbf{S}$  in  $\mathfrak{R}^{n \times q}$ . Let  $\mathbf{C}$  be an interior point of  $\mathbf{S}$  and let  $\mathbf{B}(\mathbf{C}; r) \subset \mathbf{S}$  be a ball with center  $\mathbf{C}$  and radius  $r$ . If there exists  $\mathbf{R}_{\mathbf{C}}$  and a real  $(mp) \times (nq)$  matrix  $\mathbf{A}$ , depending on  $\mathbf{C}$  but not on  $\mathbf{U}$ , such that*

$$\text{vec} \mathbf{F}(\mathbf{C} + \mathbf{U}) = \text{vec} \mathbf{F}(\mathbf{C}) + \mathbf{A}(\mathbf{C}) \text{vec} \mathbf{U} + \text{vec} \mathbf{R}_{\mathbf{C}}(\mathbf{U})$$

for all  $\mathbf{U} \in \mathfrak{R}^{n \times q}$  with  $\|\mathbf{U}\| < r$  and

$$\lim_{\mathbf{U} \rightarrow 0} \frac{\mathbf{R}_{\mathbf{C}}(\mathbf{U})}{\|\mathbf{U}\|} = 0,$$

then the function  $\mathbf{F}$  is said to be differentiable at  $\mathbf{C}$ . Let

$$d\mathbf{F}(\mathbf{C}; \mathbf{U}) = \mathbf{A}(\mathbf{C}) \text{vec} \mathbf{U}.$$

Although this is a vector of length  $(mp)$ , it can be formed into a matrix of dimension  $m \times p$ , in the usual natural way. This  $m \times p$  matrix  $d\mathbf{F}(\mathbf{C}; \mathbf{U})$  is called the first differential of  $\mathbf{F}$  at  $\mathbf{C}$  with increment  $\mathbf{U}$  and the  $(mp) \times (nq)$  matrix  $\mathbf{A}(\mathbf{C})$  is called the first derivative of  $F$  at  $\mathbf{C}$ .

From the definition, it is clear that the differential of  $\mathbf{F}$  and  $\mathbf{f}$  are related by

$$\text{vec } d\mathbf{F}(\mathbf{C}; \mathbf{U}) = d\mathbf{f}(\text{vec} \mathbf{C}; \text{vec} \mathbf{U}).$$

The *Jacobian matrix* of  $\mathbf{F}$  at  $\mathbf{C}$  is defined as

$$D\mathbf{F}(\mathbf{C}) = D\mathbf{f}(\text{vec}\mathbf{C}).$$

This is an  $(mp) \times (nq)$  matrix, whose  $ij$ th element is the partial derivative of the  $i$ th component of  $\text{vec}\mathbf{F}(\mathbf{X})$  with respect to the  $j$ th element of  $\text{vec}\mathbf{X}$ , evaluated at  $\mathbf{X} = \mathbf{C}$ . The *First Identification Theorem for matrix functions* (p.96) states that if  $\mathbf{F}$  is differentiable at  $\mathbf{C}$ , then

$$\text{vec } d\mathbf{F}(\mathbf{C}; \mathbf{U}) = D\mathbf{F}(\mathbf{C})\text{vec}\mathbf{U}.$$

So we can calculate the differential of  $\mathbf{F}$  to identify its *Jacobian matrix*.

## B.2 Definition of the second differential

We first introduce the definition of *twice differentiable* on which the definition *second differential* is based. The definitions are restricted to real valued functions as in our calculations we only need to consider second differentials of real valued functions. Then we connect the Hessian matrix with the second differential. As in the *first differential* case, we give an extension of the definition to a real valued function with a matrix argument and also show how to identify the Hessian matrix from the second differential.

**Definition B.2.1** *Let  $f : \mathbf{S} \rightarrow \mathfrak{R}$  be a real valued function defined on a set  $\mathbf{S}$  in  $\mathfrak{R}^n$ , and let  $\mathbf{c}$  be an interior point of  $\mathbf{S}$ . If  $f$  is differentiable in some  $n$ -ball  $\mathbf{B}(\mathbf{c})$  and each of the partial derivatives  $D_j f$  is differentiable at  $\mathbf{c}$ , then we say that  $f$  is twice differentiable at  $\mathbf{c}$ . If  $f$  is twice differentiable at every point of an open subset  $\mathbf{E}$  of  $\mathbf{S}$ , we say  $f$  is twice differentiable on  $\mathbf{E}$ .*

The following is the definition of the second differential.

**Definition B.2.2** *Let  $f : \mathbf{S} \rightarrow \mathfrak{R}$  be twice differentiable at an interior point  $\mathbf{c}$  of  $\mathbf{S} \subset \mathfrak{R}^n$ . Let  $\mathbf{B}(\mathbf{c})$  be an  $n$ -ball lying in  $\mathbf{S}$  such that  $f$  is differentiable at every point in  $\mathbf{B}(\mathbf{c})$ , and let  $g : \mathbf{B}(\mathbf{c}) \rightarrow \mathfrak{R}$  be defined by the equation*

$$g(\mathbf{x}) = df(\mathbf{x}; \mathbf{u}).$$

Then the differential of  $g$  at  $\mathbf{c}$  with increment  $\mathbf{u}$ , i.e.  $dg(\mathbf{c}; \mathbf{u})$ , is called the second differential of  $f$  at  $\mathbf{c}$  (with increment  $\mathbf{u}$ ), and is denoted by  $d^2f(\mathbf{c}; \mathbf{u})$ .

To calculate the second differential of  $f$ , by the definition, we just need to calculate the differential of the first differential of  $f$ , i.e.

$$d^2f = d(df).$$

We have seen that the *Jacobian matrix* can be identified from the first differential. Similarly, we can identify the Hessian matrix from the second differential. The *Second Identification Theorem* (p.107) states that if  $f$  is twice differentiable at  $\mathbf{c}$ , then

$$d^2f(\mathbf{c}; \mathbf{u}) = \mathbf{u}'(\mathbf{H}f(\mathbf{c}))\mathbf{u},$$

where  $\mathbf{H}f(\mathbf{c})$  is the  $n \times n$  symmetric Hessian matrix of  $f$  at  $\mathbf{c}$  with  $(i, j)$  entry equal to  $\partial^2 f(\mathbf{c})/(\partial \mathbf{c}_i \partial \mathbf{c}_j)$ . Therefore, once we have calculated the second differential, the Hessian matrix is obtainable.

Similarly as in Section B.1, the extension of the second differential from vector functions to matrix functions is straightforward using the *vec* operator. As we only consider real valued functions, we restrict the extension to a real valued function with a matrix argument.

We follow the notation in the definition of the first differential of matrix functions. Let the domain of  $f$  be  $\mathbf{S} \subseteq \mathfrak{R}^{n \times q}$  and let  $\mathbf{C}$  be an interior point of  $\mathbf{S}$ . Let  $\mathbf{B}(\mathbf{C}; r) \subset \mathbf{S}$  be a ball with center  $\mathbf{C}$  and radius  $r$  and let  $\mathbf{U}$  be a point in  $\mathfrak{R}^{n \times q}$  with  $\|\mathbf{U}\| < r$ , so that  $\mathbf{C} + \mathbf{U} \in \mathbf{B}(\mathbf{C}; r)$ . The second differential of  $f$  at  $\mathbf{C}$  is then defined as

$$d^2f(\mathbf{C}; \mathbf{U}) = d^2f(\text{vec}\mathbf{C}; \text{vec}\mathbf{U}).$$

The *Second Identification Theorem for matrix functions* (p.115) says if  $f$  is twice differentiable at  $\mathbf{C}$ , then

$$d^2f(\mathbf{C}; \mathbf{U}) = (\text{vec}\mathbf{U})' \mathbf{H}f(\mathbf{C}) \text{vec}\mathbf{U},$$

where  $\mathbf{H}f(\mathbf{C})$  is the  $nq \times nq$  symmetric Hessian matrix of  $f$  at  $\mathbf{C}$  defined as

$$\mathbf{H}f(\mathbf{C}) \equiv \mathbf{H}f(\text{vec}\mathbf{C}).$$

That is, the  $ij$ th element of  $\mathbf{H}f(\mathbf{C})$  is the second order derivative of  $f$  with respect to the  $i$ th and  $j$ th element of  $\text{vec}\mathbf{X}$  where  $\mathbf{X} \in \mathbf{S}$ , evaluated at  $\mathbf{C}$ .

### B.3 Matrix algebraic and differential rules

In this section, we list the matrix algebraic and differential rules (chap.8) which will be used without specific reference in our derivations. In the following, we let  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{C}$  and  $\mathbf{D}$  denote constant matrices,  $\mathbf{u}$  denote a vector function and  $\mathbf{U}$  and  $\mathbf{V}$  denote matrix functions. We let  $\otimes$  stand for the Kronecker product. The rules are the following.

- $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$ , provided  $\mathbf{AB}$  is square.
- $\text{tr}(\mathbf{A}'\mathbf{B}) = (\text{vec}\mathbf{A})'\text{vec}\mathbf{B}$ .
- $\text{tr}(\mathbf{ABCD}) = (\text{vec}\mathbf{D})'(\mathbf{A} \otimes \mathbf{C}')(\text{vec}\mathbf{B}')$ , provided  $\mathbf{ABCD}$  is defined and square.
- $d\mathbf{A} = \mathbf{0}$ .
- $d\mathbf{AU} = \mathbf{A}d\mathbf{U}$ .
- $d(\mathbf{U} + \mathbf{V}) = d\mathbf{U} + d\mathbf{V}$ .
- $d(\mathbf{UV}) = (d\mathbf{U})\mathbf{V} + \mathbf{U}(d\mathbf{V})$ .
- $d(\mathbf{U}') = (d\mathbf{U})'$ .
- $d(\ln \det \mathbf{U}) = \text{tr}\mathbf{U}^{-1}(d\mathbf{U})$ .
- $d(\mathbf{U}^{-1}) = -\mathbf{U}^{-1}(d\mathbf{U})\mathbf{U}^{-1}$ .
- $d(\text{tr}\mathbf{U}) = \text{tr}(d\mathbf{U})$ .
- $d(\text{vec}\mathbf{U}) = \text{vec}(d\mathbf{U})$ .
- $d(\mathbf{u}'\mathbf{A}\mathbf{u}) = \mathbf{u}'(\mathbf{A} + \mathbf{A}')d\mathbf{u} = (d\mathbf{u})'(\mathbf{A} + \mathbf{A}')\mathbf{u}$ .

## B.4 Calculations in Section 3.3.2

Recall the observed data log-likelihood has the expression

$$\Lambda_N = -\frac{N}{2} \ln \det (\mathbf{C}\boldsymbol{\Sigma}_x\mathbf{C}' + \boldsymbol{\Sigma}_d) - \frac{1}{2} \sum_{i=1}^N (\mathbf{W}_i - \boldsymbol{\mu}_W)' (\mathbf{C}\boldsymbol{\Sigma}_x\mathbf{C}' + \boldsymbol{\Sigma}_d)^{-1} (\mathbf{W}_i - \boldsymbol{\mu}_W).$$

We want to maximize it over  $\boldsymbol{\Sigma}_x$  while holding the other parameters fixed. In this section, holding the other parameters fixed, we calculate the first differential of  $\Lambda_N$  to obtain the first order condition (3.10) and calculate the second differential to obtain the Hessian matrix in (3.14).

As we treat  $\boldsymbol{\Sigma}_x$  as the only unknown parameter, it immediately follows from the expression of  $\boldsymbol{\Sigma}_W$  in (3.9)

$$d\boldsymbol{\Sigma}_W = \mathbf{C}d\boldsymbol{\Sigma}_x\mathbf{C}'. \quad (\text{B.1})$$

In our derivation, we will use the shorter notation  $d\boldsymbol{\Sigma}_W$  before we reach (3.10) or (3.14). We have

$$\begin{aligned} d\Lambda_N &= -\frac{N}{2} d(\ln \boldsymbol{\Sigma}_W) - \frac{1}{2} \sum_{i=1}^N (\mathbf{W}_i - \boldsymbol{\mu}_W)' (d \boldsymbol{\Sigma}_W^{-1}) (\mathbf{W}_i - \boldsymbol{\mu}_W) \\ &= -\frac{N}{2} \text{tr} [\boldsymbol{\Sigma}_W^{-1} d\boldsymbol{\Sigma}_W] + \frac{1}{2} \sum_{i=1}^N (\mathbf{W}_i - \boldsymbol{\mu}_W)' \boldsymbol{\Sigma}_W^{-1} (d \boldsymbol{\Sigma}_W) \boldsymbol{\Sigma}_W^{-1} (\mathbf{W}_i - \boldsymbol{\mu}_W) \\ &= -\frac{N}{2} \text{tr} [\boldsymbol{\Sigma}_W^{-1} d\boldsymbol{\Sigma}_W] + \frac{1}{2} \text{tr} \left[ \boldsymbol{\Sigma}_W^{-1} \sum_{i=1}^N (\mathbf{W}_i - \boldsymbol{\mu}_W)(\mathbf{W}_i - \boldsymbol{\mu}_W)' \boldsymbol{\Sigma}_W^{-1} (d \boldsymbol{\Sigma}_W) \right] \\ &= -\frac{N}{2} \text{tr} [\boldsymbol{\Sigma}_W^{-1} (\boldsymbol{\Sigma}_W - \mathbf{S}_W) \boldsymbol{\Sigma}_W^{-1} (d\boldsymbol{\Sigma}_W)] \end{aligned} \quad (\text{B.2})$$

Recalling (B.1), we now have

$$d\Lambda_N = -\frac{N}{2} \text{tr} [\mathbf{C}' \boldsymbol{\Sigma}_W^{-1} (\boldsymbol{\Sigma}_W - \mathbf{S}_W) \boldsymbol{\Sigma}_W^{-1} \mathbf{C} d\boldsymbol{\Sigma}_x].$$

By the *First Identification Theorem for matrix functions* mentioned in Sec-

tion B.1, we obtain the *Jacobian matrix* of  $\Lambda_N$  at  $\Sigma_x$  as

$$D \Lambda_N(\Sigma_x) = \text{vec} \left\{ \mathbf{C}' \Sigma_W^{-1} (\Sigma_W - \mathbf{S}_W) \Sigma_W^{-1} \mathbf{C} \right\}.$$

Equating it to zero yields

$$\mathbf{C}' \Sigma_W^{-1} (\Sigma_W - \mathbf{S}_W) \Sigma_W^{-1} \mathbf{C} = \mathbf{0}$$

which is equivalent to the first order condition (3.10).

Next we calculate  $d^2 \Lambda_N$  to identify the Hessian matrix in (3.14). Starting from (B.2), we have

$$\begin{aligned} d^2 \Lambda_N &= -\frac{N}{2} \text{tr} \left[ (d\Sigma_W^{-1})(\Sigma_W - \mathbf{S}_W) \Sigma_W^{-1} (d\Sigma_W) \right] - \frac{N}{2} \text{tr} \left[ \Sigma_W^{-1} d(\Sigma_W - \mathbf{S}_W) \Sigma_W^{-1} (d\Sigma_W) \right] \\ &\quad - \frac{N}{2} \text{tr} \left[ \Sigma_W^{-1} (\Sigma_W - \mathbf{S}_W) d\Sigma_W^{-1} (d\Sigma_W) \right] \\ &= \frac{N}{2} \text{tr} \left[ \Sigma_W^{-1} (d\Sigma_W) \Sigma_W^{-1} (\Sigma_W - \mathbf{S}_W) \Sigma_W^{-1} (d\Sigma_W) \right] - \frac{N}{2} \text{tr} \left[ \Sigma_W^{-1} (d\Sigma_W) \Sigma_W^{-1} (d\Sigma_W) \right] \\ &\quad + \frac{N}{2} \text{tr} \left[ \Sigma_W^{-1} (\Sigma_W - \mathbf{S}_W) \Sigma_W^{-1} (d\Sigma_W) \Sigma_W^{-1} (d\Sigma_W) \right]. \end{aligned}$$

The first term and the last term at the right hand side are equal, and so they can be combined into one term

$$N \text{tr} \left[ \Sigma_W^{-1} (d\Sigma_W) \Sigma_W^{-1} (\Sigma_W - \mathbf{S}_W) \Sigma_W^{-1} (d\Sigma_W) \right].$$

Then

$$d^2 \Lambda_N = N \text{tr} \left[ \Sigma_W^{-1} (d\Sigma_W) \Sigma_W^{-1} \left( \frac{1}{2} \Sigma_W - \mathbf{S}_W \right) \Sigma_W^{-1} (d\Sigma_W) \right]$$

Recall (B.1). Right hand side of the above

$$\begin{aligned} &= N \text{tr} \left[ \Sigma_W^{-1} \mathbf{C} d\Sigma_x \mathbf{C}' \Sigma_W^{-1} \left( \frac{1}{2} \Sigma_W - \mathbf{S}_W \right) \Sigma_W^{-1} \mathbf{C} d\Sigma_x \mathbf{C}' \right] \\ &= N \text{tr} \left[ \mathbf{C}' \Sigma_W^{-1} \mathbf{C} d\Sigma_x \mathbf{C}' \Sigma_W^{-1} \left( \frac{1}{2} \Sigma_W - \mathbf{S}_W \right) \Sigma_W^{-1} \mathbf{C} d\Sigma_x \right] \\ &= N (\text{vec } d\Sigma_x)' \left[ \mathbf{C}' \Sigma_W^{-1} \mathbf{C} \otimes \mathbf{C}' \Sigma_W^{-1} \left( \frac{1}{2} \Sigma_W - \mathbf{S}_W \right) \Sigma_W^{-1} \mathbf{C} \right] \text{vec}(d\Sigma_x). \end{aligned}$$



When evaluated at the critical point  $\hat{\Sigma}_x$  which satisfies the first order condition (3.10),  $d^2\Lambda_N$  is then

$$-\frac{N}{2}(\text{vec } d\Sigma_x)' \left[ \mathbf{C}' \hat{\Sigma}_W^{-1} \mathbf{C} \otimes \mathbf{C}' \hat{\Sigma}_W^{-1} \mathbf{C} \right] \text{vec}(d\Sigma_x).$$

By the *Second Identification Theorem for matrix functions* mentioned in Section B.2, we have at  $\hat{\Sigma}_x$  the Hessian matrix is equal to

$$\mathbf{H}(\hat{\Sigma}_x) = -(N/2) \left( \hat{\mathbf{D}} \otimes \hat{\mathbf{D}} \right), \quad \text{where } \hat{\mathbf{D}} = \mathbf{C}' \hat{\Sigma}_W^{-1} \mathbf{C}.$$

This is the matrix we saw in (3.14).

## B.5 Calculations in Section 3.3.4

Recall we want to maximize the log-likelihood

$$\tilde{\Lambda}_N = -\frac{N}{2} \ln(\beta' \mathbf{K} \beta + \sigma^2) - \frac{1}{2(\beta' \mathbf{K} \beta + \sigma^2)} \sum_{i=1}^N (Y_i - \beta_0 - \beta' \mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu}))^2$$

over  $\{\beta, \sigma^2\}$  to find the update while fixing the other parameters. In this section, we derive the first order conditions (3.22) and (3.23) via calculating the first differential of  $\tilde{\Lambda}_N$  with respect to  $\{\beta, \sigma^2\}$ . Calculating the second differential of  $\tilde{\Lambda}_N$  then gives us the Hessian matrix (3.26).

The following two differentials will facilitate our calculation in  $d\tilde{\Lambda}_N$  and  $d^2\tilde{\Lambda}_N$ . Recall the expression of  $\sigma_{Y|z}^2$  in (3.19). We have

$$d\sigma_{Y|z}^2 \equiv d(\beta' \mathbf{K} \beta + \sigma^2) = 2\beta' \mathbf{K} d\beta + d\sigma^2. \quad (\text{B.3})$$

Let

$$\mathbf{g}(\beta) = \sum_{i=1}^N (Y_i - \beta_0 - \beta' \mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu})) (\mathbf{z}_i - \boldsymbol{\mu})' \mathbf{G}'. \quad (\text{B.4})$$

We then obtain

$$d \left[ \sum_{i=1}^N (Y_i - \beta_0 - \beta' \mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu}))^2 \right] = -2 \sum_{i=1}^N (Y_i - \beta_0 - \beta' \mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu})) (d\beta)' \mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu})$$

$$\begin{aligned}
 &= -2 \sum_{i=1}^N (Y_i - \beta_0 - \boldsymbol{\beta}' \mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu})) (\mathbf{z}_i - \boldsymbol{\mu})' \mathbf{G}'(d\boldsymbol{\beta}) \\
 &= -2 \mathbf{g}(\boldsymbol{\beta}) d\boldsymbol{\beta}
 \end{aligned} \tag{B.5}$$

To calculate  $d\tilde{\Lambda}_N$ , we use the terms  $\sigma_{Y|z}^2$  and  $\boldsymbol{\beta}' \mathbf{K} \boldsymbol{\beta} + \sigma^2$  interchangeably.

$$\begin{aligned}
 d\tilde{\Lambda}_N &= -\frac{N}{2} d \left[ \ln(\boldsymbol{\beta}' \mathbf{K} \boldsymbol{\beta} + \sigma^2) \right] - d \left[ \frac{1}{2(\boldsymbol{\beta}' \mathbf{K} \boldsymbol{\beta} + \sigma^2)} \right] \sum_{i=1}^N (Y_i - \beta_0 - \boldsymbol{\beta}' \mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu}))^2 \\
 &\quad - \frac{1}{2(\boldsymbol{\beta}' \mathbf{K} \boldsymbol{\beta} + \sigma^2)} d \left[ \sum_{i=1}^N (Y_i - \beta_0 - \boldsymbol{\beta}' \mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu}))^2 \right]
 \end{aligned}$$

By (B.3) and (B.5), the right hand side above is equal to

$$-\frac{N}{2} \frac{2\boldsymbol{\beta}' \mathbf{K} d\boldsymbol{\beta} + d\sigma^2}{\sigma_{Y|z}^2} + \frac{2\boldsymbol{\beta}' \mathbf{K} d\boldsymbol{\beta} + d\sigma^2}{2\sigma_{Y|z}^4} \sum_{i=1}^N (Y_i - \beta_0 - \boldsymbol{\beta}' \mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu}))^2 + \frac{1}{\sigma_{Y|z}^2} \mathbf{g}(\boldsymbol{\beta}) d\boldsymbol{\beta}.$$

Let

$$\mathbf{c}(\boldsymbol{\beta}, \sigma^2) = \left[ \sum_{i=1}^N (Y_i - \beta_0 - \boldsymbol{\beta}' \mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu}))^2 - N\sigma_{Y|z}^2 \right]. \tag{B.6}$$

Then  $d\tilde{\Lambda}_N$  is equal to

$$\frac{d\sigma^2}{2\sigma_{Y|z}^4} \mathbf{c}(\boldsymbol{\beta}, \sigma^2) + \frac{1}{\sigma_{Y|z}^4} \left[ \boldsymbol{\beta}' \mathbf{K} \mathbf{c}(\boldsymbol{\beta}, \sigma^2) + \sigma_{Y|z}^2 \mathbf{g}(\boldsymbol{\beta}) \right] d\boldsymbol{\beta}. \tag{B.7}$$

By the *First Identification Theorem* mentioned in Section B.1, we obtain the first order conditions

$$\begin{aligned}
 \frac{1}{\sigma_{Y|z}^4} \left[ \boldsymbol{\beta}' \mathbf{K} \mathbf{c}(\boldsymbol{\beta}, \sigma^2) + \sigma_{Y|z}^2 \mathbf{g}(\boldsymbol{\beta}) \right] &= \mathbf{0}, \\
 \frac{1}{2\sigma_{Y|z}^4} \mathbf{c}(\boldsymbol{\beta}, \sigma^2) &= 0
 \end{aligned}$$

which lead to (3.22) and (3.23).

Calculating  $d^2\tilde{\Lambda}_N$  to identify the Hessian matrix is lengthy and tedious. In fact, we don't need the closed form of the Hessian matrix but the Hes-

sian matrix evaluated at the critical points  $\{\hat{\boldsymbol{\beta}}, \hat{\sigma}^2\}$  given in (3.26). So in our derivation, we will make use of the first order conditions to simplify calculation.

We notice, equivalently, the critical points  $\{\hat{\boldsymbol{\beta}}, \hat{\sigma}^2\}$  only need to satisfy

$$\mathbf{g}(\hat{\boldsymbol{\beta}}) = \mathbf{0} \quad (\text{B.8})$$

$$\mathbf{c}(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) = 0. \quad (\text{B.9})$$

From (B.6), using (B.3) and (B.5) we have

$$\begin{aligned} d\mathbf{c}(\boldsymbol{\beta}, \sigma^2) &= -2 \sum_{i=1}^N (Y_i - \beta_0 - \boldsymbol{\beta}'\mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu})) (\mathbf{z}_i - \boldsymbol{\mu})' \mathbf{G}' d\boldsymbol{\beta} - 2N\boldsymbol{\beta}'\mathbf{K}d\boldsymbol{\beta} - Nd\sigma^2 \\ &= -2\mathbf{g}(\boldsymbol{\beta})d\boldsymbol{\beta} - 2N\boldsymbol{\beta}'\mathbf{K}d\boldsymbol{\beta} - Nd\sigma^2, \end{aligned}$$

which is a function of  $\boldsymbol{\beta}$ . By (B.8),

$$d\mathbf{c}(\hat{\boldsymbol{\beta}}, \sigma^2) = -2N(d\boldsymbol{\beta})'\mathbf{K}\hat{\boldsymbol{\beta}} - Nd\sigma^2. \quad (\text{B.10})$$

Now we calculate  $d^2\tilde{\Lambda}_N$  starting from (B.7). We first calculate

$$d \left[ \frac{d\sigma^2}{2\sigma_{Y|z}^4} \mathbf{c}(\boldsymbol{\beta}, \sigma^2) \right]$$

which is

$$d \left( \frac{1}{2\sigma_{Y|z}^4} \right) d\sigma^2 \mathbf{c}(\boldsymbol{\beta}, \sigma^2) + \frac{d\sigma^2}{2\sigma_{Y|z}^4} d \mathbf{c}(\boldsymbol{\beta}, \sigma^2).$$

When at  $\{\hat{\boldsymbol{\beta}}, \hat{\sigma}^2\}$ , by (B.9) and (B.10), it is equal to

$$-\frac{Nd\sigma^2}{2\hat{\sigma}_{Y|z}^4} \left( 2(d\boldsymbol{\beta})'\mathbf{K}\hat{\boldsymbol{\beta}} + d\sigma^2 \right) = -\frac{N}{\hat{\sigma}_{Y|z}^4} \left( d\boldsymbol{\beta}' d\sigma^2 \right) \begin{pmatrix} \mathbf{K}\hat{\boldsymbol{\beta}} \\ 1/2 \end{pmatrix} d\sigma^2. \quad (\text{B.11})$$

Then we calculate

$$d \left[ \frac{1}{\sigma_{Y|z}^4} \left[ \boldsymbol{\beta}'\mathbf{K} \mathbf{c}(\boldsymbol{\beta}, \sigma^2) + \sigma_{Y|z}^2 \mathbf{g}(\boldsymbol{\beta}) \right] d\boldsymbol{\beta} \right]$$

which is

$$d \left( \frac{1}{\sigma_{Y|z}^4} \right) \left[ \boldsymbol{\beta}' \mathbf{K} \mathbf{c}(\boldsymbol{\beta}, \sigma^2) + \sigma_{Y|z}^2 \mathbf{g}(\boldsymbol{\beta}) \right] d\boldsymbol{\beta} + \frac{1}{\sigma_{Y|z}^4} d \left[ \boldsymbol{\beta}' \mathbf{K} \mathbf{c}(\boldsymbol{\beta}, \sigma^2) + \sigma_{Y|z}^2 \mathbf{g}(\boldsymbol{\beta}) \right] d\boldsymbol{\beta}. \quad (\text{B.12})$$

At  $\{\hat{\boldsymbol{\beta}}, \hat{\sigma}^2\}$ , the term in the first square brackets vanishes by (B.8) and (B.9).

Thus, at  $\{\hat{\boldsymbol{\beta}}, \hat{\sigma}^2\}$ , (B.12) is equal to

$$\frac{1}{\hat{\sigma}_{Y|z}^4} \left[ (d\boldsymbol{\beta})' \mathbf{K} \mathbf{c}(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) + \hat{\boldsymbol{\beta}}' \mathbf{K} d\mathbf{c}(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) + d\sigma_{Y|z}^2 \mathbf{g}(\hat{\boldsymbol{\beta}}) + \hat{\sigma}_{Y|z}^2 d\mathbf{g}(\hat{\boldsymbol{\beta}}) \right] d\boldsymbol{\beta}.$$

From (B.4), we have

$$d\mathbf{g}(\boldsymbol{\beta}) = -(d\boldsymbol{\beta})' \sum_{i=1}^N \mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu})(\mathbf{z}_i - \boldsymbol{\mu})' \mathbf{G}'. \quad (\text{B.13})$$

Again by (B.8)-(B.10) and that  $d\mathbf{c}(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2)$  is a scalar, (B.12) is equal to

$$\begin{aligned} & \frac{1}{\hat{\sigma}_{Y|z}^4} \left[ \left( -2N(d\boldsymbol{\beta})' \mathbf{K} \hat{\boldsymbol{\beta}} - Nd\sigma^2 \right) \hat{\boldsymbol{\beta}}' \mathbf{K} - \hat{\sigma}_{Y|z}^2 (d\boldsymbol{\beta})' \sum_{i=1}^N \mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu})(\mathbf{z}_i - \boldsymbol{\mu})' \mathbf{G}' \right] d\boldsymbol{\beta} \\ &= -\frac{N}{\hat{\sigma}_{Y|z}^4} (d\boldsymbol{\beta}' \ d\sigma^2) \begin{pmatrix} 2\mathbf{K} \hat{\boldsymbol{\beta}} \hat{\boldsymbol{\beta}}' \mathbf{K} + \frac{\hat{\sigma}_{Y|z}^2}{N} \sum_{i=1}^N \mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu})(\mathbf{z}_i - \boldsymbol{\mu})' \mathbf{G}' \\ \hat{\boldsymbol{\beta}}' \mathbf{K} \end{pmatrix} d\boldsymbol{\beta} \quad (\text{B.14}) \end{aligned}$$

Combining (B.11) and (B.14), eventually we get, at  $(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2)$ ,

$$d^2 \tilde{\Lambda}_N(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) = (d\boldsymbol{\beta}' \ d\sigma^2) \mathbf{H} \tilde{\Lambda}(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) \begin{pmatrix} d\boldsymbol{\beta} \\ d\sigma^2 \end{pmatrix},$$

where

$$\mathbf{H} \tilde{\Lambda}(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) = -\frac{N}{\hat{\sigma}_{Y|z}^4} \begin{pmatrix} 2\mathbf{K} \hat{\boldsymbol{\beta}} \hat{\boldsymbol{\beta}}' \mathbf{K} + \frac{\hat{\sigma}_{Y|z}^2}{N} \sum_{i=1}^N \mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu})(\mathbf{z}_i - \boldsymbol{\mu})' \mathbf{G}' & \mathbf{K} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\beta}}' \mathbf{K} & 1/2 \end{pmatrix}.$$

By the *Second Identification Theorem* mentioned in Section B.2,  $\mathbf{H} \tilde{\Lambda}(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2)$  is the Hessian matrix and we have seen it in (3.26).