The University of British Columbia

Department of Statistics

Technical Report # 235

Linear mixed models for measurement error in functional regression

Nancy Heckman and Wei Wang

November 5th, 2007

Linear mixed models for measurement error in functional regression

Nancy Heckman and Wei Wang

November 5, 2007

Abstract

Regression models with a scalar response and a functional predictor have been extensively studied. One approach is to approximate the functional predictor using eigenfunction expansion with the coefficient vector being random. The random coefficient vector is also known as random effects. In our study of this regression model, 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.

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

$$Y_i = \beta_0 + \int_a^b \beta(t) \left[Z_i(t) - \mathrm{E}(Z_i(t)) \right] \mathrm{d}t + \mathrm{e}_i. \tag{1}$$

The goal is to estimate β.

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

$$z_{ij} = Z_i(t_{ij}) + \epsilon_{ij}, \qquad \mathrm{Cov}(\epsilon_{i1}, \ldots, \epsilon_{in_i}) = \mathbf{\Sigma}_{\epsilon_i}. \tag{2}$$

Here, we consider data where the Z_{ij}'s are observed with error, and we model the function Z_i using random effects with a set of basis functions $\phi_1, \ldots, \phi_K$. In our estimation process, we approximate $Z_i(\cdot)$ as

$$Z_i(t) = \mu(t) + \sum_{k=1}^{K} x_{ik}\phi_k(t), \tag{3}$$

where μ is smooth and $\mathbf{x}_i \equiv (x_{i1}, \ldots, x_{iK})'$ are independent and normally distributed with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}_x$.

This approach was taken by James (2002), Müller (2005) and James and Silverman (2005) for data with and without measurement error. James used ϕ_k's equal to a basis for natural cubic splines and also used this basis for modelling μ. James's approach is similar to that described in Section 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 ϕ_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 μ, the ϕ_k's and the x_{ij}'s. The second part incorporates the Y_i's to estimate β_0 and β. 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 μ 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 $\phi_1, \ldots, \phi_K$ and $\hat{\lambda}_1, \ldots, \hat{\lambda}_K$. Let $\hat{x}_{ik}$ be the best linear unbiased estimate of individual i's kth PC score, $\hat{x}_{ik} = \mathrm{E}(x_{ik}|z_{i1}, \ldots, z_{in_i})$, calculated assuming normality, (2) with $\Sigma_{\epsilon_i} = \sigma^2\mathbf{I}$ and (3) with $\mathbf{\Sigma}_x = \mathrm{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_K)$. The second part of Müller's methodology involves regressing Y_i on $\hat{x}_{i1}, \ldots, \hat{x}_{iK}$ to estimate $\beta_k = \int \beta(t)\ \phi_k(t)\ \mathrm{d}t$ and then setting $\hat{\beta}(t) = \sum \hat{\beta}_k \phi_k(t)$. Müller justifies the use of this regression by showing that $\mathrm{E}(Y_i|z_{ij}, j = 1, \ldots, n_i) = \beta_0 + \sum_1^\infty \beta_k x_{ik}$, 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 β. 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 ϕ_k's. Since $\int \beta(t)[Z_i(t) - \mathrm{E}(Z_i(t))]\ \mathrm{d}t = \sum_1^K x_{ik} \int \beta(t)\ \phi_k(t)\ \mathrm{d}t$, we can only hope to estimate $\int \beta(t)\ \phi_k(t)\ \mathrm{d}t$, $k = 1, \ldots, K$. We cannot estimate $\int \beta(t)\ f(t)\ \mathrm{d}t$ for f orthogonal to the ϕ_k's. This issue was noted by James and Silverman (2005) who handled it by adding a penalty term which penalizes β's that have $\int \beta(t)f(t)\mathrm{d}t$ big when $\mathrm{Var}(\int Z(t)f(t)\mathrm{d}t)$ is small.

Müller's approach has the disadvantage that it does not fully use the Y_i's: the x_{ik}'s in (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 $\mathrm{E}(Y_i|z_{ij}, j = 1, \ldots, n_i) = \beta_0 + \sum_1^\infty \beta_k x_{ik}$ does not hold if the eigenvalues or eigenfunctions are incorrect. In particular, the calculation relies on $\mathrm{Cov}(x_{ij}, x_{ik}) = 0$ for $j \neq k$.

We consider a hybrid approach. Like Müller, we use $\phi_1, \ldots, \phi_K$ equal to the first K estimated eigenfunctions of the Z_i process. Thus we not only improve on James's choice of ϕ_k's but also focus on "estimable directions" of β. We then treat these ϕ_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 $\mathbf{\Sigma}_x$. Thus we improve on Müller's procedure, where the x_{ik}'s are estimated using only the z_{ij}'s, and $\mathbf{\Sigma}_x$ 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. In this work, we also propose test statistics for hypothesis testing of $\beta(\cdot)$. In Section 4.1, we consider testing the nullity of the function

β, i.e. testing $H_o: \ \beta(t) = 0, \ \text{for all } t \in [a, b]$. In the two sample situation, we consider testing the equality of a selection β^s and a control β^c, i.e. testing $H_o: \ \beta^s(t) = \beta^c(t), \ \text{for all } t \in [a, b]$. We propose a new integrated t-statistic and three test statistics that are more standard. In Section 5, we derive expressions of the residuals of the model fit and discuss model diagnostics based on the analysis of residuals. In Section 6.2, we apply the model to analyze a biological data set. In Section 7, via a simulation study, we compare our ECME estimate of β 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 Section 8.

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 ϕ_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 (1), (2) and (3) above hold. Let $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{in})'$, $\boldsymbol{\mu} = (\mu(t_1), \ldots, \mu(t_n))'$, $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{in})'$ represent errors and $\mathbf{z}_i = (z_{i1}, \ldots, z_{in})'$ be the observed values. Write

$$\mathbf{z}_i = \mathbf{Z}_i + \boldsymbol{\epsilon}_i \equiv \boldsymbol{\mu} + \mathbf{A}\mathbf{x}_i + \boldsymbol{\epsilon}_i, \tag{4}$$

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

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

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

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 β 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

$$Y_i = \beta_0 + \boldsymbol{\beta}'\mathbf{T}\mathbf{x}_i + \mathrm{e}_i, \quad \text{where} \quad \mathbf{T}_{jk} = \int_a^b \psi_j(t)\ \phi_k(t)\ \mathrm{d}t \quad \text{and} \quad \mathrm{e}_i \sim \mathrm{N}(0, \sigma^2). \tag{5}$$

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

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

and

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

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

$$\text{and} \quad \mathrm{Cov}\,(Y_i, \mathbf{x}_i') \equiv \boldsymbol{\Sigma}_{Y,x} = \boldsymbol{\beta}'\mathbf{T}\boldsymbol{\Sigma}_x. \tag{6}$$

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

$$\mathbf{C} = \begin{pmatrix} \mathbf{A} \\ \boldsymbol{\beta}'\mathbf{T} \end{pmatrix} \tag{7}$$

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

$$\boldsymbol{\Sigma}_d = \mathrm{diag}(\sigma_\epsilon^2\mathbf{R}, \sigma^2). \tag{8}$$

Then

$$\mathrm{Cov}(\mathbf{W}_i) \equiv \mathbf{\Sigma}_W = \mathbf{C}\mathbf{\Sigma}_x\mathbf{C}' + \mathbf{\Sigma}_d \tag{9}$$

and the log-likelihood of the observed data is

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

up to an additive constant term. Unknown model parameters are $\boldsymbol{\theta} = (\boldsymbol{\mu}, \beta_0, \mathbf{\Sigma}_x, \sigma_\epsilon^2, \boldsymbol{\beta}, \sigma^2)$. Directly maximizing Λ_N over $\boldsymbol{\theta}$ does not give us closed forms of the parameter estimates except for $\boldsymbol{\mu}$ and β_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 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 β_0 are $\bar{\mathbf{z}}$ and $\bar{Y}$ respectively. So throughout we take

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

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 $(\mathbf{z}_i, Y_i, \mathbf{x}_i),\ i = 1, \ldots, N$, as the complete data. We update $\sigma_\epsilon^{2(t)}$ 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 Λ_N directly.

Throughout our ECME procedure in Sections 3.2-3.4, 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 $i = 1, \ldots, n$, $\mathbf{z}_i = \mathbf{u} + \mathbf{v}'\mathbf{x}_i$;

(d) there exists no $\mathbf{v}$ such, for all $i = 1, \ldots, n$, $Y_i = \bar{Y} + \mathbf{v}'(\mathbf{z}_i - \bar{\mathbf{z}})$.

The restrictions on $\mathbf{A}$ and $\mathbf{T}$ are easily satisfied. Assumption (b) requires J, the number of the ψ_j basis functions to be no larger than K, the number

of the ϕ_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 (7) is also of full column rank.

3.1 Initial estimates of parameters other than $\boldsymbol{\mu}$ and β_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 σ^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 σ^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 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 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 $\lambda_1, \ldots, \lambda_n$ 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 $\lambda_1, \ldots, \lambda_K$. 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 $\boldsymbol{\Sigma}_x$, σ_ϵ^2, $\boldsymbol{\beta}$, and σ^2, as described

below.

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

We update $\boldsymbol{\Sigma}_x^{(t)}$ by maximizing Λ_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 8.4, differentiating Λ_N with respect to $\boldsymbol{\Sigma}_x$ and equating to zero yields the first order condition

$$\mathbf{C}'\boldsymbol{\Sigma}_W^{-1}\mathbf{C} = \mathbf{C}'\boldsymbol{\Sigma}_W^{-1}\ \mathbf{S}_W\ \boldsymbol{\Sigma}_W^{-1}\mathbf{C}. \tag{10}$$

Here, $\mathbf{C}$ depends on $\boldsymbol{\beta}^{(t)}$ and $\boldsymbol{\Sigma}_W = \mathbf{C}\boldsymbol{\Sigma}_x\mathbf{C}' + \boldsymbol{\Sigma}_d^{(t)}$ from (9). Equation (10) holds provided $\boldsymbol{\Sigma}_W$ is invertible at the critical value of $\boldsymbol{\Sigma}_x$. Since we assume that $\sigma^{2(t)}$ and $\sigma_\epsilon^{2(t)}$ are positive, $\boldsymbol{\Sigma}_d^{(t)}$ is positive definite. So $\boldsymbol{\Sigma}_W$ will be invertible provided $\boldsymbol{\Sigma}_x$ is non-negative definite.

We now solve (10) for $\boldsymbol{\Sigma}_x$, first deriving two useful identities, (11) and (12). For ease, we drop the hats and superscript t's on the parameter estimates that are being held fixed, that is, on $\hat{\boldsymbol{\mu}}_W, \sigma_\epsilon^{2(t)}, \boldsymbol{\beta}^{(t)}$, and $\sigma^{2(t)}$. Direct multiplication and some manipulation of the left hand side of the following shows that

$$\left(\mathbf{C}'\boldsymbol{\Sigma}_d^{-1}\mathbf{C}\right) \times \left[\left(\mathbf{C}'\boldsymbol{\Sigma}_d^{-1}\mathbf{C}\right)^{-1} + \boldsymbol{\Sigma}_x\right] \mathbf{C}'\boldsymbol{\Sigma}_W^{-1} = \mathbf{C}'\boldsymbol{\Sigma}_d^{-1}.$$

Solving this for $\mathbf{C}\boldsymbol{\Sigma}_W^{-1}$ yields

$$\mathbf{C}'\boldsymbol{\Sigma}_W^{-1} = \left[\left(\mathbf{C}'\boldsymbol{\Sigma}_d^{-1}\mathbf{C}\right)^{-1} + \boldsymbol{\Sigma}_x\right]^{-1} \left(\mathbf{C}'\boldsymbol{\Sigma}_d^{-1}\mathbf{C}\right)^{-1} \mathbf{C}'\boldsymbol{\Sigma}_d^{-1}. \tag{11}$$

Postmultiplying both sides of identity (11) by $\mathbf{C}$ yields

$$\mathbf{C}'\boldsymbol{\Sigma}_W^{-1}\mathbf{C} = \left(\left(\mathbf{C}'\boldsymbol{\Sigma}_d^{-1}\mathbf{C}\right)^{-1} + \boldsymbol{\Sigma}_x\right)^{-1}. \tag{12}$$

Substituting (11) into the right side of (10) and (12) into the left side of (10) yields

$$\left(\mathbf{C}'\boldsymbol{\Sigma}_d^{-1}\mathbf{C}\right)^{-1} + \boldsymbol{\Sigma}_x = \mathbf{F}\ \mathbf{S}_W\ \mathbf{F}',$$

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

$$\hat{\boldsymbol{\Sigma}}_x = \mathbf{F}\ \mathbf{S}_W\ \mathbf{F}' - \left(\mathbf{C}'\boldsymbol{\Sigma}_d^{-1}\mathbf{C}\right)^{-1} = \mathbf{F}\left(\mathbf{S}_W - \boldsymbol{\Sigma}_d\right)\mathbf{F}', \tag{13}$$

which is strictly positive definite. And so, clearly we have $\boldsymbol{\Sigma}_W$ invertible at the critical point.

To see that the updated $\hat{\boldsymbol{\Sigma}}_x$ leads to an increase in Λ_N, we show that the Hessian matrix, $\mathbf{H}(\boldsymbol{\Sigma}_x)$ evaluated at $\hat{\boldsymbol{\Sigma}}_x$ is negative definite. The ijth element of $\mathbf{H}(\boldsymbol{\Sigma}_x)$ is the second order partial derivative of Λ_N with respect to the ith and jth elements of the vectorized $\boldsymbol{\Sigma}_x$. From calculations in Section 8.4, we have

$$\mathbf{H}(\hat{\boldsymbol{\Sigma}}_x) = -(N/2)\left(\hat{\mathbf{D}} \otimes \hat{\mathbf{D}}\right), \quad \text{where } \hat{\mathbf{D}} = \mathbf{C}'\hat{\boldsymbol{\Sigma}}_W^{-1}\mathbf{C}, \tag{14}$$

which is clearly negative definite.

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 $\sigma^{2(t)}$ and $\sigma_\epsilon^{2(t)}$ are positive and if $\boldsymbol{\Sigma}_x^{(t)} > \mathbf{0}$, 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 $(\mathbf{z}_i, Y_i, \mathbf{x}_i)$, $i = 1, \ldots, N$, are our complete data and $\mathbf{W}_i \equiv (\mathbf{z}_i', Y_i)'$, $i = 1, \ldots, N$, 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 $\mathrm{E}_{\boldsymbol{\theta}^{(t)}}\left(\sum_{i=1}^N \ln f(\mathbf{z}_i, Y_i, \mathbf{x}_i) \mid \cdot\right)$ and the M-step maximizes this conditional expectation over σ_ϵ^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 σ_ϵ^2, we can ignore the last two terms and obtain $\sigma_\epsilon^{2(t+1)}$ via maximizing $\mathrm{E}_{\boldsymbol{\theta}^{(t)}}\left(\sum_{i=1}^N \ln f(\mathbf{z}_i|\mathbf{x}_i) \mid \cdot\right)$ over σ_ϵ^2.

From (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 (6), we have

$$\mathrm{Cov}(\mathbf{W}_i, \mathbf{x}_i) = \mathbf{C}\boldsymbol{\Sigma}_x \tag{15}$$

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

$$\mathrm{E}[\mathbf{x}_i|\mathbf{W}_i] \equiv \boldsymbol{\mu}_{x_i|W_i} = \boldsymbol{\Sigma}_x \mathbf{C}'\boldsymbol{\Sigma}_W^{-1}(\mathbf{W}_i - \boldsymbol{\mu}_W), \tag{16}$$

$$\mathrm{Cov}[\mathbf{x}_i|\mathbf{W}_i] \equiv \boldsymbol{\Sigma}_{x|W} = \boldsymbol{\Sigma}_x - \boldsymbol{\Sigma}_x \mathbf{C}'\boldsymbol{\Sigma}_W^{-1}\mathbf{C}\boldsymbol{\Sigma}_x. \tag{17}$$

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

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

Differentiating this conditional mean with respect to σ_ϵ^2 and equating the derivative to zero yields

$$\sigma_\epsilon^{2(t+1)} = \frac{1}{nN}\tilde{s} + \frac{1}{n}\mathrm{tr}[\mathbf{R}^{-1}\mathbf{A}\boldsymbol{\Sigma}_{x|W}^{(t)}\mathbf{A}']. \tag{18}$$

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

$$\boldsymbol{\Sigma}_{x|W}^{(t)} = \left(\boldsymbol{\Sigma}_x^{(t)^{-1}} + \mathbf{C}^{(t)\prime}\boldsymbol{\Sigma}_d^{(t)^{-1}}\mathbf{C}^{(t)}\right)^{-1}$$

which is positive definite. Given $\boldsymbol{\Sigma}_{x|W}^{(t)} > \mathbf{0}$, assumption (a) then implies that $\mathbf{A}\boldsymbol{\Sigma}_{x|W}^{(t)}\mathbf{A}' \geq \mathbf{0}$. Together with the fact that $\mathbf{R} > \mathbf{0}$, the second term in (18) is thus nonnegative.

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

The updates of $\boldsymbol{\beta}^{(t)}$ and $\sigma^{2(t)}$ maximize Λ_N over $\boldsymbol{\beta}$ and σ^2, holding the other parameters fixed. Suppose that $\sigma_\epsilon^{2(t)} > 0$ and $\boldsymbol{\Sigma}_x^{(t)} > \mathbf{0}$. 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 σ^2. We also note given $\mathbf{z}_i$, y_i is normal with mean

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

and variance

$$\sigma^2_{Y|z} \equiv \mathrm{Var}(Y_i|\mathbf{z}_i) = \boldsymbol{\beta}'\mathbf{K}\boldsymbol{\beta} + \sigma^2 \tag{19}$$

where

$$\mathbf{G} = \mathbf{T}\boldsymbol{\Sigma}_x\mathbf{A}'\boldsymbol{\Sigma}_z^{-1} \tag{20}$$

and

$$\mathbf{K} = \mathbf{T}\boldsymbol{\Sigma}_x\mathbf{T}' - \mathbf{T}\boldsymbol{\Sigma}_x\mathbf{A}'\boldsymbol{\Sigma}_z^{-1}\mathbf{A}\boldsymbol{\Sigma}_x\mathbf{T}'.$$

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

$$\tilde{\Lambda}_N = -\frac{N}{2}\ln(\boldsymbol{\beta}'\mathbf{K}\boldsymbol{\beta} + \sigma^2) - \frac{1}{2(\boldsymbol{\beta}'\mathbf{K}\boldsymbol{\beta} + \sigma^2)}\sum_{i=1}^{N}\left(Y_i - \beta_0 - \boldsymbol{\beta}'\mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu})\right)^2. \tag{21}$$

With detailed derivation in Section 8.5, equating $\partial\tilde{\Lambda}_N/\partial\boldsymbol{\beta}$ and $\partial\tilde{\Lambda}_N/\partial\sigma^2$ to zero yields respectively

$$\frac{1}{\boldsymbol{\beta}'\mathbf{K}\boldsymbol{\beta} + \sigma^2}\sum_{i=1}^{N}\left(Y_i - \beta_0 - \boldsymbol{\beta}'\mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu})\right)\mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu}) = 0 \tag{22}$$

$$\frac{1}{(\boldsymbol{\beta}'\mathbf{K}\boldsymbol{\beta} + \sigma^2)^2}\left[\sum_{i=1}^{N}\left(Y_i - \beta_0 - \boldsymbol{\beta}'\mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu})\right)^2 - N(\boldsymbol{\beta}'\mathbf{K}\boldsymbol{\beta} + \sigma^2)\right] = 0. \tag{23}$$

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 (22) for $\boldsymbol{\beta}$ and (23) for σ^2 gives

$$\begin{aligned}\boldsymbol{\beta}^{(t+1)} &= \hat{\boldsymbol{\beta}} = \mathbf{M}^{-1}\mathbf{G}\sum_{i=1}^{N}(\mathbf{z}_i - \boldsymbol{\mu})(Y_i - \beta_0) \\ \sigma^{2(t+1)} &= \hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \beta_0 - \boldsymbol{\beta}'\mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu})\right)^2 - \boldsymbol{\beta}'\mathbf{K}\boldsymbol{\beta}. \qquad (24)\end{aligned}$$

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 σ^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 (24) implies

$$\hat{\sigma}^2_{Y|z} \equiv \hat{\boldsymbol{\beta}}'\mathbf{K}\hat{\boldsymbol{\beta}} + \hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \beta_0 - \hat{\boldsymbol{\beta}}'\mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu})\right)^2, \qquad (25)$$

which is positive by assumption (d). With detailed calculation in Section 8.5, the Hessian matrix $\mathbf{H}\tilde{\Lambda}(\boldsymbol{\beta}, \sigma^2)$ when evaluated at $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ equals

$$\mathbf{H}\tilde{\Lambda}(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) = -\frac{N}{(\hat{\sigma}^2_{Y|z})^2}\begin{pmatrix} 2\mathbf{K}\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}'\mathbf{K} + \frac{\hat{\sigma}^2_{Y|z}}{N}\mathbf{M} & \mathbf{K}\hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\beta}}'\mathbf{K} & \frac{1}{2} \end{pmatrix}. \qquad (26)$$

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

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

4 Inference for β

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

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

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

We estimate $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ 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 (26). We take $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}$ 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.

4.1 Hypothesis testing for β

4.1.1 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

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

The unrestricted supremum of Λ_N is achieved at the ECME estimates described in Section 3. We can also use the ECME procedure to calculate the first supremum. To obtain the first supermum, we observe under the restriction $\boldsymbol{\beta} = \mathbf{0}$, $\mathbf{z}_i$ 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 σ^2 as known. We update $\boldsymbol{\Sigma}_x$ by maximizing Λ_N directly while holding σ_ϵ^2 fixed. We update σ_ϵ^2 by finding its EM estimate $\sigma_\epsilon^{2(t+1)}$ while holding $\boldsymbol{\Sigma}_x$ fixed. We iterate untill convergence occurs.

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

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

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

The third statistic is the integrated t-statistic

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

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 $\{1, \ldots, N\}$. To simulate the null distribution of a test statistic, we make Q new data sets via permuting the Y_i's: the qth data set is simply $\{\mathbf{z}_1, Y_{i_1}, \ldots, \mathbf{z}_N, Y_{i_N}\}$ where $i_1, \ldots, i_N$ is a random permutation of $1, \ldots, N$. The p-value is then calculated as the proportion of the resulting Q statistic values larger than the original observed value.

4.1.2 Testing equality of two β's

Often we want to know if the β'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, $\Lambda^s_{N^s}$ and $\Lambda^c_{N^c}$, each have the same expressions as in (21) but with possibly different parameter values, superscripted by s or c.

The first test statistic is from a likelihood ratio test

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

Each of the last two suprema is calculated separately, using the ECME estimates from Section 3.

We can also apply this ECME procedure to calculate the first supremum. Under the restriction $\boldsymbol{\beta}^s = \boldsymbol{\beta}^c$, $\Lambda^s_{N^s}$ and $\Lambda^c_{N^c}$ have a common parameter $\boldsymbol{\beta} \equiv \boldsymbol{\beta}^s = \boldsymbol{\beta}^c$. Following the same argument as in Section 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 3.2. To update each of $\sigma_\epsilon^{c2(t)}$ and $\sigma_\epsilon^{s2(t)}$, we follow the steps outlined in Section 3.3.

To update $\boldsymbol{\beta}^{(t)}$, $\sigma^{s2(t)}$ and $\sigma^{c2(t)}$, we proceed as in Section 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 $\tilde{\Lambda}^s_{N^s}$ and $\tilde{\Lambda}^c_{N^c}$ in a manner analagous to $\tilde{\Lambda}_N$ in (21). By an argument similar to that in Section 3.4, we must find $\boldsymbol{\beta}$, σ^{s2} and σ^{c2} to maximize $\tilde{\Lambda}^s_{N^s} + \tilde{\Lambda}^c_{N^c}$. Differentiating $\tilde{\Lambda}^s_{N^s} + \tilde{\Lambda}^c_{N^c}$ with respect to σ^{s2} and σ^{c2} and setting equal to zero yields equations analagous to the equation for $\sigma^{2(t+1)}$ in (24).

Unfortunately, differentiating $\tilde{\Lambda}^s_{N^s} + \tilde{\Lambda}^c_{N^c}$ 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 σ^{s2} and σ^{c2}. Instead of maximizing $\tilde{\Lambda}^s_{N^s} + \tilde{\Lambda}^c_{N^c}$ with respect to $\boldsymbol{\beta}$ we maximize $\tilde{\tilde{\Lambda}}^{\,s}_{N^s} + \tilde{\tilde{\Lambda}}^{\,c}_{N^c}$ 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 $\boldsymbol{\Sigma}_x^{s(t+1)}$ 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 $\sigma^{s2(t)}_{Y|z}$, no longer contains the parameters $\boldsymbol{\beta}$ and σ^{s2}. Let

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

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

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

$$\begin{aligned}\boldsymbol{\beta}^{(t+1)} &= \left(\frac{1}{\sigma^{s2(t)}_{Y|z}}\mathbf{M}^s + \frac{1}{\sigma^{c2(t)}_{Y|z}}\mathbf{M}^c\right)^{-1} \\ &\quad \left(\frac{1}{\sigma^{s2(t)}_{Y|z}}\mathbf{G}^s\sum(\mathbf{z}^s_i - \boldsymbol{\mu}^s)(Y^s_i - \beta^s_0) + \frac{1}{\sigma^{c2(t)}_{Y|z}}\mathbf{G}^c\sum(\mathbf{z}^c_i - \boldsymbol{\mu}^c)(Y^c_i - \beta^c_0)\right).\end{aligned}$$

where

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

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 $\hat{\boldsymbol{\Sigma}}^s_{\boldsymbol{\beta}}$ and define $\hat{\boldsymbol{\beta}}^c$ and $\hat{\boldsymbol{\Sigma}}^c_{\boldsymbol{\beta}}$ similarly. The two sample statistic is

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

We consider a two sample statistic based on function evaluations, rather than on basis coefficients, recalling the notation in Section 4.1.1. Let $\tilde{\boldsymbol{\beta}}^s = \boldsymbol{\Psi}^s\hat{\boldsymbol{\beta}}^s$ be the vector of function evaluations of $\hat{\beta}^s$ at a sequence of time points. Let $\hat{\boldsymbol{\Sigma}}^s_{\tilde{\boldsymbol{\beta}}} = \boldsymbol{\Psi}^s\hat{\boldsymbol{\Sigma}}^s_{\boldsymbol{\beta}}\boldsymbol{\Psi}^{s\prime}$ be the estimate of $\tilde{\boldsymbol{\beta}}^s$'s covariance matrix. Similarly define $\tilde{\boldsymbol{\beta}}^c$ and $\hat{\boldsymbol{\Sigma}}^c_{\tilde{\boldsymbol{\beta}}}$. 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 4.1.1, 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 $\mathbf{\Psi}^s = \mathbf{\Psi}^c$ and both of them are of full column rank.

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

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

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 $\beta^s(t) = \beta^c(t)$ 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 qth "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.

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 (4) and (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 (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 (9). We get

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

and thus

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

It then follows

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

The last element of $\hat{\mathbf{W}}_i$ gives the fitted value of Y_i and the last element of $\boldsymbol{r}_i$ 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.

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 $\phi_1, \ldots, \phi_K$. Then we conduct the data analysis.

6.1 Choosing the basis functions

We choose the basis functions, the ϕ_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 $\hat{\mathbf{\Sigma}}$ be the sample covariance matrix of the $\mathbf{z}_i$'s. We find a sequence of orthogonal vectors $\tilde{\mathbf{v}}_j \in \boldsymbol{R}^n$ to maximize

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

where λ 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 $t_1, \ldots, t_n$ 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 λ, the vectors $\tilde{\mathbf{v}}_j$ are eigenvectors of the matrix $\mathbf{G}^{-1/2} \hat{\mathbf{\Sigma}} \mathbf{G}^{-1/2}$, 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 λ can be done by cross-validation, but for simplicity here, we select λ by examining the smoothness of the resulting $\tilde{\mathbf{v}}_j$'s. In the data analysis, we chose $\lambda = 100$.

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 take the averaged wheel running distance over weeks 5 to 60 as the response Y. The predictor $Z(\cdot)$ is the log transformed body mass. We want to know if any of the groups have β non-zero and if there is any difference between the selection β^s and the control β^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 1 and 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.

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 ith mouse's average log body mass and

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

the ith 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 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 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 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.

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 (4) and (5) within each group, using $R = \mathbf{I}$, and choose the same basis functions for β as for the log body mass, calculated in Section 6.3. Within each group, we plot the Y_i residuals against the fitted Y_i's to detect outliers as outlined in Section 5. Residual plots of each group are in Figure 5. In the control males, we see an outlier at the left side. This outlier may be influential in the estimation of β. So we remove this outlier and refit the model to the male control group. The first panel of Figure 6 is the same as the plot in the second panel of Figure 5, showing the outlier in the control males. The second panel of Figure 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.

6.5 Inference for $\beta(\cdot)$

In each of the four groups, we estimated β and calculated standard errors from both the Hessian matrix and the bootstrap, as described in Section 4. These are shown in Figure 7.

To determine if $\beta(t) = 0$ for all $t \in [-1, 60]$, we can study the plots in Figure 7 and we can conduct a formal hypothesis test. From the plots in

Figure 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 4.1.1. We compute the standard errors using the Hessian matrix (26). The results are in Table 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-likehood 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.

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

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 8 and see that U_l is an increasing function of U_w. This may be due to the normality assumption in the model.

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 9 shows the pointwise difference between the selection $\hat{\beta}$ and the control $\hat{\beta}$ 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 β and the control β.

Table 2 gives results of the test $H_o : \ \beta^s(t) = \beta^c(t), \ \text{for all } t \in [-1, 60]$ 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-likehood 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.

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

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.

7 Simulation study

In the simulation study, we compare the pointwise mean squared errors of our ECME estimate of β 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 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 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 β^s and β^c.

In the one-sample comparison in Section 7.2, we simulate data using parameter values estimated from the male selected group data analyzed in Section 6.2. In the two-sample comparison in Section 7.3, we simulate selection group data using estimates from the male selected group data and the control group data using estimates from the male control group data without the outlier.

7.1 Two stage estimate

Recall from Section 1 that the calculation of Müller's (2005) estimate of β 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 β. We suspect that ignoring the y_i's in the first part of the procedure will lead to poorer estimation of β.

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 ϕ_k's, see (3). We then construct the matrix $\mathbf{T}$ and fit model (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 ϵ_i's to calculate our predictors, the BLUPs of $\mathbf{x}_i$ given $\mathbf{z}_i$, that is, to calculate $\mathrm{E}(\mathbf{x}_i|\mathbf{z}_i)$. For the second part of the analysis, we find the least squares estimate of $\boldsymbol{\beta}$ according to the regression model

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

In our fit of (4), we force $\boldsymbol{\Sigma}_x$ to be diagonal. This is in keeping with Múller, 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.

7.2 One sample comparison

We simulate data based on parameter and eigenfunction estimates from the data analysis of the male selected group in Section 6.2. Let $\boldsymbol{\mu}^s, \boldsymbol{\Sigma}_x^s, \sigma_\epsilon^{s2}, \sigma^{s2}, \beta_0^s$, $\boldsymbol{\beta}^s$ 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 $\mathrm{N}\left(\boldsymbol{\mu}^s, \mathbf{A}^s\boldsymbol{\Sigma}_x^s\mathbf{A}^{s\prime}\right)$ and let the observed $\mathbf{z}_i$ be $\mathbf{Z}_i$ plus $\mathrm{N}(\mathbf{0}, \sigma_\epsilon^{s2}\mathbf{I})$ noise. We consider four possible β functions to describe the relationship between Y_i and Z_i:

$$Y_i = \beta_0^s + \int_{-1}^{60} \beta(t)\left[Z_i(t) - \bar{Z}_i(t)\right]\mathrm{d}t + \mathrm{e}_i.$$

The integral is calculated using the R function "sintegral" and we simulate e_i from $\mathrm{N}(0, \sigma^{s2})$. 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 6.2, we found that β^s was significantly different from zero.

Thus, for each value of γ, we can generate an observed data sets $(\mathbf{z}_i, Y_i^\gamma)$, $i = 1, \ldots, 39$.

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

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

at each observation point t. The results are in Figure 10 where the first panel shows the pattern of the true β's, the next two panels show the pointwise MSE's of our two estimates of β and the last panel shows the differences of these pointwise MSE's. In the plots, the MSE's increase as γ increases and both estimates of β 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 β 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 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 γ increases.

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

To test $H_o : \ \beta(t) = 0, \ \text{for all } t \in [-1, 60]$, we use the test statistics U_w and U_f with standard errors calculated using the Hessian matrix (26) as described in Section 4. For each data set, we run 300 permutations to calculate the p-values. We simulate 100 data sets for each value of γ and choose levels $\alpha = 0.01$ and $\alpha = 0.05$. Figure 11 and Figure 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 γ increases. There is little difference in power between the two statistics.

7.3 Two sample comparison

To simulate data from two independent samples, we choose model parameters using the male selected group data and the male control group data analysed

in Section 6.2. We simulate data for the selection group and, separately, for the control group using the same methodology as in Section 7.2 but with different true β's.

Let β^s and β^c be the estimates of β from the original male selected 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 β 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 β is $2\gamma\Delta\beta = \gamma(\beta^s - \beta^c)$.

To estimate β^s and β^c, we fit the selection data and control data separately. For each value of γ we simulate 100 data sets and calculate the MSE as in Section 7.2 in each group. Figures 13 and 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 10.

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

$$\sum_{1}^{100} \left(\hat{\beta}^s(t) - \hat{\beta}^c(t) - \beta^s(t) + \beta^c(t))^2\right) /100$$

at each observation point t. Figure 15 shows the results. The patterns of the two MSE's are similar to their counterparts in Figure 10. The MSE's of the two estimates are comparable but in general the MSE of the ECME estimate is smaller.

To test $H_o : \ \beta^s(t) = \beta^c(t), \ \text{for all } t \in [-1, 60]$, we compare the four test statistics, U_l, U_w, U_e and U_f, with standard errors calculated using the Hessian matrix (26). For each data set, we run 300 permutations to calculate the p-values. We simulate 100 data sets for each value of γ and choose levels $\alpha = 0.01$ and $\alpha = 0.05$. Figures 16 and 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 γ is large, but U_f and U_w are comparable.

7.4 Edge effect discussion in one-sample MSE comparison

In Figure 10, we see that the MSE's of the ECME estimate of β 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 β before week 10 as shown at the first panel of Figure 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 selected 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 7.2. The simulation analysis result is in Figure 18. t The edge effect still exists but this time the MSE's of ECME and the two stage method are comparable at the edges.

8 Appendix

In this appendix, we provide the calculations in Sections 3.2 and 3.4 where we find the updates of $\mathbf{\Sigma}_x$ and $\{\boldsymbol{\beta}, \sigma^2\}$ in the ECME procedure.

In Section 8.4, we derive the first order condition (10) and the Hessian matrix (14) of Section 3.2 where we maximize the log-likelihood Λ_N over $\mathbf{\Sigma}_x$ while holding the other parameters fixed.. In Section 8.5, we derive the first order conditions (22) and (23), and the Hessian matrix (26) of Section 3.4 where we maximize $\tilde{\Lambda}_N$ over $\{\boldsymbol{\beta}, \sigma^2\}$ 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 8.1-8.3, 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).

8.1 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 8.1 *Let* $\mathbf{f} : \mathbf{S} \to \Re^m$ *be a function defined on a set* $\mathbf{S}$ *in* $\Re^n$*. 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 $\mathbf{u} \in \Re^n$ *with* $||\mathbf{u}|| < r$ *and*

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

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*

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

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 $\boldsymbol{D}\mathbf{f}$ be the $m \times n$ *Jacobian matrix* of $\mathbf{f}$ whose ijth element is $\boldsymbol{D}_j f_i$: the partial derivative of the ith component function f_i of $\mathbf{f}$ with respect to the jth coordinate. The *First Identification Theorem* (p.87) states that the first derivative $\mathbf{A}(\mathbf{c})$ is $\boldsymbol{D}\mathbf{f}(\mathbf{c})$ 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

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

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

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

Let $\mathbf{F} : \mathbf{S} \to \Re^{m\times p}$ be a matrix function defined on a set $\mathbf{S}$ in $\Re^{n\times q}$. 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 $\mathbf{f} : \mathrm{vec}\ \mathbf{S} \to \Re^{m\times p}$ defined by

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

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

Definition 8.2 *Let $\mathbf{F} : \mathbf{S} \to \Re^{m\times p}$ be a matrix function defined on a set $\mathbf{S}$ in $\Re^{n\times q}$. 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*

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

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

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

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

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

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* $\mathrm{d}\mathbf{F}(\mathbf{C}; \mathbf{U})$ *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

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

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

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

This is an $(mp) \times (nq)$ matrix, whose ijth element is the partial derivative of the ith component of $\mathrm{vec}\mathbf{F}(\mathbf{X})$ with respect to the jth element of $\mathrm{vec}\mathbf{X}$, 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

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

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

8.2 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 8.3 *Let $f : \mathbf{S} \to \Re$ be a real valued function defined on a set $\mathbf{S}$ in $\Re^n$, 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 $\boldsymbol{D}_j f$ 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 8.4 *Let $f : \mathbf{S} \to \Re$ be twice differentiable at an interior point $\mathbf{c}$ of $\mathbf{S} \subset \Re^n$. 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 $g : \mathbf{B}(\mathbf{c}) \to \Re$ be defined by the equation*

$$g(\mathbf{x}) = \mathrm{d}f(\mathbf{x}; \mathbf{u}).$$

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

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.

$$\mathrm{d}^2 f = \mathrm{d}(\mathrm{d}f).$$

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

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

where $\boldsymbol{H}f(\mathbf{c})$ 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 8.1, 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 $\mathbf{S} \subseteq \Re^{n \times q}$ 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 $\Re^{n \times q}$ 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

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

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

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

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

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

That is, the ijth element of $\boldsymbol{H}f(\mathbf{C})$ is the second order derivative of f with respect to the ith and jth element of vec$\mathbf{X}$ where $\mathbf{X} \in \mathbf{S}$, evaluated at $\mathbf{C}$.

8.3 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.

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

8.4 Calculations in Section 3.2

Recall the observed data log-likehood has the expression

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

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 Λ_N to obtain the first order condition (10) and calculate the second differential to obtain the Hessian matrix in (14).

As we treat $\boldsymbol{\Sigma}_x$ as the only unknown parameter, it immediately follows from the expression of $\boldsymbol{\Sigma}_W$ in (9)

$$\mathrm{d}\boldsymbol{\Sigma}_W = \mathbf{C}\mathrm{d}\boldsymbol{\Sigma}_x\mathbf{C}'. \tag{27}$$

In our derivation, we will use the shorter notation $\mathrm{d}\Sigma_W$ before we reach (10) or (14). We have

$$\begin{aligned}
\mathrm{d}\Lambda_N &= -\frac{N}{2}\mathrm{d}\left(\ln\boldsymbol{\Sigma}_W\right) - \frac{1}{2}\sum_{i=1}^{N}(\mathbf{W}_i - \boldsymbol{\mu}_W)'\left(\mathrm{d}\ \boldsymbol{\Sigma}_W^{-1}\right)(\mathbf{W}_i - \boldsymbol{\mu}_W) \\
&= -\frac{N}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}_W^{-1}\mathrm{d}\boldsymbol{\Sigma}_W\right] + \frac{1}{2}\sum_{i=1}^{N}(\mathbf{W}_i - \boldsymbol{\mu}_W)'\boldsymbol{\Sigma}_W^{-1}\left(\mathrm{d}\ \boldsymbol{\Sigma}_W\right)\boldsymbol{\Sigma}_W^{-1}(\mathbf{W}_i - \boldsymbol{\mu}_W) \\
&= -\frac{N}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}_W^{-1}\mathrm{d}\boldsymbol{\Sigma}_W\right] + \frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}_W^{-1}\sum_{i=1}^{N}(\mathbf{W}_i - \boldsymbol{\mu}_W)(\mathbf{W}_i - \boldsymbol{\mu}_W)'\boldsymbol{\Sigma}_W^{-1}\left(\mathrm{d}\ \boldsymbol{\Sigma}_W\right)\right] \\
&= -\frac{N}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}_W^{-1}\left(\boldsymbol{\Sigma}_W - \mathbf{S}_W\right)\boldsymbol{\Sigma}_W^{-1}\left(\mathrm{d}\boldsymbol{\Sigma}_W\right)\right] \qquad (28)
\end{aligned}$$

Recalling (27), we now have

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

By the *First Identification Theorem for matrix functions* mentioned in Section 8.1, we obtain the *Jacobian matrix* of Λ_N at $\boldsymbol{\Sigma}_x$ as

$$\boldsymbol{D}\ \Lambda_N(\boldsymbol{\Sigma}_x) = \mathrm{vec}\left\{\mathbf{C}'\boldsymbol{\Sigma}_W^{-1}\left(\boldsymbol{\Sigma}_W - \mathbf{S}_W\right)\boldsymbol{\Sigma}_W^{-1}\mathbf{C}\right\}.$$

Equating it to zero yields

$$\mathbf{C}'\boldsymbol{\Sigma}_W^{-1}\left(\boldsymbol{\Sigma}_W - \mathbf{S}_W\right)\boldsymbol{\Sigma}_W^{-1}\mathbf{C} = \mathbf{0}$$

which is equivalent to the first order condition (10).

Next we calculate $\mathrm{d}^2\Lambda_N$ to identify the Hessian matrix in (14). Starting from (28), we have

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

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

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

Then

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

Recall (27). Right hand side of the above

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

When evaluated at the critical point $\hat{\mathbf{\Sigma}}_x$ which satisfies the first order condition (10), $\mathrm{d}^2\Lambda_N$ is then

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

By the *Second Identification Theorem for matrix functions* mentioned in Section 8.2, we have at $\hat{\mathbf{\Sigma}}_x$ the Hessian matrix is equal to

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

This is the matrix we saw in (14).

8.5 Calculations in Section 3.4

Recall we want to maximize the log-likelihood

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

over $\{\boldsymbol{\beta}, \sigma^2\}$ to find the update while fixing the other parameters. In this section, we derive the first order conditions (22) and (23) via calculating the first differential of $\tilde{\Lambda}_N$ with respect to $\{\boldsymbol{\beta}, \sigma^2\}$. Calculating the second differential of $\tilde{\Lambda}_N$ then gives us the Hessian matrix (26).

The following two differentials will facilitate our calculation in $\mathrm{d}\tilde{\Lambda}_N$ and $\mathrm{d}^2\tilde{\Lambda}_N$. Recall the expression of $\sigma^2_{Y|z}$ in (19). We have

$$\mathrm{d}\sigma^2_{Y|z} \equiv \mathrm{d}\left(\boldsymbol{\beta}'\mathbf{K}\boldsymbol{\beta} + \sigma^2\right) = 2\boldsymbol{\beta}'\mathbf{K}\mathrm{d}\boldsymbol{\beta} + \mathrm{d}\sigma^2. \tag{29}$$

Let

$$\mathbf{g}(\boldsymbol{\beta}) = \sum_{i=1}^{N}\left(Y_i - \beta_0 - \boldsymbol{\beta}'\mathbf{G}(\mathbf{z}_i - \boldsymbol{\mu})\right)(\mathbf{z}_i - \boldsymbol{\mu})'\mathbf{G}'. \tag{30}$$

We then obtain

$$\begin{aligned}
\mathrm{d}\left[\sum_{i=1}^{N}\left(Y_i-\beta_0-\boldsymbol{\beta}'\mathbf{G}(\mathbf{z}_i-\boldsymbol{\mu})\right)^2\right] &= -2\sum_{i=1}^{N}\left(Y_i-\beta_0-\boldsymbol{\beta}'\mathbf{G}(\mathbf{z}_i-\boldsymbol{\mu})\right)(\mathrm{d}\boldsymbol{\beta})'\mathbf{G}(\mathbf{z}_i-\boldsymbol{\mu}) \\
&= -2\sum_{i=1}^{N}\left(Y_i-\beta_0-\boldsymbol{\beta}'\mathbf{G}(\mathbf{z}_i-\boldsymbol{\mu})\right)(\mathbf{z}_i-\boldsymbol{\mu})'\mathbf{G}'(\mathrm{d}\boldsymbol{\beta}) \\
&= -2\mathbf{g}(\boldsymbol{\beta})\mathrm{d}\boldsymbol{\beta} \qquad (31)
\end{aligned}$$

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

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

By (29) and (31), the right hand side above is equal to

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

Let

$$\mathbf{c}(\boldsymbol{\beta},\sigma^2)=\left[\sum_{i=1}^{N}\left(Y_i-\beta_0-\boldsymbol{\beta}'\mathbf{G}(\mathbf{z}_i-\boldsymbol{\mu})\right)^2-N\sigma^2_{Y|z}\right]. \qquad (32)$$

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

$$\frac{\mathrm{d}\sigma^2}{2\sigma^4_{Y|z}}\,\mathbf{c}(\boldsymbol{\beta},\sigma^2)+\frac{1}{\sigma^4_{Y|z}}\left[\boldsymbol{\beta}'\mathbf{K}\;\mathbf{c}(\boldsymbol{\beta},\sigma^2)+\sigma^2_{Y|z}\mathbf{g}(\boldsymbol{\beta})\right]\mathrm{d}\boldsymbol{\beta}. \qquad (33)$$

By the *First Identification Theorem* mentioned in Section 8.1, we obtain the first order conditions

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

which lead to (22) and (23).

Calculating $\mathrm{d}^2\tilde{\Lambda}_N$ to identify the Hessian matrix is lengthy and tedious. In fact, we don't need the closed form of the Hessian matrix but the Hessian matrix evaluated at the critical points $\{\hat{\boldsymbol{\beta}}, \hat{\sigma}^2\}$ given in (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

$$\mathbf{g}(\hat{\boldsymbol{\beta}}) = \mathbf{0} \tag{34}$$

$$\mathbf{c}(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) = 0. \tag{35}$$

From (32), using (29) and (31) we have

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

which is a function of $\boldsymbol{\beta}$. By (34),

$$\mathrm{d}\mathbf{c}(\hat{\boldsymbol{\beta}}, \sigma^2) = -2N(\mathrm{d}\boldsymbol{\beta})'\mathbf{K}\boldsymbol{\beta} - N\mathrm{d}\sigma^2. \tag{36}$$

Now we calculate $\mathrm{d}^2\tilde{\Lambda}_N$ starting from (33). We first calculate

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

which is

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

When at $\{\hat{\boldsymbol{\beta}}, \hat{\sigma}^2\}$, by (35) and (36), it is equal to

$$-\frac{N\mathrm{d}\sigma^2}{2\hat{\sigma}^4_{Y|z}}\left(2(\mathrm{d}\boldsymbol{\beta})'\mathbf{K}\hat{\boldsymbol{\beta}} + \mathrm{d}\sigma^2\right) = -\frac{N}{\hat{\sigma}^4_{Y|z}}\left(\mathrm{d}\boldsymbol{\beta}'\ \mathrm{d}\sigma^2\right)\begin{pmatrix}\mathbf{K}\hat{\boldsymbol{\beta}} \\ 1/2\end{pmatrix}\mathrm{d}\sigma^2. \tag{37}$$

Then we calculate

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

which is

$$\mathrm{d}\left(\frac{1}{\sigma^4_{Y|z}}\right)\left[\boldsymbol{\beta}'\mathbf{K}\ \mathbf{c}(\boldsymbol{\beta},\sigma^2)+\sigma^2_{Y|z}\mathbf{g}(\boldsymbol{\beta})\right]\mathrm{d}\boldsymbol{\beta}+\frac{1}{\sigma^4_{Y|z}}\mathrm{d}\left[\boldsymbol{\beta}'\mathbf{K}\ \mathbf{c}(\boldsymbol{\beta},\sigma^2)+\sigma^2_{Y|z}\mathbf{g}(\boldsymbol{\beta})\right]\mathrm{d}\boldsymbol{\beta}. \tag{38}$$

At $\{\hat{\boldsymbol{\beta}},\hat{\sigma}^2\}$, the term in the first square brackets vanishes by (34) and (35). Thus, at $\{\hat{\boldsymbol{\beta}},\hat{\sigma}^2\}$, (38) is equal to

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

From (30), we have

$$\mathrm{d}\mathbf{g}(\boldsymbol{\beta})=-(\mathrm{d}\boldsymbol{\beta})'\sum_{i=1}^{N}\mathbf{G}(\mathbf{z}_i-\boldsymbol{\mu})(\mathbf{z}_i-\boldsymbol{\mu})'\mathbf{G}'. \tag{39}$$

Again by (34)-(36) and that $\mathrm{d}\mathbf{c}(\hat{\boldsymbol{\beta}},\sigma^2)$ is a scalar, (38) is equal to

$$\begin{aligned}&\frac{1}{\hat{\sigma}^4_{Y|z}}\left[\left(-2N(\mathrm{d}\boldsymbol{\beta})'\mathbf{K}\hat{\beta}-N\mathrm{d}\sigma^2\right)\hat{\boldsymbol{\beta}}'\mathbf{K}-\hat{\sigma}^2_{Y|z}(\mathrm{d}\boldsymbol{\beta})'\sum_{i=1}^{N}\mathbf{G}(\mathbf{z}_i-\boldsymbol{\mu})(\mathbf{z}_i-\boldsymbol{\mu})'\mathbf{G}'\right]\mathrm{d}\boldsymbol{\beta}\\ =\ &-\frac{N}{\hat{\sigma}^4_{Y|z}}\left(\mathrm{d}\boldsymbol{\beta}'\ \mathrm{d}\sigma^2\right)\begin{pmatrix}2\mathbf{K}\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}'\mathbf{K}+\frac{\hat{\sigma}^2_{Y|z}}{N}\sum_{i=1}^{N}\mathbf{G}(\mathbf{z}_i-\boldsymbol{\mu})(\mathbf{z}_i-\boldsymbol{\mu})'\mathbf{G}'\\ \hat{\boldsymbol{\beta}}'\mathbf{K}\end{pmatrix}\mathrm{d}\boldsymbol{\beta}\end{aligned} \tag{40}$$

Combining (37) and (40), eventually we get, at $(\hat{\boldsymbol{\beta}},\hat{\sigma}^2)$,

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

where

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

By the *Second Identification Theorem* mentioned in Section 8.2, $\boldsymbol{H}\tilde{\Lambda}(\hat{\boldsymbol{\beta}},\hat{\sigma}^2)$ is the Hessian matrix and we have seen it in (26).

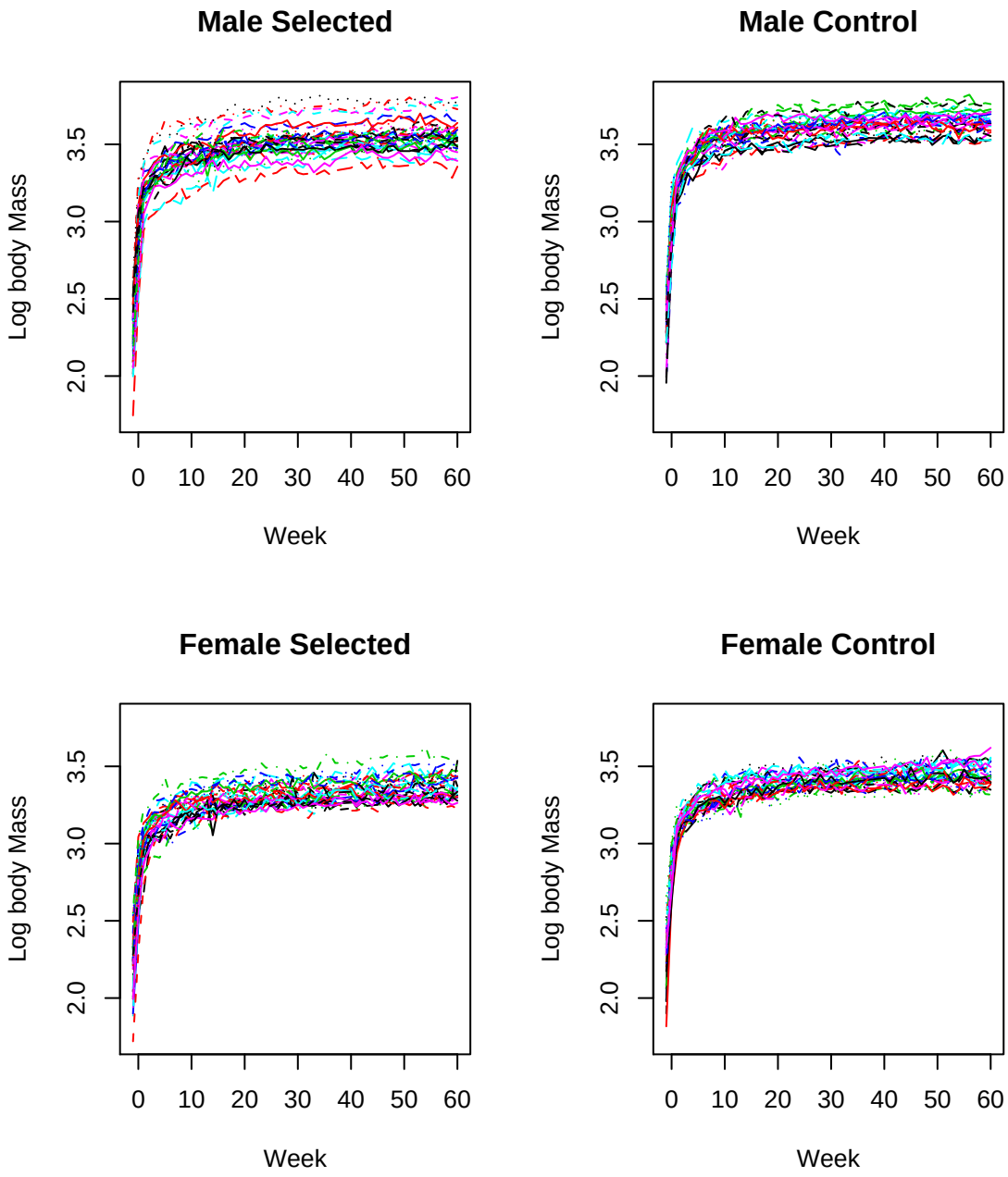

Figure 1: Predictor: log transformed body mass.

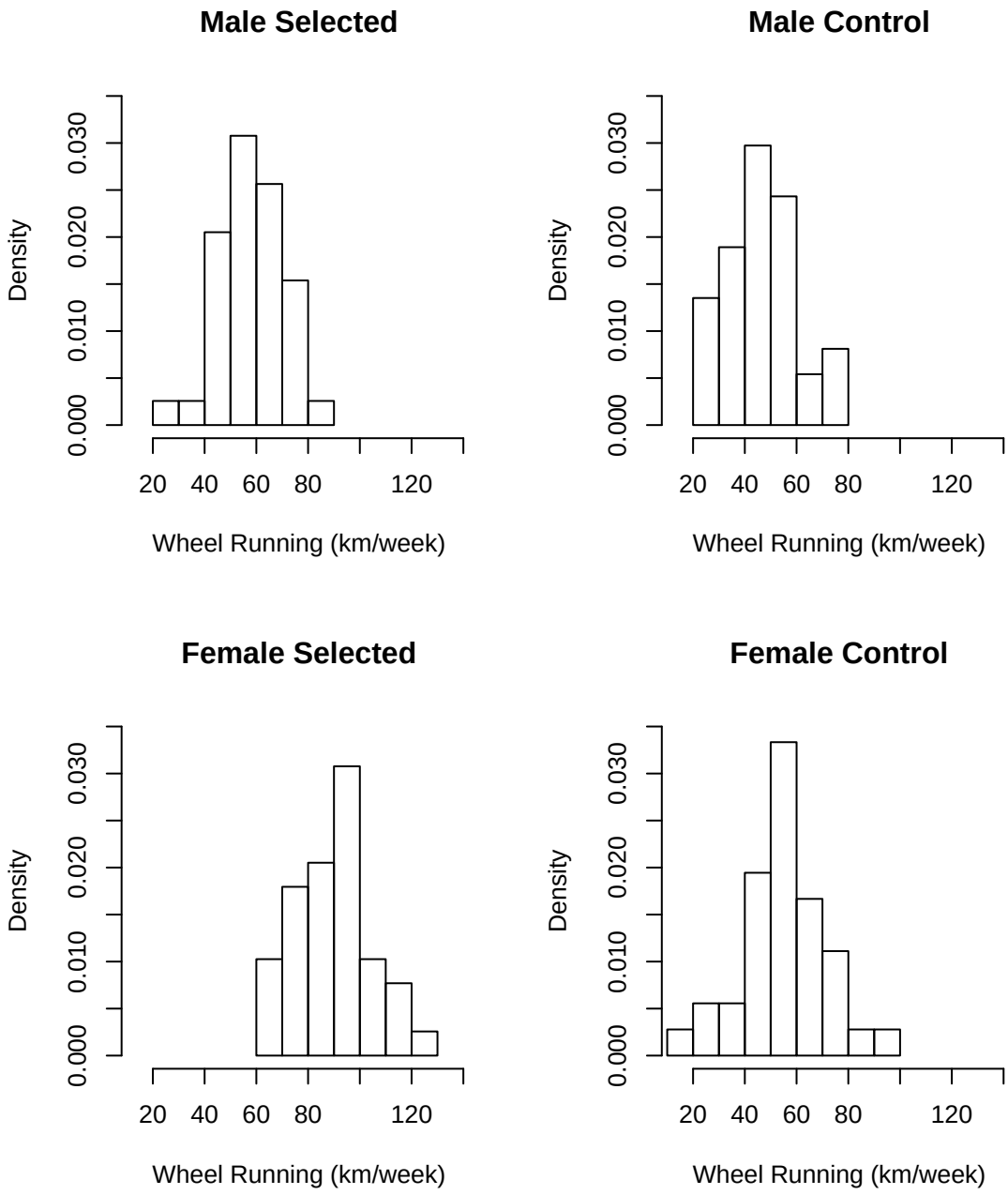

Figure 2: Histogram of the response: averaged wheel running.

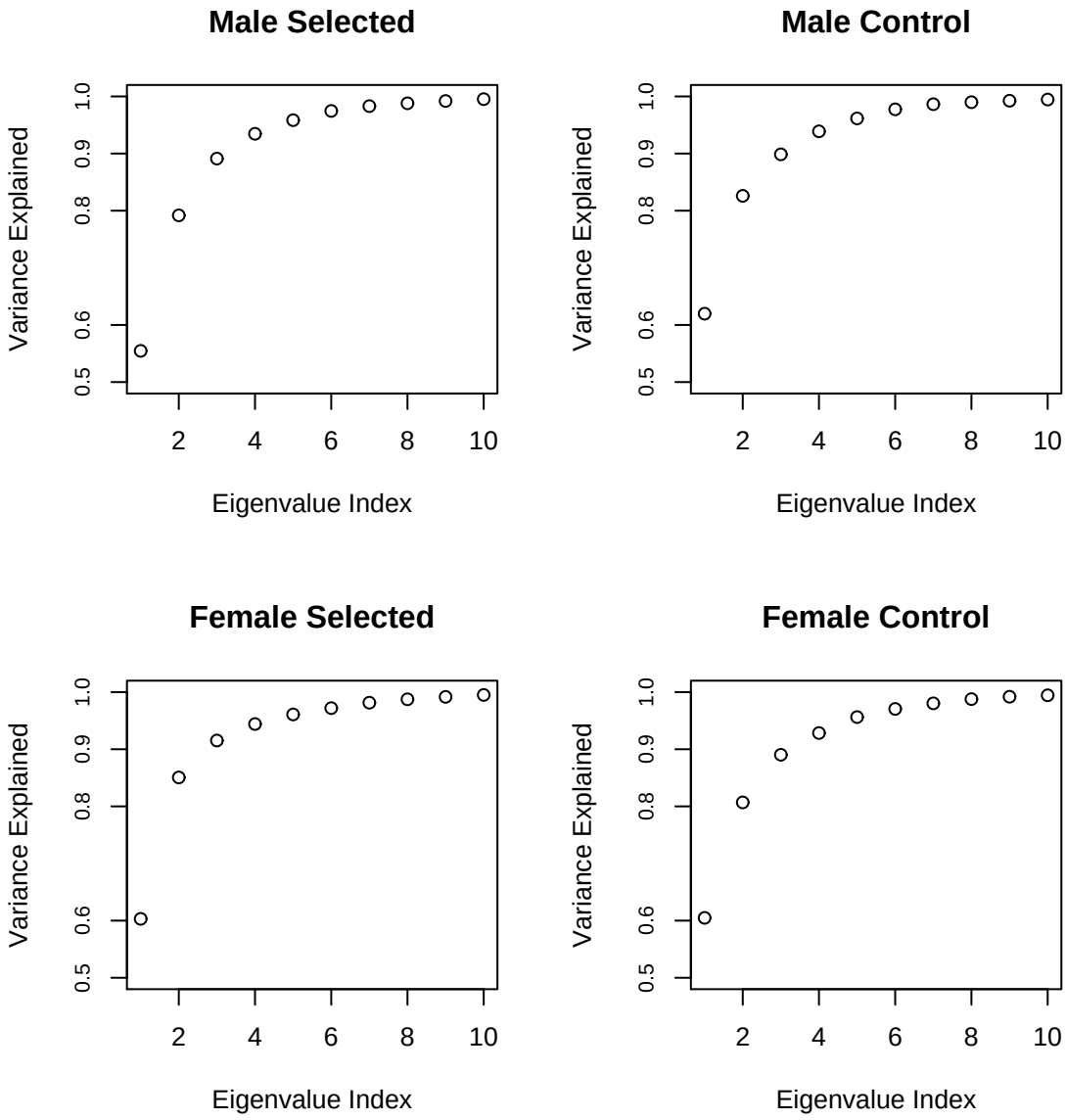

Figure 3: Plots of the proportion of cumulative variance of the centered log body mass explained by the first ten principal components in each group, after the individual average log body mass has been removed.

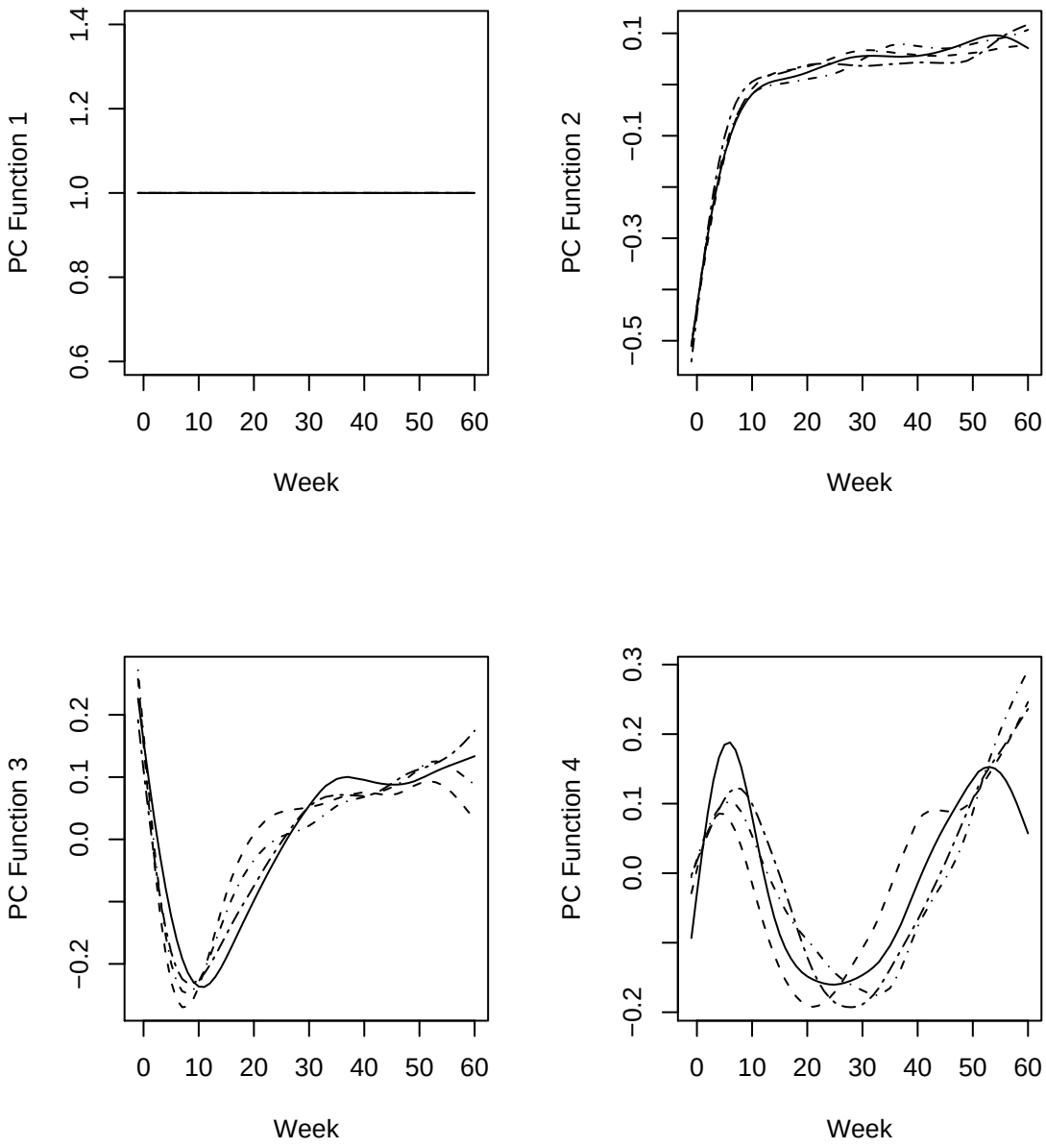

Figure 4: The constant function and the first three smoothed eigenfunctions of the covariance of centered log body mass.

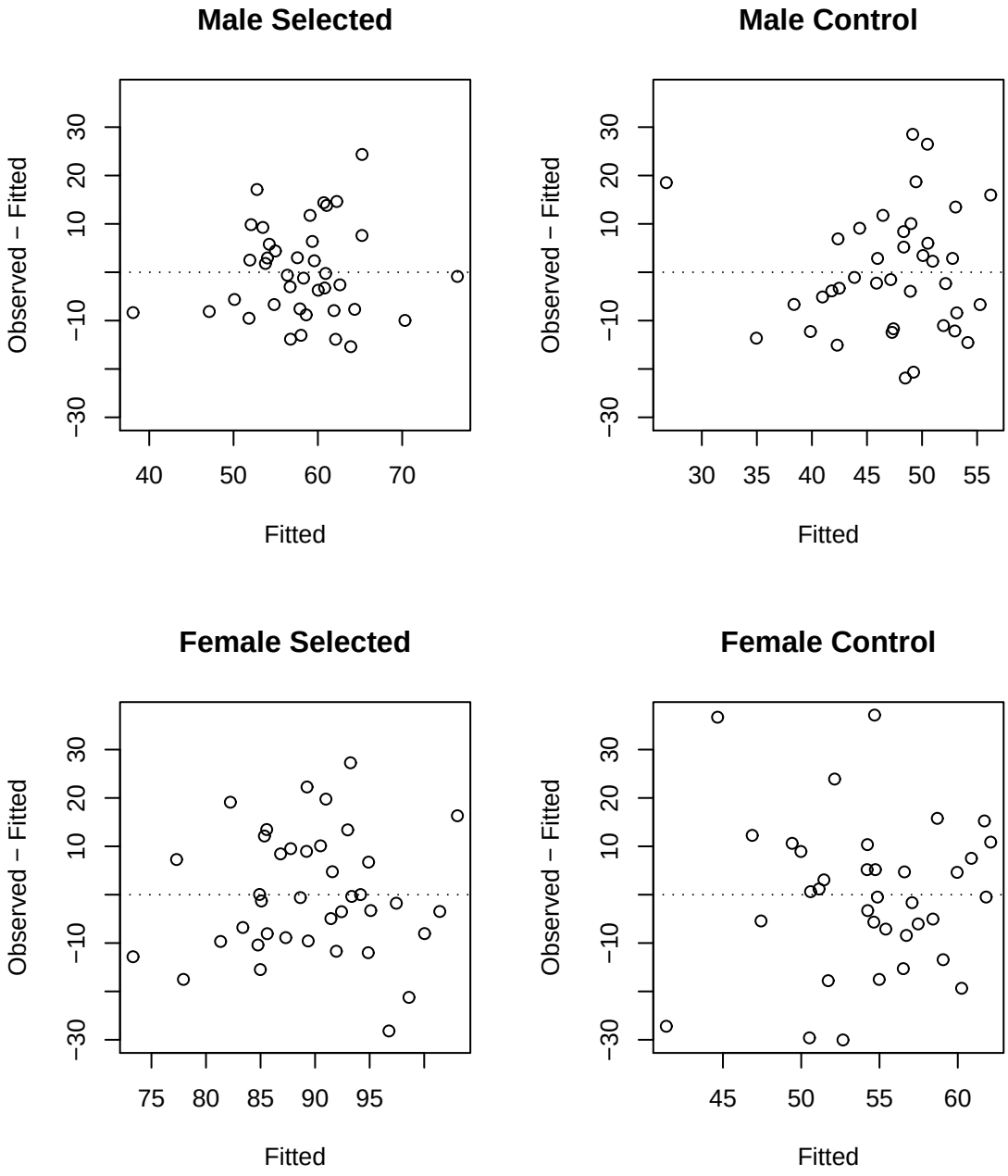

Figure 5: Residuals of the fit of the Y_i.

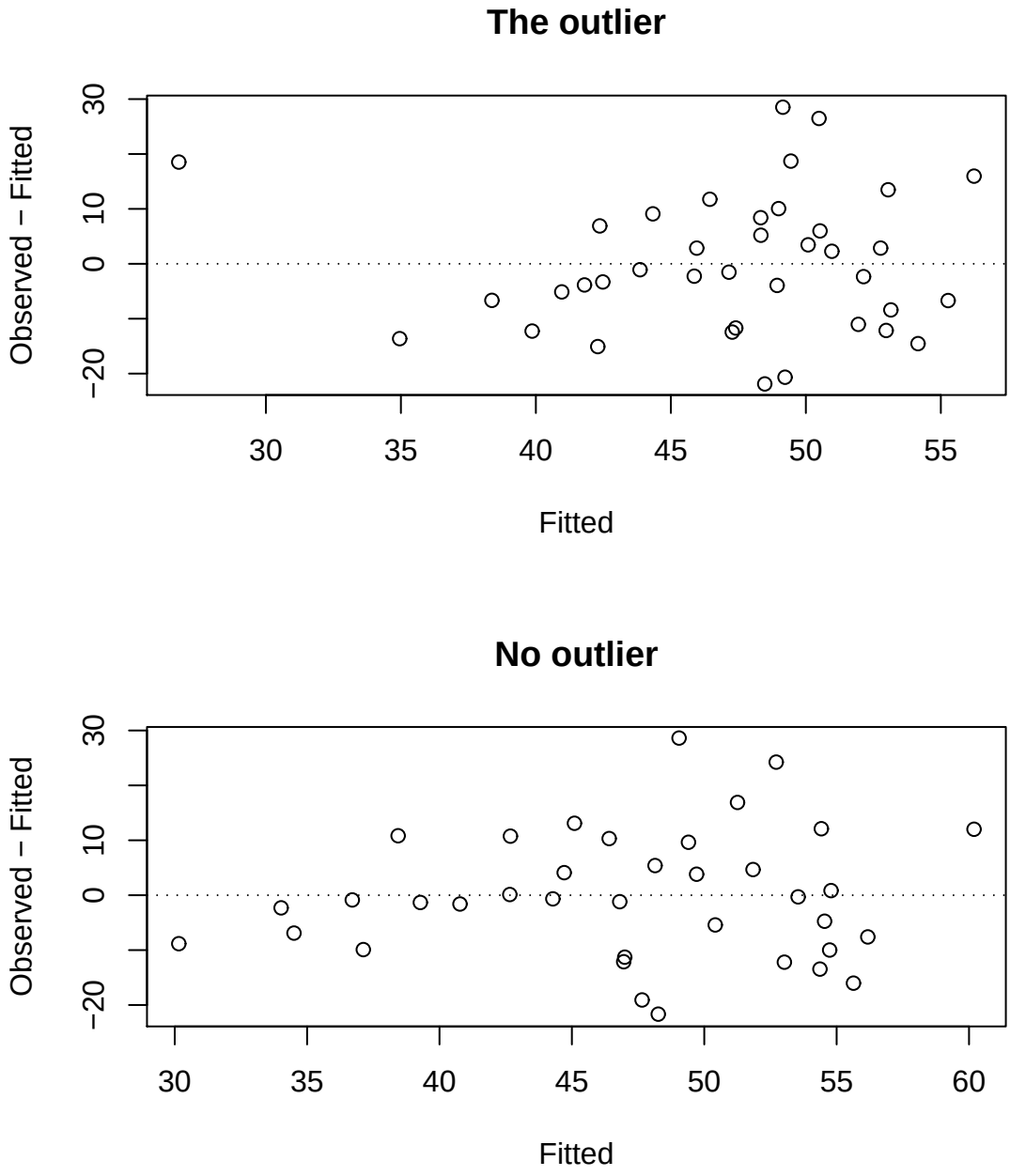

Figure 6: Y_i residuals in the male control group before and after removing the outlier.

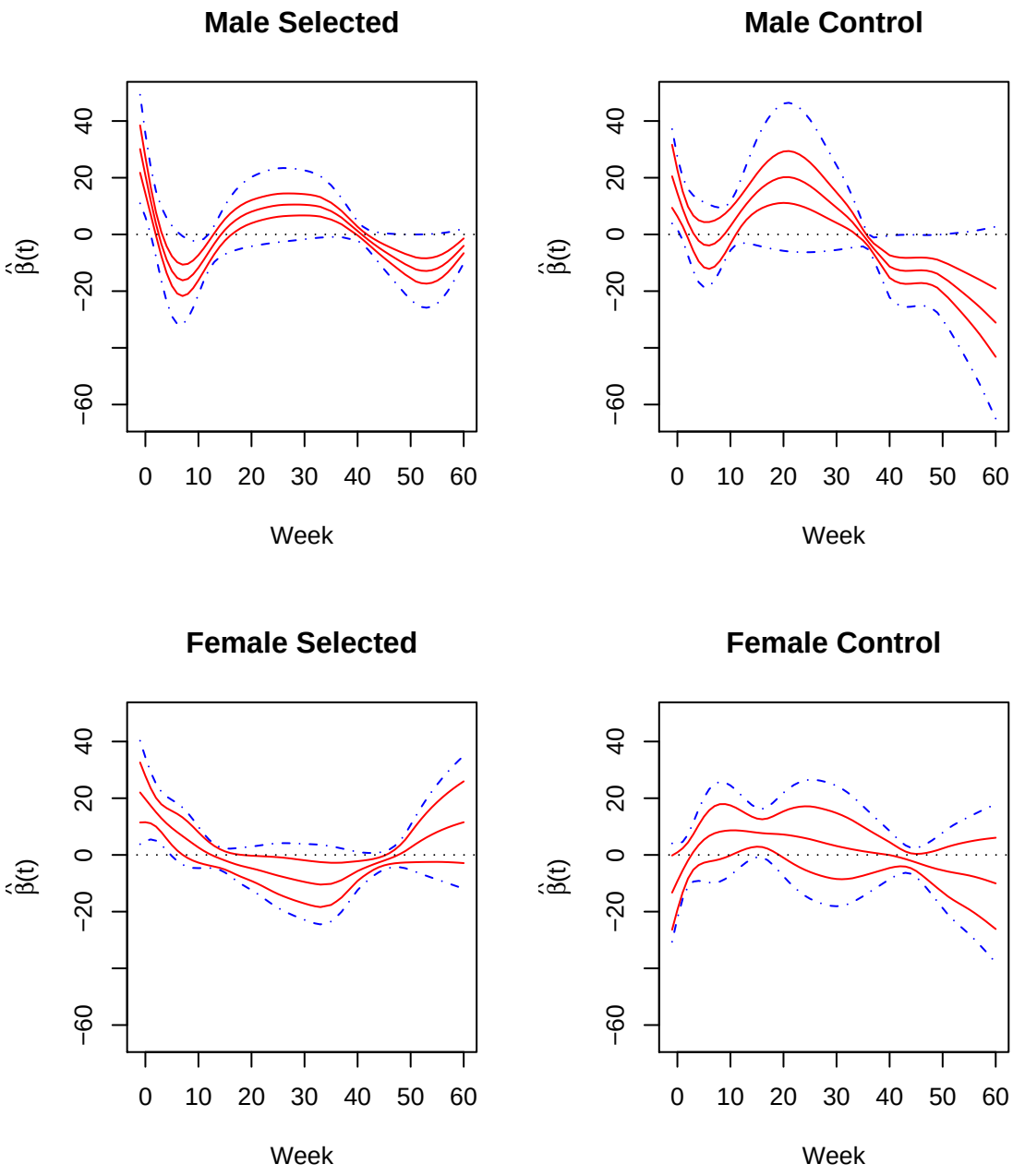

Figure 7: Plots of $\hat{\beta}$ and its standard errors computed from the Hessian matrix (solid) and from the bootstrap (dash-dot).

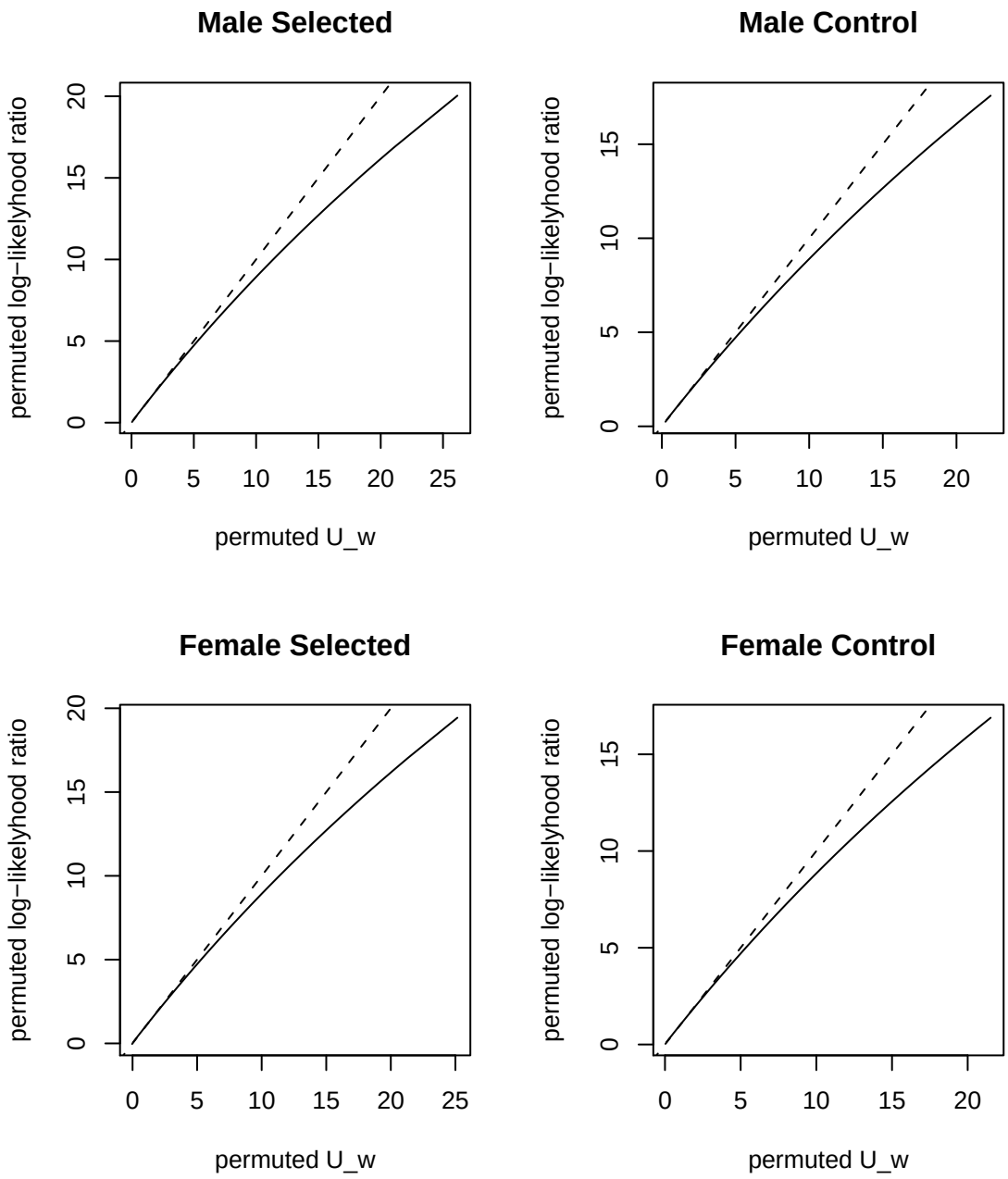

Figure 8: Comparing the permuted values of the generalized likelihood ratio statistic U_l with the Wald statistic U_w.

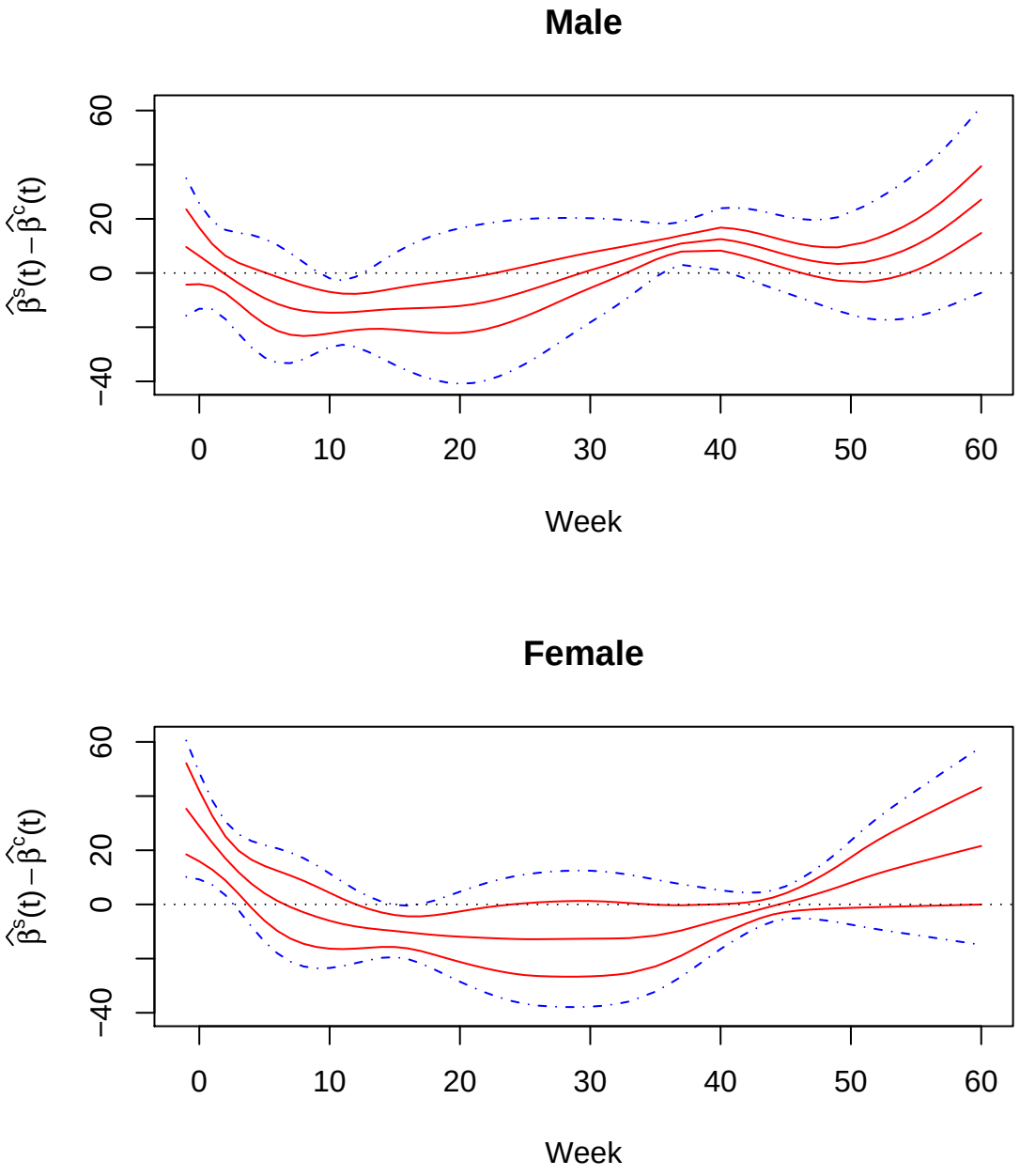

Figure 9: Plots of $\hat{\beta}^s - \hat{\beta}^c$ and the standard errors of the difference computed from the Hessian matrix (solid) and from the bootstrap (dash-dot).

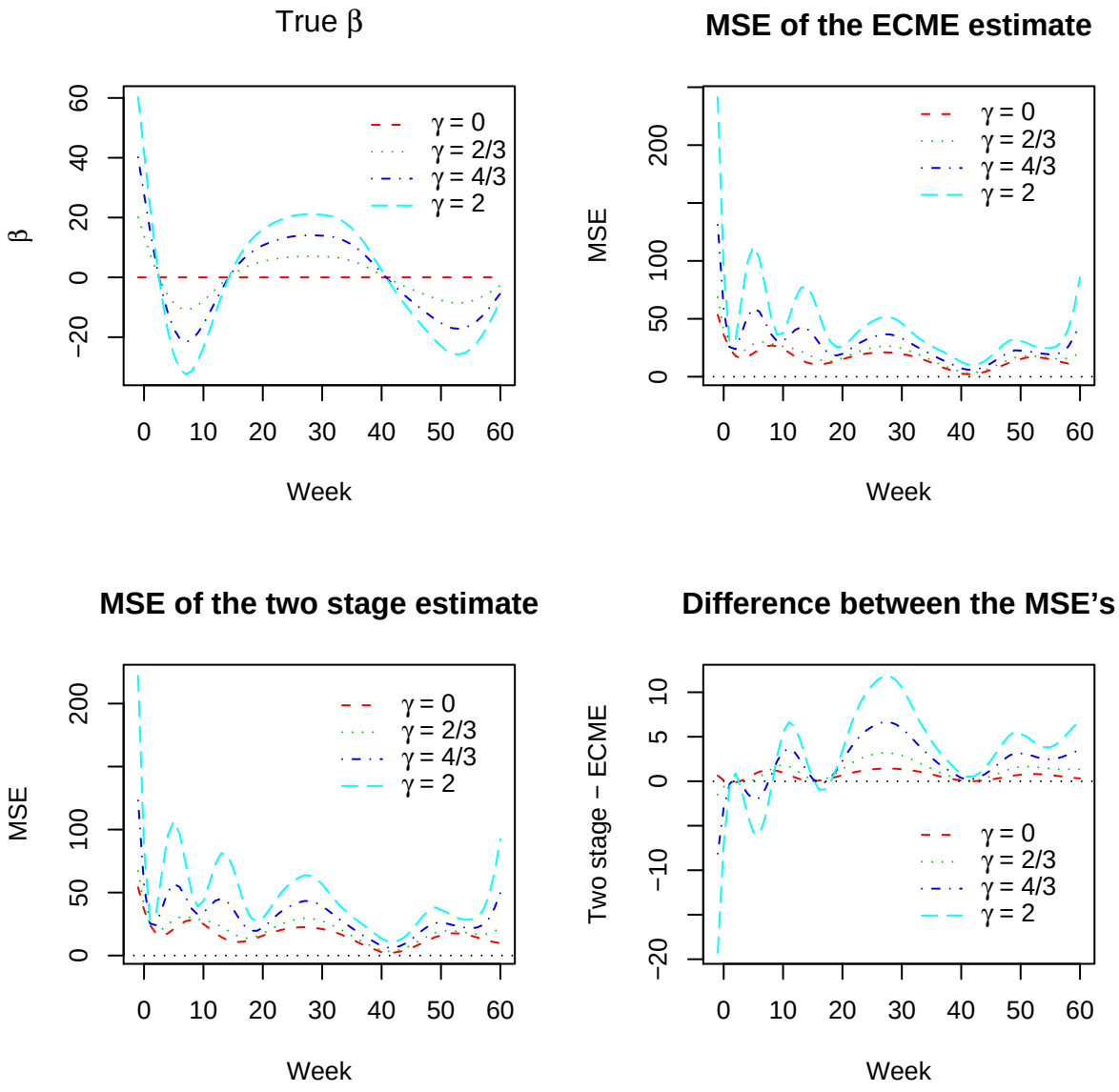

Figure 10: MSE of the estimate of β for each γ value. Compare the ECME method with the two stage method.

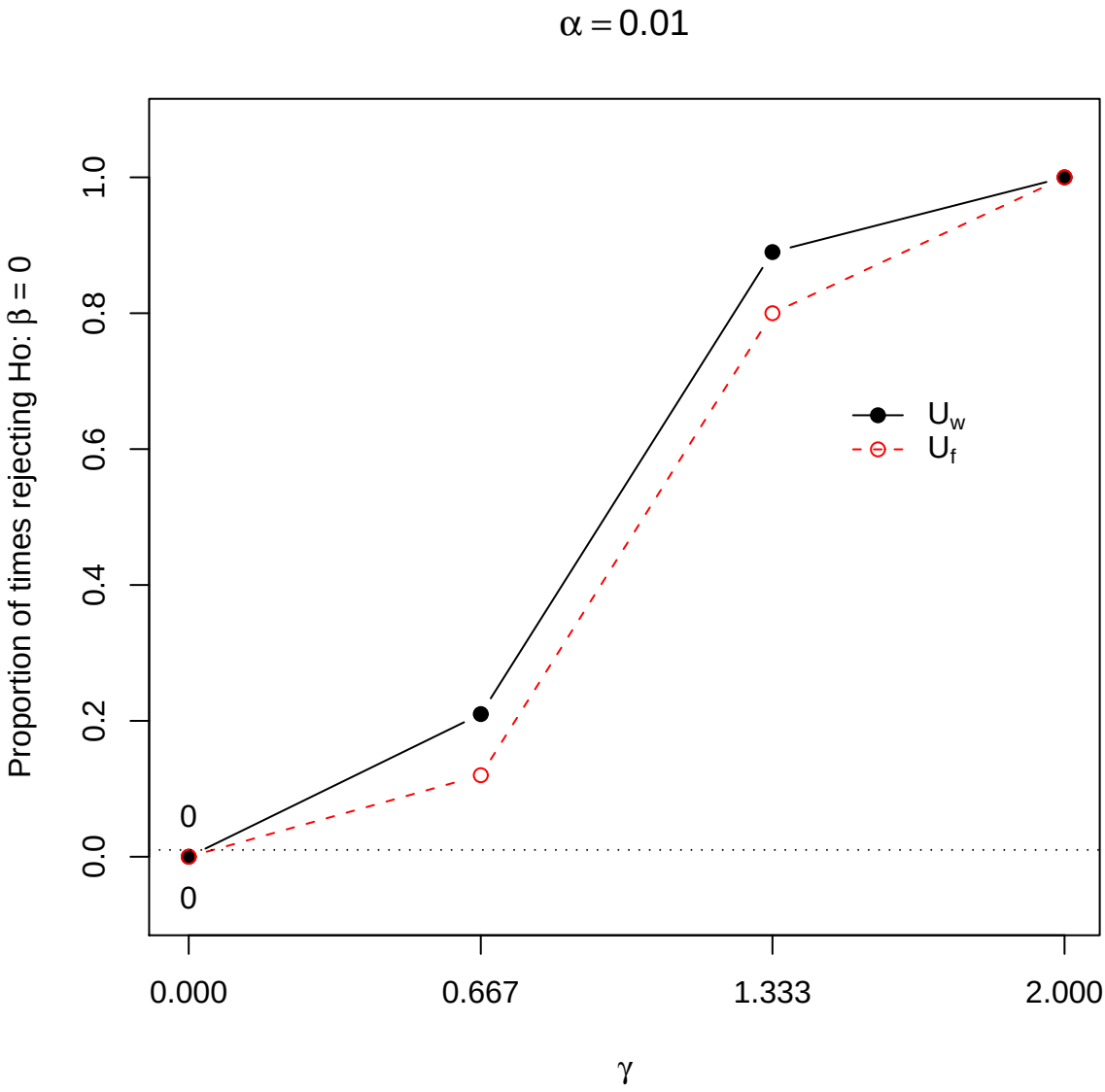

Figure 11: Proportion of times H_o is rejected using level $\alpha = 0.01$, where $H_o : \ \beta(t) = 0, \ \text{for all } t \in [-1, 60]$. Two test statistics are considered, U_w and U_f.

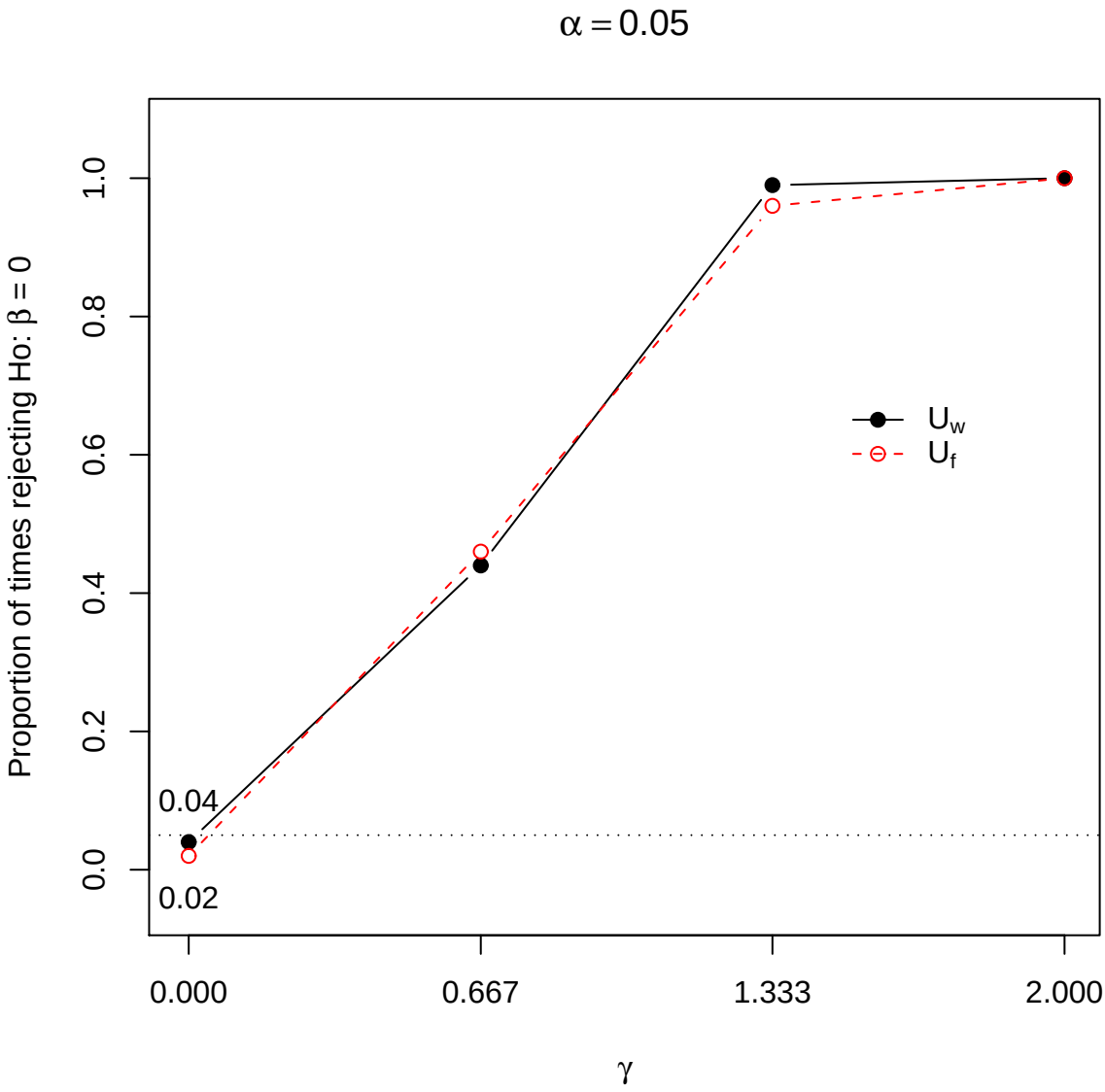

Figure 12: Proportion of times H_o is rejected using level $\alpha = 0.05$, where $H_o : \ \beta(t) = 0, \ \text{for all } t \in [-1, 60]$. Two test statistics are considered, U_w and U_f.

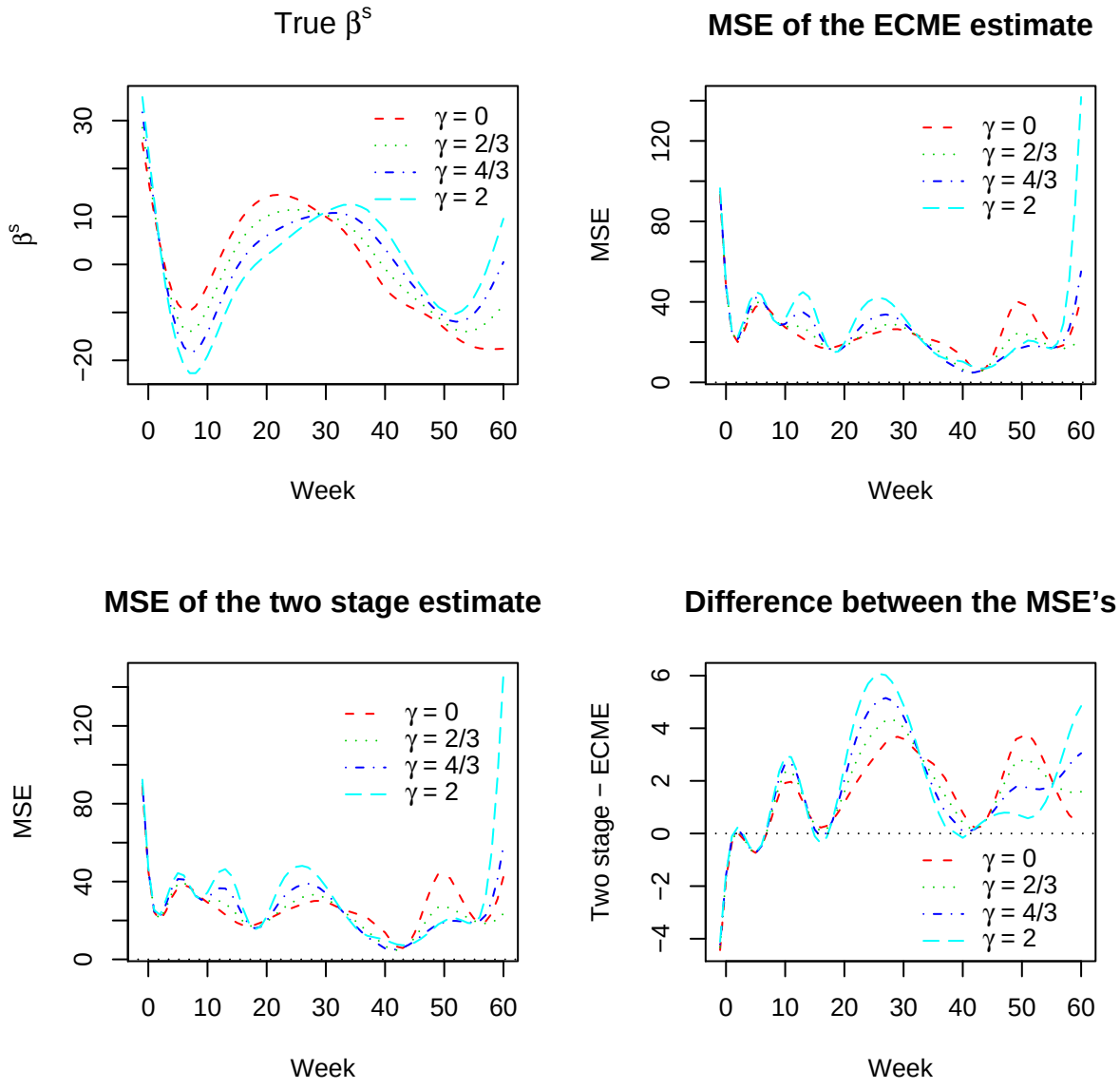

Figure 13: MSE of the estimate of β^s for each γ value. Compare the ECME method with the two stage method.

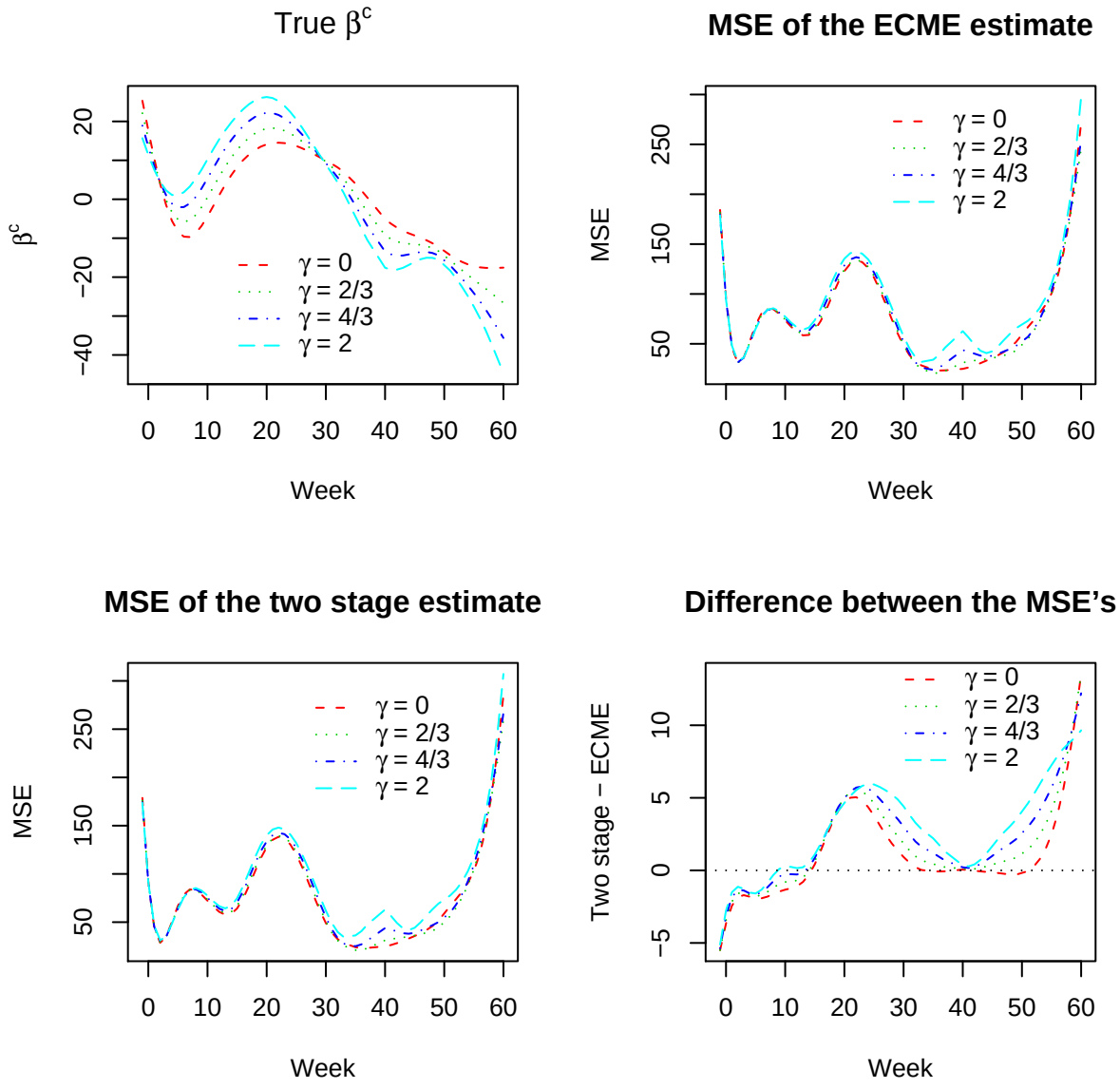

Figure 14: MSE of the estimate of β^c for each γ value. Compare the ECME method with the two stage method.

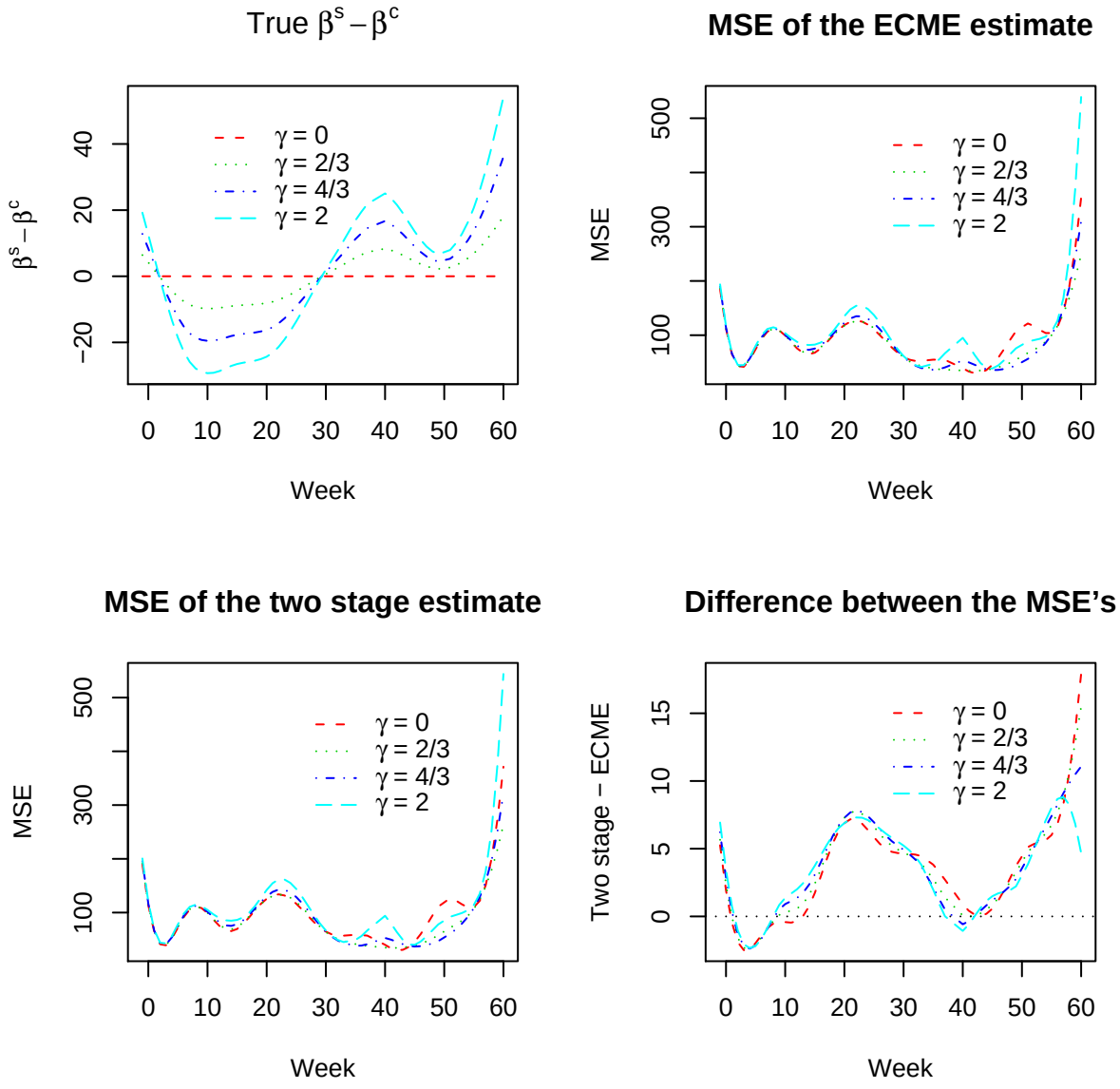

Figure 15: MSE of the estimate of $\beta^s - \beta^c$ for each γ value. Compare the ECME method with the two stage method.

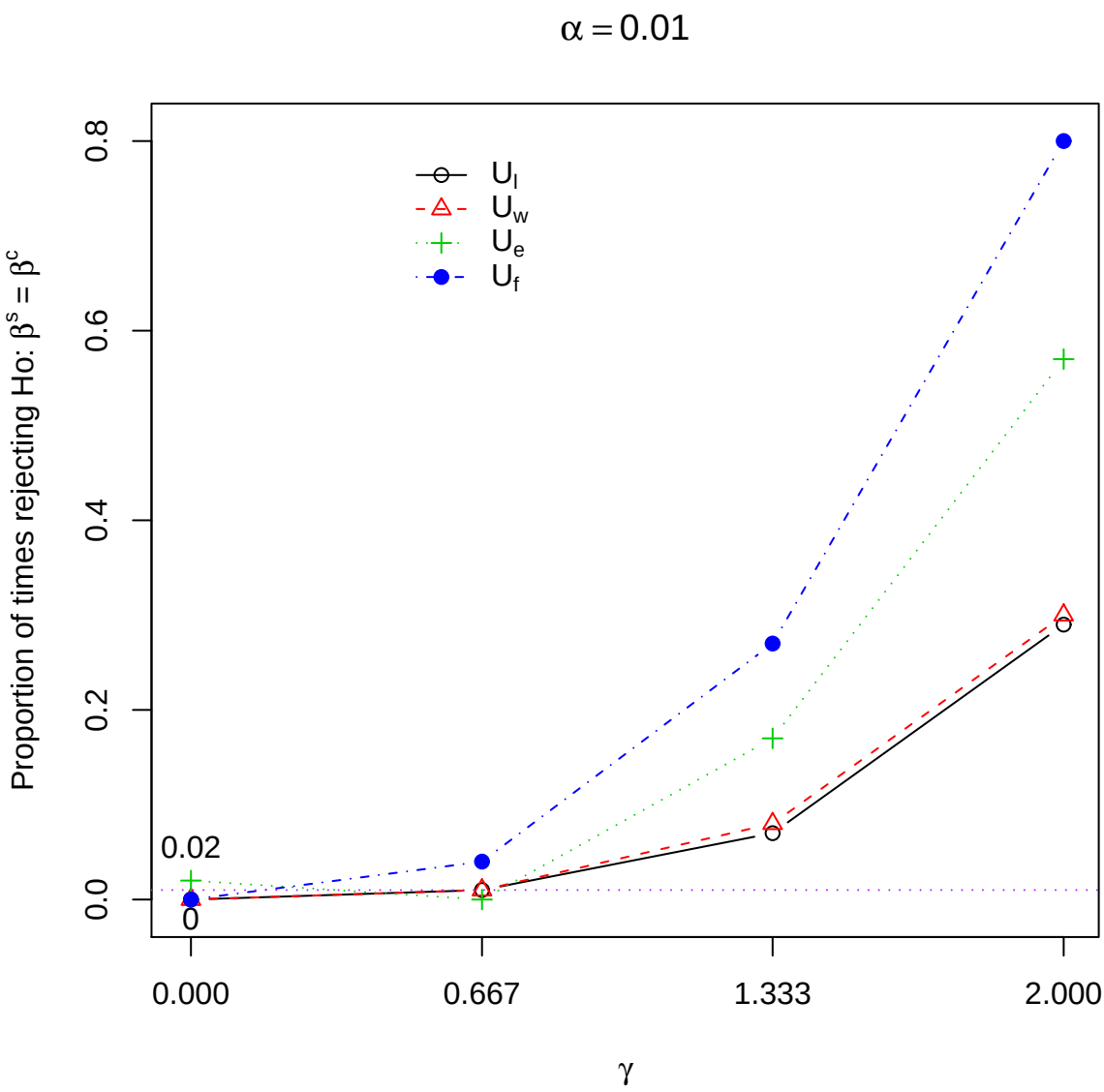

Figure 16: Proportion of times H_o is rejected using level $\alpha = 0.01$, where $H_o: \ \beta^s = \beta^c$, for all $t \in [-1, 60]$. Four test statistics are considered U_l, U_w, U_e and U_f.

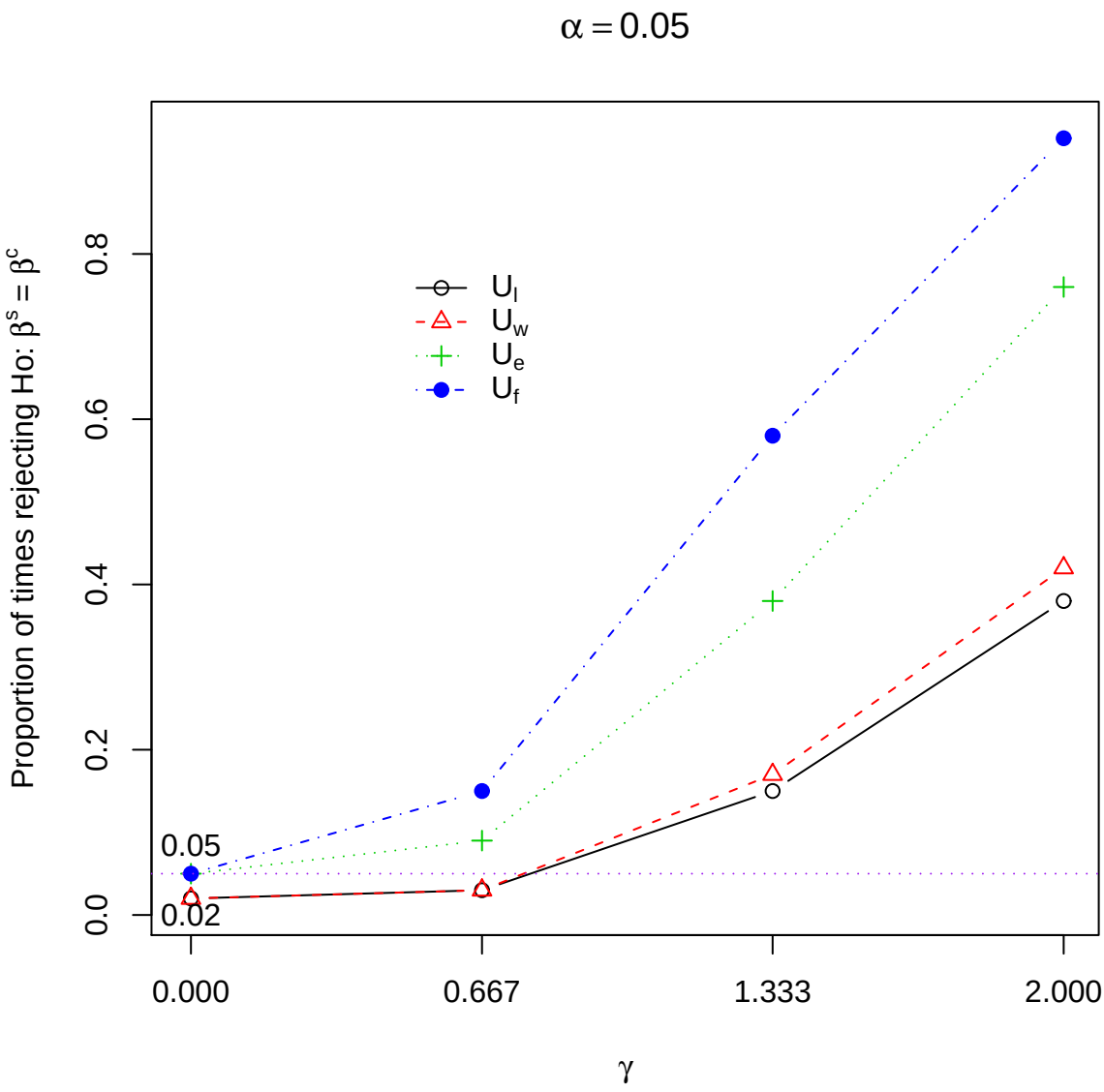

Figure 17: Proportion of times H_o is rejected using level $\alpha = 0.05$, where $H_o: \ \beta^s = \beta^c, \ \text{for all } t \in [-1, 60]$. Four test statistics are considered U_l, U_w, U_e and U_f.

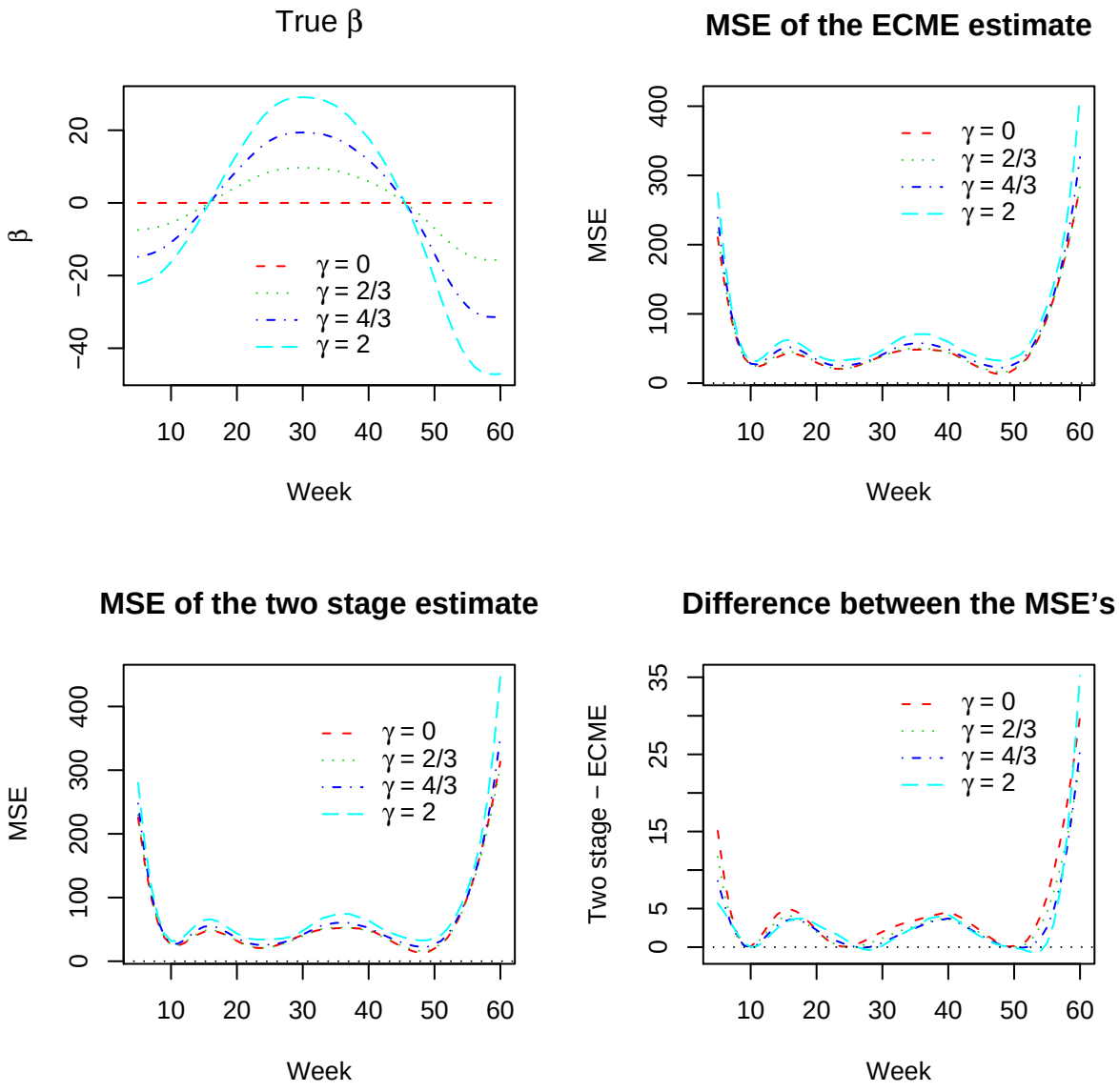

Figure 18: MSE of the estimate of β for each γ value using truncated log body mass. Compare the ECME method with the two stage method.

References

[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 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.