UNIVERSITY OF CALIFORNIA
RIVERSIDE

Semiparametric Mixture Regression for Asynchronous Longitudinal Data Using Multivariate Functional Principal Component Analysis

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Applied Statistics

by

Ruihan Lu

June 2023

Dissertation Committee:

Dr. Yehua Li, Chairperson
Dr. Weixin Yao
Dr. Esra Kurum

The Dissertation of Ruihan Lu is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I am deeply grateful to my esteemed advisor, Dr. Yehua Li, and my invaluable co-advisor, Dr. Weixin Yao. Without their unwavering guidance and support, I would not have reached the point I am at today. Furthermore, I extend my heartfelt thanks to Dr. Esra Kurum for her invaluable professional suggestions, which have greatly contributed to the development of my work.

To my family for all the support.

ABSTRACT OF THE DISSERTATION

Semiparametric Mixture Regression for Asynchronous Longitudinal Data Using Multivariate Functional Principal Component Analysis

by

Ruihan Lu

Doctor of Philosophy, Graduate Program in Applied Statistics
University of California, Riverside, June 2023
Dr. Yehua Li, Chairperson

The transitional phase of menopause induces significant hormonal fluctuations, exerting a profound influence on women's long-term well-being. In the extensive longitudinal investigation of women's health during mid-life and beyond, known as the Study of Women's Health Across the Nation (SWAN), hormonal biomarkers are repeatedly assessed. However, these measurements follow an asynchronous schedule compared to other variables prone to errors, such as physical and cardiovascular measurements. To gain deeper insights into the diverse characteristics within the study population, we conducted a subgroup analysis employing a semiparametric mixture regression model. This approach allows us to explore how the relationship between hormonal responses and other time-varying or time-invariant covariates varies across subgroups. To address the challenges posed by asynchronous scheduling and measurement errors, we propose a novel strategy involving the modeling of time-varying covariate trajectories as functional data. This is achieved through the utilization of reduced rank Karhunen-Loève expansions, where splines are employed to capture the mean and eigenfunctions. Additionally, we introduce an Expectation-Maximization (EM) algo-

rithm to effectively fit a joint model. This model simultaneously incorporates the mixture regression for the hormonal response and the functional principal component (FPC) model for the asynchronous, time-varying covariates. Importantly, we treat the latent subgroup membership and FPC scores as missing data in this framework. Furthermore, we explore data-driven methods to determine the optimal number of subgroups within the population. Through our comprehensive analysis of the SWAN data, we unveil a crucial subgroup structure within the aging female population, shedding light on important distinctions and patterns among women undergoing menopause.

Contents

List of Figures **x**

List of Tables **xii**

1 Motivation and Literature Review **1**
1.1 Motivation 1
1.2 Literature Review 8
1.2.1 Functional principal component analysis 8
1.2.2 The heterogeneous linear fixed/mixed model 11

2 Model Structure and Model Fitting for Emerald with Fixed Effect **16**
2.1 Data structure 16
2.2 Mixture regression model 17
2.3 Functional modeling of the asynchronous longitudinal covariates 18
2.4 Observed likelihood function 21
2.5 EM algorithm 22
2.5.1 E-step 24
2.5.2 M-step 25

3 Model Structure and Model Fitting for Emerald with Random Effect **27**
3.1 Mixture regression 27
3.2 Observed likelihood function 29
3.3 EM algorithm 29
3.3.1 E-step 30
3.3.2 M-step 31
3.4 Stopping criterion 32
3.5 Practical implementations 33

4 Simulation Study **35**
4.1 Simulation setting 36
4.2 Generate the response variable from a linear fixed model 38
4.3 Generate the response variable from a linear mixed model 39

4.4 Label switching 41
4.5 Initial values 41
4.6 Results for fixed effect mixture component 43
4.7 Results for mixed effect mixture components 50
4.8 Results for model selection 56
4.9 Model implementation 57

5 Real Data Analysis **59**
5.1 Pre-analysis 60
5.2 Results for SWAN data 62

6 Conclusions **75**
6.1 Summary 75
6.2 Future work 76
Appendix A: Details on the observed data likelihood function in Chapter 2 . . . 84
Appendix B: Details on the observed data likelihood function in Chapter 3 . . . 87
Appendix C: Technical details of the EM step in Chapter 2 89
Appendix D: Technical details of the EM step in Chapter 3 97
Appendix E: Additional results for real data analysis 106
Appendix F: Additional simulation results with sample size equal to 500 110

List of Figures

1.1 Visit time for 25 randomly selected subjects, where each row corresponds to one randomly selected patient labeled by her ID, and the points mark the visiting times for various types of measurements. 4
1.2 DHEAS, glucose, systolic blood pressure, and triglycerides trajectories for a randomly selected subject. 5
1.3 Total number of records in hormone measurement 7
1.4 Total number of records in physical measurement 8
1.5 Total number of records in cardiovascular measurement 8

4.1 (a) True mean curves for $X_1(t)$; (b) True mean curve for $X_2(t)$ 37
4.2 (a) Two eigenfunctions for $\boldsymbol{X}_1$; (b) Two eigenfunctions for $\boldsymbol{X}_2$. In both graphs, the solid black line plots the first eigenfunction and the dashed blue line plots the second eigenfunction. 37
4.3 Y value labelled by subgroup 39
4.4 Y value labelled by subgroup 40
4.5 Fitted mean function for time-varying covariates $\boldsymbol{X}_1$ via 200 simulation runs with fixed effect mixture components in **Emerald** and **fdapace**. The solid red line plots the true function and dashed blue line plots out the average of the fitted function and the light gray lines are 200 fitted lines. 45
4.6 Fitted mean function for time-varying covariates $\boldsymbol{X}_2$ via 200 simulation runs with fixed effect mixture components in **Emerald** and **fdapace**. The solid red line plots the true function and dashed blue line plots out the average of the fitted function and the light gray lines are 200 fitted lines. 46
4.7 Fitted eigenfunction for time-varying covariates $\boldsymbol{X}_1$ via 200 simulation runs with fixed effect mixture components in **Emerald** and **fdapace**. The solid red line plots the true function and dashed blue line plots out the average of the fitted function and the light gray lines are 200 fitted lines. 47
4.8 Fitted eigenfunction for time-varying covariates $\boldsymbol{X}_2$ via 200 simulation runs with fixed effect mixture components in **Emerald** and **fdapace**. The solid red line plots the true function and dashed blue line plots out the average of the fitted function and the light gray lines are 200 fitted lines. 48

5.1 Density plot of original DHEAS 61
5.2 Boxplot of DHEAS colored by subgroups 63
5.3 Boxplot of DHEAS colored by subgroups via 8 visits 63
5.4 Partial effect plots for glucose versus DHEAS. The red line plots our fitted coefficients. 69
5.5 Partial effect plots for triglycerides versus DHEAS. The red line plots our fitted coefficients. 70
5.6 Partial effect plots for systolic blodd pressure versus DHEAS. The red line plots our fitted coefficients. 70
5.7 SWAN data: fitted mean for glucose, triglycerides, and systolic BP. Plots show the fitted mean functions (solid red), pointwise 95% bootstrap confidence intervals (dashed blue), as well as the scatter plot of the observations. 71
5.8 SWAN data: fitted eigenfunction for glucose, triglycerides, and systolic BP. Plots show the top two eigenfunctions as solid black and dashed blue, respectively. 72

A.1 SWAN data: comparison of the fitted and eigenfunction for glucose. The mean plot shows the fitted mean functions (solid red), pointwise 95% bootstrap confidence intervals (dotted brown) as well as the fitted mean curve from **fdapace** (dashed blue). The eigenfunction plot shows the fitted eigenfunctions from **Emerald** (solid red) and from **fdapace** (dashed blue). . . . 107
A.2 SWAN data: comparison of the fitted mean and eigenfunction for triglycerides. The mean plot shows the fitted eigenfunctions from **Emerald** (solid red) and from **fdapace** (dashed blue). The eigenfunction plot shows the fitted eigenfunctions from **Emerald** (solid red) and from **fdapace** (dashed blue). 108
A.3 SWAN data: comparison of the fitted mean and eigenfunction for systolic BP. The mean plot shows the fitted eigenfunctions from **Emerald** (solid red) and from **fdapace** (dashed blue). The eigenfunction plot shows the fitted eigenfunctions from **Emerald** (solid red) and from **fdapace** (dashed blue). 109
A.4 Fitted mean for time-varying covariates $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ in 3-subgroup simulation study with $n = 500$. The solid red line plots the true function and dashed blue line plots out the average of the fitted function and the light gray lines are 200 fitted lines. 112
A.5 Fitted eigenfunction for time-varying covariates $\boldsymbol{X}_2$ in 3-subgroup simulation study with $n = 500$. The solid red line plots the true function and dashed blue line plots out the average of the fitted function and the light gray lines are 200 fitted lines. 113

List of Tables

1.1 Observed date for physical, hormonal, cardiovascular measurements for two randomly selected subjects 6

4.1 Summary of FPCA estimators in the simulation study with fixed effect mixture components 43
4.2 Bias (SD) for the logistic regression parameters for the subgroup labels with fixed effect mixture components. 49
4.3 Bias (SD) for the mixture regression coefficients with fixed effect mixture components. 49
4.4 Bias (SD) for the variance components and Mean (SD) for ARI in the 3-subgroup simulation study with fixed effect mixture components 50
4.5 Summary of FPCA estimators in the simulation study with mixed effect mixture components 52
4.6 Bias (SD) for the logistic regression parameters for the subgroup labels . . . 54
4.7 Bias (SD) for the mixture regression coefficients 55
4.8 Bias(SD) for the variance parameters in the 3-subgroup simulation study . 55

5.1 AWE and BIC value for SWAN data 62
5.2 Subjects in Each subgroup by Race 64
5.3 Percentile of Time-invariant Variable 65
5.4 Estimate (SD) of the coefficients for logistic regression: $\widehat{\boldsymbol{\gamma}}$ in Equation (2.1) 66
5.5 Estimate (SD) of the coefficients for mixture regression: $\widehat{\boldsymbol{\beta}}$ in Equation (3.1) 68

A.1 Summary of FPCA estimators in the simulation study with $n = 500$ 111
A.2 Bias (SD) for the logistic regression parameters for the subgroup labels in the 3-subgroup simulation study with $n = 500$ 112
A.3 Bias (SD) for the mixture regression coefficients in the 3-subgroup simulation study with $n = 500$ 114
A.4 Bias(SD) for the variance parameters in the 3-subgroup simulation study with $n = 500$ 114

Chapter 1

Motivation and Literature Review

1.1 Motivation

Dehydroepiandrosterone sulfate (DHEAS) is a steroid hormone that occurs naturally in the body and is primarily synthesized by the adrenal glands, gonads, and the brain. It plays a crucial role in the production of testosterone, the male sex hormone, and estrogen, the female sex hormone. The concentration of DHEAS in the bloodstream varies throughout life and is influenced by factors such as age, sex, and overall health [17, 43]. Extensive research has established significant connections between DHEAS deficiency and various degenerative disorders of the nervous system. For example, individuals with Addison's disease, a condition characterized by adrenal insufficiency, often exhibit low levels of DHEAS [18]. Additionally, studies have investigated the link between DHEAS and neurodegenerative diseases such as Alzheimer's disease, where lower levels of DHEAS have been observed in affected individuals [38]. Moreover, imbalances in DHEAS have been associated with mental health conditions, including depression and schizophrenia [53, 49]. Increasing DHEAS levels

has been shown to improve mood states in individuals with clinical depression and enhance psychological well-being among the elderly [56, 35]. Furthermore, DHEAS supplementation can help restore hormone levels in patients with adrenal deficiencies. Conversely, abnormally high levels of DHEAS can be indicative of certain inherited adrenal gland disorders [25]. Understanding the role of DHEAS and its implications in various physiological and psychological processes is essential for advancing our knowledge of hormone-related disorders and potentially developing targeted interventions for their management and treatment.

Our research focuses on investigating the levels of DHEAS in women during the transitional period of menopause using data obtained from the Study of Women's Health Across the Nation (SWAN). SWAN is a large-scale longitudinal study funded by the National Institutes of Health (NIH) that aims to comprehensively examine various aspects of women's health during the menopausal transition.

The SWAN study enrolled 3,302 premenopausal or early perimenopausal women between the ages of 42 and 53. These participants were recruited from seven different cities across the United States, including Boston, Pittsburgh, Oakland, Los Angeles, Detroit, Newark, and Chicago. The geographic diversity of the study population allows for capturing the experiences and characteristics of women undergoing the menopausal transition in different regions. Upon enrollment, the participants underwent baseline measurements on a designated day referred to as day 0. Subsequently, they were followed up annually for a period of 10 years until 2006. However, it is important to note that the timing of follow-up visits varied randomly for each individual and was subject-specific. This variation in follow-up intervals was influenced by factors such as participant availability, missed appointments,

and random dropouts. To account for these variations and ensure an accurate representation of the longitudinal nature of the study, the recorded follow-up times are expressed as the number of days since enrollment rather than following a fixed calendar-based schedule.

In our analysis, we specifically aim to explore the relationship between DHEAS levels and various factors, including cardiovascular biomarkers, physical health indicators, and demographic backgrounds. By utilizing the extensive SWAN dataset, we seek to contribute to the existing knowledge on menopause and gain valuable insights into the hormonal dynamics and associated health implications during this critical transitional phase in women's lives.

One challenge in analyzing the SWAN data is the asynchronous design, where the response variable (DHEAS levels) and time-varying covariates are measured at varying times for different subjects. This asynchrony is illustrated in Figure 1.1, which displays the observation times of physical, hormonal, and cardiovascular measurements for a random sample of 25 subjects. As shown, the measurement times differ not only between different variables but also across subjects for the same variable. Managing and accounting for this asynchrony is essential in the data analysis process and requires appropriate statistical methods to handle the variability in measurement times effectively. Figure 1.2 showcases spaghetti plots that effectively illustrate the longitudinal trajectories of various variables for a randomly selected subject. The variables depicted include DHEAS (hormonal), glucose, triglycerides (cardiovascular), and systolic blood pressure (physical). These plots provide a visual representation of how these variables evolve over time for the chosen individual and provide a evidence that the measurements are also variable-specific. It is important to

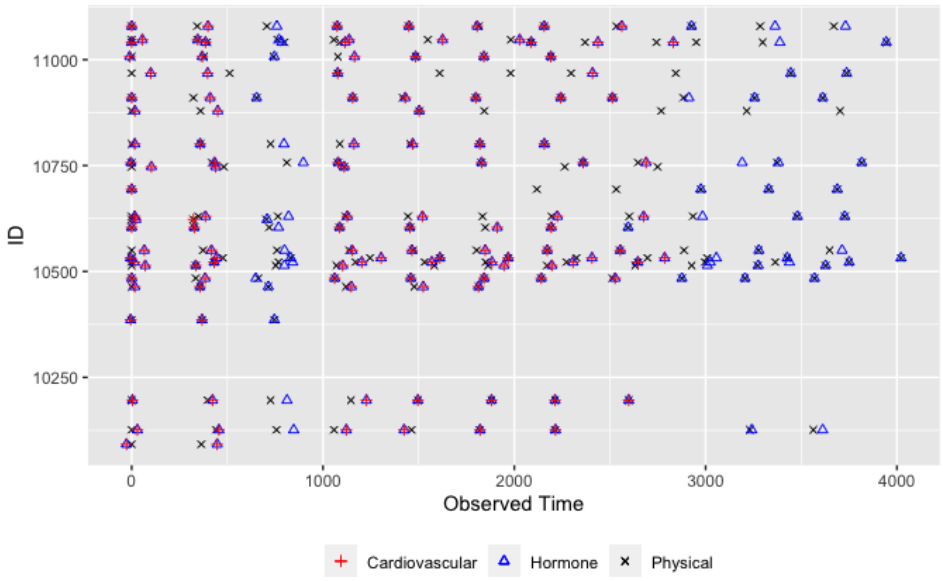


Figure 1.1: Visit time for 25 randomly selected subjects, where each row corresponds to one randomly selected patient labeled by her ID, and the points mark the visiting times for various types of measurements.

note that the longitudinal trajectories exhibit local fluctuations, which can be indicative of measurement errors. These fluctuations highlight the inherent variability and noise present in the data, emphasizing the need to account for such errors when conducting statistical analysis and drawing conclusions from the study. [8]. Additional evidence supporting the asynchronous design can be found in Table 1.1. The table presents a portion of the data for a randomly selected subject, displaying the visit times for DHEAS (hormonal), glucose, triglycerides (cardiovascular), and systolic blood pressure (physical). This information further underscores the asynchronous nature of the study, as the measurement times for these variables are not synchronized across visits. The table serves as a valuable resource for analyzing and understanding the temporal patterns and potential variations in these measurements for the selected subject. Both Figure 1.1 and Figure 1.2, along with Table 1.1,

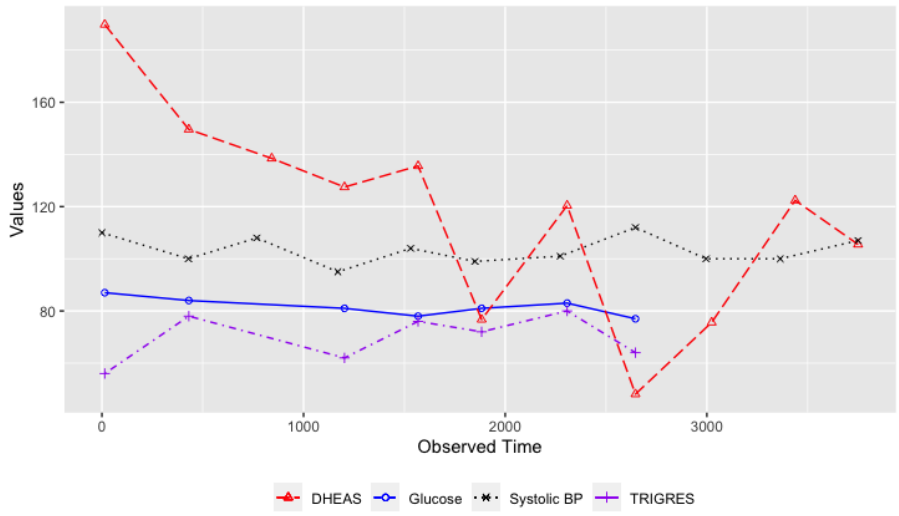


Figure 1.2: DHEAS, glucose, systolic blood pressure, and triglycerides trajectories for a randomly selected subject.

provide clear visual and tabular evidence of the asynchronous nature of physical and cardiovascular measurements in comparison to DHEAS measurements.

Furthermore, it is notable that DHEAS, being a hormonal measurement, has a median of 9 repeated measurements. In contrast, physical measurements have a median of 10 repeated measures, and cardiovascular measurements have a median of 6 repeated measures. This discrepancy in the number of repeated measures indicates variations in the frequency of data collection for these different types of measurements. To provide a comprehensive overview of the distribution and frequency of visits for physical, hormonal, and cardiovascular measurements within the SWAN study, we present Figures 1.3, 1.4, and 1.5. These figures offer valuable insights into the total number of observations and the variation in visit frequencies across different types of measurements. Figure 1.3 displays the distribution of visits for physical measurements. It shows the number of observations made at each visit throughout the study duration. From the figure, we can observe the varying

Table 1.1: Observed date for physical, hormonal, cardiovascular measurements for two randomly selected subjects

ID	Physical	Hormonal	Cardiovascular
10046	4	0	0
10046	420	433	433
10046	811	897	NA
10046	1077	1079	1079
10046	1509	NA	NA
10046	1827	1929	1929
10046	2359	2360	2360
10046	2647	2689	2689
10046	NA	3191	NA
10046	3374	3387	NA
10046	3816	3861	NA
10464	0	16	16
10464	357	359	359
10464	714	714	NA
10464	1120	1148	1148
10464	1477	1524	1524

visit frequencies for physical measurements, with some visits having a higher number of observations compared to others. This variation is primarily due to the asynchronous design of the study and the subject-specific follow-up intervals. Figure 1.4 presents the distribution of visits for hormonal measurements, including DHEAS levels. Similar to Figure 1.3, it illustrates the number of observations made at each visit for hormonal measurements. From the figure, we can observe that the visit frequencies for hormonal measurements differ from those of physical measurements. This discrepancy can be attributed to the specific

measurement protocols and the nature of the variables being captured. Finally, Figure 1.5 focuses on the distribution of visits for cardiovascular measurements. It provides insights into the number of observations obtained at each visit for cardiovascular-related variables. As with the previous figures, we can observe variability in visit frequencies for cardiovascular measurements, reflecting the asynchronous and subject-specific nature of data collection.

It is important to note that the total number of observations for physical and cardiovascular measurements may differ from that of DHEAS measurements. This difference arises due to the varying availability and timing of data collection for different variables within the SWAN study. Analyzing and accounting for these variations in the number of observations is crucial in conducting a comprehensive and accurate analysis of the relationships between DHEAS levels and other factors within the dataset.

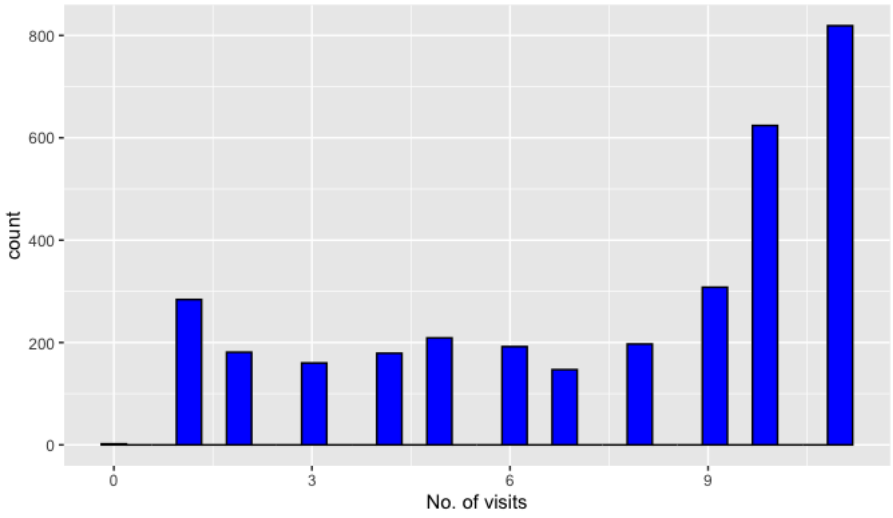


Figure 1.3: Total number of records in hormone measurement

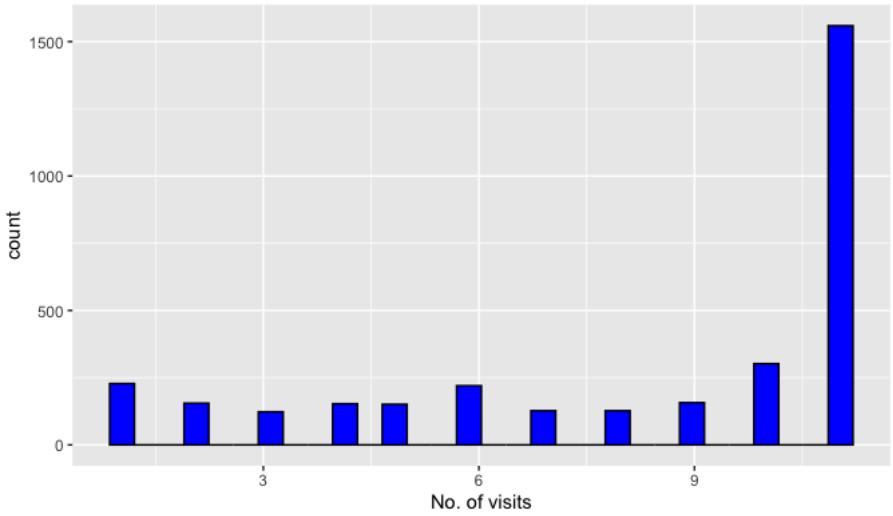


Figure 1.4: Total number of records in physical measurement

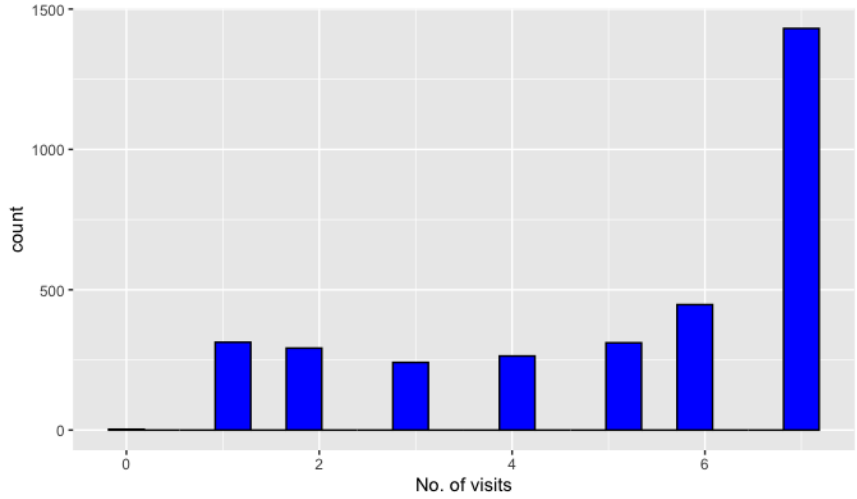


Figure 1.5: Total number of records in cardiovascular measurement

1.2 Literature Review

1.2.1 Functional principal component analysis

The current state-of-the-art method for effectively modeling asynchronous and error-prone longitudinal data, such as the physical and cardiovascular trajectories observed

in the SWAN dataset, involves treating them as sparse functional data [42, 61, 31]. In functional data analysis, each observation is represented as a curve rather than a single point or finite-dimensional vector. Among the techniques used in functional data analysis, functional principal component analysis (FPCA) has emerged as a widely utilized approach. FPCA allows for the exploration of the dominant modes of variation present in functional data and facilitates the transformation of these inherently infinite-dimensional curves into a finite-dimensional vector of random scores. This approach relies on the Karhunen-Loève decomposition and operates under mild assumptions. By employing FPCA, the underlying stochastic process is represented as a sequence of uncorrelated random variables, referred to as functional principal components (FPCs) or scores. These FPCs can be truncated to a finite vector, making them practical and applicable in real-world scenarios. While the concept of principal components analysis (PCA) was independently introduced by Pearson in 1901 and Hotelling in 1933, the application of PCA to functional data was first attempted by Rao and Tucker in 1958. They utilized multivariate PCA without modifying the observed function values [45]. By leveraging FPCA, researchers gain a powerful tool for effectively modeling longitudinal data collected at various time points. FPCA enables the identification of dominant modes of variation within functional data and facilitates their representation as finite-dimensional vectors of random scores. Specifically, let X be a random process defined on an interval $\mathcal{T}$ with the properties that $\mathrm{E}(\|X\|)^2_{L_2} < \infty$ and $\mathrm{E}(X) = 0$. The primary objective of functional principal component analysis (FPCA) is to decompose the latent stochastic process into a linear combination of functional principal components (FPCs) that maximize the variation observed in the curves. By utilizing FPCs, which represent

the dominant modes of variation in the data, it becomes possible to effectively approximate the infinite-dimensional longitudinal trajectories. In practice, the number of FPCs is often truncated to a finite vector, as the top few FPCs capture the majority of the variability in the underlying stochastic process. Therefore, $X(t)$ can be represented as follows:

$$X(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_k \phi_k(t), \tag{1.1}$$

where $\mu(t)$ is the mean function of $X(t)$, ξ_k are centered uncorrelated random variables, $\phi_k(\cdot)$ are normalized eigenfunctions of the covariance function $\mathrm{Cov}(s,t) = \mathrm{Cov}\{X(s), X(t)\}$, and $\mathrm{Var}(\xi_k)$, often denoted by λ_k, is the eigenvalue of covariance corresponding to ϕ_k. It is also conventionally assumed that $\lambda_1 > \lambda_2 > \ldots > 0$. The eigenfunctions $\phi_1(\cdot), \phi_2(\cdot), \ldots$, referred to as principal components in the context of FPCA, form an orthonormal constraint[6]. Effective dimensional reduction is achieved by choosing the top few FPCs that cumulatively explain a large proportion of the variation.

In summary, to model the true underlying functional trajectories, a reduced-rank Karhunen-Loève expansion is employed, treating them as realizations of a stochastic process [21]. The observed data consists of discrete observations on these random functions, which are subject to errors and are recorded at subject-specific time points. Functional data analysis approaches, in contrast to classical linear mixed models, offer nonparametric and data-driven advantages. In the context of modeling longitudinal or sparse functional data, functional principal component analysis (FPCA) has been widely adopted, with the unknown eigenfunctions represented as splines [26]. In the field of longitudinal data analysis, [32] have proposed data-driven methods for selecting the number of functional principal

components (FPCs). These methods have also been extended to handle the joint modeling of multiple longitudinal/functional trajectories [66, 22]. Theoretical properties of multivariate functional principal component analysis (mFPCA) have been investigated by [11, 19, 58]. However, it should be noted that these existing methods are limited in their ability to handle heterogeneous data, which restricts their applicability in certain scenarios.

1.2.2 The heterogeneous linear fixed/mixed model

In addition to the presence of error-contaminated and asynchronous time-varying covariates, conflicting findings exist in the literature regarding the relationship between DHEAS and certain variables, such as triglycerides and systolic blood pressure. For instance, [44] found significant relationships between DHEAS and systolic blood pressure levels. However, contrasting results were reported by [55], who suggested that DHEAS had a neutral effect on blood pressure (BP), indicating no significant association. Considering the discrepancies in these findings, we believe that the lack of significant relationships observed in some studies could potentially be attributed to the absence of subgroup analysis. It is possible that one subgroup may demonstrate a positive association between DHEAS and systolic blood pressure, while another subgroup may exhibit a negative association. Failing to investigate the subgroup effect could result in overlooking important information and potentially masking meaningful associations between DHEAS and systolic blood pressure.

In the context of a large longitudinal dataset with complex structures like SWAN, it is essential to investigate the presence of subgroups within the study population that exhibits distinct dynamics of DHEAS during the menopause transition. To achieve this,

mixture regression models play a critical role and have been widely used in diverse fields, including business, biology, social sciences, and medicine [20, 34, 46, 37]. These models offer a valuable tool for exploring heterogeneity in the data. In recent years, significant efforts have been made to extend mixture regression models to longitudinal data settings. In this article, we propose two models: one is a finite mixture model with fixed effects, and the other is a finite mixture model with both random and fixed effects. These models enable the examination of subgroup-specific dynamics in the context of longitudinal data, allowing for a more comprehensive understanding of the heterogeneity in the study population during the menopause transition.

let $Y_i = (Y_{i1,\ldots,Y_{in_i}})$ be the response vector for the n_i measurements of the subject i with $i = 1, \ldots, N$. The linear fixed effect model for the response vector Y_i can be defined as

$$Y_i = X_i\beta + \epsilon_i, \tag{1.2}$$

and a linear mixed model for the response vector Y_i, which extends the fixed effect model with random effects, can be defined as

$$Y_i = X_i\beta + Z_i u_i + \epsilon_i, \tag{1.3}$$

where X_i is a $n_i \times p$ design matrix for the $p-$vector of fixed effects β, and Z_i is a $n_i \times q$ design matrix associated to the $q-$vector of random effects u_i which represents the subject-specific regression coefficients. The errors ϵ_i are assumed to be normally distributed with mean zero

and covariance matrix $\sigma \boldsymbol{I}_{n_i}$, and are assumed to be independent of the vector of random effects u_i. The standard definition of a C-component mixture (or C latent classes) model is

$$\sum_{c=1}^{C} \pi_c f(Y_i | X_i, \beta, u_i), \tag{1.4}$$

where $0 < \pi_c < 1$ and $\sum_{c=1}^{C} \pi_c = 1$. In our discussion, we extend the probability of class membership π_c to be subject-specific, which is π_{ic} and will be estimated by logistic regression. More mixture models for longitudinal data examples, can be found in [51], [23], [40], and [24]. However, these methods require the time-varying covariates to be synchronized in time with the response measurements, and they cannot accommodate measurement errors in the covariate processes.

To our knowledge, there is currently no existing method that simultaneously addresses the three main challenges present in the SWAN data:

1. *asynchronous* design between the longitudinal response variable and time-varying covariates;

2. the presence of *measurement errors* in the time-varying covariates;

3. the existence of a *heterogeneous* population with latent subgroup structures.

In order to fill this gap, we propose a novel method called **Emerald** (sEmiparametric MixturE Regression for Asynchronous Longitudinal Data). **Emerald** combines mixture regression with multivariate functional principal component analysis (mFPCA) to handle the complex nature of the data. It accommodates the heterogeneity of asynchronous and error-contaminated longitudinal covariate processes. The proposed method models the re-

lationship between DHEAS levels and other time-varying (physical and cardiovascular) as well as time-invariant (demographic and other baseline traits) covariates using a mixture of linear mixed models. This approach allows for different associations between the variables across subgroups. The latent subgroup membership is modeled using a multi-categorical logistic regression. To capture the unknown eigenfunctions of the time-varying covariate processes, we employ B-splines. The latent functional principal component (FPC) scores and the latent subgroup membership are treated as missing data. We propose utilizing an EM algorithm to fit the proposed model and estimate the parameters. Through simulation studies, we demonstrate the advantages of the proposed method over existing approaches, such as the **lcmm** method introduced by [40]. **Emerald** provides a comprehensive solution to address the challenges posed by the SWAN data, offering improved accuracy and performance in modeling the dynamics of DHEAS during the menopause transition.

The remainder of the paper is organized as follows. In Chapter 2, we introduce the data structure and model assumptions, including the proposed mixture regression model with a fixed effect for the DHEAS response and the multivariate functional data model for asynchronous longitudinal covariate processes. In addition, Chapter 2 outlines the EM algorithm used to fit the semiparametric mixture regression linear fixed model. Chapter 3 outlines the data structure and model assumptions, including the proposed mixture regression model with a random effect for the DHEAS response, the EM algorithm used to fit the semiparametric mixture regression linear mixed model and discusses practical considerations such as stopping rules and model selection methods. We present simulation studies in Chapter 4 to illustrate the numerical performance of the proposed methods, followed

by the analysis of the SWAN data in Chapter 5. Chapter 6 provides concluding remarks. Technical details on the algorithm used to fit the proposed method can be found in the Appendix, and additional numerical results are included in the supporting materials.

Chapter 2

Model Structure and Model Fitting for Emerald with Fixed Effect

2.1 Data structure

Let $\{Y_i(t), \boldsymbol{X}_i(t), \boldsymbol{Z}_i(t)\}$, $i = 1, \ldots, n$, be independent realizations of a multivariate longitudinal process $\{Y(t), \boldsymbol{X}(t), \boldsymbol{Z}(t)\}$ defined in a time interval $\mathcal{T}$. In the SWAN data, $Y_i(t)$ is the DHEAS level of the ith subject at time t; $\boldsymbol{X}_i(t) = (X_{i1}, \ldots, X_{id_x})^\top(t)$ is a d_x-dimensional time-varying covariates that are *asynchronous* with $Y_i(t)$, such as the physical and cardiovascular measurements; and $\boldsymbol{Z}_i(t)$ is a d_z-dimensional vector that includes either time-invariant covariates or time-varying covariates observed at the same time as $Y_i(t)$ and without error. Let $Y_{ij} = Y_i(t_{ij})$ and $\boldsymbol{Z}_{ij} = \boldsymbol{Z}_i(t_{ij})$, $j = 1, \ldots, m_{y,i}$, be the discrete observations on $Y_i(t)$ and $\boldsymbol{Z}_i(t)$. The observed, error-contaminated measurements on $X_{iv}(t)$ are $W_{ivl} = X_{iv}(s_{ivl}) + U_{ivl}$, $l = 1, \ldots, m_{x,iv}$, where U_{ivl} are independent zero-mean, Gaussian

errors with variance σ_v^2 for $v = 1, 2, \ldots, d_x$. The observation times $\{s_{ivl},\ l = 1, \ldots, m_{x,iv}\}$ of W are irregular and subject-specific, may differ across different index v, and are different from the observation times $\{t_{ij}, j = 1, \ldots, m_{y,i}\}$ of Y and $\boldsymbol{Z}$. Put $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{im_{y,i}})^\top$, $\boldsymbol{Z}_i = (\boldsymbol{Z}_{i1}, \ldots, \boldsymbol{Z}_{im_{y,i}})^\top$, $\boldsymbol{W}_{iv} = (W_{iv1}, \ldots, W_{ivm_{x,iv}})^\top$, and let $\boldsymbol{W}_i = (\boldsymbol{W}_{i1}^\top, \ldots, \boldsymbol{W}_{id_x}^\top)^\top$ be a vector of dimension $m_{x,i} = \sum_{v=1}^{d_x} m_{x,iv}$ pooling all W values in subject i together.

2.2 Mixture regression model

We assume the study population has C latent subgroups and will discuss its selection in Chapter 3.5. We define $\boldsymbol{L}_i = (L_{i1}, L_{i2}, \ldots, L_{iC})^\top$ to be a latent vector of subgroup membership indicators for the ith subject, such that $L_{ic} = 1$ if the ith subject belongs to the cth subgroup and 0 otherwise, $c = 1, \ldots, C$. And we assume that the ith subject's subgroup membership depends on a vector of time-independent covariates $\mathbb{Z}_i$, a sub-vector of $\boldsymbol{Z}_i$, including demographic information and baseline measurements. Given $\mathbb{Z}_i$, we assume $\boldsymbol{L}_i \sim$ Multinomial$(\boldsymbol{\pi}_i)$, where $\boldsymbol{\pi}_i = (\pi_{i1}, \ldots, \pi_{iC})^\top$ is modeled by a logistic model

$$\pi_{ic}(\mathbb{Z}_i) = \frac{\exp(\widetilde{\mathbb{Z}}_i^\top \boldsymbol{\gamma}_c)}{1 + \sum_{c'=1}^{C-1} \exp(\widetilde{\mathbb{Z}}_i^\top \boldsymbol{\gamma}_{c'})}, \quad \text{for } c = 1, 2, \ldots, C-1, \tag{2.1}$$

and $\pi_{iC} = 1 - \sum_{c=1}^{C-1} \pi_{ic}$. Here, for simplicity, we write $\pi_{iC}(\mathbb{Z}_i)$ as π_{iC}, $\widetilde{\mathbb{Z}}_i = (1, \mathbb{Z}_i^\top)^\top$, and $\boldsymbol{\gamma}_c = (\gamma_{0c}, \boldsymbol{\gamma}_{z,c}^\top)^\top$.

In our analysis, we assume that the variation in the response variable $\boldsymbol{Y}$ can be adequately explained by the measurement error. Given $L_{ic} = 1$, the response can be

modeled by a linear fixed effect model

$$Y_i(t) = \beta_{0,c} + \boldsymbol{X}_i(t)^\top \boldsymbol{\beta}_{x,c} + \boldsymbol{Z}_i(t)^\top \boldsymbol{\beta}_{z,c} + \epsilon_i(t), \tag{2.2}$$

where $\boldsymbol{\beta}_c = (\beta_{0,c}, \boldsymbol{\beta}_{x,c}^\top, \boldsymbol{\beta}_{z,c}^\top)^\top$ is a subgroup-specific coefficient vector and $\epsilon_i(t)$ is a zero-mean Gaussian error with variance σ_ϵ^2.

2.3 Functional modeling of the asynchronous longitudinal covariates

We model the process $\boldsymbol{X}_i$(t) as multivariate functional data [11, 19] with mean function

$$\boldsymbol{\mu}_x(t) = \mathrm{E}\{\boldsymbol{X}(t)\} = (\mu_{x,1}, \ldots, \mu_{x,d_x})^\top(t)$$

and covariance function as

$$\boldsymbol{R}(t_1, t_2) = \mathrm{E}[\{\boldsymbol{X}(t_1) - \boldsymbol{\mu}_x(t_1)\}\{\boldsymbol{X}(t_2) - \boldsymbol{\mu}_x(t_2)\}^\top] = \{R_{v,v'}(t_1, t_2)\}_{v,v'=1}^{d_x},$$

which is a matrix of covariance and cross-covariance functions between different components in $\boldsymbol{X}$.

Suppose the covariance of X_v has the spectrum decomposition $R_{v,v}(t_1, t_2) = \sum_{k=1}^{p_v} \omega_{vk}\psi_{vk}(t_1)\psi_{vk}(t_2)$, where $\omega_{v1} \geq \omega_{v2} \geq \ldots \geq \omega_{vp_v} > 0$ are the eigenvalues, the eigenfunctions $\boldsymbol{\psi}_v(t) = (\psi_{v1}, \psi_{v2}, \ldots, \psi_{vp_v})^\top(t)$ are a set of orthonormal functions with

$\int_{\mathcal{T}} \boldsymbol{\psi}_v(t)\boldsymbol{\psi}_v^\top(t)dt = I$. Theoretically, the number of principal components p_v can be infinite but is often assumed to be finite for sparse functional data [66, 32]. By the standard Karhunen-Loéve (KL) expansion [21],

$$X_{iv}(t) = \mu_{x,v}(t) + \sum_{k=1}^{p_v} \psi_{vk}(t)\xi_{ivk}, \quad \text{for } v = 1, 2, \ldots, d_x, \tag{2.3}$$

where the principal component scores $\xi_{ivk} = \int_{\mathcal{T}}\{X_{iv}(t) - \mu_{x,v}(t)\}\psi_{vk}(t)dt$ are zero-mean latent variables such that $\mathrm{Cov}(\xi_{ivk}, \xi_{ivk'}) = \omega_{vk}I(k = k')$.

The cross-covariance between different components of $\boldsymbol{X}$ can also be represented by the eigenfunctions, $R_{v,v'}(t_1, t_2) = \mathrm{Cov}\{X_v(t_1), X_{v'}(t_2)\} = \sum_{k=1}^{p_v}\sum_{k'=1}^{p_{v'}} \omega_{vv',kk'}\psi_{vk}(t_1)\psi_{v'k'}(t_2)$, for $v \neq v'$, where $\omega_{vv',kk'} = \mathrm{Cov}(\xi_{ivk}, \xi_{iv'k'})$. We collect the FPC scores of the ith subject into a $p = \sum_{v=1}^{d_x} p_v$ dimensional vector $\boldsymbol{\xi}_i = \left(\boldsymbol{\xi}_{i1}^\top, \boldsymbol{\xi}_{i2}^\top, \ldots, \boldsymbol{\xi}_{id_x}^\top\right)^\top$, where $\boldsymbol{\xi}_{iv} = (\xi_{iv1}, \xi_{iv2}, \ldots, \xi_{ivp_v})^\top$. We assume

$$\boldsymbol{\xi}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\xi), \tag{2.4}$$

where $\boldsymbol{\Sigma}_\xi = (\boldsymbol{\Sigma}_{\xi,vv'})_{v,v'=1}^{d_x}$ with $\boldsymbol{\Sigma}_{\xi,vv} = \mathrm{diag}(\omega_{v1}, \ldots, \omega_{vp_v})$, $v = 1, \ldots, d_x$, and $\boldsymbol{\Sigma}_{\xi,vv'} = \left(\omega_{vv',kk'}\right)_{k,k'=1}^{p_v,p_{v'}}$ for $v \neq v'$.

Next, we model the unknown functions as splines in (2.3). To accommodate the orthonormal constraints on the eigenfunctions, we will use orthogonalized B-spline basis and the procedure can be summarized below:

To orthonormalize the spline basis, we can choose $\mathbb{B}(t) = (\mathbb{B}_1(t), \ldots, \mathbb{B}_q(t))^\top$ be our initial spline basis, and it is not necessary to be an orthonormal basis such as the B-spline basis. We suppose that there exists a matrix $\boldsymbol{D}$ such that $\boldsymbol{B}(t) = \mathbb{B}(t)\boldsymbol{D}$ and

this matrix $\boldsymbol{D}$ could be obtained as follows. Taking eigenvalue decomposition of $\frac{1}{g}\mathbb{B}^\top\mathbb{B}$, where we write $\mathbb{B} = \{\mathbb{B}(t_1), \ldots, \mathbb{B}(t_g)\}^\top$. Here we take t_g discrete observed points. Then we can have $\boldsymbol{R}$ to be the square matrix whose kth column is the eigenvector R_k of our matrix and $\boldsymbol{Q}$ be to the diagonal matrix whose diagonal elements are the corresponding eigenvalues. Let $\boldsymbol{D} = \boldsymbol{R}\boldsymbol{Q}^{-\frac{1}{2}}$, then we can get $\boldsymbol{B}^\top\boldsymbol{B} = \boldsymbol{D}^\top\mathbb{B}^\top\mathbb{B}\boldsymbol{D} = \boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{R}^\top\mathbb{B}^\top\mathbb{B}\boldsymbol{R}\boldsymbol{Q}^{-\frac{1}{2}} = \boldsymbol{Q}^{-\frac{1}{2}}\boldsymbol{R}^\top\boldsymbol{R}\boldsymbol{Q}\boldsymbol{R}^{-1}\boldsymbol{R}\boldsymbol{Q}^{-\frac{1}{2}} = \boldsymbol{I}$.

We then allow different components of $\boldsymbol{X}$ to have different basis functions, possibly defined on different sets of knots. Let $\boldsymbol{B}_v(t) = \{B_{v1}(t), \ldots, B_{vq_v}(t)\}^\top$ be the spline basis function for modeling $\mu_{x,v}(t)$ and $\psi_{v,k}(t)$'s, and

$$\mu_{x,v}(t) = \boldsymbol{B}_v^\top(t)\boldsymbol{\theta}_{\mu v}, \qquad \psi_{vk}(t) = \boldsymbol{B}_v^\top(t)\boldsymbol{\theta}_{\psi v,k}, \tag{2.5}$$

where $\boldsymbol{\theta}_{\mu,v}$ and $\boldsymbol{\theta}_{\psi v,k}$ are q_v-dim, unknown coefficient vectors. Let $\boldsymbol{\Theta}_{\psi v} = (\boldsymbol{\theta}_{\psi v,1}, \ldots, \boldsymbol{\theta}_{\psi v,p_v})$ be the $q_v \times p_v$ coefficient matrix for $\boldsymbol{\psi}_v(t)$. Assuming that $\boldsymbol{B}_v$ has been orthogonalized such that $\int_{\mathcal{T}} \boldsymbol{B}_v(t)\boldsymbol{B}_v^\top(t)dt = I$, then the orthonormal constraint on $\boldsymbol{\psi}_v(t)$ can be translated to

$$\boldsymbol{\Theta}_{\psi v}^\top\boldsymbol{\Theta}_{\psi v} = I, \quad v = 1, \ldots, d_x. \tag{2.6}$$

Let $\boldsymbol{B}_{iv}^* = \{\boldsymbol{B}_v(s_{iv1}), \boldsymbol{B}_v(s_{iv2}), \ldots, \boldsymbol{B}_v(s_{ivm_{x,iv}})\}^\top$, $\boldsymbol{B}_{iv} = \{\boldsymbol{B}_v(t_{i1}), \boldsymbol{B}_v(t_{i2}), \ldots, \boldsymbol{B}_v(t_{im_{y,i}})\}^\top$ be $m_{x,iv} \times q_v$ and $m_{y,i} \times q_v$ matrices, respectively, which evaluating the spline basis functions on the observational time of $\boldsymbol{W}_{iv}$ and $\boldsymbol{Y}_i$, respectively. Then, the conditional distribution

of $\boldsymbol{W}_{iv}$ given $\boldsymbol{\xi}_{iv}$ is

$$[\boldsymbol{W}_{iv}|\boldsymbol{\xi}_{iv}] \sim N(\boldsymbol{B}_{iv}^{*}\boldsymbol{\theta}_{\mu v} + \textstyle\sum_{k=1}^{p_v}\xi_{ivk}\boldsymbol{B}_{iv}^{*}\boldsymbol{\theta}_{\psi v,k}, \sigma_v^2\boldsymbol{I}), \tag{2.7}$$

$v = 1, 2, \ldots, d_x$, $i = 1, \ldots, n$.

Combining Models (2.1), (2.2), (2.3) and (2.5), we have the following semiparametric mixture regression model with fixed effects:

$$[\boldsymbol{Y}_i|\boldsymbol{Z}_i, \boldsymbol{\xi}_i] \sim \sum_{c=1}^{C}\pi_{ic}(\mathbb{Z}_i)N\Big\{\beta_{0,c} + \boldsymbol{Z}_i\boldsymbol{\beta}_{z,c} + \sum_{v=1}^{d_x}\beta_{x,cv}(\boldsymbol{B}_{iv}\boldsymbol{\theta}_{\mu v} + \boldsymbol{\Psi}_{iv}\boldsymbol{\xi}_{iv}), \sigma_\epsilon^2\boldsymbol{I}\Big\}, \tag{2.8}$$

where $\pi_{ic}(\mathbb{Z}_i)$ is given in (2.1), $\boldsymbol{\Psi}_{iv} = \boldsymbol{B}_{iv}\boldsymbol{\Theta}_{\psi v}$ is the matrix containing the eigenfunctions $\boldsymbol{\psi}_v$ evaluated on $\mathbf{t}_i = (t_{i1}, \ldots, t_{im_{y,i}})^\top$.

2.4 Observed likelihood function

We are focused on the distribution of $\boldsymbol{Y}$ and $\boldsymbol{W}$, so conditional on $\boldsymbol{Z}_i$, the observed data likelihood is

$$\mathcal{L} = \prod_{i=1}^{n} f(\boldsymbol{Y}_i, \boldsymbol{W}_i|\boldsymbol{Z}_i), \tag{2.9}$$

with

$$\begin{aligned} f(\boldsymbol{Y}_i, \boldsymbol{W}_i|\boldsymbol{Z}_i) &= \int f(\boldsymbol{\xi}_i, \boldsymbol{Y}_i, \boldsymbol{W}_i|\boldsymbol{Z}_i)d\boldsymbol{\xi}_i \\ &= \sum_{c=1}^{C}\frac{\pi_{ic}}{(2\pi)^{(m_{y,i}+m_{x,i})/2}\sigma_\epsilon}|\boldsymbol{\Lambda}_i|^{-\frac{1}{2}}|\boldsymbol{\Sigma}_\xi|^{-\frac{1}{2}}|\boldsymbol{\Omega}_{ic}|^{\frac{1}{2}} \times \exp\Big\{-\frac{1}{2}(\widetilde{\boldsymbol{Y}}_{ic}^\top\widetilde{\boldsymbol{Y}}_{ic}/\sigma_\epsilon^2 \end{aligned}$$

$$+\sum_{v=1}^{d_x}(\boldsymbol{W}_{iv}-\boldsymbol{B}_{iv}^{*}\boldsymbol{\theta}_{\mu v})^{\otimes 2}/\sigma_v^2-\widetilde{\boldsymbol{\xi}}_{ic}^{\top}\boldsymbol{\Omega}_{ic}^{-1}\widetilde{\boldsymbol{\xi}}_{ic})\}, \tag{2.10}$$

where $\widetilde{\boldsymbol{Y}}_{ic} = \boldsymbol{Y}_i - \beta_{0,c}\mathbf{1}_{m_{y,i}} - \sum_{v=1}^{d_x}\beta_{x,cv}\boldsymbol{B}_{iv}\boldsymbol{\theta}_{\mu v} - \boldsymbol{Z}_i\boldsymbol{\beta}_{z,c}$, $m_{x,i} = \sum_{v=1}^{d_x} m_{x,iv}$, $\boldsymbol{\Omega}_{ic} = (\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{\Psi}}_{ic}^{\top}\widetilde{\boldsymbol{\Psi}}_{ic} + \boldsymbol{\Psi}_i^{*\top}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\Psi}_i^{*} + \boldsymbol{\Sigma}^{-1})^{-1}, \widetilde{\boldsymbol{\xi}}_{ic} = \boldsymbol{\Omega}_{ic}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{\Psi}}_{ic}^{\top}\widetilde{\boldsymbol{Y}}_{ic} + \boldsymbol{\Psi}_i^{*\top}\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i)$,$\widetilde{\boldsymbol{W}}_i = \boldsymbol{W}_i - \boldsymbol{B}_i^{*}\boldsymbol{\Theta}_\mu$ with $\boldsymbol{B}_i^{*} = \mathrm{diag}(\boldsymbol{B}_{i1}^{*},\ldots,\boldsymbol{B}_{id_x}^{*})$ be a $m_{x,i}\times q$ matrix with diagonal block $\boldsymbol{B}_{iv}^{*}$ and $\boldsymbol{\Theta}_\mu = (\boldsymbol{\theta}_{\mu 1}^{\top},\ldots,\boldsymbol{\theta}_{\mu d_x}^{\top})^{\top}$ is a $q\times 1$ vector that contains all the coefficients for splines in the mean function, where $q = \sum_{v=1}^{d_X} q_v$. Also, $\boldsymbol{\Psi}_i^{*} = \mathrm{diag}(\boldsymbol{B}_{i1}^{*}\boldsymbol{\Theta}_{\psi 1},\ldots,\boldsymbol{B}_{id_x}^{*}\boldsymbol{\Theta}_{\psi d_x})$ is a $q\times p$ matrix. Moreover, $\widetilde{\boldsymbol{\Psi}}_{ic} = (\widetilde{\boldsymbol{\Psi}}_{i1c},\widetilde{\boldsymbol{\Psi}}_{i2c},\ldots,\widetilde{\boldsymbol{\Psi}}_{id_xc})$ with each $\widetilde{\boldsymbol{\Psi}}_{ivc} = \beta_{x,cv}\boldsymbol{B}_{iv}\boldsymbol{\Theta}_{\psi v}$. Additionally, $\boldsymbol{\Lambda}_i$ is a diagonal square matrix with diagonal entry $\sigma_v^2\boldsymbol{I}_{m_{x,iv}}$. Additionally, the detailed derivation for Equation (2.10) is given in Appendix A.

2.5 EM algorithm

To overcome the challenges associated with the complicated covariance structures and orthonormal constraints (2.6) in maximizing the marginal likelihood (2.10) directly, we employ an EM (Expectation-Maximization) algorithm to fit the proposed model. In the EM algorithm, we treat the latent principal component (PC) scores $\boldsymbol{\xi}$ and the latent subgroup membership $\boldsymbol{L}$ as missing data.

The EM algorithm is an iterative optimization method that alternates between the E-step and the M-step. In the E-step, we estimate the expected values of the missing data given the current parameter estimates. Specifically, we compute the conditional expectation of the latent PC scores and subgroup membership based on the observed data and the current parameter values. In the M-step, we maximize the expected complete-data log-

likelihood with respect to the model parameters. This involves updating the parameter estimates by maximizing the expected log-likelihood or Newton-Raphson method, taking into account the completed data that includes the estimated values of the missing data from the E-step. By iteratively performing the E-step and M-step, the algorithm aims to find the best estimates of the model parameters that maximize the marginal likelihood.

The conditional distribution of $\boldsymbol{L}, \boldsymbol{Y}, \boldsymbol{W}$ and $\boldsymbol{\xi}$ given $\boldsymbol{Z}$ is expressed as following

$$f(\boldsymbol{L}, \boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{\xi} | \boldsymbol{Z}) = f(\boldsymbol{Y} | \boldsymbol{L}, \boldsymbol{\xi}, \boldsymbol{Z}) \times f(\boldsymbol{L} | \boldsymbol{Z}) \times f(\boldsymbol{W} | \boldsymbol{\xi}) \times f(\boldsymbol{\xi}),$$

where $[\boldsymbol{Y} | L_c = 1, \boldsymbol{Z}, \boldsymbol{\xi}] \sim N\{\beta_{0,c}\mathbf{1} + \sum_{v=1}^{d_x} \beta_{x,cv}(\boldsymbol{\mu}_v + \boldsymbol{\Psi}_v \boldsymbol{\xi}) + \boldsymbol{Z}^\top \boldsymbol{\beta}_{z,c}, \sigma_\epsilon^2 \boldsymbol{I}\}$, $[L_c = 1 | \boldsymbol{Z}]$ is a multinomial distribution that has C categories with each category is π_c probability, $(\boldsymbol{W}_v | \boldsymbol{\xi}_v) \sim N(\boldsymbol{\mu}_v^* + \boldsymbol{\Psi}_v^* \boldsymbol{\xi}_v, \sigma_v^2 \boldsymbol{I})$ and $\boldsymbol{\xi} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\xi)$. So the conditional distribution function can be expressed as

$$\begin{aligned} f(\boldsymbol{L}, \boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{\xi} | \boldsymbol{Z}) &= \prod_{c=1}^{C} \{ \frac{\pi_c}{(2\pi\sigma_\epsilon^2)^{\frac{m_{y,i}}{2}}} \exp[-\frac{1}{2\sigma_\epsilon^2} \{\boldsymbol{Y} - \beta_{0c}\mathbf{1} - \sum_{v=1}^{d_x} \beta_{x,cv}(\boldsymbol{\mu}_v + \boldsymbol{\Psi}_v \boldsymbol{\xi}_v) - \boldsymbol{Z}\boldsymbol{\beta}_{z,c}\}^{\otimes 2}] \\ &\times \prod_{v=1}^{d_x} [\frac{1}{(2\pi\sigma_v^2)^{\frac{m_{x,iv}}{2}}} \exp\{-\frac{1}{2\sigma_v^2} (\boldsymbol{W}_v - \boldsymbol{\mu}_v^* - \boldsymbol{\Psi}_v^* \boldsymbol{\xi}_v)^{\otimes 2}\}] \\ &\times \frac{1}{(2\pi)^{\frac{p}{2}}} |\Sigma|^{-\frac{1}{2}} \exp(-\frac{1}{2} \boldsymbol{\xi}^\top \Sigma^{-1} \boldsymbol{\xi})\}^{L_c}, \end{aligned} \tag{2.11}$$

where for a matrix A, we define $A^{\otimes 2} = A^\top A$.

If two latent variables $\boldsymbol{\xi}$ and $\boldsymbol{L}$ are known, and collect all parameters in the model into $\boldsymbol{\Theta}$, including $\{\boldsymbol{\gamma}_c;\ c = 1, \ldots, C-1\}$, $\{\boldsymbol{\theta}_{\mu,v}, \boldsymbol{\Theta}_{\psi v}; v = 1, \ldots, d_x\}$, $\{\boldsymbol{\beta}_c; c = 1, \ldots, C\}$, $\boldsymbol{\Sigma}_\xi$,

and σ_ϵ^2. our complete data likelihood is given as

$$\mathcal{L}(\boldsymbol{\Theta};\boldsymbol{Y},\boldsymbol{W},\boldsymbol{\xi},\boldsymbol{L}|\boldsymbol{Z}) = \prod_{i=1}^{n}\Bigg[\prod_{c=1}^{C}\{\pi_{ic}f(\boldsymbol{Y}_i|\boldsymbol{Z}_i,\boldsymbol{\xi}_i,L_{ic}=1;\boldsymbol{\Theta})\}^{L_{ic}}\Bigg]f(\boldsymbol{W}_i|\boldsymbol{\xi}_i;\boldsymbol{\Theta})f(\boldsymbol{\xi}_i|\boldsymbol{\Theta}).$$

Then taking a logarithm of our likelihood and multiplying it by a constant, it will be

$$\begin{aligned}
-2 \quad\times\quad \log\mathcal{L}(\boldsymbol{\Theta};\boldsymbol{Y},\boldsymbol{W},\boldsymbol{\xi},\boldsymbol{L}|\boldsymbol{Z}) &= -2\sum_{i=1}^{n}\sum_{c=1}^{C}L_{ic}\{\log(\pi_{ic})+\log f_c(\boldsymbol{Y}_i|\boldsymbol{Z}_i,\boldsymbol{\xi}_i,L_{ic}=1;\boldsymbol{\Theta})\}\\
&\quad -2\sum_{i=1}^{n}\sum_{v=1}^{d_x}\{\log f(\boldsymbol{W}_{iv}|\boldsymbol{\xi}_{iv};\boldsymbol{B}_{iv},\boldsymbol{\theta}_{\mu v},\boldsymbol{\Theta}_{\psi v},\sigma_v^2)+\log f(\boldsymbol{\xi}_{iv})\}\\
&= \sum_{i=1}^{n}\sum_{c=1}^{C}\Bigg[L_{ic}m_{y,i}\log(\sigma_\epsilon^2)+\frac{L_{ic}}{\sigma_\epsilon^2}\{\boldsymbol{Y}_i-\boldsymbol{\beta}_{0,c}\mathbf{1}-\sum_{v=1}^{d_x}\beta_{x,cv}(\boldsymbol{B}_{iv}\boldsymbol{\theta}_{\mu v}+\boldsymbol{B}_{iv}\boldsymbol{\Theta}_{\psi v}\boldsymbol{\xi}_{iv})-\boldsymbol{Z}_i\boldsymbol{\beta}_{z,c}\}^{\otimes 2}\\
&\quad -2L_{ic}\log(\pi_{ic})\Bigg]+\sum_{i=1}^{n}\sum_{v=1}^{d_x}\{m_{xv,i}\log(\sigma_v^2)+\frac{1}{\sigma_v^2}(\boldsymbol{W}_{iv}-\boldsymbol{B}_{iv}^{*}\boldsymbol{\theta}_{\mu v}-\boldsymbol{B}_{iv}^{*}\boldsymbol{\Theta}_{\psi v}\boldsymbol{\xi}_{iv})^{\otimes 2}\}\\
&\quad +\{n\log(|\Sigma|)+\sum_{i=1}^{n}\boldsymbol{\xi}_i^{\top}\Sigma^{-1}\boldsymbol{\xi}_i\}. \qquad (2.12)
\end{aligned}$$

2.5.1 E-step

Let $\boldsymbol{\Theta}^{(\kappa)}$ be the vector containing current parameter values at the end of Step κ of the EM iteration. The E-step of the $(\kappa+1)$th step, defines the EM loss function $Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) = -2\mathrm{E}\Big[\{\log\mathcal{L}(\boldsymbol{\Theta};\boldsymbol{Y},\boldsymbol{W},\boldsymbol{\xi},\boldsymbol{L}|\boldsymbol{Z})\}|\boldsymbol{Y},\boldsymbol{W},\boldsymbol{Z};\boldsymbol{\Theta}^{(\kappa)}\Big]$, which is the conditional expectation of the complete data likelihood given the observed data and the current parameter value.

$$\begin{aligned}
Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) &= -2\mathrm{E}\Big[\{\log\mathcal{L}(\boldsymbol{\Theta};\boldsymbol{Y},\boldsymbol{W},\boldsymbol{\xi},\boldsymbol{L}|\boldsymbol{Z})\}|\boldsymbol{Y},\boldsymbol{W},\boldsymbol{Z};\boldsymbol{\Theta}^{(\kappa)}\Big]\\
&= \sum_{i=1}^{n}\sum_{c=1}^{C}\Bigg(\mathrm{E}(L_{ic}|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)})\Big\{-2\log(\pi_{ic})+m_{y,i}\times\log(\sigma_\epsilon^2)+\frac{\widetilde{\boldsymbol{Y}}_{ic}^{\top}\widetilde{\boldsymbol{Y}}_{ic}}{\sigma_\epsilon^2}\Big\}\\
&\quad -\frac{1}{\sigma_\epsilon^2}\Big[\widetilde{\boldsymbol{Y}}_{ic}^{\top}\widetilde{\boldsymbol{\Psi}}_{ic}\mathrm{E}(L_{ic}\boldsymbol{\xi}_i|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)})-\mathrm{E}(L_{ic}\boldsymbol{\xi}_i^{\top}|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)})\widetilde{\boldsymbol{\Psi}}_{ic}^{\top}\widetilde{\boldsymbol{Y}}_{ic}
\end{aligned}$$

$$
\begin{aligned}
&+\operatorname{tr}\{\mathrm{E}(L_{ic}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^\top|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)})\widetilde{\boldsymbol{\Psi}}_{ic}^\top\widetilde{\boldsymbol{\Psi}}_{ic}\}\Big]\Big) + n\log(|\boldsymbol{\Lambda}_i|) + n\log(|\boldsymbol{\Sigma}|)\\
&+\sum_{i=1}^{n}\Big[\widetilde{\boldsymbol{W}}_i^\top\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i - \widetilde{\boldsymbol{W}}_i^\top\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\Psi}_i^*\mathrm{E}(\boldsymbol{\xi}_i|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)})\\
&-\mathrm{E}(\boldsymbol{\xi}_i^\top|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)})\boldsymbol{\Psi}_i^{*\top}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{W}_i\\
&+\operatorname{tr}\{\mathrm{E}(\boldsymbol{\xi}_i\boldsymbol{\xi}_i^\top|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)})(\boldsymbol{\Psi}_i^{*\top}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\Psi}_i^* + \boldsymbol{\Sigma}^{-1})\}\Big] \qquad (2.13)
\end{aligned}
$$

with $\widetilde{\boldsymbol{Y}}_{ic}, \widetilde{\boldsymbol{\Psi}}_{ic}, \widetilde{\boldsymbol{W}}_i,$ and $\boldsymbol{\Psi}_i^*$ defined in 2.10.

Then the desired predictions required by the EM algorithm are given as

$$
\begin{aligned}
\mathrm{E}(L_{ic}|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)}) &= \widetilde{\pi}_{ic}\\
&= \frac{\pi_{ic}|\boldsymbol{\Omega}_{ic}|^{\frac{1}{2}}\exp\{-\frac{1}{2}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic}^\top\widetilde{\boldsymbol{Y}}_{ic} - \widetilde{\boldsymbol{\xi}}_{ic}^\top\boldsymbol{\Omega}_{ic}^{-1}\widetilde{\boldsymbol{\xi}}_{ic})\}}{\sum_{c'=1}^{C}\pi_{ic'}|\boldsymbol{\Omega}_{ic'}|^{\frac{1}{2}}\exp\{-\frac{1}{2}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic'}^\top\widetilde{\boldsymbol{Y}}_{ic'} - \widetilde{\boldsymbol{\xi}}_{ic'}^\top\boldsymbol{\Omega}_{ic'}^{-1}\widetilde{\boldsymbol{\xi}}_{ic'})\}},\\
\mathrm{E}(L_{ic}\boldsymbol{\xi}_i|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)}) &= \widetilde{\pi}_{ic}\widetilde{\boldsymbol{\xi}}_{ic},\\
\mathrm{E}(L_{ic}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^\top|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)}) &= \widetilde{\pi}_{ic}\widetilde{\boldsymbol{\Sigma}}_{ic} = \widetilde{\pi}_{ic}(\widetilde{\boldsymbol{\xi}}_{ic}\widetilde{\boldsymbol{\xi}}_{ic}^\top + \boldsymbol{\Omega}_{ic}),\\
\mathrm{E}(\boldsymbol{\xi}_i|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)}) &= \sum_{c=1}^{C}\widetilde{\pi}_{ic}\widetilde{\boldsymbol{\xi}}_{ic} = \widetilde{\boldsymbol{\xi}}_i,\\
\mathrm{E}(\boldsymbol{\xi}_i\boldsymbol{\xi}_i^\top|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)}) &= \sum_{c=1}^{C}\widetilde{\pi}_{ic}(\widetilde{\boldsymbol{\xi}}_{ic}\widetilde{\boldsymbol{\xi}}_{ic}^\top + \boldsymbol{\Omega}_{ic}) = \sum_{c=1}^{C}\widetilde{\pi}_{ic}\widetilde{\boldsymbol{\Sigma}}_{ic} = \widetilde{\boldsymbol{\Sigma}}_i, \qquad (2.14)
\end{aligned}
$$

where $\boldsymbol{\Omega}_{ic}$ and $\widetilde{\boldsymbol{\xi}}_{ic}$ are defined in (2.10). The deriving process of these formulas is given in Appendix C.

2.5.2 M-step

It is clear that the unknown parameters are separated in the Q function and therefore separate optimization is feasible. Substitute expectation of unknown variables

and separate our Q function as

$$
\begin{aligned}
Q_1(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) &= \sum_{i=1}^{n}\sum_{c=1}^{C} -2 \times \widetilde{\pi}_{ic} \times \log(\pi_{ic}), \\
Q_2(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) &= \sum_{i=1}^{n}\sum_{c=1}^{C} \widetilde{\pi}_{ic}\Bigg[m_{y,i}\log(\sigma_\epsilon^2) + \frac{1}{\sigma_\epsilon^2}\Bigg\{ \sum_{v=1}^{d_x}\sum_{v'=1}^{d_x} \beta_{x,cv}\beta_{x,cv'}\,\mathrm{tr}(\boldsymbol{\Theta}_{\psi v}^\top \boldsymbol{B}_{iv}^\top \boldsymbol{B}_{iv'}\boldsymbol{\Theta}_{\psi v'}\boldsymbol{\Omega}_{ic,v'v}) \\
&\quad + (\boldsymbol{Y}_i - \beta_{0,c}\mathbf{1} - \sum_{v=1}^{d_x}\beta_{x,cv}\boldsymbol{B}_{iv}\boldsymbol{\theta}_{\mu v} - \sum_{v=1}^{d_x}\beta_{x,cv}\boldsymbol{B}_{iv}\boldsymbol{\Theta}_{\psi v}\widetilde{\boldsymbol{\xi}}_{iv,c} - \boldsymbol{Z}_i\boldsymbol{\beta}_{z,c})^{\otimes 2}\Bigg\}\Bigg] \\
Q_3(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) &= \sum_{i=1}^{n}\sum_{v=1}^{d_x}\Bigg[\frac{1}{\sigma_v^2}\Bigg\{ (\boldsymbol{W}_{iv} - \boldsymbol{B}_{iv}^*\boldsymbol{\theta}_{\mu v} - \boldsymbol{B}_{iv}^*\boldsymbol{\Theta}_{\psi v}\widetilde{\boldsymbol{\xi}}_{iv})^{\otimes 2} - \widetilde{\boldsymbol{\xi}}_{iv}^\top \boldsymbol{\Theta}_{\psi v}^\top \boldsymbol{B}_{iv}^{*\top}\boldsymbol{B}_{iv}^*\boldsymbol{\Theta}_{\psi v}\widetilde{\boldsymbol{\xi}}_{iv} \\
&\quad + \mathrm{tr}(\boldsymbol{\Theta}_{\psi v}^\top \boldsymbol{B}_{iv}^{*\top}\boldsymbol{B}_{iv}^*\boldsymbol{\Theta}_{\psi v}\widetilde{\boldsymbol{\Sigma}}_{i,vv})\Bigg\} + m_{xv,i}\log(\sigma_v^2)\Bigg], \\
Q_4(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) &= n\log(|\boldsymbol{\Sigma}|) + \sum_{i=1}^{n}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\widetilde{\boldsymbol{\Sigma}}_i).
\end{aligned} \tag{2.15}
$$

Since different components of $Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)})$ in Equation (3.5) depend on different parameters, we can update various components of $\boldsymbol{\Theta}$ separately. Some parameters may require updating using Newton-Raphson iterations for more accurate estimation. For a detailed description of the algorithm, please refer to Section Appendix C. By updating the parameters of $\boldsymbol{\Theta}$ in separate steps while satisfying the orthogonal constraints, we ensure that each component is optimized effectively. This iterative process helps refine the estimation and improve the overall performance of the model. Please refer to Section Appendix C for a comprehensive explanation of the algorithm used in the M-step.

Chapter 3

Model Structure and Model Fitting for Emerald with Random Effect

3.1 Mixture regression

If the measurement errors alone, defined in Chapter 2, are not sufficient to explain the majority of the variation in the response variable $\boldsymbol{Y}$, it may be necessary to introduce a random effect into our linear model. This random effect accounts for additional unobserved factors that contribute to the variability in $\boldsymbol{Y}$. By incorporating a random effect, we can capture the individual-specific variation and address the potential sources of unexplained variability in the data. The inclusion of a random effect allows for the modeling of both systematic and random components of the variation in $\boldsymbol{Y}$. The systematic components are captured by the fixed effects, representing the overall population-level trends and associations, while the random effect captures the individual-specific deviations from

the population-level patterns. The random effect accounts for unobserved factors that may influence the response variable but are not explicitly measured or accounted for by the covariates. By incorporating a random effect into the linear model, we can better capture the complex structure and heterogeneity in the data. This approach provides a more comprehensive and accurate representation of the underlying processes and helps improve the understanding of the relationships between the variables of interest.

Note that in this section, the definitions for the logistic regression model, functional model, mixture model, and observed data remain the same as presented in Chapter 2. The methodology for functional modeling and the logistic regression model also remains consistent with what was described in Chapter 2. By maintaining the same definitions and methodologies, we ensure continuity and consistency throughout the analysis, allowing for a clear comparison and evaluation of the proposed methods.

Given $L_{ic} = 1$, the response can be modeled by a linear mixed model

$$Y_i(t) = \beta_{0,c} + \boldsymbol{X}_i(t)^\top \boldsymbol{\beta}_{x,c} + \boldsymbol{Z}_i(t)^\top \boldsymbol{\beta}_{z,c} + b_{i,c} + \epsilon_i(t), \tag{3.1}$$

where $\boldsymbol{\beta}_c$ and $\epsilon_i(t)$ share the same definition as Equation 2.2. $b_{i,c} \sim N(0, \sigma^2_{b,c})$ is a subject-specific random effect with a subgroup-specific variance.

Therefore, by combining the models (2.1), (3.1), (2.3) and (2.5) and a semiparametric mixture regression model with random effects:

$$[\boldsymbol{Y}_i|\boldsymbol{Z}_i, \boldsymbol{\xi}_i] \sim \sum_{c=1}^{C} \pi_{ic}(\mathbb{Z}_i) N\left\{ \beta_{0,c} + \boldsymbol{Z}_i \boldsymbol{\beta}_{z,c} + \sum_{v=1}^{d_x} \beta_{x,cv}(\boldsymbol{B}_{iv}\boldsymbol{\theta}_{\mu v} + \boldsymbol{\Psi}_{iv}\boldsymbol{\xi}_{iv}), \boldsymbol{V}_{ic} \right\}, \tag{3.2}$$

where $\pi_{ic}(\mathbb{Z}_i)$ is given in (2.1), $\boldsymbol{\Psi}_{iv} = \boldsymbol{B}_{iv}\boldsymbol{\Theta}_{\psi v}$ is the matrix containing the eigenfunctions $\boldsymbol{\psi}_v$ evaluated on $\mathbf{t}_i = (t_{i1}, \ldots, t_{im_{y,i}})^\top$, $\boldsymbol{V}_{ic} = \sigma^2_{b,c}\boldsymbol{J} + \sigma^2_\epsilon \boldsymbol{I}$ is the compound symmetry conditional covariance matrix for $\boldsymbol{Y}_i$ in subgroup c, and $\boldsymbol{J}$ is a matrix with 1's in all entries.

3.2 Observed likelihood function

When we model each mixture component with a linear mixed model, the marginal distribution of $\boldsymbol{Y}$ and $\boldsymbol{W}$ can be written as

$$\begin{aligned} f(\boldsymbol{Y}_i, \boldsymbol{W}_i|\boldsymbol{Z}_i) &= \int f(\boldsymbol{\xi}_i, \boldsymbol{Y}_i, \boldsymbol{W}_i|\boldsymbol{Z}_i) d\boldsymbol{\xi}_i \\ &= \sum_{c=1}^{C} \frac{\pi_{ic}}{(2\pi)^{(m_{y,i}+m_{x,i})/2}} |\boldsymbol{V}_{ic}|^{-\frac{1}{2}} |\boldsymbol{\Lambda}_i|^{-\frac{1}{2}} |\boldsymbol{\Sigma}_\xi|^{-\frac{1}{2}} |\boldsymbol{\Omega}_{ic}|^{\frac{1}{2}} \times \exp\Big\{ -\frac{1}{2} (\widetilde{\boldsymbol{Y}}_{ic}^\top \boldsymbol{V}_{ic}^{-1} \widetilde{\boldsymbol{Y}}_{ic} \\ &\quad + \sum_{v=1}^{d_x} (\boldsymbol{W}_{iv} - \boldsymbol{B}^*_{iv}\boldsymbol{\theta}_{\mu v})^{\otimes 2}/\sigma_v^2 - \widetilde{\boldsymbol{\xi}}_{ic}^\top \boldsymbol{\Omega}_{ic}^{-1} \widetilde{\boldsymbol{\xi}}_{ic}) \Big\}, \end{aligned} \tag{3.3}$$

where the definition of $\widetilde{\boldsymbol{Y}}_{ic}, m_{x,i}, \widetilde{\boldsymbol{W}}_i, \boldsymbol{B}^*_i, \boldsymbol{\Theta}_\mu, \boldsymbol{\Psi}^*_i, \widetilde{\boldsymbol{\Psi}}_{ic}$, and $\boldsymbol{\Lambda}_i$ is the same as Equation 2.10. While $\boldsymbol{\Omega}_{ic} = (\widetilde{\boldsymbol{\Psi}}_{ic}^\top \boldsymbol{V}_{ic}^{-1} \widetilde{\boldsymbol{\Psi}}_{ic} + \boldsymbol{\Psi}_i^{*\top} \boldsymbol{\Lambda}_i^{-1} \boldsymbol{\Psi}^*_i + \boldsymbol{\Sigma}_\xi^{-1})^{-1} = (\boldsymbol{\Omega}_{ic,vv'})_{v,v'=1}^{d_x}$, $\widetilde{\boldsymbol{\xi}}_{ic} = \boldsymbol{\Omega}_{ic}(\widetilde{\boldsymbol{\Psi}}_{ic}^\top \boldsymbol{V}_{ic}^{-1} \widetilde{\boldsymbol{Y}}_{ic} + \boldsymbol{\Psi}_i^{*\top} \boldsymbol{\Lambda}_i^{-1} \widetilde{\boldsymbol{W}}_i)$. Additionally, the detailed derivation for Equation (3.3) is given in Appendix B.

3.3 EM algorithm

The unknown parameters in this mixture model with mixed effects contain all the unknown parameters in the fixed effect and also include the parameters in the random effects. Let's gathering all parameters in the model into $\boldsymbol{\Theta}$, including $\{\boldsymbol{\gamma}_c;\ c = 1, \ldots, C-1\}$, $\{\boldsymbol{\theta}_{\mu,v}, \boldsymbol{\Theta}_{\psi v}; v = 1, \ldots, d_x\}$, $\{\boldsymbol{\beta}_c,\ \sigma^2_{b,c}; c = 1, \ldots, C\}$, $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}$, and σ^2_ϵ. If we treat the latent

principal component (PC) scores $\boldsymbol{\xi}$ and the latent subgroup membership $\boldsymbol{L}$ as missing data, the complete data likelihood is

$$\mathcal{L}(\boldsymbol{\Theta};\boldsymbol{Y},\boldsymbol{W},\boldsymbol{\xi},\boldsymbol{L}|\boldsymbol{Z}) = \prod_{i=1}^{n}\left[\prod_{c=1}^{C}\{\pi_{ic}f(\boldsymbol{Y}_i|\boldsymbol{Z}_i,\boldsymbol{\xi}_i,L_{ic}=1;\boldsymbol{\Theta})\}^{L_{ic}}\right]f(\boldsymbol{W}_i|\boldsymbol{\xi}_i;\boldsymbol{\Theta})f(\boldsymbol{\xi}_i|\boldsymbol{\Theta}).$$

Specifically, a detailed expression for the complete data log-likelihood is

$$\begin{aligned}-2\times\log\mathcal{L}(\boldsymbol{\Theta};\boldsymbol{Y},\boldsymbol{W},\boldsymbol{\xi},\boldsymbol{L}|\boldsymbol{Z}) \quad\propto\quad & \sum_{i=1}^{n}\sum_{c=1}^{C}\big\{L_{ic}(\widetilde{\boldsymbol{Y}}_{ic}-\sum_{v=1}^{d_x}\widetilde{\boldsymbol{\Psi}}_{ivc}\boldsymbol{\xi}_{iv})^{\top}\boldsymbol{V}_{ic}^{-1}(\widetilde{\boldsymbol{Y}}_{ic}-\sum_{v=1}^{d_x}\widetilde{\boldsymbol{\Psi}}_{ivc}\boldsymbol{\xi}_{iv})\\ & +L_{ic}\log(|\boldsymbol{V}_{ic}|)-2L_{ic}\log(\pi_{ic})\big\}+\sum_{i=1}^{n}\sum_{v=1}^{d_x}\{m_{x,iv}\log(\sigma_v^2)\\ & +\frac{1}{\sigma_v^2}(\boldsymbol{W}_{iv}-\boldsymbol{B}_{iv}^{*}\boldsymbol{\theta}_{\mu v}-\boldsymbol{B}_{iv}^{*}\boldsymbol{\Theta}_{\psi v}\boldsymbol{\xi}_{iv})^{\otimes 2}\}+\{n\log(|\boldsymbol{\Sigma}_{\xi}|)+\sum_{i=1}^{n}\boldsymbol{\xi}_i^{\top}\boldsymbol{\Sigma}_{\xi}^{-1}\boldsymbol{\xi}_i\}. \quad (3.4)\end{aligned}$$

3.3.1 E-step

Let $\boldsymbol{\Theta}^{(\kappa)}$ be the vector containing current parameter values at the end of Step κ of the EM iteration. The E-step of the (κ+1)th step, defines the EM loss function $Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) = -2\mathrm{E}\Big[\{\log\mathcal{L}(\boldsymbol{\Theta};\boldsymbol{Y},\boldsymbol{W},\boldsymbol{\xi},\boldsymbol{L}|\boldsymbol{Z})\}|\boldsymbol{Y},\boldsymbol{W},\boldsymbol{Z};\boldsymbol{\Theta}^{(\kappa)}\Big]$, which is the conditional expectation of the complete data likelihood given the observed data and the current parameter value. By detailed calculation in Appendix D, we have

$$Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) = Q_1(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) + Q_2(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) + Q_3(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) + Q_4(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)})$$

where

$$
\begin{aligned}
Q_1(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) &= \sum_{i=1}^{n}\sum_{c=1}^{C} -2 \times \widetilde{\pi}_{ic} \times \log(\pi_{ic}), \\
Q_2(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) &= \sum_{i=1}^{n}\sum_{c=1}^{C}\Bigg(\widetilde{\pi}_{ic}\Bigg[\Bigg\{\boldsymbol{Y}_i - \beta_{0,c} - \boldsymbol{Z}_i\boldsymbol{\beta}_{z,c} - \sum_{v=1}^{d_x}\beta_{x,cv}\boldsymbol{B}_{iv}(\boldsymbol{\theta}_{\mu v} + \boldsymbol{\Theta}_{\psi v}\widetilde{\boldsymbol{\xi}}_{iv,c})\Bigg\}^{\otimes 2}_{\boldsymbol{V}_{ic}} \\
&\quad + \sum_{v=1}^{d_x}\sum_{v'=1}^{d_x}\beta_{x,cv}\beta_{x,cv'}\,\mathrm{tr}\{\boldsymbol{\Theta}_{\psi v}^{\top}\boldsymbol{B}_{iv}^{\top}\boldsymbol{V}_{ic}^{-1}\boldsymbol{B}_{iv'}\boldsymbol{\Theta}_{\psi v'}\boldsymbol{\Omega}_{ic,v'v}\} + \log(|\boldsymbol{V}_{ic}|)\Bigg]\Bigg), \\
Q_3(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) &= \sum_{i=1}^{n}\sum_{v=1}^{d_x}\Bigg[m_{x,iv}\log(\sigma_v^2) + \frac{1}{\sigma_v^2}\Bigg\{(\boldsymbol{W}_{iv} - \boldsymbol{B}_{iv}^{*}\boldsymbol{\theta}_{\mu v} - \boldsymbol{B}_{iv}^{*}\boldsymbol{\Theta}_{\psi v}\widetilde{\boldsymbol{\xi}}_{iv})^{\otimes 2} \\
&\quad - \widetilde{\boldsymbol{\xi}}_{iv}^{\top}\boldsymbol{\Theta}_{\psi v}^{\top}(\boldsymbol{B}_{iv}^{*})^{\top}\boldsymbol{B}_{iv}^{*}\boldsymbol{\Theta}_{\psi v}\widetilde{\boldsymbol{\xi}}_{iv} + \mathrm{tr}(\boldsymbol{\Theta}_{\psi v}^{\top}(\boldsymbol{B}_{iv}^{*})^{\top}\boldsymbol{B}_{iv}^{*}\boldsymbol{\Theta}_{\psi v}\widetilde{\boldsymbol{\Sigma}}_{i,vv})\Bigg\}\Bigg], \\
Q_4(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) &= n\log(|\boldsymbol{\Sigma}_{\xi}|) + \sum_{i=1}^{n}\mathrm{tr}(\boldsymbol{\Sigma}_{\xi}^{-1}\widetilde{\boldsymbol{\Sigma}}_i). \qquad (3.5)
\end{aligned}
$$

Here, $\widetilde{\pi}_{ic} = \mathrm{E}(L_{ic}|\boldsymbol{Y}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i)$, $\widetilde{\boldsymbol{\xi}}_{iv,c} = \mathrm{E}(L_{ic}\boldsymbol{\xi}_{iv}|\boldsymbol{Y}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i)/\widetilde{\pi}_{ic}$ with $\widetilde{\boldsymbol{\xi}}_{ic} = (\widetilde{\boldsymbol{\xi}}_{i1,c}^{\top}, \widetilde{\boldsymbol{\xi}}_{i2,c}^{\top}, \ldots, \widetilde{\boldsymbol{\xi}}_{id_x,c}^{\top})^{\top}$, $\widetilde{\boldsymbol{\xi}}_{iv} = \mathrm{E}(\boldsymbol{\xi}_{iv}|\boldsymbol{Y}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i) = \sum_{c=1}^{C}\widetilde{\pi}_{ic}\widetilde{\boldsymbol{\xi}}_{iv,c}$, and $\widetilde{\boldsymbol{\Sigma}}_i = \mathrm{E}(\boldsymbol{\xi}_i\boldsymbol{\xi}_i^{\top}|\boldsymbol{Y}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i) = (\widetilde{\boldsymbol{\Sigma}}_{i,vv'})_{v,v'=1}^{d_x}$. For any matrix $\boldsymbol{A}$ and a positive definite matrix $\boldsymbol{V}$, denote $\boldsymbol{A}_{\boldsymbol{V}}^{\otimes 2} = \boldsymbol{A}^{\top}\boldsymbol{V}^{-1}\boldsymbol{A}$. The detailed expressions for the quantities above are derived in Appendix D.

3.3.2 M-step

In the M-step of the $(\kappa+1)$th iteration, we update the estimates as $\widehat{\boldsymbol{\Theta}}^{(\kappa+1)} = \mathrm{argmin} Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)})$ subject to the orthogonal constraints in (2.6). Since different parts of $Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)})$ in (3.5) depend on different parameters, we can update various components of $\boldsymbol{\Theta}$ separately, and some of the parameters need to be updated with Newton-Raphson iterations. The detailed algorithm is provided in Appendix D.

3.4 Stopping criterion

The algorithm iteratively updates the parameter estimates based on the conditional expectations of the latent variables until convergence is achieved. The specific updates for each parameter depend on the structure of the model and the optimization techniques used. By incorporating the orthogonal constraints and employing appropriate numerical optimization methods, the algorithm ensures that the estimated parameters satisfy the necessary conditions. A widely used criterion to stop the iteration is when

$$\max_l \left| \frac{\Theta_l^{(\kappa+1)} - \Theta_l^{(\kappa)}}{\Theta_l^{(\kappa)} + \delta_1} \right| < \delta_2,$$

$\Theta_l^{(\kappa+1)}$ is the lth component in $\mathbf{\Theta}^{(\kappa+1)}$. Here, δ_1 and δ_2 are predetermined constants which can be set at δ_1=0.001 and δ_2=0.005 following the suggestions in [22]. One can also stop the EM iterations when the improvement in the log-likelihood is below a threshold. [1] proposed to stop the algorithm when

$$|\log \mathcal{L}^{(\kappa+1)} - \log \mathcal{L}^{(\kappa)}| < 10^{-8},$$

where $\mathcal{L}$ is the observed data log-likelihood function defined in (3.3). In our numerical studies, we stop the iteration when either stopping criterion is met.

The choice of initial value in EM algorithms can have a significant impact on the results, a bad initial value can potentially preventing the attainment of the global optimum. To get the algorithm started appropriately, we first initialize the FPCA estimators, including the mean function, eigenvalues, and eigenfunctions using existing packages, such as the

fdapace package of [61] and the **fpca** package of [39]. Then, we can use the imputed trajectories of $X_{iv}(t)$ by the FPCA estimates to fit a mixture regression model to get the initial value of the other parameters using existing packages such as the **lcmm** by [40].

3.5 Practical implementations

To implement the proposed model, several parameters need to be determined, including the number of clusters (C), the number of principal components (p_v) for the asynchronous time-varying covariate processes, and the number of spline basis functions (q_v).

To select these tuning parameters, a multi-dimensional grid search approach can be used, where a model selection criterion is minimized. However, this approach can be computationally expensive. Alternatively, we suggest choosing the tuning parameters separately, starting with the least sensitive ones. For the number of spline basis functions, denoted as q_v, we can follow the guidance provided by [39, 31] on functional principal component analysis (FPCA) for sparse functional data. Specifically, we can set q_v as the ceiling value of $n^{1/(2r_v+1)}$, where n is the sample size and r_v is the order of spline functions used to model process X_v. The selection of the number of principal components, p_v, can be done using the Aikaike information criterion (AIC) proposed in [32]. By fitting the initial FPCA model, one can obtain the estimated FPCs. Then, different values of p_v can be tested, and the corresponding AIC values can be calculated. The value of p_v that yields the lowest AIC can be chosen to fix the number of principal components at this stage. Our method requires a starting value for the FPCs, which can be obtained using existing software, as mentioned

in the previous subsections. One can use the Aikaike information criterion (AIC) proposed in [32] to choose and fix the number of PCs, p_v, at this stage.

To address the most important model selection issue in a mixture regression, i.e. the number of sub-groups, we propose to choose C by either the Bayesian information criterion (BIC) or the Approximate Weight of Evidence (AWE) criterion [2]:

$$\text{BIC(C)} = -2\log\mathcal{L} + K_C \times \log n \quad \text{and} \quad \text{AWE(C)} = -2\log\mathcal{L} + 2K_C(\frac{3}{2} + \log n),$$

where $\mathcal{L}$ is the observed data likelihood defined in (3.3), $K_C = (1+d_z)\times(2C-1) + d_x \times C + \sum_{v=1}^{d_x}(q_v + p_v + q_v \times p_v - \frac{p_v(p_v+1)}{2}) + \sum_{v=1}^{d_x}\sum_{v'=1,v>v'}^{d_x} p_v \times p_{v'} + C$ is the total number of parameters in the model with C subgroups taking into consideration of the orthonormal constraints (2.6). Both criteria work well in our simulation studies.

Chapter 4

Simulation Study

We will now conduct simulation studies to evaluate the numerical performance of the two proposed methods. In these simulation studies, we generated data for two scenarios: one with $C = 2$ subgroups and another with a more complex scenario involving $C = 3$ subgroups. However, due to space limitations, we will primarily present the results for the scenario with $C = 3$ subgroups, which exhibit greater complexity. The comprehensive results for the simpler scenario can be found in the Supplemental Material accompanying this article. By focusing on the more intricate scenario with $C = 3$ subgroups, we aim to provide a detailed assessment of the performance and capabilities of the proposed methods. These results will shed light on the methods' effectiveness in accurately estimating parameters and identifying subgroup patterns within the data.

4.1 Simulation setting

Suppose the study is conducted in a time domain $\mathcal{T} = [0, 1]$, and we use the same approach to generate $\boldsymbol{X}$, $\boldsymbol{W}$, and $\boldsymbol{Z}$ for both models.

For each subject, denoted by i, we observe two time-invariant covariates, $\boldsymbol{Z}_i = (Z_{i1}, Z_{i2})^\top$, where Z_{i1} follows a standard normal distribution, $Z_{i1} \sim N(0, 1)$, and Z_{i2} follows a Bernoulli distribution with a success probability of $p = 0.5$, i.e., $Z_{i2} \sim \text{Bernoulli}(p = 0.5)$.

To generate the time-varying covariate processes, we utilize a specific procedure:

$$X_{i1}(t) = \mu_1(t) + \xi_{i11}\psi_{11}(t) + \xi_{i12}\psi_{12}(t), \quad X_{i2}(t) = \mu_2(t) + \xi_{i21}\psi_{21}(t) + \xi_{i22}\psi_{22}(t).$$

We set the mean functions to be

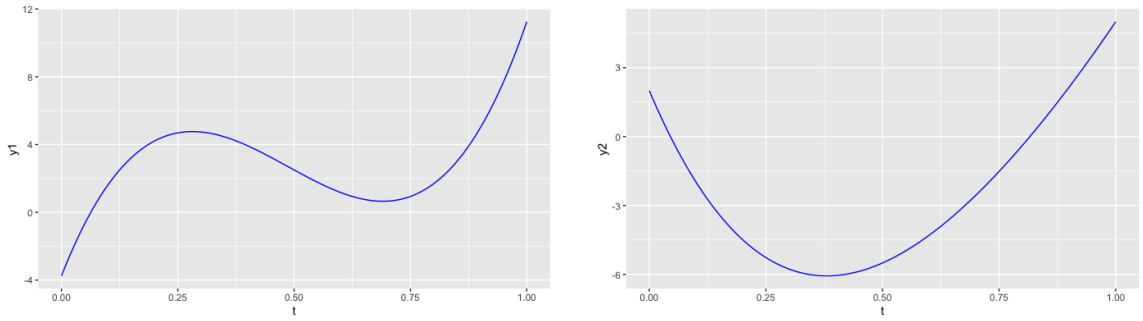
$$\mu_1(t) = 120(t - 0.5)^3 - 5(t - 0.5)^2 - 15t + 10, \quad \mu_2(t) = 12(t - 1)^4 + 15t^2 - 10;$$

and plotted as Figure 4.1. Also, we set the eigenfunctions as

$$\psi_{11}(t) = \sqrt{2}\sin(2\pi t), \quad \psi_{12}(t) = \sqrt{2}\cos(2\pi t),$$
$$\psi_{21}(t) = \sqrt{2}\sin(4\pi t), \quad \psi_{22}(t) = \sqrt{2}\cos(4\pi t).$$

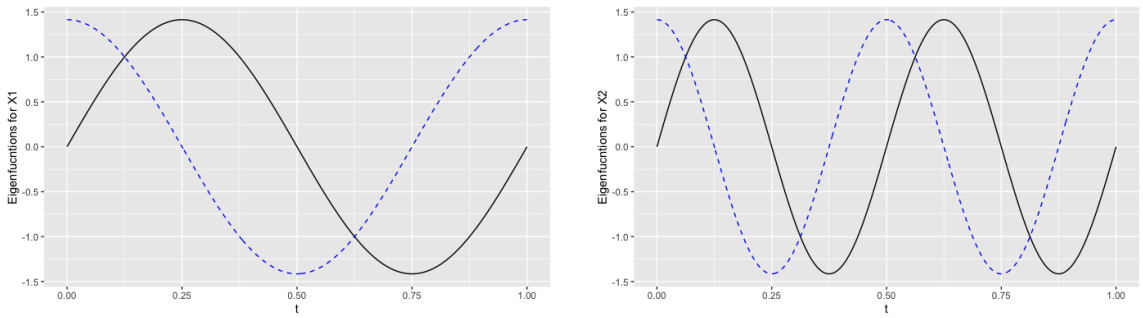
and visualize the eigenfunctions in the Figure 4.2.

Then we generated the principal component scores $\boldsymbol{\xi}_i$ from a multivariate normal distribution with a zero mean and covariance matrix $\boldsymbol{\Sigma}_\xi$ as described in equation (2.4).



(a) $\boldsymbol{X}_1$ (b) $\boldsymbol{X}_2$

Figure 4.1: (a) True mean curves for $X_1(t)$; (b) True mean curve for $X_2(t)$



(a) $\boldsymbol{X}_1$ (b) $\boldsymbol{X}_2$

Figure 4.2: (a) Two eigenfunctions for $\boldsymbol{X}_1$; (b) Two eigenfunctions for $\boldsymbol{X}_2$. In both graphs, the solid black line plots the first eigenfunction and the dashed blue line plots the second eigenfunction.

Specifically, we set the eigenvalues for $\boldsymbol{X}_1$ to be 4.9 and 3.6, and for $\boldsymbol{X}_2$ to be 3 and 2. The cross-covariance between the PC scores was set as follows: $\omega_{12,11} = 2.5$, $\omega_{12,12} = 1.4$, $\omega_{12,21} = 1$, and $\omega_{12,22} = 1.2$. Due to the unavailability of complete covariate trajectories,

we obtained discrete, error-contaminated observations as $W_{ivl} = X_{iv}(s_{ivl}) + U_{ivl}$, where $l = 1, \ldots, m_{x,iv}$, $v = 1, 2$, and $m_{x,iv}$ follows a discrete uniform distribution between 8 and 10 for each i and v. We assumed that all covariates were observed at baseline ($s_{iv1} = 0$ for all v and i), and the remaining follow-up times followed a uniform distribution on the interval [0,1]. The measurement error U_{ivl} was generated from a normal distribution with mean 0 and variance σ_v^2, where we set σ_1^2 to be 1 and σ_2^2 to be 1.5.

We assume that $Y_i(t)$ is observed at baseline so that $t_{i1} = 0$, the followups are jittered around 0.3, 0.5, 0.7, and 0.9 with a local normal distribution with mean 0 and standard deviation 0.05. And a vector of latent subgroup membership indicators $\boldsymbol{L}_i$ is generated from a multinomial distribution with probabilities $\boldsymbol{\pi}_i = (\pi_{i1}, \pi_{i2}, \pi_{i3})^\top$ following the logistic model (2.1), we use all $\boldsymbol{Z}_i$ as covariates in the model, which is $\mathbb{Z}_i = \boldsymbol{Z}_i$. We set $\boldsymbol{\gamma}_1 = (\gamma_{01}, \gamma_{z11}, \gamma_{z21})^\top = (0.2, 0.8, -1.0)^\top$ and $\boldsymbol{\gamma}_2 = (\gamma_{02}, \gamma_{z12}, \gamma_{z22})^\top = (0.6, -1.3, -0.5)^\top$.

4.2 Generate the response variable from a linear fixed model

Given the latent subgroup membership indicators $\boldsymbol{L}_i$, we can generate the response $Y_i(t)$ from (2.2) using the coefficients $\boldsymbol{\beta}_c = (\beta_{0c}, \beta_{xc1}, \beta_{xc2}, \beta_{zc1}, \beta_{zc2})^\top$ for the group-specific effects. We generate $\boldsymbol{\beta}$ from a multivariate normal distribution with a mean of the zero vector and a variance from a diagonal matrix with 3 on the diagonals. Then we can set the true values for $\boldsymbol{\beta}$ as follows:

- For subgroup 1: $\boldsymbol{\beta}_1 = (3.02, 2.23, 1.67, -2.72, -1.65)^\top$;
- For subgroup 2: $\boldsymbol{\beta}_2 = (3.11, -0.71, -1.23, 0.48, 2.28)^\top$;

- For subgroup 3: $\boldsymbol{\beta}_3 = (0.59, 2.60, -0.70, 1.85, 1.68)^\top$.

Additionally, the random error term ϵ_i is generated from a normal distribution with a mean of 0 and a variance of 1.5.

To visualize the values of $\boldsymbol{Y}$, we can create a scatter plot, as shown in Figure 4.3, which provides a graphical representation of the simulated response values.

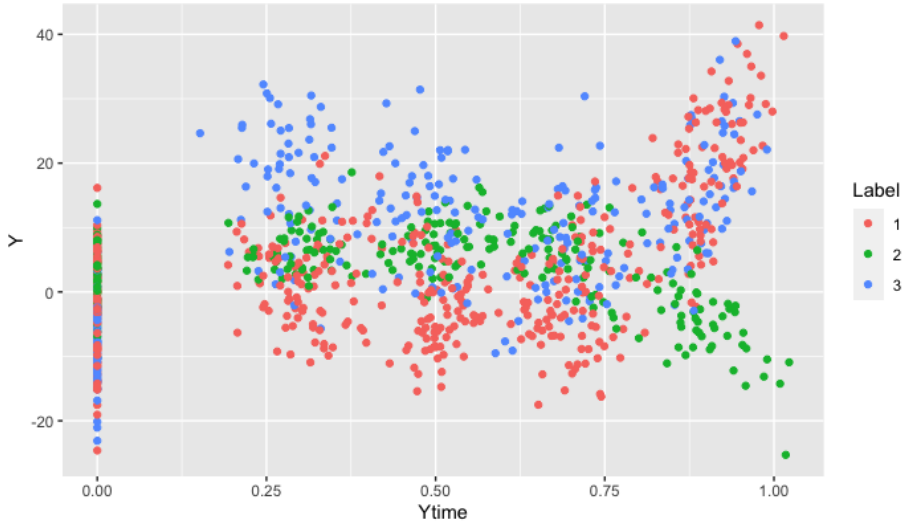


Figure 4.3: Y value labelled by subgroup

4.3 Generate the response variable from a linear mixed model

In Chapter 3, we extended the linear fixed effect model to a linear mixed effect model, which incorporates random effects and provides a more comprehensive explanation of the variability in the response vector $\boldsymbol{Y}$. To demonstrate the potential benefits of this extended model, we also conducted a simulation study. Given the latent subgroup membership indicators $\boldsymbol{L}_i$, the response $Y_i(t)$ was generated from the mixture regression model

(3.1). The coefficients $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, and $\boldsymbol{\beta}_3$ were re-generated from the same distribution as in Section 4.2, with values of

- For subgroup 1: $\boldsymbol{\beta}_1 = (3.18, -1.79, 2.52, 1.48, -1.17)^\top$;
- For subgroup 2: $\boldsymbol{\beta}_2 = (-0.42, 2.68, -0.69, 1.02, 0.54)^\top$;
- For subgroup 3: $\boldsymbol{\beta}_3 = (-2.66, 2.83, 2.32, 1.14, 0.76)^\top$.

For the random effects, the variances $\sigma^2_{b,c}$ were set to be 2, 1.5, and 1 for the three subgroups, respectively. The error variance σ^2_ϵ was set to be 4. These parameter values were chosen to reflect the variability in the response within and between subgroups. We plot $\boldsymbol{Y}$ colored by subgroup labels as Figure 4.4.

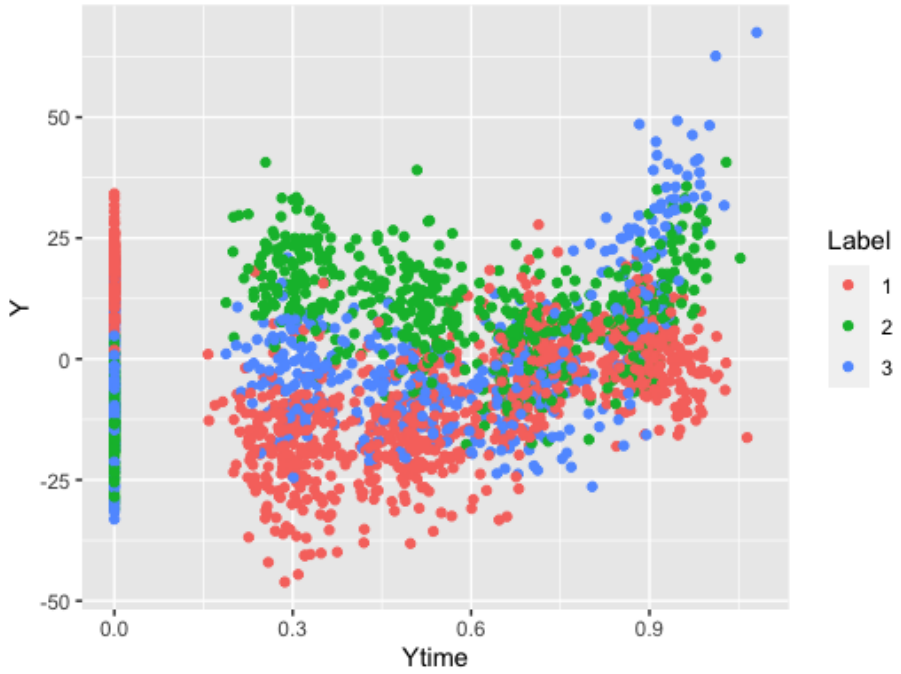


Figure 4.4: Y value labelled by subgroup

4.4 Label switching

We apply these two proposed methods for each simulated dataset and repeat the simulation 200 times with a sample size of $n = 200$.

Before assessing the estimation accuracy of component-specific parameters using the 200 replicates, it is important to address the label-switching problem [50, 62]. This issue arises due to the permutation invariance of the mixture likelihood, which can lead to inconsistent component labels across different replicates. Resolving this problem typically involves imposing constraints on the parameter estimates to ensure consistent labeling across replicates. However, different order constraints can result in significantly different labeling outcomes in the context of multivariate mixture regression. In our simulation studies, we employ an approach proposed by [63] to address the label-switching problem. This method identifies labels by maximizing the complete likelihood, helping mitigate the issue, and providing consistent and meaningful component labels. By applying this approach, we aim to obtain reliable and interpretable results while maintaining consistency across the replicates.

4.5 Initial values

To initialize the EM algorithm, we follow the procedure outlined in Chapter 3. Initially, we set the initial values of q_v and p_v as described in Chapter 3. The estimation of the mean function $\mu_v(t)$, eigenfunctions $\psi_{vk}(t)$, and error variance σ_v^2 is performed using the R package **fdapace**. We then fit regression splines to these estimated functions to obtain initial values for the spline coefficients $\boldsymbol{\Theta}_\mu$ and $\boldsymbol{\Theta}_\psi$. For the initial values of $\boldsymbol{\beta}$,

$\sigma^2_{b,c}$, σ^2_ϵ, and $\boldsymbol{\gamma}$, we utilize the R package **lcmm**. However, it should be noted that the **lcmm** package requires all time-varying covariates to be measured simultaneously as the response. To address this, we employ the last observation carried forward (LOCF) method to impute the missing covariates. This enables us to obtain the necessary initial values for the aforementioned parameters in the EM algorithm.

In our method, we employ multivariate functional principal component analysis (mFPCA) to jointly estimate the time-varying covariate processes. We summarize the mFPCA estimators obtained using our approach and compare them with the initial FPCA estimators obtained from the **fdapace** package. It is important to note that **fdapace** performs univariate FPCA separately for $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ using kernel smoothing. However, it does not provide estimators for the cross-covariance parameters between the two covariate processes. For any nonparametric function f, the Integral squared error (ISE) for its estimator $\widehat{f}$ is defined as

$$\text{ISE}(\widehat{f}) = \int_0^1 \{\widehat{f}(t) - f(t)\}^2 dt.$$

The posterior mean $\widetilde{\boldsymbol{\xi}}_i = \sum_{c=1}^{C} \widetilde{\boldsymbol{\xi}}_{ic}$, as defined in (2.10) or (3.3), can be used as an estimate of the principal component scores. We denote the final value of $\widetilde{\boldsymbol{\xi}}_i$ at the convergence of the EM algorithm as $\widehat{\boldsymbol{\xi}}_i$. Since the principal component scores of different orders can have different variances, we summarize the estimation results of the principal component scores using the relative mean squared error (RMSE) defined as

$$\text{RMSE}(\xi_{vk}) = \frac{\frac{1}{n}\sum_{i=1}^{n}(\widehat{\xi}_{i,vk} - \xi_{i,vk})^2}{\text{Var}(\xi_{vk})} \quad k = 1, \ldots, p_v, \quad v = 1, \ldots, d_x.$$

4.6 Results for fixed effect mixture component

Table 4.1 provides a summary of the mean and standard deviations (SD) for the integrated squared error (ISE) and relative mean squared error (RMSE) of the functional estimators, including the mean, eigenfunctions, and principal component scores for $X_1(t)$ and $X_2(t)$. It also includes the bias and SD of the eigenvalues, cross-covariance parameters, and error variance of $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$.

Table 4.1: Summary of FPCA estimators in the simulation study with fixed effect mixture components

X_1	$\mu_1(t)$	$\psi_{11}(t)$	$\psi_{12}(t)$	ξ_{11}	ξ_{12}
Emerald	0.182 (0.649)	-0.004 (0.236)	0.008 (0.249)	0.051 (0.022)	0.088 (0.031)
fdapace	-0.838 (0.081)	0.008 (0.121)	-0.023 (0.123)	0.056 (0.010)	0.280 (0.056)
X_2	$\mu_2(t)$	$\psi_{21}(t)$	$\psi_{22}(t)$	ξ_{21}	ξ_{22}
Emerald	0.001 (0.153)	-0.014 (0.302)	-0.012 (0.313)	0.087 (0.047)	0.109 (0.073)
fdapace	-0.156 (0.064)	0.161 (0.197)	-0.077 (0.148)	0.116 (0.023)	0.164 (0.050)

(a) Mean (SD) for the ISE of the functional estimators and RMSE of the FPCA scores

Eigenvalues	ω_{11}	ω_{12}	ω_{21}	ω_{22}
Emerald	0.195 (0.204)	0.672 (0.269)	0.074 (0.277)	0.125 (0.253)
fdapace	-1.747 (0.232)	-0.882 (0.173)	-2.016 (0.105)	-1.366 (0.054)

(b) Bias (SD) for the eigenvalues

Variance	σ_1^2	σ_2^2
Emerald	0.052 (0.078)	0.042 (0.070)
fdapace	0.460 (0.398)	2.381 (0.239)

(c) Bias (SD) for the error term in $\boldsymbol{W}$

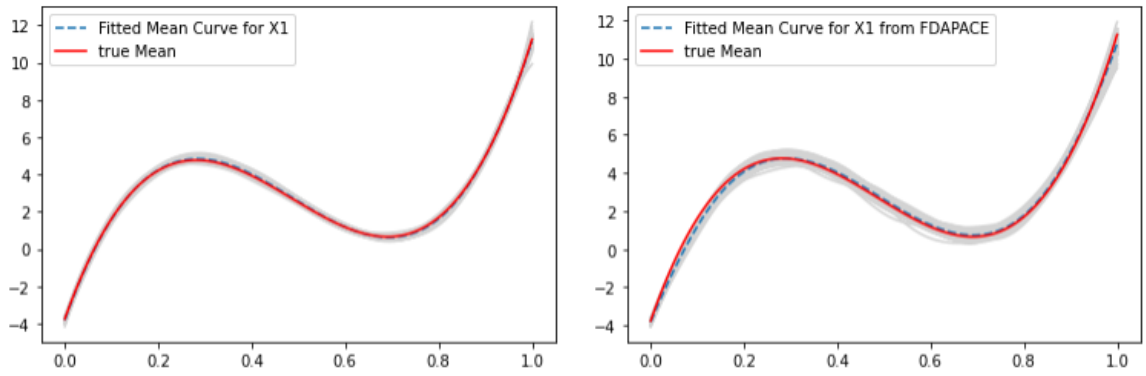
Cross-Covariance	$\omega_{12,11}$	$\omega_{12,11}$	$\omega_{12,21}$	$\omega_{12,22}$
Emerald	0.335 (0.212)	0.354 (0.234)	0.001 (0.206)	0.161 (0.115)

(d) Bias (SD) of the cross-covariance parameters

Whenever available, we provide counterparts of the **fdapace** package. The **fdapace** package includes its own built-in bandwidth selectors, and we utilize its default settings. One notable advantage of our approach is the simultaneous modeling of multiple cross-correlated longitudinal covariates and the response process. This enables information sharing among data from different sources, thereby improving the efficiency of FPCA estimation. The benefits of our algorithm are summarized in Table 4.1, clearly demonstrating its superior performance compared to **fdapace**. Our algorithm achieves higher accuracy and lower standard deviation values, as evidenced by the numerical results obtained in this simulation study. Furthermore, visual representations of our algorithm's performance and that of **fdapace** are presented in Figures 4.5, 4.6, 4.7, and 4.8, respectively. These plots depict the mean and eigenfunctions across 200 simulation runs, showcasing the excellent fit between our algorithm's predictions and the true functions.

Next, we summarize the mixture regression results. For comparison, we also compare our method with **lcmm** package by [40], which can fit mixture regression models to longitudinal data but require the covariates measured on the same time points as the response without measurement error. We consider two versions of **lcmm**:

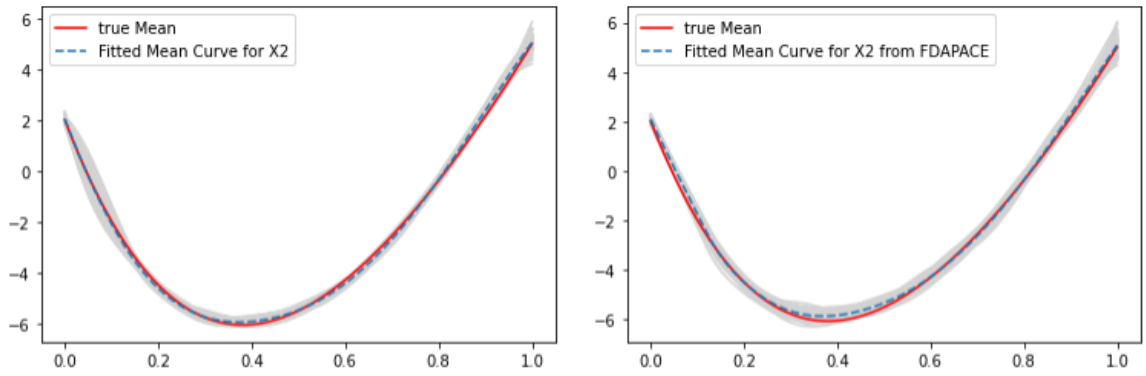
- *oracle* version, denoted as $\mathbf{lcmm}_O$, where we assume that the synchronized true covariates $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are observed without errors;
- *LOCF* version, denoted as $\mathbf{lcmm}_L$, where the missing true values of $\boldsymbol{X}$ are imputed using the last observed values.



(a) $\boldsymbol{X}_1$ from **Emerald** (b) $\boldsymbol{X}_1$ from **fdapace**

Figure 4.5: Fitted mean function for time-varying covariates $\boldsymbol{X}_1$ via 200 simulation runs with fixed effect mixture components in **Emerald** and **fdapace**. The solid red line plots the true function and dashed blue line plots out the average of the fitted function and the light gray lines are 200 fitted lines.

In addition, our method provides posterior probabilities $\widetilde{\pi}_{ic}$, which can be used to cluster subjects into subgroups. Similarly, the **lcmm** package also produces its own clustering results. We use the Adjusted Rand Index (ARI) [64] to compare the clustering results from different methods. It's important to note that both our algorithm and the **lcmm** package may flip the labels during the latent class modeling process. However, as long as the subjects are correctly divided into the correct subgroups, the ARI is an appropriate metric to compare the clustering results. The Rand Index (RI) considers all pairs of samples and counts pairs that are assigned to the same or different clusters in the predicted and true clusterings, while the adjusted Rand Index establishes a baseline by using the expected similarity of all pairwise comparisons between clusterings specified by a random model. By



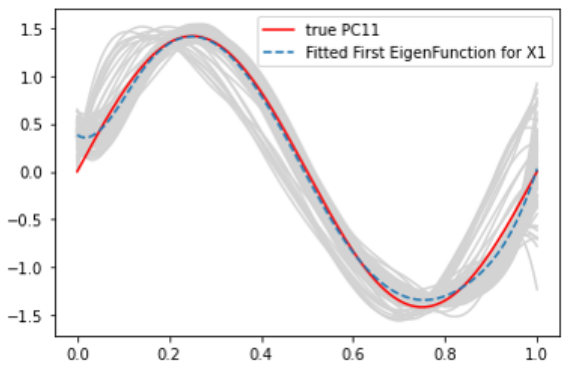
(a) $\boldsymbol{X}_2$ from **Emerald** (b) $\boldsymbol{X}_2$ from **fdapace**

Figure 4.6: Fitted mean function for time-varying covariates $\boldsymbol{X}_2$ via 200 simulation runs with fixed effect mixture components in **Emerald** and **fdapace**. The solid red line plots the true function and dashed blue line plots out the average of the fitted function and the light gray lines are 200 fitted lines.

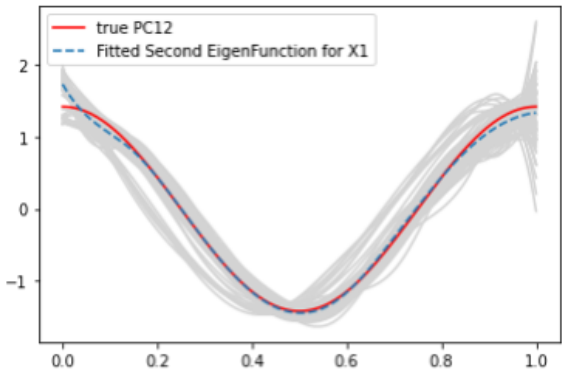
comparing the ARI values and the clustering results, we can assess the performance of our method relative to the **lcmm** package. The equation to calculate the adjusted rand index is

$$ARI = \frac{\text{RI} - \text{Expected RI}}{\max(\text{RI}) - \text{Expected RI}},$$

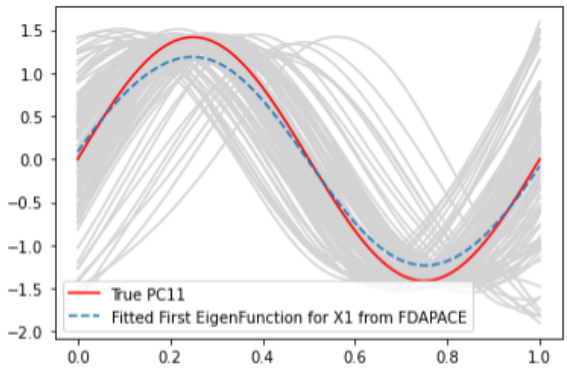
where RI represents the raw rand index score and the adjusted Rand index is thus ensured to have a value close to 0.0 for random labeling independent of the number of clusters and samples and exactly 1.0 when the clustering is identical. In other words, the ARI is a number between 0 and 1, with a higher ARI indicating a higher level of similarity between the clustering result with the true subgroup membership.



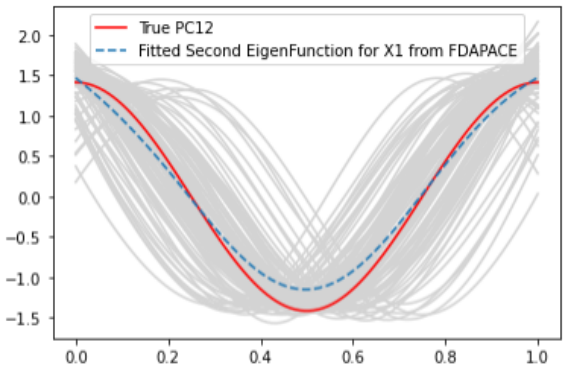
(a) $\boldsymbol{X}_1$ from **Emerald**



(b) $\boldsymbol{X}_1$ from **Emerald**



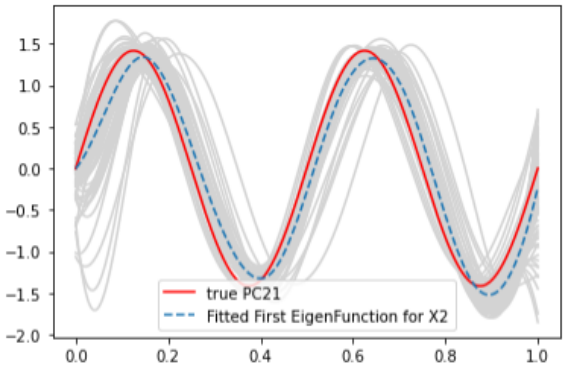
(c) $\boldsymbol{X}_1$ from **fdapace**



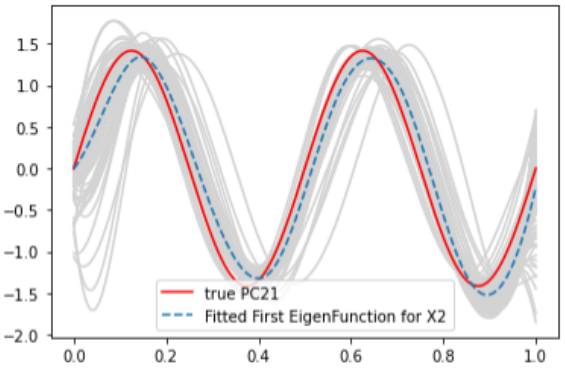
(d) $\boldsymbol{X}_1$ from **fdapace**

Figure 4.7: Fitted eigenfunction for time-varying covariates $\boldsymbol{X}_1$ via 200 simulation runs with fixed effect mixture components in **Emerald** and **fdapace**. The solid red line plots the true function and dashed blue line plots out the average of the fitted function and the light gray lines are 200 fitted lines.

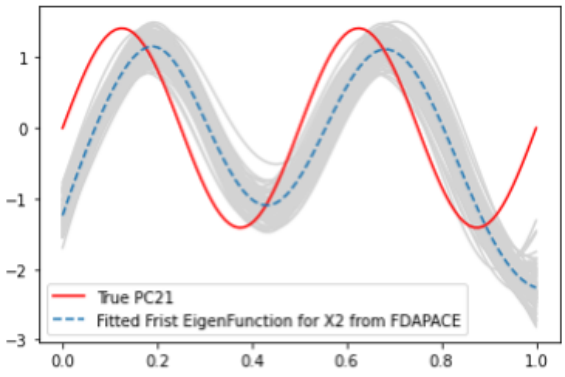
Results for the bias and standard deviations of various coefficients, including the logistic regression and mixture regression, are summarized in Table 4.2 and Table 4.3.



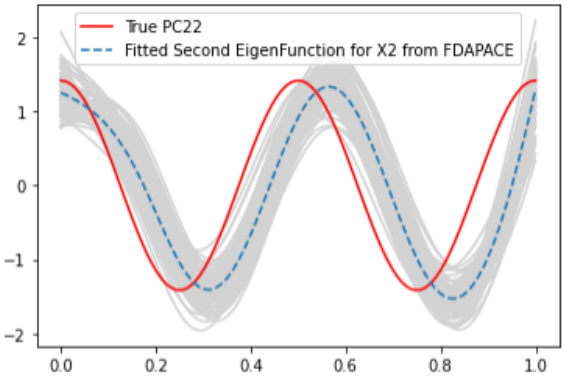
(a) $\boldsymbol{X}_2$ from **Emerald**



(b) $\boldsymbol{X}_2$ from **Emerald**



(c) $\boldsymbol{X}_2$ from **fdapace**



(d) $\boldsymbol{X}_2$ from **fdapace**

Figure 4.8: Fitted eigenfunction for time-varying covariates $\boldsymbol{X}_2$ via 200 simulation runs with fixed effect mixture components in **Emerald** and **fdapace**. The solid red line plots the true function and dashed blue line plots out the average of the fitted function and the light gray lines are 200 fitted lines.

A careful examination of Table 4.2 and 4.3 reveals that our method demonstrates superior performance compared to the *LOCF* version of **lcmm** in terms of the coefficients in the mixture model. In most cases, our method provides results that closely approximate

Table 4.2: Bias (SD) for the logistic regression parameters for the subgroup labels with fixed effect mixture components.

	lcmm$_O$	**lcmm**$_L$	**Emerald**
γ_{01}	0.229 (0.380)	0.410 (0.208)	-0.060 (0.358)
γ_{z11}	-0.148 (0.295)	-0.201(0.327)	0.103(0.572)
γ_{z21}	-1.037(1.170)	-1.597(2.880)	-0.020(0.248)
γ_{02}	0.937(1.173)	1.175(1.109)	-0.066(0.388)
γ_{z12}	0.199(0.302)	0.261(0.364)	0.138(0.841)
γ_{z22}	-0.235(0.379)	-0.284(0.421)	-0.012(0.332)

Table 4.3: Bias (SD) for the mixture regression coefficients with fixed effect mixture components.

	lcmm$_O$	**lcmm**$_L$	**Emerald**
$\beta_{0,1}$	-0.017 (0.166)	0.237 (0.540)	-0.033 (0.262)
$\beta_{x,11}$	0.002 (0.006)	-0.044 (0.064)	-0.021 (0.021)
$\beta_{x,12}$	-0.001 (0.005)	0.023 (0.063)	-0.025 (0.024)
$\beta_{z,11}$	0.018 (0.205)	-1.881 (0.518)	0.039 (0.273)
$\beta_{z,12}$	-0.020 (0.154)	1.676 (0.102)	-0.006 (0.126)
$\beta_{0,2}$	0.016 (0.172)	-1.155 (0.396)	-0.104 (0.358)
$\beta_{x,21}$	-0.006 (0.008)	1.322 (0.397)	-0.016 (0.043)
$\beta_{x,22}$	0.004 (0.009)	0.453 (0.262)	-0.031 (0.025)
$\beta_{z,21}$	-0.201 (1.391)	0.793 (0.366)	0.138 (0.812)
$\beta_{z,22}$	0.040 (0.169)	-0.436 (0.678)	-0.103 (0.065)
$\beta_{0,3}$	0.028 (0.236)	-0.316 (0.199)	0.079 (0.402)
$\beta_{x,31}$	-0.001 (0.006)	-0.143 (0.405)	-0.017 (0.041)
$\beta_{x,32}$	-0.001 (0.007)	-0.350 (0.227)	0.016 (0.050)
$\beta_{z,31}$	-0.011 (0.295)	0.159 (0.175)	-0.086 (0.082)
$\beta_{z,32}$	-0.001 (0.165)	0.007 (0.104)	0.097 (0.069)

those of the *oracle* version. Furthermore, our method achieves the highest classification accuracy, even surpassing the performance of the *oracle* version in **lcmm**. These findings indicate the excellent performance and robustness of our algorithm in estimating the mixture regression parameters and accurately classifying subjects into subgroups.

Table 4.4 presents a summary of the bias and standard deviations of the variance parameters, along with the ARI values. The bias and standard deviations reflect the accuracy and variability of the estimated variance components in our model, which demonstrate superior performance with smaller bias and standard deviation compared to the **lcmm** package. The ARI values further indicate that our model outperforms **lcmm**, even when considering *oracle* data, in terms of classification accuracy.

Table 4.4: Bias (SD) for the variance components and Mean (SD) for ARI in the 3-subgroup simulation study with fixed effect mixture components

Method	σ_ϵ^2	ARI
$\mathbf{lcmm}_O$	-0.100 (0.337)	0.974 (0.021)
$\mathbf{lcmm}_L$	1.158 (1.067)	0.750 (0.071)
Emerald	0.056 (0.313)	0.993 (0.001)

4.7 Results for mixed effect mixture components

We also conducted a simulation study for the mixed effect mixture components using datasets generated based on Section 4.3. In Table 4.5, we present a comprehensive summary of the simulation study results, including the mean and standard deviations (SD) for the ISE of the functional estimators, encompassing the mean and eigenfunctions for

$X_1(t)$ and $X_2(t)$, respectively. The ISE values reflect the accuracy of the estimated mean and eigenfunctions. The table also provides the bias and SD of the scalar parameters, such as the eigenvalues, cross-covariance parameters, and error variance of the covariate processes. These values offer insights into the accuracy and variability of the estimated scalar parameters. Additionally, the Relative Mean Squared Error (RMSE) of the Functional Principal Component (FPC) scores is reported, which indicates the precision of the estimated scores. These results allow for a comprehensive evaluation of the performance of our method in estimating the functional parameters and capturing the underlying variability in the covariate processes within the context of mixed effect mixture components.

The advantages of our algorithm are clearly demonstrated by the results summarized in Table 4.5. Our algorithm outperforms **fdapace** in terms of accuracy and stability when estimating the functional and scalar parameters. This is evident from the smaller Mean Integrated Squared Errors (ISE) (or biases) and smaller standard deviations of our estimators compared to **fdapace**. Furthermore, our algorithm provides more accurate predictions of the Principal Component (PC) scores, as indicated by the lower RMSE values. This implies that our method captures the underlying variability in the data more effectively and produces more reliable estimates of the PC scores. Additionally, our algorithm performs well in estimating the cross-covariance parameters, with reasonable performance indicated by the results. This further highlights the effectiveness of our approach in capturing the relationships between the different covariate processes.

Overall, the results presented in Table 4.5 demonstrate the superior performance of our algorithm in terms of accuracy, stability, and prediction of functional and scalar

Table 4.5: Summary of FPCA estimators in the simulation study with mixed effect mixture components

X_1	$\mu_1(t)$	$\psi_{11}(t)$	$\psi_{12}(t)$	ξ_{11}	ξ_{12}
Emerald	0.003(0.003)	0.031(0.073)	0.030(0.069)	0.046(0.037)	0.057(0.044)
fdapace	0.033(0.054)	0.268(0.551)	0.267(0.509)	0.258(0.378)	0.359(0.455)
X_2	$\mu_2(t)$	$\psi_{21}(t)$	$\psi_{22}(t)$	ξ_{21}	ξ_{22}
Emerald	0.014(0.014)	0.109(0.302)	0.117(0.249)	0.135(0.157)	0.151(0.225)
fdapace	0.021(0.027)	0.702(0.867)	0.642(0.651)	0.538(0.116)	0.764(0.144)

(a) Mean (SD) for the ISE of the functional estimators and RMSE of the FPC scores

Eigenvalues	ω_{11}	ω_{12}	ω_{21}	ω_{22}
Emerald	-0.103 (0.150)	-0.060 (0.580)	-0.031 (0.354)	0.019 (0.304)
fdapace	-1.661 (1.630)	-0.958 (1.331)	-2.134 (0.437)	-1.415 (0.296)

(b) Bias (SD) for the eigenvalues

Variance	σ_1^2	σ_2^2
Emerald	0.060 (0.063)	0.036 (0.063)
fdapace	0.024 (0.375)	2.680 (0.214)

(c) Bias (SD) for the error term in $\boldsymbol{W}$

Cross-Covariance	$\omega_{12,11}$	$\omega_{12,11}$	$\omega_{12,21}$	$\omega_{12,22}$
Emerald	0.130 (0.540)	-0.217 (0.285)	-0.075 (0.374)	0.139 (0.175)

(d) Bias (SD) of the cross-covariance parameters

parameters compared to the **fdapace** package. Figures W1 and W2 in the Supporting Materials present visual comparisons of the estimated mean and eigenfunctions using both **fdapace** and our proposed method. These plots allow us to visually assess the performance of the functional estimators by comparing them to the true functions and by showing the pointwise 2.5% and 97.5% percentiles of the estimators. The graphs provide additional evidence supporting the results presented in Table 4.5. Specifically, they demonstrate that

our proposed method produces functional estimators with less bias and variation compared to **fdapace**. By closely aligning with the true functions and exhibiting narrower confidence intervals, our method achieves higher accuracy and provides more reliable estimates of the mean and eigenfunctions. The visual comparisons in Figures W1 and W2 serve to reinforce the findings from the quantitative analysis in Table 4.5, further confirming the superior performance of our method in terms of functional estimation.

Table 4.6 offers a comprehensive summary of the bias and standard deviations of the logistic regression coefficients. The results clearly demonstrate the superior performance of our algorithm, **Emerald**, compared to **lcmm**. Our method outperforms **lcmm**L in terms of bias and standard deviation. Furthermore, even without time limitations, **Emerald** exhibits smaller bias and standard deviation compared to **lcmm**O. These findings highlight the robustness and effectiveness of our algorithm in accurately estimating the logistic regression coefficients. In terms of classification accuracy, our simulation studies demonstrate that **Emerald** achieves an average Adjusted Rand Index (ARI) of 0.9916, with a small standard deviation of 0.01 over 200 runs. In comparison, the average ARIs for **lcmm**O and **lcmm**L are 0.9895 and 0.9393, respectively, with larger standard deviations of 0.067 and 0.064.

Table 4.7 and 4.8 provides a summary of the bias and standard deviations of the various model parameters, including the mixture regression coefficients, and variances for measurement error and group-specific random intercept. The results demonstrate that our algorithm, **Emerald**, consistently outperforms the "last observation carried forward" approach used in **lcmm**, as evidenced by smaller bias and standard deviations. Notably,

Table 4.6: Bias (SD) for the logistic regression parameters for the subgroup labels

	lcmm$_O$	**lcmm**$_L$	**Emerald**
γ_{01}	0.229(0.380)	0.409(0.208)	-0.060(0.358)
γ_{z11}	-0.148(0.295)	-0.201(0.327)	0.103(0.572)
γ_{z21}	-1.037(1.170)	-1.597(2.880)	-0.020(0.248)
γ_{02}	0.937(1.173)	1.175(1.109)	-0.066(0.388)
γ_{z12}	0.199(0.302)	0.261(0.364)	0.138(0.841)
γ_{z22}	-0.235(0.379)	-0.284(0.421)	-0.012(0.332)

without time point limitations, our algorithm achieves comparable performance to **lcmm** with oracle data.

Table 4.7: Bias (SD) for the mixture regression coefficients

	lcmm$_O$	**lcmm**$_L$	**Emerald**
$\beta_{0,1}$	-0.017(0.166)	0.237(0.536)	-0.033(0.262)
$\beta_{x,11}$	0.002(0.006)	-0.044(0.064)	-0.021(0.021)
$\beta_{x,12}$	-0.001(0.005)	0.023(0.063)	-0.025(0.025)
$\beta_{z,11}$	0.018(0.205)	-1.88(0.518)	0.039(0.273)
$\beta_{z,12}$	-0.020(0.154)	1.676(0.102)	-0.006(0.126)
$\beta_{0,2}$	0.016(0.172)	-1.155(0.396)	-0.104(0.358)
$\beta_{x,21}$	-0.006(0.008)	1.322(0.397)	0.016(0.043)
$\beta_{x,22}$	0.004(0.009)	0.453(0.262)	-0.031(0.025)
$\beta_{z,21}$	-0.201(1.391)	0.793(0.366)	0.138(0.812)
$\beta_{z,22}$	0.040(0.169)	-0.436(0.678)	-0.103(0.065)
$\beta_{0,3}$	0.028(0.236)	-0.316(0.199)	0.079(0.402)
$\beta_{x,31}$	-0.001(0.006)	-0.143(0.405)	-0.017(0.041)
$\beta_{x,32}$	-0.001(0.007)	-0.350(0.227)	0.016(0.050)
$\beta_{z,31}$	-0.011(0.295)	0.159(0.175)	-0.086(0.082)
$\beta_{z,32}$	-0.010(0.165)	0.007(0.104)	0.097(0.069)

Table 4.8: Bias(SD) for the variance parameters in the 3-subgroup simulation study

Method	σ^2_{b1}	σ^2_{b2}	σ^2_{b3}	σ^2_{ϵ}
lcmm$_O$	-0.236 (0.615)	-0.077 (2.473)	1.036 (2.530)	1.466 (1.324)
lcmm$_L$	-1.465 (0.973)	0.143 (2.461)	-0.318 (2.045)	5.143 (0.300)
Emerald	-0.158(0.467)	-0.172(0.547)	-0.085(0.506)	0.968(0.943)

Furthermore, the results from Table 4.8 highlight the accuracy and stability of our algorithm in estimating these variances, further emphasizing its robust performance in capturing the underlying data characteristics. These findings indicate that **Emerald** out-

performs **lcmm** in terms of classification error rate, providing more accurate and consistent clustering results.

In conclusion, the results presented in Table 4.6, 4.7, and Table 4.8 confirm the superiority of our algorithm, **Emerald**, in estimating model parameters, capturing the variability in the data, and achieving accurate classification in comparison to the **lcmm** package. We can confident claim that **Emerald** is superior competitive.

4.8 Results for model selection

The model selection performance of **Emerald** is evaluated using two criteria: the Bayesian Information Criterion (BIC) and Akaike's Weight of Evidence (AWE). The results demonstrate that **Emerald** exhibits a high accuracy in correctly selecting the number of subgroups. Using BIC, **Emerald** achieves a selection accuracy of 76%, while using AWE, the accuracy improves to 77%. When **Emerald** fails to select the correct model, it tends to overestimate the number of subgroups rather than underestimate it. This tendency to err on the side of caution is generally considered preferable, as it avoids missing important findings. In contrast, the $\mathbf{lcmm}_L$ method, which employs the "last observation carried forward" approach, only achieves a correct model selection rate of 46%. This indicates that $\mathbf{lcmm}_L$ has a higher propensity for making incorrect model selections compared to **Emerald**.

In an additional simulation study presented in Section Appendix F, where the sample size is increased to 500, the probability of **Emerald** selecting the correct number of subgroups further improves to 89% using BIC and 90% using AWE. This highlights the

robustness and effectiveness of our method in selecting the appropriate number of subgroups as the sample size increases.

In summary, when the correctness of model selection is concerned, the results underscore the superior model selection performance of **Emerald** compared to $\mathbf{lcmm}_L$, with higher accuracy in identifying the correct number of subgroups and mitigating the risk of underestimating the complexity of the data.

4.9 Model implementation

Equations (2.2) and (3.1) demonstrate that fixed effect mixture components can be viewed as a special case of the mixed effect mixture model. By setting the variance components for the random effects to nearly zero, the mixed effect mixture components can be reduced into fixed effect mixture components. This implies that mixed-effect mixture models offer additional flexibility by incorporating random effects. In the real data analysis, we have chosen to apply a linear mixed effect model to the mixture components instead of a fixed effect model. The motivation behind this decision is to capture and explain more variation in the response variable by considering the random effects. By incorporating random effects, the model accounts for the potential correlations and heterogeneity among the observations, which can lead to a better understanding and interpretation of the data. Using a linear mixed effect model allows for the estimation of both fixed effects, which capture the population-level relationships between covariates and the response, and random effects, which capture the subject-specific deviations from the population-level trends. By including random effects, you are able to capture the individual-specific variations and

dependencies that may exist within the data. Overall, the decision to apply a linear mixed effect model to the mixture components in the real data analysis is driven by the goal of capturing additional variation and accounting for the complexities present in the data through the incorporation of random effects.

Chapter 5

Real Data Analysis

We now proceed with the analysis of the SWAN data, as described in Chapter 1. The objective of our analysis is to investigate the relationship between DHEAS levels and several cardiovascular and physical biomarkers, while also considering demographic backgrounds. To achieve this, we will conduct a comprehensive examination of the SWAN dataset, employing our proposed method to explore the associations between DHEAS levels and the selected biomarkers. Additionally, we will take into account demographic factors to better understand potential variations in these relationships across different subgroups. By investigating the relationship between DHEAS levels and cardiovascular and physical biomarkers, as well as considering demographic backgrounds, we aim to contribute valuable insights into the complex interplay between these variables in the SWAN population.

5.1 Pre-analysis

The time-varying covariates of interest include triglycerides (TGs), glucose (GLU), and systolic blood pressure (SBP). These measurements follow different schedules, as illustrated in Figure 1.1 and 1.2. TG and GLU measurements align with the schedule of cardiovascular biomarkers, while SBP measurements align with the schedule of physical biomarkers. Importantly, all these measurements are asynchronous with respect to DHEAS. In addition to the time-varying covariates, we consider baseline age, BMI, and race as time-invariant covariates. Race is categorized into four groups: White, Black, Asian, and Hispanic, accounting for 47.39%, 28.08%, 16.67%, and 7.87% of the SWAN subjects, respectively. To incorporate race as a covariate, we encode it using three dummy variables, with White as the reference category. These dummy variables serve as indicators for Black, Asian, and Hispanic, respectively, allowing us to examine the impact of race on the relationship between DHEAS levels and the biomarkers of interest.

It is important to note that the hormonal and physical biomarkers were measured over a period of up to 10 years. However, due to the study design, the cardiovascular biomarkers were only measured up to the 7th visit. Therefore, our analysis focuses on the data observed from baseline up to 3086 days, which corresponds to the completion of the 7th visit for all subjects. To facilitate the analysis, we rescale this time period to the interval [0,1], where 0 represents the baseline and 1 represents the completion of the 7th visit. Regarding the normal range of DHEAS levels for premenopausal or early perimenopausal women, [52] states that it typically falls between 13 and 240 $\mu g/dL$. However, it's worth mentioning that the SWAN data includes some abnormally high DHEAS levels. This is

evident from the visual representation of the underlying distribution of DHEAS in Figure 5.1, which shows a heavy right skew. To address the presence of extreme outliers in the

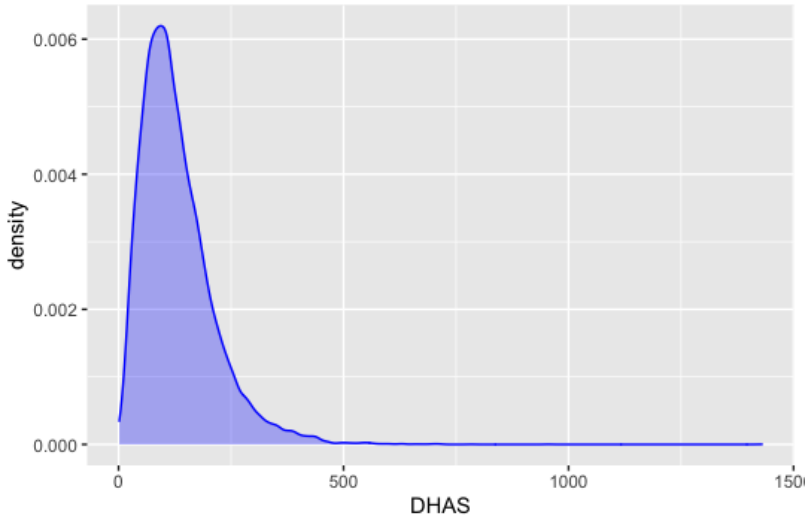


Figure 5.1: Density plot of original DHEAS

DHEAS data, a criterion based on the interquartile range (IQR) is applied. Observations with DHEAS values that exceed 3 times the IQR above the 3rd quartile are identified and removed from the dataset. This helps to mitigate the impact of these extreme values on the analysis. After applying this criterion, DHEAS values exceeding 457 are removed, which accounts for 3.81% of the total observations. Additionally, the variables TG, GLU, and SBP are standardized. This involves centering each variable around its marginal mean and dividing the centered variable by its marginal standard deviation. Standardization ensures that these variables are on a comparable scale, making interpretation and analysis easier.

5.2 Results for SWAN data

In accordance with the procedure described in Section 3.5, a univariate Functional Principal Component Analysis (FPCA) is performed on the three time-varying covariate processes using the **fdapace** package separately. This analysis allows for obtaining initial estimates of the functional parameters. To determine the number of principal components to retain for each covariate process, the AIC criterion, as proposed by [32], is employed. In this particular analysis, 2 principal components are selected for each covariate process.

Moving forward, the focus shifts to determining the number of subgroups present in the SWAN data. Table 5.1 provides the values of the AWE and BIC for different numbers of subgroups. Both criteria suggest that there are 2 subgroups present in the SWAN data. Based on these results, the assumption of 2 subgroups is adopted for subsequent analyses.

Table 5.1: AWE and BIC value for SWAN data

	AWE	BIC
1-subgroup	14730.69	14730.69
2-subgroup	14109.91	13910.98
3-subgroup	14216.09	14116.62
4-subgroup	14280.13	13981.73
5-subgroup	14452.06	14062.34

Utilizing the posterior probabilities $\widetilde{\pi}_{ic}$, a classification is performed, assigning a total of 2802 subjects to Group 1 and 336 subjects to Group 2. The average DHEAS level for Group 1 is 115, while for Group 2 it is 252. This distinction is illustrated in Figure 5.2, which displays a boxplot visualization of the DHEAS levels for the two groups. The average

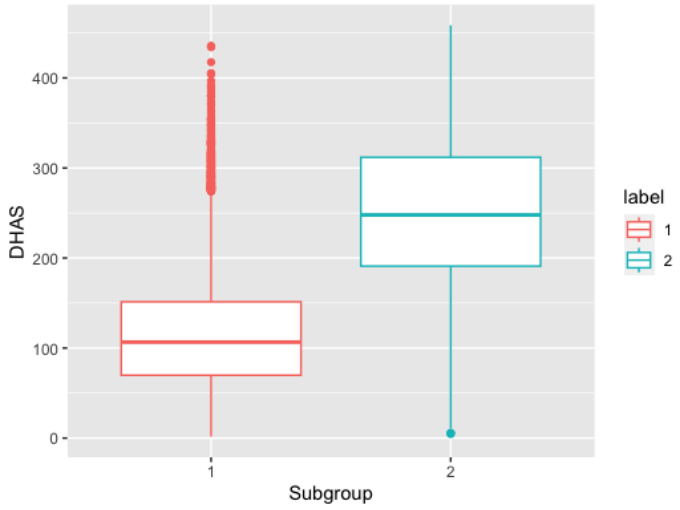


Figure 5.2: Boxplot of DHEAS colored by subgroups

value of DHEAS across the eight visits ranges from 113 to 118 for Group 1 and from 240 to 259 for Group 2. The boxplot in Figure 5.3 displays the distribution of DHEAS levels for the two groups across the eight visits. The results indicate that Group 2 consistently exhibits higher levels of DHEAS compared to Group 1. Based on the DHEAS level and

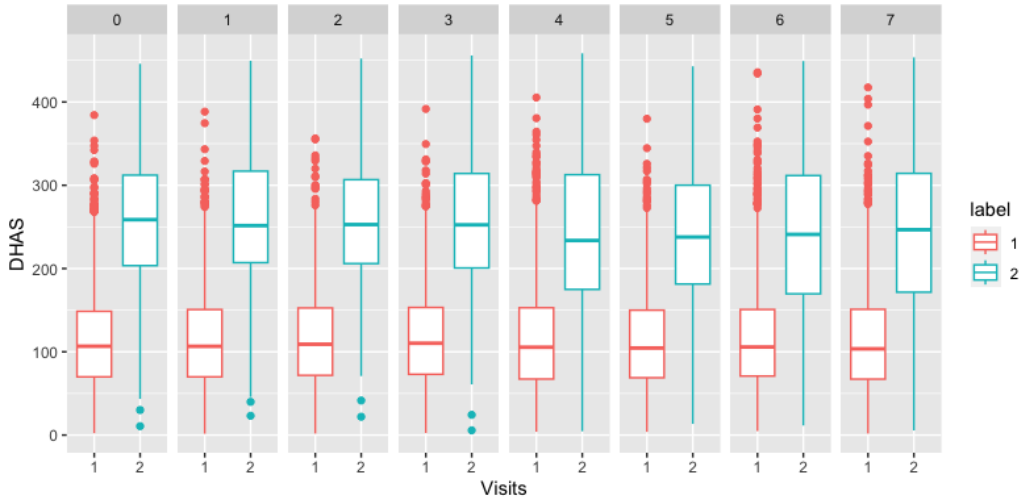


Figure 5.3: Boxplot of DHEAS colored by subgroups via 8 visits

number of memberships, we can interpret Group 1 as the majority subgroup consisting of menopausal women with normal DHEAS levels, while Group 2 represents a minority subgroup characterized by high DHEAS levels. Notably, the identification of Group 2 as a distinct subgroup is a unique finding provided by our new method.

In terms of racial composition, Group 1 largely reflects the overall racial distribution of the entire SWAN study population. However, Group 2 shows a lower proportion of Black individuals (17.9%) and Hispanics (3.3%) compared to the study population. For a detailed breakdown of the frequency of subjects in each cluster by race, please refer to Table 5.2. The median age for Group 1 is 46, while for Group 2 it is 45. The median BMI

Table 5.2: Subjects in Each subgroup by Race

Race	Subgroup 1	Subgroup 2
White	1291	196
Black	821	60
Asian	454	69
Hispanic	236	11

is 26.2 for Group 1 and 25.1 for Group 2.

Table 5.3 provides percentiles of age and BMI for the two subgroups, allowing for a more detailed understanding of the distribution of these variables within each subgroup. To assess the relationship between race and subgroup membership, we conducted a chi-square test for independence and found a significant correlation, indicating that race is associated with subgroup membership. This suggests that race is a potential factor that contributes to the differentiation of subgroups. Furthermore, we performed a t-test to compare the age

Table 5.3: Percentile of Time-invariant Variable

Percentile	Age	BMI
0%	42	15
25%	43	22.5
50%	45	25.1
75%	47	30.1
100%	52	57

(a) Subgroup 1

Percentile	Age	BMI
0%	42	15.4
25%	44	22.6
50%	46	26.2
75%	48	31.5
100%	53	64.8

(b) Subgroup 2

difference between the subgroups and found a significant difference. Similarly, for BMI, there was a significant difference in BMI levels between Group 1 and Group 2. These statistical tests provide evidence that there are demographic differences, specifically in terms of age and BMI, across the identified subgroups.

Table 5.4 presents the estimated coefficients of the logistic regression with their corresponding standard errors. To obtain the standard errors, we employed a parametric bootstrap procedure. In this procedure, we used the observed $\mathbf{Z}$ and the observed time points t_{ij} and s_{ivl} to regenerate $\mathbf{L}$, ξ, $\mathbf{W}$, and $\mathbf{Y}$ based on the estimated model. The proposed method was then applied to each bootstrap sample using the same tuning parameters as in the real data analysis. The standard errors reported in Table 5.4 and 5.5 were computed

based on 200 bootstrap samples, providing a measure of the variability of the estimated coefficients.

Table 5.4: Estimate (SD) of the coefficients for logistic regression: $\widehat{\boldsymbol{\gamma}}$ in Equation (2.1)

	Est (SE)
Intercept	1.574 (0.040)
Age	0.343 (0.015)
BMI	0.0599 (0.021)
Black	0.700 (0.039)
Asian	0.004 (0.081)
Hispanic	0.726 (0.110)

The logistic regression coefficients presented in Table 5.4 provide valuable insights into the factors which can influence subgroup membership. The results reveal that baseline age, BMI level, and race significantly contribute to the classification into Group 1, representing the normal subgroup characterized by relatively lower DHEAS levels. Specifically, the findings indicate that Black and Hispanic women, older women, and women with higher BMI are more likely to be classified into this subgroup. These results align with previous research on DHEAS levels in different racial and ethnic groups. Studies such as [29] have consistently reported lower DHEAS levels among African Americans and Hispanics compared to white and Asian women. The observed inverse relationship between age and DHEAS levels, as demonstrated by studies conducted by [43, 3, 10], further supports the association found in our analysis. Additionally, the association between higher BMI or obesity and lower DHEAS levels has been well-documented in the literature. Taken together, these findings reinforce the influence of age, BMI level, and race on subgroup membership

and contribute to our understanding of the associations between these factors and DHEAS levels in women.

Table 5.5 presents the estimated coefficients of the mixture regression model, along with their corresponding standard errors. These standard errors were computed using the same bootstrap procedure defined before. When we focus on examining the coefficients for Group 1, which represents the majority of middle-aged women and serves as the health norm, the results show that there is a positive association between DHEAS levels and GLU and SBP, while there is a negative association with TG, Age, and BMI. These findings are consistent with the existing literature on DHEAS. Furthermore, our analysis reveals racial disparities in DHEAS levels. Women of color, particularly black women, have lower DHEAS levels compared to white women. This finding aligns with previous studies that have consistently reported racial differences in DHEAS levels. On the other hand, for women classified into Group 2, characterized by high DHEAS levels, the associations with GLU and BMI are more pronounced. Specifically, there is a stronger positive association with GLU and a stronger negative association with BMI compared to Group 1. However, the associations with TG and SBP exhibit opposite signs compared to Group 1. These results provide valuable insights into the associations between DHEAS levels and other covariates within each subgroup, highlighting the distinct patterns observed in Group 1 and Group 2.

There are new findings regarding time-varying variables and in Figures 5.4 to 5.6, we display the partial effect plots illustrating the relationship between DHEAS and the time-varying covariates GLU, TG, and SBP, respectively, within each subgroup. These plots provide visualizations of the different effects of these variables on DHEAS levels in

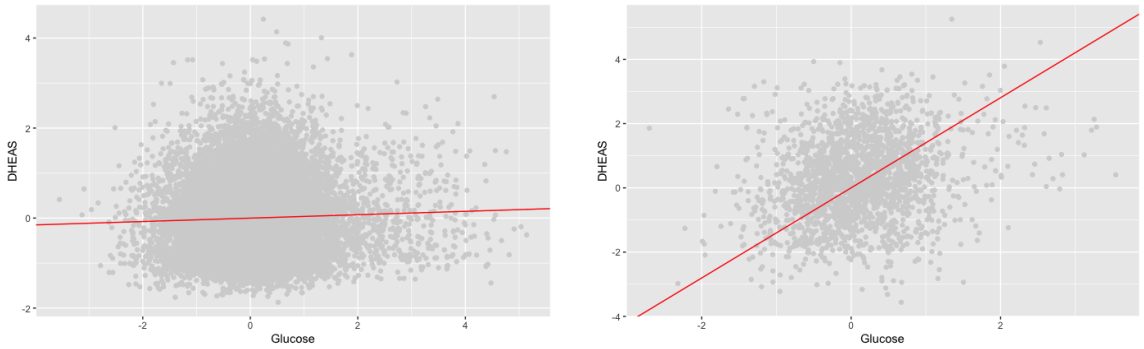
Table 5.5: Estimate (SD) of the coefficients for mixture regression: $\widehat{\boldsymbol{\beta}}$ in Equation (3.1)

	Group 1	Group 2
Intercept	-0.169 (0.013)	1.261 (0.051)
Glucose	0.038 (0.008)	1.403 (0.166)
Triglycerides	-0.052 (0.008)	0.578 (0.232)
Systolic Blood Pressure	0.020 (0.009)	-0.859 (0.225)
Age	-0.063 (0.005)	-0.012 (0.021)
BMI	-0.043 (0.006)	-0.183 (0.028)
Black	-0.255 (0.012)	0.232 (0.047)
Asian	-0.134 (0.017)	-0.371 (0.050)
Hispanic	-0.134 (0.019)	-0.206 (0.119)

Group 1 and Group 2. By analyzing these partial effect plots, we can effectively ascertain the direction of the relationship between time-varying variables and the response variable.

The concept of partial effect allows us to examine how a one-unit change in a specific independent variable influences the value of the response variable (DHEAS) while keeping all other variables constant. By examining the partial effect plots, we can observe the distinct patterns of association between DHEAS and the time-varying covariates within each subgroup. Furthermore, these plots enable us to make a visual comparison of the effects of GLU, TG, and SBP on DHEAS levels between Group 1 and Group 2, thereby highlighting the differences between these two subgroups. To assess the partial effect of a particular variable, such as glucose, it is necessary to calculate the change in the dependent variable (in this case, DHEAS) that results from the change in the independent variable of interest. In detail, for any time-varying covariate $X_{ivj}(t_{ij})$, its partial effect plot in subgroup c is done by plotting $Y_{ij} - \beta_{0,c} - \sum_{v' \neq v} \widehat{X}_{iv'j}\widehat{\beta}_{x,cv'} - \boldsymbol{Z}_{ij}^{\top}\beta_{z,c}$ against $\widehat{X}_{ivj}$, where

$\widehat{X}_{ivj} = \widehat{\mu}_{x,v}(t_{ij}) + \sum_{k=1}^{p_v} \widehat{\xi}_{ivk}\widehat{\psi}_{vk}(t_{ij})$ is the posterior mean of the missing, synchronized time-varying covariate value. These plots provide a visual representation of how the relationship between DHEAS and each time-varying covariate varies across the subgroups. The differing slopes highlight the different effects and associations of these covariates with DHEAS levels in Group 1 and Group 2. This visual analysis further supports the identification of subgroup-specific patterns and reinforces the importance of considering the subgroups when examining the relationship between DHEAS and the time-varying covariates.

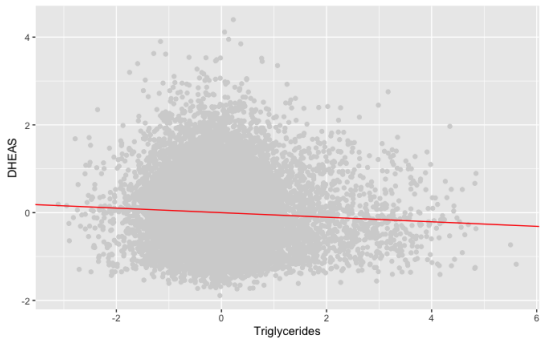


(a) Glucose v.s DHEAS in Group 1

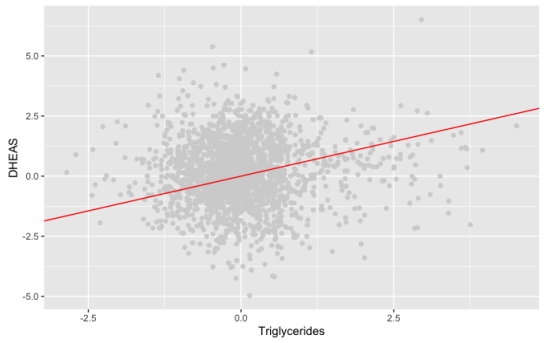
(b) Glucose v.s DHEAS in Group 2

Figure 5.4: Partial effect plots for glucose versus DHEAS. The red line plots our fitted coefficients.

To further visualize the time-varying covariates, we present Figure 5.7 and Figure 5.8. These figures display the mean curves with observations for glucose, triglycerides, and systolic blood pressure, along with their principal component effect plots. In the plots,

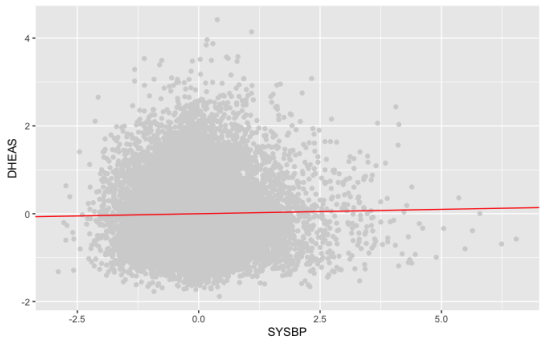


(a) Triglycerides v.s DHEAS in Group 1

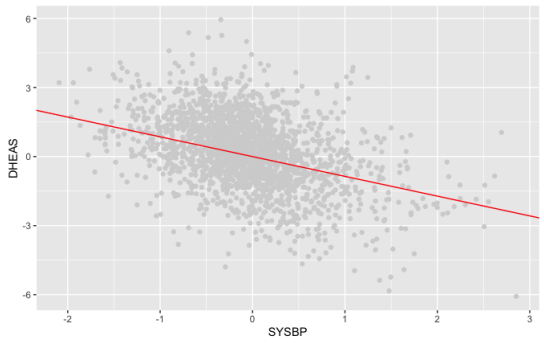


(b) Triglycerides v.s DHEAS in Group 2

Figure 5.5: Partial effect plots for triglycerides versus DHEAS. The red line plots our fitted coefficients.

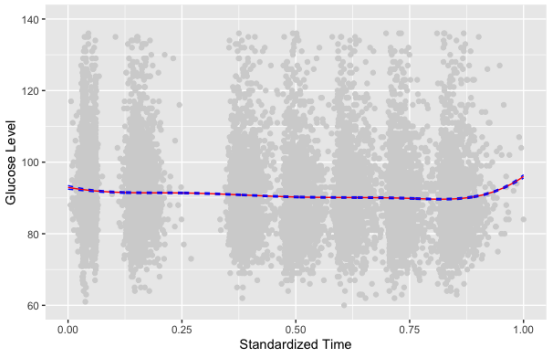


(a) Systolic BP v.s DHEAS in Group 1

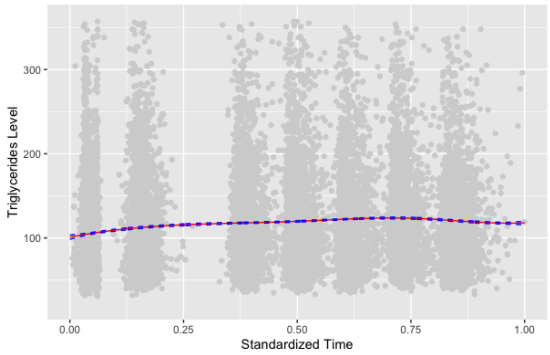


(b) Systolic BP v.s DHEAS in Group 2

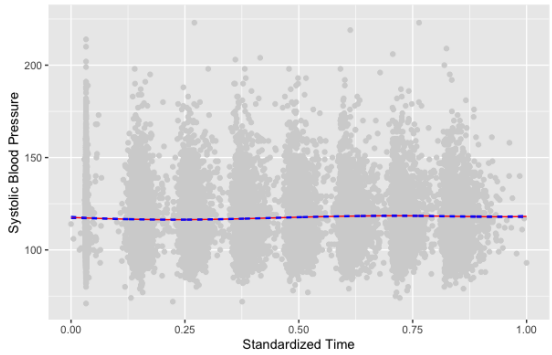
Figure 5.6: Partial effect plots for systolic blodd pressure versus DHEAS. The red line plots our fitted coefficients.



(a) Mean curve of glucose



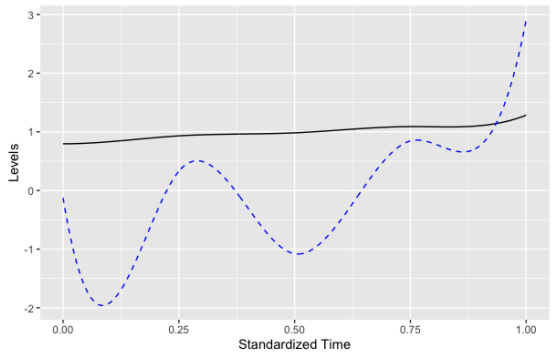
(b) Mean curve of triglycerides



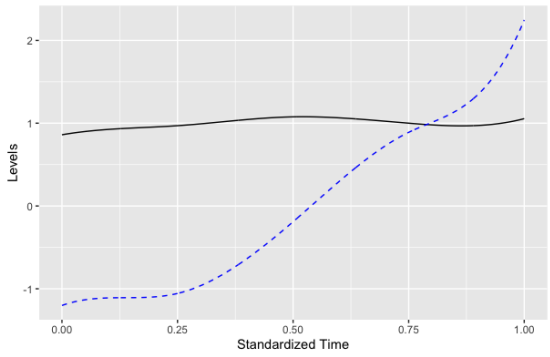
(c) Mean curve of systolic BP

Figure 5.7: SWAN data: fitted mean for glucose, triglycerides, and systolic BP. Plots show the fitted mean functions (solid red), pointwise 95% bootstrap confidence intervals (dashed blue), as well as the scatter plot of the observations.

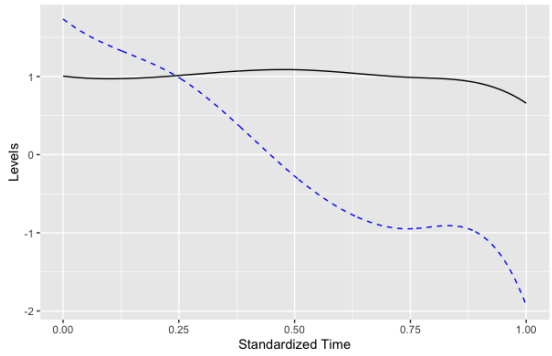
we can observe that the confidence bands for the fitted mean curves are relatively narrow. This is due to the large sample size in our analysis, which provides more precise estimates. Regarding the principal component effect plots, the first eigenfunction appears to show a constant trend, indicating a random intercept component that captures the variations in the



(a) Eigenfunctions of glucose



(b) Eigenfunctions of triglycerides



(c) Eigenfunctions of systolic BP

Figure 5.8: SWAN data: fitted eigenfunction for glucose, triglycerides, and systolic BP. Plots show the top two eigenfunctions as solid black and dashed blue, respectively.

time-varying variables across the observations. On the other hand, the second eigenfunction exhibits an approximately linear trend, representing a random slope component that accounts for the varying rates of change in the time-varying variables. These eigenfunctions contribute to the interpretation of the variations observed in the time-varying covariates.

Overall, these visualizations allow us to explore and interpret the relationships between DHEAS and the time-varying covariates, shedding light on the patterns and trends within the data.

The **fdapace** package provides insights into the cumulative variance explained by the principal components for each covariate. For glucose, the first four principal components explain 90.73%, 95.37%, 98.85%, and 99.66% of the total variation. For triglycerides, the first three principal components explain 92.03%, 96.93%, and 99.31% of the total variation. Lastly, for systolic blood pressure, the first three principal components explain 92.63%, 97.94%, and 99.88% of the total variation. The first principal component accounts for 94.59%, 94.78%, and 93.57% of the total variation in glucose, triglycerides, and systolic blood pressure, respectively, in the **Emerald** dataset. To compare the results obtained from **Emerald** with those from the **fdapace** package, the mean and principal component effect plots are provided in Appendix E. These plots offer a visual comparison of the patterns and trends observed in the **Emerald** dataset and the results obtained using the **fdapace** package.

The results reported in Table 5.5 confirm previous research findings that age and BMI are negatively correlated with DHEAS levels [65, 3]. This aligns with the existing literature, which suggests that as women age increases or have higher BMI, their DHEAS levels tend to decrease. The negative correlation indicates that older age and higher BMI are associated with lower DHEAS levels in the study population.

Indeed, the time-varying covariates, baseline information, and demographic findings presented in this study align with the existing understanding that DHEAS production

decreases with age and can be influenced by various factors, including body composition and hormonal changes.

Chapter 6

Conclusions

6.1 Summary

The previous contradictory and inconclusive studies regarding the relationship between DHEAS and other cardiovascular and physical biomarkers highlight the need to consider potential subgroups within the general population. However, conducting a subgroup analysis of DHEAS levels in menopausal women using a large longitudinal database from SWAN poses challenges due to asynchronous and error-prone time-varying covariates. To overcome these challenges, we propose an innovative semiparametric mixture regression model called **Emerald**. In the **Emerald** model, we address these complexities by simultaneously modeling multiple asynchronous, error-prone, cross-correlated, time-varying covariate processes using a multivariate Functional Principal Component Analysis (FPCA) approach. Additionally, we establish a connection between the longitudinal response process and the covariates by employing a linear mixed model with coefficients and random effects dependent on latent subgroup memberships. To capture subgroup membership, we

utilize a multinomial logistic regression that incorporates time-invariant covariates. To facilitate flexible semiparametric modeling of the longitudinal processes, we employ reduced rank polynomial spline approximation for estimating the unknown mean and eigenfunctions of the covariate processes. The unknown parameters are estimated using an EM algorithm, and the number of subgroups is determined using data-driven methods such as BIC and AWE.

Numerical studies demonstrate the superior performance of **Emerald** compared to existing methods. In the analysis of the SWAN data, we identify two important subgroups among the women in the study. The dominant group exhibits DHEAS dynamics consistent with existing literature, while the minority group displays elevated DHEAS levels that have more pronounced associations with cardiovascular biomarkers such as glucose and triglycerides.

6.2 Future work

To further enhance the capabilities of **Emerald**, we are currently focusing on integrating variable selection techniques. The SWAN dataset contains a substantial number of covariates, making it crucial to identify the most informative ones. To accomplish this, we employ coordinate descent, enabling us to iteratively update the mixture coefficients individually. This approach facilitates efficient optimization and enhances the accuracy of variable selection. By identifying the most relevant covariates contributing to subgroup differentiation, we improve the overall interpretability of the model. Additionally, to account for the randomness in the response variables, we propose jointly modeling the time-varying

covariates with the response variable using multivariate Functional Principal Component Analysis (mFPCA). This allows us to capture the nonlinear relationships and dependencies between the covariates and the response variables, further enhancing the model's ability to uncover meaningful insights.

By incorporating variable selection techniques and jointly modeling the time-varying covariates with the response variables using mFPCA, we aim to enhance the predictive power, interpretability, and robustness of **Emerald** in analyzing the SWAN dataset. These advancements will provide valuable insights into the relationships between DHEAS levels, cardiovascular biomarkers, and subgroup differentiation among menopausal women.

Bibliography

[1] Revlin Abbi, Elia El-Darzi, Christos Vasilakis, and Peter Millard. Analysis of stopping criteria for the EM algorithm in the context of patient grouping according to length of stay. *4th International IEEE Conference Intelligent Systems*, 1:3–9, 2008.

[2] Serkan Akogul and Erisoglu Murat. An approach for determining the number of clusters in a model-based cluster analysis. *Entropy*, 19:452, 2017.

[3] Rowyda N Al-Harithy. Dehydroepiandrosterone sulfate levels in women. Relationships with body mass index, insulin and glucose levels. *Saudi Medical Journal*, 24(8):837–841, 2003.

[4] Kazutaka Aoki and Yasuo Terauchi. Effect of dehydroepiandrosterone (DHEA) on diabetes mellitus and obesity. *Vitamins and Hormones*, 108:355–365, 2018.

[5] X Bai, K Chen, and W Yao. Mixture of linear mixed models using multivariate t distribution. *Journal of Statistical Computation and Simulation*, 86:771–787, Mar 2016.

[6] Michal Benko, Wolfgang Härdle, and Alois Kneip. Common functional principal components. *Annals of Statistics*, 37(1):1–34, 2009.

[7] D Carroll, AC Phillips, JM Lord, W Arlt, and GD Batty. Cortisol, dehydroepiandrosterone sulphate, their ratio and hypertension: evidence of associations in male veterans from the vietnam experience study. *Journal of Human Hypertension*, 25(7):418–424, 2011.

[8] Raymond J Carroll, David Ruppert, Leonard A Stefanski, and Ciprian M Crainiceanu. *Measurement Error in Nonlinear Models: A Modern Perspective*. Chapman and Hall/CRC, Boca Raton, FL, 2nd edition edition, 2006.

[9] G. Celeux, M. Hurn, and C.P. Robert. Computational and inferential difficulties with mixture posterior distributions. *Journal of American Statistical Association*, 95:957–970, 2000.

[10] Mei-Jou Chen, Chin-Der Chen, Jehn-Hsiahn Yang, Chi-Ling Chen, Hong-Nerng Ho,

Wei-Shiung Yang, and et al. High serum dehydroepiandrosterone sulfate is associated with phenotypic acne and a reduced risk of abdominal obesity in women with polycystic ovary syndrome. *Human Reproduction*, 26:227–234, 2011.

[11] Jeng Min Chiou, Yu Ting Chen, and Ya Fang Yang. Multivariate functional principal component analysis: A normalization approach. *Statistica Sinica*, 24(4):1571–1596, 2014.

[12] Ciprian M. Crainiceanu, Ana-Maria Staicu, and Chong-Zhi Di. Generalized multilevel functional regression. *Journal of the American Statistical Association*, 104(488):1550–1561, 2009.

[13] A. J. Cron and M. West. Efficient classification-based relabeling in mixture models. *Journal of the American Statistical Association*, 65:16–20, 2011.

[14] V.M. de Bruin, M.C. Vieira, M.N. Rocha, and G.S. Viana. Cortisol and dehydroepiandosterone sulfate plasma levels and their relationship to aging, cognitive function, and dementia. *Brain and Cognition*, 50(2):316–323, 2002.

[15] J.J. Dziak, R. Li, S. Tan, X. Shiffman, and M.P. Shiyko. Modeling intensive longitudinal data with mixtures of nonparametric trajectories and time-varying effects. *Psychological Methods*, 20:444–469, 2015.

[16] S. Frühwirth-Schnatter. *Finite Mixture and Markov Switching Models*. Springer, 2006.

[17] Noreen Goldman and Dana A.Glei. Sex differences in the relationship between dheas and health. *Experimental Gerontology*, 42(42):979–987, 2007.

[18] Eleanor M. Gurnell, Penelope J. Hunt, Suzanne E. Curran, Catherine L. Conway, Eleanor M. Pullenayegum, Felicia A. Huppert, and et al. Long-term DHEA replacement in primary adrenal insufficiency: a randomized, controlled trial. *Journal of Clinical Endocrinology and Metabolism*, 93(2):400–409, 2008.

[19] Clara Happ and Sonja Greven. Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association*, 113(522):649–659, 2018.

[20] Christian Hennig. Identifiablity of models for clusterwise linear regression. *Journal of Classification*, 17(2):273–296, 2000.

[21] Tailen Hsing and Randall Eubank. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. John Wiley and Sons, Chichester, West Sussex, UK, 2015.

[22] Hui Huang, Yehua Li, and Yongtao Guan. Joint modeling and clustering paired generalized longitudinal trajectories with application to cocaine abuse treatment data. *Journal of the American Statistical Association*, 508(109):1412–1424, 2014.

[23] M. Huang, R. Li, H. Wang, and W. Yao. Estimating mixture of Gaussian processes by kernel smoothing. *Journal of Business and Economic Statistics*, 32:259–270, 2014.

[24] M. Huang, S. Wang, W. Yao, and Y. Chen. Statistical inference and applications of mixture of varying coefficient models. *Scandinavian Journal of Statistics*, 45(3):618–643, 2018.

[25] Penelope J. Hunt, Eleanor M. Gurnell, Felicia A. Huppert, Christine Richards, A. Toby Prevost, John A. H. Wass, and et al. Improvement in mood and fatigue after dehydroepiandrosterone replacement in Addison's disease in a randomized, double blind trial. *Journal of Clinical Endocrinology and Metabolism*, 85(12):4650–4656, 2000.

[26] Gareth M. James, Trevor J. Hastie, and Catherine A. Sugar. Principal component models for sparse functional data. *Biometrika*, 87(3):587–602, 2000.

[27] A. Jasra, C. C. Holmes, and Stephens D. A. Markov chain monte carlo methods and the label switching problem in bayesian mixture modeling. *Statistical Science*, 20:50–67, 2005.

[28] Onyedikachi O John. Robustness of quantile regression to outliers. *American Journal of Applied Mathematics and Statistics*, 3(2):86–88, 2015.

[29] Bill L. Lasley, Nanette Santoro, John F. Randolf, Ellen B. Gold, Sybil Crawford, Gerson Weiss, and et al. The relationship of circulating dehydroepiandrosterone, testosterone, and estradiol to stages of the menopausal transition and ethnicity. *Journal of Clinical Endocrinology and Metabolism*, 87(8):3760–3767, 2002.

[30] Richard A. Levine and Juanjuan Fan. An automated (markov chain) monte carlo em algorithm. *Journal of Statistical Computation and Simulation*, 74(5):349–360, 2004.

[31] Yehua Li and Tailen Hsing. Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Annals of Statistics*, 38(6):3321–3351, 2010.

[32] Yehua Li, Naisyin Wang, and Raymond J Carroll. Selecting the number of principal components in functional data. *Journal of the American Statistical Association*, 108(504):1284–1294, 2013.

[33] Nicole Maninger, Owen M. Wolkowitz, Victor I. Reus, Elissa S. Epel, and Synthia H.

Mellon. Neurobiological and neuropsychiatric effects of dehydroepiandrosterone (dhea) and dhea sulfate (dheas). *Frontiers in Neuroendocrinology*, 30(1):65–91, 2009.

[34] G. J. McLachlan and D. Peel. *Finite Mixture Models*. John Wiley and Sons, New York, 2000.

[35] A. J. Morales, R. H. Haubrich, J. Y. Hwang, H. Asakura, and S. S. C. Yen. The effect of six months treatment with a 100 mg daily dose of dehydroepiandrosterone (dhea) on circulating sex steroids, body composition and muscle strength in age-advanced men and women. *Clinical Endocrinology*, 49(4):421–432, 1998.

[36] Charles A Morgan, Steve Southwick, Gary Hazlett, Ann Rasmusson, Gary Hoyt, Zoran Zimolo, and et al. Relationships among plasma dehydroepiandrosterone sulfate and cortisol levels, symptoms of dissociation, and objective performance in humans exposed to acute stress. *Archives of General Psychiatry*, 61:819–825, 2004.

[37] Hien Duy Nguyen and Faicel Chamroukhi. Practical and theoretical aspects of mixture-of-experts modeling: An overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4):1246, 2018.

[38] Xiongfeng Pan, Xinyin Wu, Atipatsa C. Kaminga, Shi wu Wen, and Aizhong Liu. Dehydroepiandrosterone and Dehydroepiandrosterone sulfate in Alzheimer's disease: A systematic review and meta-analysis. *Frontiers in Aging Neuroscience*, 11:61, 2019.

[39] Debashis Paul and Jie Peng. Consistency of restricted maximum likelihood estimators of principal components. *Annals of Statistics*, 37(3):1229 – 1271, 2009.

[40] Cecil Proust-Lima, Viviane Philipps, and Benoit Liquet. Estimation of extended mixed models using latent classes and latent processes: The r package lcmm. *Journal of Statistical Software*, 78(2):1–56, 2017.

[41] Ying Qin, Heitor O. Santos, Vahid Khani, Shing Cheng Tan, and Yingpeng Zhi. Effects of dehydroepiandrosterone (dhea) supplementation on the lipid profile: A systematic review and dose-response meta-analysis of randomized controlled trials. *Nutrition, Metabolism and Cardiovascular Diseases*, 30(9):1465–1475, 2020.

[42] James O Ramsay and Bernhard W Silverman. *Functional Data Analysis*. Springer, New York, 2nd edition, 2005.

[43] Nikolaos Samaras, Dimitrios Samaras, Emilia Frangos, Alexandre Froster, and Jacques Philippe. A review of age-related dehydroepiandrosterone decline and its association with well-known geriatric syndromes: is treatment beneficial? *Rejuvenation Research*, 16(4):285–294, 2013.

[44] Heribert Schunkert, Hans-Werner Hense, Tilo Andus, Günter A.J. Riegger, and Rainer H. Straub. Relation between dehydroepiandrosterone sulfate and blood pressure levels in a population-based sample. *American Journal of Hypertension*, 12:1140–1143, 1999.

[45] Martin Sewell. *Principal Component Analysis*. CiteSeer, 2008.

[46] A. Skrondal and S. Rabe-Hesketh. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Chapman and Hall/CRC, London, UK, 1st edition, 2004.

[47] M. Sperrin, T. Jaki, and E. Wit. Probabilistic relabelling strategies for the label switching problem in bayesian mixture models. *Statistics and Computing*, 20(3):357–366, 2010.

[48] Manivannan Srinivasan, Brian A. Irving, Ketan Dhatariya, Katherine A. Klaus, Stacy J. Hartman, Joseph P. McConnell, and K. Sreekumaran Nair. Effect of dehydroepiandrosterone replacement on lipoprotein profile in hypoadrenal women. *Journal of Clinical Endocrinology and Metabolism*, 94(3):761–764, 2009.

[49] Luboslav Stárka, Michaela Dušková, and Martin Hill. Dehydroepiandrosterone: A neuroactive steroid. *Journal of Steroid Biochemistry and Molecular Biology*, 145:254–260, 2015.

[50] Matthew Stephens. Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B*, 62(4):795–809, 2000.

[51] X. Tan, M.P. Shiyko, R. Li, Y. Li, and L. Dierker. A time-varying effect model for intensive longitudinal data. *Psychological Methods*, 17:61–77, 2012.

[52] Rathna Kumari Udayakumar and K Padma. Dehydroepiandrosterone levels in type 2 diabetes. *International Journal of Medical Research and Health Sciences*, 3:411, 2014.

[53] Theodore B. Vanltallie. Stress: a risk factor for serious illness. *Metabolism*, 51:40–45, 2002.

[54] D. Von Mühlen, G. A. Laughlin, D. Kritz-Silverstein, J. Bergstrom, and R. Bettencourt. Effect of dehydroepiandrosterone supplementation on bone mineral density, bone markers, and body composition in older adults: the dawn trial. *Osteoporosis International*, 19(5):699–707, 2008.

[55] Furong Wang, Yanfei He, Heitor O. Santos, Brijesh Sathian, James C. Price, and Jianjun Diao. The effects of dehydroepiandrosterone (DHEA) supplementation on

body composition and blood pressure: A meta-analysis of randomized clinical trials. *Steroids*, 163:108710, 2020.

[56] Hung Tsung Wang, Shu Man Chen, Shin Da Lee, Mei Chich Hsu, Kun Ning Chen, Yuh Feng Liou, and et al. The role of DHEA-S in the mood adjustment against negative competition outcome in golfers. *Journal of Sports Sciences*, 27(3):291–297, 2009.

[57] S. Wang, M. Huang, X. Wu, and W. Yao. Mixture of functional linear models and its application to CO2-GDP functional data. *Computational Statistics and Data Analysis*, 97:1–15, 2016.

[58] Raymond KW Wong, Yehua Li, and Zhengyuan Zhu. Partially linear functional additive models for multivariate functional data. *Journal of the American Statistical Association*, 114(525):406–418, 2019.

[59] Sijia Xiang, Weixin Yao, and Guangren Yang. An Overview of Semiparametric Extensions of Finite Mixture Models. *Statistical Science*, 34(3):391 – 404, 2019.

[60] Fang Yao, Yuejiao Fu, and Thomas C.M.Lee. Functional mixture regression. *Biostatistics*, 12:341–353, 2011.

[61] Fang Yao, Hans-Georg Müller, and Jane-Ling Wang. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590, 2005.

[62] W. Yao and B. G. Lindsay. Bayesian mixture labeling by highest posterior density. *Journal of American Statistical Association*, 104:758–767, 2009.

[63] Weixin Yao. Label switching and its solutions for frequentist mixture models. *Journal of Statistical Computation and Simulation*, 85(5):1000–1012, 2015.

[64] Ka Yee Yeung and Walter Ruzzo. Principal component analysis for clustering gene expression data. *Bioinformatics*, 17(9):763–774, 2001.

[65] D G Young, G Skibinski, J I Mason, and K James. The influence of age and gender on serum dehydroepiandrosterone sulphate (DHEA-S), IL-6, IL-6 soluble receptor (IL-6 sR) and transforming growth factor beta 1 (TGF-beta1) levels in normal healthy blood donors. *Clinical and Experimental Immunology*, 117(3):476–481, 1999.

[66] Lan Zhou, Jianhua Z. Huang, and Raymond J. Carroll. Joint modelling of paired sparse functional data using principal components. *Biometrika*, 95:601–619, 2008.

Appendix A: Details on the observed data likelihood function in Chapter 2

The conditional distribution of $\boldsymbol{L}, \boldsymbol{Y}, \boldsymbol{W}$ and $\boldsymbol{\xi}$ given $\boldsymbol{Z}$ is expressed as following

$$f(\boldsymbol{L}, \boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{\xi} | \boldsymbol{Z}) \quad = \quad f(\boldsymbol{Y} | \boldsymbol{L}, \boldsymbol{\xi}, \boldsymbol{Z}) \times f(\boldsymbol{L} | \boldsymbol{Z}) \times f(\boldsymbol{W} | \boldsymbol{\xi}) \times f(\boldsymbol{\xi}),$$

where $[\boldsymbol{Y} | L_c = 1, \boldsymbol{Z}, \boldsymbol{\xi}] \sim N\{\beta_{0,c}\mathbf{1} + \sum_{v=1}^{d_x} \beta_{x,cv}(\boldsymbol{\mu}_v + \boldsymbol{\Psi}_v \boldsymbol{\xi}) + \boldsymbol{Z}^\top \boldsymbol{\beta}_{z,c}, \sigma_\epsilon^2 \boldsymbol{I}\}$, $[L_c = 1 | \boldsymbol{Z}]$ is a multinomial distribution that has C categories with each category is π_c probability, $(\boldsymbol{W}_v | \boldsymbol{\xi}_v) \sim N(\boldsymbol{\mu}_v^* + \boldsymbol{\Psi}_v^* \boldsymbol{\xi}_v, \sigma_v^2 \boldsymbol{I})$ and $\boldsymbol{\xi} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\xi)$.

Hence, the joint distribution of $\boldsymbol{L}, \boldsymbol{\xi}, \boldsymbol{Y}$, and $\boldsymbol{W}$ conditional on $\boldsymbol{Z}$ can be written as

$$\begin{aligned} f(\boldsymbol{L}, \boldsymbol{\xi}, \boldsymbol{Y}, \boldsymbol{W} | \boldsymbol{Z}) \quad = \quad & \prod_{c=1}^{C} \{ \frac{\pi_c}{(2\pi\sigma_\epsilon^2)^{\frac{m_{y,i}}{2}}} \exp[-\frac{1}{2\sigma_\epsilon^2} \{\boldsymbol{Y} - \beta_{0c}\mathbf{1} - \sum_{v=1}^{d_x} \beta_{x,cv}(\boldsymbol{\mu}_v + \boldsymbol{\Psi}_v \boldsymbol{\xi}_v) - \boldsymbol{Z}\boldsymbol{\beta}_{z,c}\}^{\otimes 2}] \\ & \times \prod_{v=1}^{d_x} [\frac{1}{(2\pi\sigma_v^2)^{\frac{m_{xv,i}}{2}}} \exp\{-\frac{1}{2\sigma_v^2} (\boldsymbol{W}_v - \boldsymbol{\mu}_v^* - \boldsymbol{\Psi}_v^* \boldsymbol{\xi}_v)^{\otimes 2}\}] \\ & \times \frac{1}{(2\pi)^{\frac{p}{2}}} |\Sigma|^{-\frac{1}{2}} \exp(-\frac{1}{2} \boldsymbol{\xi}^\top \boldsymbol{\Sigma}_\xi^{-1} \boldsymbol{\xi})\}^{L_c}, \end{aligned} \tag{A.1}$$

where for a matrix A, we define $A^{\otimes 2} = A^\top A$. Also, with the smoothing spline expression, we can have

$$\begin{aligned} f(\boldsymbol{Y}_i, \boldsymbol{W}_i | \boldsymbol{Z}_i) \quad = \quad & \int f(\boldsymbol{\xi}_i, \boldsymbol{Y}_i, \boldsymbol{W}_i | \boldsymbol{Z}_i) d\boldsymbol{\xi}_i = \int \sum_{c=1}^{C} f(L_{ic} = 1, \boldsymbol{\xi}_i, \boldsymbol{Y}_i, \boldsymbol{W}_i | \boldsymbol{Z}_i) d\boldsymbol{\xi}_i \\ = \quad & \int \sum_{c=1}^{C} f(L_{ic} = 1 | \boldsymbol{Z}_i) f(\boldsymbol{Y}_i | \boldsymbol{\xi}_i, \boldsymbol{Z}_i, L_{ic}) f(\boldsymbol{W}_i | \boldsymbol{\xi}_i) f(\boldsymbol{\xi}_i) d\boldsymbol{\xi}_i \\ = \quad & \int \sum_{c=1}^{C} \frac{\pi_{ic}}{(2\pi)^{(m_{y,i}+m_{x,i}+p)/2} \sigma_\epsilon^{m_{y,i}}} |\boldsymbol{\Lambda}_i|^{-\frac{1}{2}} |\boldsymbol{\Sigma}_\xi|^{-\frac{1}{2}} \exp\{-\frac{1}{2\sigma_\epsilon^2} (\widetilde{\boldsymbol{Y}}_{ic} - \widetilde{\boldsymbol{\Psi}}_{ic} \boldsymbol{\xi}_i)^{\otimes 2} \end{aligned}$$

$$
\begin{aligned}
&-\frac{1}{2}(\widetilde{\boldsymbol{W}}_i-\boldsymbol{\Psi}_i^*\boldsymbol{\xi}_i)^\top\boldsymbol{\Lambda}_i^{-1}(\widetilde{\boldsymbol{W}}_i-\boldsymbol{\Psi}_i^*\boldsymbol{\xi}_i)-\frac{1}{2}\boldsymbol{\xi}_i^\top\boldsymbol{\Sigma}_\xi^{-1}\boldsymbol{\xi}_i\}d\boldsymbol{\xi}_i \\
=&\sum_{c=1}^{C}\frac{\pi_{ic}}{(2\pi)^{(m_{y,i}+m_{x,i}+p)/2}\sigma_\epsilon^{m_{y,i}}}|\boldsymbol{\Lambda}_i|^{-\frac{1}{2}}|\boldsymbol{\Sigma}_\xi|^{-\frac{1}{2}}\exp(-\frac{1}{2\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic}^\top\widetilde{\boldsymbol{Y}}_{ic}-\frac{1}{2}\widetilde{\boldsymbol{W}}_i^\top\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i) \\
&\times\int\exp\Bigg[-\frac{1}{2}\Big\{\boldsymbol{\xi}_i^\top(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{\Psi}}_{ic}^\top\widetilde{\boldsymbol{\Psi}}_{ic}+\boldsymbol{\Psi}_i^{*\top}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\Psi}_i^*+\boldsymbol{\Sigma}_\xi^{-1})\boldsymbol{\xi}_i-(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic}^\top\widetilde{\boldsymbol{\Psi}}_{ic}+\widetilde{\boldsymbol{W}}_i^\top\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\Psi}_i^*)\boldsymbol{\xi}_i \\
&-\boldsymbol{\xi}_i^\top(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{\Psi}}_{ic}^\top\widetilde{\boldsymbol{Y}}_{ic}+\boldsymbol{\Psi}_i^{*\top}\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i)\Big\}\Bigg]d\boldsymbol{\xi}_i \\
=&\sum_{c=1}^{C}\frac{\pi_{ic}}{(2\pi)^{(m_{y,i}+m_{x,i}+p)/2}\sigma_\epsilon^{m_{y,i}}}|\boldsymbol{\Lambda}_i|^{-\frac{1}{2}}|\boldsymbol{\Sigma}_\xi|^{-\frac{1}{2}}\exp(-\frac{1}{2\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic}^\top\widetilde{\boldsymbol{Y}}_{ic}-\frac{1}{2}\widetilde{\boldsymbol{W}}_i^\top\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i) \\
&\times(2\pi)^{\frac{p}{2}}|\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{\Psi}}_{ic}^\top\widetilde{\boldsymbol{\Psi}}_{ic}+\boldsymbol{\Psi}_i^{*\top}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\Psi}_i^*+\boldsymbol{\Sigma}_\xi^{-1}|^{-\frac{1}{2}}\exp\{\frac{1}{2}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic}^\top\widetilde{\boldsymbol{\Psi}}_{ic}+\widetilde{\boldsymbol{W}}_i^\top\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\Psi}_i^*) \\
&\times(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{\Psi}}_{ic}^\top\widetilde{\boldsymbol{\Psi}}_{ic}+\boldsymbol{\Psi}_i^{*\top}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\Psi}_i^*+\boldsymbol{\Sigma}_\xi^{-1})(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic}^\top\widetilde{\boldsymbol{\Psi}}_{ic}+\widetilde{\boldsymbol{W}}_i^\top\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\Psi}_i^*)^\top\} \\
=&\sum_{c=1}^{C}\frac{\pi_{ic}}{\sigma_\epsilon^{m_{y,i}}(2\pi)^{\frac{m_{y,i}+m_{x,i}}{2}}}|\boldsymbol{\Lambda}_i|^{-\frac{1}{2}}|\boldsymbol{\Sigma}_\xi|^{-\frac{1}{2}}|\boldsymbol{\Omega}_{ic}|^{\frac{1}{2}}\exp\Big\{-\frac{1}{2}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic}^\top\widetilde{\boldsymbol{Y}}_{ic}+\widetilde{\boldsymbol{W}}_i^\top\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i \\
&-\widetilde{\boldsymbol{\xi}}_{ic}^\top\boldsymbol{\Omega}_{ic}^{-1}\widetilde{\boldsymbol{\xi}}_{ic})\Big\}
\end{aligned}
\tag{A.2}
$$

with $\boldsymbol{\Omega}_{ic}$ and $\widetilde{\boldsymbol{\xi}}_{ic}$ defined as

$$
\begin{aligned}
\boldsymbol{\Omega}_{ic} &= (\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{\Psi}}_{ic}^\top\widetilde{\boldsymbol{\Psi}}_{ic}+\boldsymbol{\Psi}_i^{*\top}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\Psi}_i^*+\boldsymbol{\Sigma}_\xi^{-1})^{-1}; \\
\widetilde{\boldsymbol{\xi}}_{ic} &= \boldsymbol{\Omega}_{ic}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{\Psi}}_{ic}^\top\widetilde{\boldsymbol{Y}}_{ic}+\boldsymbol{\Psi}_i^{*\top}\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i).
\end{aligned}
\tag{A.3}
$$

And $m_{x,i}=\sum_{v=1}^{d_x}m_{x,iv}, p=\sum_{v=1}^{d_x}p_v$, $\widetilde{\boldsymbol{Y}}_{ic}=\boldsymbol{Y}_i-\beta_{0,c}\mathbf{1}_{m_{y,i}}-\sum_{v=1}^{d_x}\beta_{x,cv}\boldsymbol{B}_{iv}\boldsymbol{\theta}_{\mu v}-\boldsymbol{Z}_i\boldsymbol{\beta}_{z,c}$ models all the fixed effect in $\boldsymbol{Y}_i$ except for the random function, and $\widetilde{\boldsymbol{W}}_i=\boldsymbol{W}_i-\boldsymbol{B}_i^*\boldsymbol{\Theta}_\mu$ models all the fixed effect in $\boldsymbol{W}_i$ except for the random function. Here, $\boldsymbol{W}_i$ is a long vector that contains all $\boldsymbol{W}$ variables for subject i. $\boldsymbol{B}_i^*=\mathrm{diag}(\boldsymbol{B}_{i1}^*,\ldots,\boldsymbol{B}_{id_x}^*)$ is a $m_{x,i}\times q$ matrix with diagonal block $\boldsymbol{B}_{iv}^*$ and $\boldsymbol{\Theta}_\mu=(\boldsymbol{\theta}_{\mu 1}^\top,\ldots,\boldsymbol{\theta}_{\mu d_x}^\top)^\top$ is a $q\times 1$ vector that contains all the coefficients for splines in the mean function, where $q=\sum_{v=1}^{d_X}q_v$. $\boldsymbol{B}_{iv}^*$ is the matrix of

spline basis, with each row representing the basis function at $\boldsymbol{X}_{iv}$'s observed time. Also denote $\boldsymbol{\Psi}_i^* = \boldsymbol{B}_i^* \boldsymbol{\Theta}_\psi$, with $\boldsymbol{\Theta}_\psi = \text{diag}(\boldsymbol{\Theta}_{\psi 1}, \ldots, \boldsymbol{\Theta}_{\psi d_x})$ be the $q \times p$ matrix with diagonal block $\boldsymbol{\Theta}_{\psi v}$. Here $\boldsymbol{B}_{iv}$ is the matrix of spline basis, with each row representing the basis function at $\boldsymbol{Y}_i$'s observed time, which is different from $\boldsymbol{B}_{iv}^*$. Last, denoting the discrete matrix of eigenfunction as $\widetilde{\boldsymbol{\Psi}}_{ic} = (\widetilde{\boldsymbol{\Psi}}_{i1c}, \widetilde{\boldsymbol{\Psi}}_{i2c}, \ldots, \widetilde{\boldsymbol{\Psi}}_{id_xc})$ with $\widetilde{\boldsymbol{\Psi}}_{ivc} = \beta_{x,cv} \boldsymbol{B}_{iv} \boldsymbol{\Theta}_{\psi v}$. Last, $\boldsymbol{\Lambda}_i$ is a diagonal square matrix with diagonal entry $\sigma_v^2 \boldsymbol{I}_{m_{x,iv}}$.

Appendix B: Details on the observed data likelihood function in Chapter 3

The likelihood for the observed data on the ith subject is

$$
\begin{aligned}
f(\boldsymbol{Y}_i, \boldsymbol{W}_i|\boldsymbol{Z}_i) &= \int \sum_{c=1}^{C} f(L_{ic}=1, \boldsymbol{\xi}_i, \boldsymbol{Y}_i, \boldsymbol{W}_i|\boldsymbol{Z}_i) d\boldsymbol{\xi}_i \\
&= \int \sum_{c=1}^{C} f(L_{ic}=1|\boldsymbol{Z}_i) f(\boldsymbol{Y}_i|\boldsymbol{\xi}_i, \boldsymbol{Z}_i, L_{ic}=1) f(\boldsymbol{W}_i|\boldsymbol{\xi}_i) f(\boldsymbol{\xi}_i) d\boldsymbol{\xi}_i \\
&= \int \sum_{c=1}^{C} \frac{\pi_{ic}}{(2\pi)^{(m_{y,i}+m_{x,i}+p)/2}} |\boldsymbol{V}_{ic}|^{-\frac{1}{2}} |\boldsymbol{\Lambda}_i|^{-\frac{1}{2}} |\boldsymbol{\Sigma}_\xi|^{-\frac{1}{2}} \times \exp\Bigg[-\frac{1}{2} \boldsymbol{\xi}_i^\top \boldsymbol{\Sigma}_\xi^{-1} \boldsymbol{\xi}_i \\
&\quad -\frac{1}{2}\big\{ (\widetilde{\boldsymbol{Y}}_{ic} - \widetilde{\boldsymbol{\Psi}}_{ic}\boldsymbol{\xi}_i)^\top \boldsymbol{V}_{ic}^{-1} (\widetilde{\boldsymbol{Y}}_{ic} - \widetilde{\boldsymbol{\Psi}}_{ic}\boldsymbol{\xi}_i) + (\widetilde{\boldsymbol{W}}_i - \boldsymbol{\Psi}_i^* \boldsymbol{\xi}_i)^\top \boldsymbol{\Lambda}_i^{-1} (\widetilde{\boldsymbol{W}}_i - \boldsymbol{\Psi}_i^* \boldsymbol{\xi}_i) \big\} \Bigg] d\boldsymbol{\xi}_i \\
&= \sum_{c=1}^{C} \frac{\pi_{ic}}{(2\pi)^{(m_{y,i}+m_{x,i}+p)/2}} |\boldsymbol{V}_{ic}|^{-\frac{1}{2}} |\boldsymbol{\Lambda}_i|^{-\frac{1}{2}} |\boldsymbol{\Sigma}_\xi|^{-\frac{1}{2}} \times \exp(-\frac{1}{2} \widetilde{\boldsymbol{Y}}_{ic}^\top \boldsymbol{V}_{ic}^{-1} \widetilde{\boldsymbol{Y}}_{ic} - \frac{1}{2} \widetilde{\boldsymbol{W}}_i^\top \boldsymbol{\Lambda}_i^{-1} \widetilde{\boldsymbol{W}}_i) \\
&\quad \times \int \exp\Bigg[-\frac{1}{2} \Big\{ \boldsymbol{\xi}_i^\top (\widetilde{\boldsymbol{\Psi}}_{ic}^\top \boldsymbol{V}_{ic}^{-1} \widetilde{\boldsymbol{\Psi}}_{ic} + \boldsymbol{\Psi}_i^{*\top} \boldsymbol{\Lambda}_i^{-1} \boldsymbol{\Psi}_i^* + \boldsymbol{\Sigma}_\xi^{-1}) \boldsymbol{\xi}_i - (\widetilde{\boldsymbol{Y}}_{ic}^\top \boldsymbol{V}_{ic}^{-1} \widetilde{\boldsymbol{\Psi}}_{ic} + \widetilde{\boldsymbol{W}}_i^\top \boldsymbol{\Lambda}_i^{-1} \boldsymbol{\Psi}_i^*) \boldsymbol{\xi}_i \\
&\qquad\qquad -\boldsymbol{\xi}_i^\top (\widetilde{\boldsymbol{\Psi}}_{ic}^\top \boldsymbol{V}_{ic}^{-1} \widetilde{\boldsymbol{Y}}_{ic} + \boldsymbol{\Psi}_i^{*\top} \boldsymbol{\Lambda}_i^{-1} \widetilde{\boldsymbol{W}}_i) \Big\} \Bigg] d\boldsymbol{\xi}_i, \qquad \text{(B.1)}
\end{aligned}
$$

where $\boldsymbol{V}_{ic} = \sigma_{b,c}^2 \mathbf{1}_{m_{y,i}} \mathbf{1}_{m_{y,i}}^\top + \sigma_\epsilon^2 \boldsymbol{I}$ represents the covariance matrix for $\boldsymbol{Y}_i$ in cluster c, and the definition of $\widetilde{\boldsymbol{Y}}_{ic}, \widetilde{\boldsymbol{W}}_i, \boldsymbol{B}_i^*, \boldsymbol{\Theta}_\mu, \boldsymbol{\Psi}_i^*, \widetilde{\boldsymbol{\Psi}}_{ic}$, and $\boldsymbol{\Lambda}_i$ are the same as (A.2). Define

$$
\begin{aligned}
\boldsymbol{\Omega}_{ic} &= (\widetilde{\boldsymbol{\Psi}}_{ic}^\top \boldsymbol{V}_{ic}^{-1} \widetilde{\boldsymbol{\Psi}}_{ic} + \boldsymbol{\Psi}_i^{*\top} \boldsymbol{\Lambda}_i^{-1} \boldsymbol{\Psi}_i^* + \boldsymbol{\Sigma}_\xi^{-1})^{-1}, \\
\widetilde{\boldsymbol{\xi}}_{ic} &= \boldsymbol{\Omega}_{ic} (\widetilde{\boldsymbol{\Psi}}_{ic}^\top \boldsymbol{V}_{ic}^{-1} \widetilde{\boldsymbol{Y}}_{ic} + \boldsymbol{\Psi}_i^{*\top} \boldsymbol{\Lambda}_i^{-1} \widetilde{\boldsymbol{W}}_i). \qquad \text{(B.2)}
\end{aligned}
$$

Given the new notation of posterior mean and covariance matrix, $\widetilde{\boldsymbol{\xi}}_{ic}$ and $\boldsymbol{\Omega}_{ic}$, we can simplify our conditional distribution as below

$$
\begin{aligned}
f(\boldsymbol{Y}_i, \boldsymbol{W}_i|\boldsymbol{Z}_i) &= \int \sum_{c=1}^{C} f(L_{ic}=1, \boldsymbol{\xi}_i, \boldsymbol{Y}_i, \boldsymbol{W}_i|\boldsymbol{Z}_i) d\boldsymbol{\xi}_i \\
&= \int \sum_{c=1}^{C} f(L_{ic}=1|\boldsymbol{Z}_i) f(\boldsymbol{Y}_i|\boldsymbol{\xi}_i, \boldsymbol{Z}_i, L_{ic}=1) f(\boldsymbol{W}_i|\boldsymbol{\xi}_i) f(\boldsymbol{\xi}_i) d\boldsymbol{\xi}_i \qquad \text{(B.3)} \\
&= \sum_{c=1}^{C} \frac{\pi_{ic}}{(2\pi)^{\frac{m_{y,i}+m_{x,i}}{2}}} |\boldsymbol{V}_{ic}|^{-\frac{1}{2}} |\boldsymbol{\Lambda}_i|^{-\frac{1}{2}} |\boldsymbol{\Sigma}_\xi|^{-\frac{1}{2}} |\boldsymbol{\Omega}_{ic}|^{\frac{1}{2}} \exp\Big\{ -\frac{1}{2} (\widetilde{\boldsymbol{Y}}_{ic}^\top \boldsymbol{V}_{ic}^{-1} \widetilde{\boldsymbol{Y}}_{ic} \\
&\quad + \widetilde{\boldsymbol{W}}_i^\top \boldsymbol{\Lambda}_i^{-1} \widetilde{\boldsymbol{W}}_i - \widetilde{\boldsymbol{\xi}}_{ic}^\top \boldsymbol{\Omega}_{ic}^{-1} \widetilde{\boldsymbol{\xi}}_{ic}) \Big\}.
\end{aligned}
$$

Appendix C: Technical details of the EM step in Chapter 2

The E-step of the $(\kappa+1)$th step, the EM loss function can be written as

$$
\begin{aligned}
Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) &= -2\mathrm{E}\Big[\{\log \mathcal{L}(\boldsymbol{\Theta};\boldsymbol{Y},\boldsymbol{W},\boldsymbol{\xi},\boldsymbol{L}|\boldsymbol{Z})\}|\boldsymbol{Y},\boldsymbol{W},\boldsymbol{Z};\boldsymbol{\Theta}^{(\kappa)}\Big] \\
&= \sum_{i=1}^{n}\sum_{c=1}^{C}\Bigg(\mathrm{E}(L_{ic}|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)})\Big\{-2\log(\pi_{ic})+m_{y,i}\times\log(\sigma_\epsilon^2)+\frac{\widetilde{\boldsymbol{Y}}_{ic}^{\top}\widetilde{\boldsymbol{Y}}_{ic}}{\sigma_\epsilon^2}\Big\} \\
&\quad -\frac{1}{\sigma_\epsilon^2}\Big[\widetilde{\boldsymbol{Y}}_{ic}^{\top}\widetilde{\boldsymbol{\Psi}}_{ic}\mathrm{E}(L_{ic}\boldsymbol{\xi}_i|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)})-\mathrm{E}(L_{ic}\boldsymbol{\xi}_i^{\top}|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)})\widetilde{\boldsymbol{\Psi}}_{ic}^{\top}\widetilde{\boldsymbol{Y}}_{ic} \\
&\quad +\mathrm{tr}\{\mathrm{E}(L_{ic}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^{\top}|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)})\widetilde{\boldsymbol{\Psi}}_{ic}^{\top}\widetilde{\boldsymbol{\Psi}}_{ic}\}\Big]\Bigg)+n\log(|\boldsymbol{\Lambda}_i|)+n\log(|\boldsymbol{\Sigma}|) \\
&\quad +\sum_{i=1}^{n}\Big[\widetilde{\boldsymbol{W}}_i^{\top}\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i-\widetilde{\boldsymbol{W}}_i^{\top}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\Psi}_i^{*}\mathrm{E}(\boldsymbol{\xi}_i|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)}) \\
&\quad -\mathrm{E}(\boldsymbol{\xi}_i^{\top}|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)})\boldsymbol{\Psi}_i^{*\top}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{W}_i \\
&\quad +\mathrm{tr}\{\mathrm{E}(\boldsymbol{\xi}_i\boldsymbol{\xi}_i^{\top}|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)})(\boldsymbol{\Psi}_i^{*\top}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\Psi}_i^{*}+\boldsymbol{\Sigma}^{-1})\}\Big] \qquad \text{(C.1)}
\end{aligned}
$$

with $\widetilde{\boldsymbol{Y}}_{ic},\widetilde{\boldsymbol{\Psi}}_{ic},\widetilde{\boldsymbol{W}}_i,$ and $\boldsymbol{\Psi}_i^{*}$ defined in (A.2). For the conditional probability mass function of L_{ic} given $\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i$, it could be expressed as

$$
\begin{aligned}
&P(L_{ic}=1|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i)=\int f(L_{ic},\boldsymbol{\xi}_i|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i)d\boldsymbol{\xi}_i \\
&=\int\frac{f(L_{ic}|\boldsymbol{Z}_i)f(\boldsymbol{Y}_i|L_{ic},\boldsymbol{\xi}_i,\boldsymbol{Z}_i)f(\boldsymbol{W}_i|\boldsymbol{\xi}_i)f(\boldsymbol{\xi}_i)}{f(\boldsymbol{Y}_i,\boldsymbol{W}_i|\boldsymbol{Z}_i)}d\boldsymbol{\xi}_i \\
&=\frac{\pi_{ic}\exp\{-\frac{1}{2}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic}^{\top}\widetilde{\boldsymbol{Y}}_{ic}+\widetilde{\boldsymbol{W}}_i^{\top}\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i)\}}{\sum_{c'=1}^{C}\pi_{ic'}(2\pi)^{\frac{p}{2}}|\boldsymbol{\Omega}_{ic'}|^{\frac{1}{2}}\exp\{-\frac{1}{2}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic'}^{\top}\widetilde{\boldsymbol{Y}}_{ic'}+\widetilde{\boldsymbol{W}}_i^{\top}\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i-\widetilde{\boldsymbol{\xi}}_{ic'}^{\top}\boldsymbol{\Omega}_{ic'}^{-1}\widetilde{\boldsymbol{\xi}}_{ic'})\}} \\
&\times\int\exp[-\frac{1}{2}\{\boldsymbol{\xi}_i^{\top}(\boldsymbol{\Psi}_i^{*\top}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\Psi}_i^{*}+\boldsymbol{\Sigma}_\xi^{-1}+\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{\Psi}}_{ic}^{\top}\widetilde{\boldsymbol{\Psi}}_{ic})\boldsymbol{\xi}_i-(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic}^{\top}\widetilde{\boldsymbol{\Psi}}_{ic}+\widetilde{\boldsymbol{W}}_i^{\top}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\Psi}_i^{*})\boldsymbol{\xi}_i \\
&-\boldsymbol{\xi}_i^{\top}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{\Psi}}_{ic}^{\top}\widetilde{\boldsymbol{Y}}_{ic}+\boldsymbol{\Psi}_i^{*\top}\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i)\}]d\boldsymbol{\xi}_i
\end{aligned}
$$

$$= \frac{\pi_{ic}|\boldsymbol{\Omega}_{ic}|^{\frac{1}{2}} \exp\{-\frac{1}{2}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic}^\top\widetilde{\boldsymbol{Y}}_{ic} - \widetilde{\boldsymbol{\xi}}_{ic}^\top\boldsymbol{\Omega}_{ic}^{-1}\widetilde{\boldsymbol{\xi}}_{ic}\}}{\sum_{c'=1}^{C}\pi_{ic'}|\boldsymbol{\Omega}_{ic'}|^{\frac{1}{2}} \exp\{-\frac{1}{2}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic'}^\top\widetilde{\boldsymbol{Y}}_{ic'} - \widetilde{\boldsymbol{\xi}}_{ic'}^\top\boldsymbol{\Omega}_{ic'}^{-1}\widetilde{\boldsymbol{\xi}}_{ic'})\}} = \widetilde{\pi}_{ic}. \tag{C.2}$$

For conditional distribution of $\boldsymbol{\xi}_i$ given $\boldsymbol{Y}_i, \boldsymbol{W}_i$ and $\boldsymbol{Z}_i$, it could be expressed as

$$\begin{aligned}
f(\boldsymbol{\xi}_i|\boldsymbol{Y}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i) &= \sum_{c=1}^{C} f(L_{ic}=1, \boldsymbol{\xi}_i|\boldsymbol{Y}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i) \\
&= \sum_{c=1}^{C} \frac{f(L_{ic}=1, \boldsymbol{\xi}_i, \boldsymbol{Y}_i, \boldsymbol{W}_i|\boldsymbol{Z}_i)}{f(\boldsymbol{Y}_i, \boldsymbol{W}_i|\boldsymbol{Z}_i)} \\
&= \sum_{c=1}^{C} \frac{\pi_{ic}\exp\{-\frac{1}{2\sigma_\epsilon^2}(\widetilde{\boldsymbol{Y}}_{ic} - \widetilde{\boldsymbol{\Psi}}_{ic}\boldsymbol{\xi}_i)^{\otimes 2} - \frac{1}{2}(\widetilde{\boldsymbol{W}}_i - \boldsymbol{\Psi}_i^*\boldsymbol{\xi}_i)^\top\boldsymbol{\Lambda}_i^{-1}(\widetilde{\boldsymbol{W}}_i - \boldsymbol{\Psi}_i^*\boldsymbol{\xi}_i) - \frac{1}{2}\boldsymbol{\xi}_i^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}_i\}}{\sum_{c'=1}^{C}\pi_{ic'}(2\pi)^{\frac{p}{2}}|\boldsymbol{\Omega}_{ic'}|^{\frac{1}{2}}\exp\{-\frac{1}{2}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic'}^\top\widetilde{\boldsymbol{Y}}_{ic'} + \widetilde{\boldsymbol{W}}_i^\top\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i - \widetilde{\boldsymbol{\xi}}_{ic'}^\top\boldsymbol{\Omega}_{ic'}^{-1}\widetilde{\boldsymbol{\xi}}_{ic'})\}}.
\end{aligned}$$

Then we can have the joint conditional expectation of $L_{ic}\boldsymbol{\xi}_i$ as

$$\begin{aligned}
&\mathrm{E}(L_{ic}\boldsymbol{\xi}_i|\boldsymbol{Y}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i) = \int L_{ic}\boldsymbol{\xi}_i \frac{f(L_{ic}|\boldsymbol{Z}_i)f(\boldsymbol{Y}_i|L_{ic}, \boldsymbol{\xi}_i, \boldsymbol{Z}_i)f(\boldsymbol{W}_i|\boldsymbol{\xi}_i)f(\boldsymbol{\xi}_i)}{f(\boldsymbol{Y}_i, \boldsymbol{W}_i|\boldsymbol{Z}_i)} d\boldsymbol{\xi}_i \\
&= \int \boldsymbol{\xi}_i f(L_{ic}=1, \boldsymbol{\xi}_i|\boldsymbol{Y}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i) d\boldsymbol{\xi}_i \\
&= \int \boldsymbol{\xi}_i \frac{\pi_{ic}\exp\{-\frac{1}{2\sigma_\epsilon^2}(\widetilde{\boldsymbol{Y}}_{ic} - \widetilde{\boldsymbol{\Psi}}_{ic}\boldsymbol{\xi}_i)^{\otimes 2} - \frac{1}{2}(\widetilde{\boldsymbol{W}}_i - \boldsymbol{\Psi}_i^*\boldsymbol{\xi}_i)^\top\boldsymbol{\Lambda}_i^{-1}(\widetilde{\boldsymbol{W}}_i - \boldsymbol{\Psi}_i^*\boldsymbol{\xi}_i) - \frac{1}{2}\boldsymbol{\xi}_i^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}_i\}}{\sum_{c'=1}^{C}\pi_{ic'}(2\pi)^{\frac{p}{2}}|\boldsymbol{\Omega}_{ic'}|^{\frac{1}{2}}\exp\{-\frac{1}{2}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic'}^\top\widetilde{\boldsymbol{Y}}_{ic'} + \widetilde{\boldsymbol{W}}_i^\top\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i - \widetilde{\boldsymbol{\xi}}_{ic'}^\top\boldsymbol{\Omega}_{ic'}^{-1}\widetilde{\boldsymbol{\xi}}_{ic'})\}} d\boldsymbol{\xi}_i \\
&= \frac{\pi_{ic}\exp\{-\frac{1}{2}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic}^\top\widetilde{\boldsymbol{Y}}_{ic} + \widetilde{\boldsymbol{W}}_i^\top\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i)\}}{\sum_{c'=1}^{C}\pi_{ic'}(2\pi)^{\frac{p}{2}}|\boldsymbol{\Omega}_{ic'}|^{\frac{1}{2}}\exp\{-\frac{1}{2}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic'}^\top\widetilde{\boldsymbol{Y}}_{ic'} + \widetilde{\boldsymbol{W}}_i^\top\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i - \widetilde{\boldsymbol{\xi}}_{ic'}^\top\boldsymbol{\Omega}_{ic'}^{-1}\widetilde{\boldsymbol{\xi}}_{ic'})\}} \\
&\quad \times \int \boldsymbol{\xi}_i \exp[-\frac{1}{2}\{\boldsymbol{\xi}_i^\top(\boldsymbol{\Psi}_i^{*\top}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\Psi}_i^* + \boldsymbol{\Sigma}^{-1} + \frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{\Psi}}_{ic}^\top\widetilde{\boldsymbol{\Psi}}_{ic})\boldsymbol{\xi}_i - (\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic}^\top\widetilde{\boldsymbol{\Psi}}_{ic} + \widetilde{\boldsymbol{W}}_i^\top\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\Psi}_i^*)\boldsymbol{\xi}_i \\
&\quad -\boldsymbol{\xi}_i^\top(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{\Psi}}_{ic}^\top\widetilde{\boldsymbol{Y}}_{ic} + \boldsymbol{\Psi}_i^{*\top}\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i)\}] d\boldsymbol{\xi}_i \\
&= \frac{\pi_{ic}|\boldsymbol{\Omega}_{ic}|^{\frac{1}{2}}\exp\{-\frac{1}{2}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic}^\top\widetilde{\boldsymbol{Y}}_{ic} - \widetilde{\boldsymbol{\xi}}_{ic}^\top\boldsymbol{\Omega}_{ic}^{-1}\widetilde{\boldsymbol{\xi}}_{ic}\}\widetilde{\boldsymbol{\xi}}_{ic}}{\sum_{c'=1}^{C}\pi_{ic'}|\boldsymbol{\Omega}_{ic'}|^{\frac{1}{2}}\exp\{-\frac{1}{2}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic'}^\top\widetilde{\boldsymbol{Y}}_{ic'} - \widetilde{\boldsymbol{\xi}}_{ic'}^\top\boldsymbol{\Omega}_{ic'}^{-1}\widetilde{\boldsymbol{\xi}}_{ic'}))\}} = \widetilde{\pi}_{ic}\widetilde{\boldsymbol{\xi}}_{ic},
\end{aligned} \tag{C.3}$$

And

$$
\begin{aligned}
&\mathrm{E}(L_{ic}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^\top|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i) = \int L_{ic}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^\top \frac{f(L_{ic}|\boldsymbol{Z}_i)f(\boldsymbol{Y}_i|L_{ic},\boldsymbol{\xi}_i,\boldsymbol{Z}_i)f(\boldsymbol{W}_i|\boldsymbol{\xi}_i)f(\boldsymbol{\xi}_i)}{f(\boldsymbol{Y}_i,\boldsymbol{W}_i|\boldsymbol{Z}_i)}d\boldsymbol{\xi}_i \\
&= \int \boldsymbol{\xi}_i\boldsymbol{\xi}_i^\top f(L_{ic},\boldsymbol{\xi}_i|\boldsymbol{Y}_i,\boldsymbol{W}_i,Z_i)d\boldsymbol{\xi}_i \\
&= \frac{\pi_{ic}\exp\{-\frac{1}{2\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic}^\top\widetilde{\boldsymbol{Y}}_{ic})\}}{\sum_{c'=1}^C \pi_{ic'}(2\pi)^{\frac{p}{2}}|\boldsymbol{\Omega}_{ic'}|^{\frac{1}{2}}\exp\{-\frac{1}{2}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic'}^\top\widetilde{\boldsymbol{Y}}_{ic'} - \widetilde{\boldsymbol{\xi}}_{ic'}^\top\boldsymbol{\Omega}_{ic'}^{-1}\widetilde{\boldsymbol{\xi}}_{ic'})\}} \\
&\quad\times\int \boldsymbol{\xi}_i\boldsymbol{\xi}_i^\top\exp[-\frac{1}{2}\{\boldsymbol{\xi}_i^\top(\boldsymbol{\Psi}_i^{*\top}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\Psi}_i^* + \boldsymbol{\Sigma}_\xi^{-1} + \frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{\Psi}}_{ic}^\top\widetilde{\boldsymbol{\Psi}}_{ic})\boldsymbol{\xi}_i - (\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic}^\top\widetilde{\boldsymbol{\Psi}}_{ic} \\
&\quad+\widetilde{\boldsymbol{W}}_i^\top\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\Psi}_i^*)\boldsymbol{\xi}_i - \boldsymbol{\xi}_i^\top(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{\Psi}}_{ic}^\top\widetilde{\boldsymbol{Y}}_{ic} + \boldsymbol{\Psi}_i^{*\top}\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i)\}d\boldsymbol{\xi}_i \\
&= \frac{\pi_{ic}|\boldsymbol{\Omega}_{ic}|^{\frac{1}{2}}\exp\{-\frac{1}{2}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic}^\top\widetilde{\boldsymbol{Y}}_{ic} - \widetilde{\boldsymbol{\xi}}_{ic}^\top\boldsymbol{\Omega}_{ic}^{-1}\widetilde{\boldsymbol{\xi}}_{ic})\}(\widetilde{\boldsymbol{\xi}}_{ic}\widetilde{\boldsymbol{\xi}}_{ic}^\top + \boldsymbol{\Omega}_{ic})}{\sum_{c'=1}^C \pi_{ic'}|\boldsymbol{\Omega}_{ic'}|^{\frac{1}{2}}\exp\{-\frac{1}{2}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic'}^\top\widetilde{\boldsymbol{Y}}_{ic'} - \widetilde{\boldsymbol{\xi}}_{ic'}^\top\boldsymbol{\Omega}_{ic'}^{-1}\widetilde{\boldsymbol{\xi}}_{ic'}))\}} \\
&= \widetilde{\pi}_{ic}(\widetilde{\boldsymbol{\xi}}_{ic}\widetilde{\boldsymbol{\xi}}_{ic}^\top + \boldsymbol{\Omega}_{ic}) = \widetilde{\pi}_{ic}\widetilde{\boldsymbol{\Sigma}_{\xi}}_{ic}.
\end{aligned} \tag{C.4}
$$

Furthermore

$$
\begin{aligned}
&\mathrm{E}(\boldsymbol{\xi}_i|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i) = \int \boldsymbol{\xi}_i f(\boldsymbol{\xi}_i|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i)d\boldsymbol{\xi}_i \\
&= \sum_{c=1}^C \frac{\pi_{ic}\exp\{-\frac{1}{2}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic}^\top\widetilde{\boldsymbol{Y}}_{ic} + \widetilde{\boldsymbol{W}}_i^\top\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i)\}}{\sum_{c'=1}^C \pi_{ic'}(2\pi)^{\frac{p}{2}}|\boldsymbol{\Omega}_{ic'}|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic'}^\top\widetilde{\boldsymbol{Y}}_{ic'} + \widetilde{\boldsymbol{W}}_i^\top\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i - \widetilde{\boldsymbol{\xi}}_{ic'}^\top\boldsymbol{\Omega}_{ic'}\widetilde{\boldsymbol{\xi}}_{ic'}\} } \\
&\quad\times\int \boldsymbol{\xi}_i\exp[-\frac{1}{2}\{\boldsymbol{\xi}_i^\top(\widetilde{\boldsymbol{\Psi}}_i^\top\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{\Psi}}_i + \boldsymbol{\Sigma}^{-1} + \frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{\Psi}}_{ic}^{*\top}\widetilde{\boldsymbol{\Psi}}_{ic}^*)\boldsymbol{\xi}_i - (\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic}^\top\widetilde{\boldsymbol{\Psi}}_{ic}^* + \widetilde{\boldsymbol{W}}_i^\top\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{\Psi}}_i)\boldsymbol{\xi}_i \\
&\quad-\boldsymbol{\xi}_i^\top(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{\Psi}}_{ic}^{*\top}\widetilde{\boldsymbol{Y}}_{ic} + \widetilde{\boldsymbol{\Psi}}_i^\top\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i)\}]d\boldsymbol{\xi}_i \\
&= \sum_{c=1}^C \frac{\pi_{ic}\exp\{-\frac{1}{2}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic}^\top\widetilde{\boldsymbol{Y}}_{ic} + \widetilde{\boldsymbol{W}}_i^\top\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i)\}|\boldsymbol{\Omega}_{ic}|^{\frac{1}{2}}\exp(\frac{1}{2}\widetilde{\boldsymbol{\xi}}_{ic}^\top\boldsymbol{\Omega}_{ic}^{-1}\widetilde{\boldsymbol{\xi}}_{ic})\widetilde{\boldsymbol{\xi}}_{ic}}{\sum_{c'=1}^C \pi_{ic'}|\boldsymbol{\Omega}_{ic'}|^{\frac{1}{2}}\exp\{-\frac{1}{2}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic'}^\top\widetilde{\boldsymbol{Y}}_{ic'} + \widetilde{\boldsymbol{W}}_i^\top\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i - \widetilde{\boldsymbol{\xi}}_{ic'}^\top\boldsymbol{\Omega}_{ic'}^{-1}\widetilde{\boldsymbol{\xi}}_{ic'})\}} \\
&= \sum_{c=1}^C \widetilde{\pi}_{ic}\widetilde{\boldsymbol{\xi}}_{ic} = \widetilde{\boldsymbol{\xi}}_i.
\end{aligned} \tag{C.5}
$$

And the second moment of $\boldsymbol{\xi}_i$ given $\boldsymbol{Y}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i$ is

$$
\begin{aligned}
&\mathrm{E}(\boldsymbol{\xi}_i\boldsymbol{\xi}_i^\top|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i) = \int \boldsymbol{\xi}_i\boldsymbol{\xi}_i^\top f(\boldsymbol{\xi}_i|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i)d\boldsymbol{\xi}_i \\
&= \sum_c^C \int \boldsymbol{\xi}_i\boldsymbol{\xi}_i^\top f(L_{ic}=1,\boldsymbol{\xi}_i|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i)d\boldsymbol{\xi}_i \\
&= \sum_{c=1}^C \frac{\pi_{ic}\exp\{-\frac{1}{2}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic}^\top\widetilde{\boldsymbol{Y}}_{ic} + \widetilde{\boldsymbol{W}}_i^\top\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i)\}|\boldsymbol{\Omega}_{ic}|^{\frac{1}{2}}\exp(\frac{1}{2}\widetilde{\boldsymbol{\xi}}_{ic}^\top\boldsymbol{\Omega}_{ic}^{-1}\widetilde{\boldsymbol{\xi}}_{ic})(\widetilde{\boldsymbol{\xi}}_{ic}\widetilde{\boldsymbol{\xi}}_{ic}^\top+\boldsymbol{\Omega}_{ic})}{\sum_{c'=1}^C \pi_{ic'}|\boldsymbol{\Omega}_{ic'}|^{\frac{1}{2}}\exp\{-\frac{1}{2}(\frac{1}{\sigma_\epsilon^2}\widetilde{\boldsymbol{Y}}_{ic'}^\top\widetilde{\boldsymbol{Y}}_{ic'} + \widetilde{\boldsymbol{W}}_i^\top\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i - \widetilde{\boldsymbol{\xi}}_{ic'}^\top\boldsymbol{\Omega}_{ic'}^{-1}\widetilde{\boldsymbol{\xi}}_{ic'})\}} \\
&= \sum_{c=1}^C \widetilde{\pi}_{ic}(\widetilde{\boldsymbol{\xi}}_{ic}\widetilde{\boldsymbol{\xi}}_{ic}^\top+\boldsymbol{\Omega}_{ic}) = \sum_{c=1}^C \widetilde{\pi}_{ic}\widetilde{\boldsymbol{\Sigma}}_{ic} = \widetilde{\boldsymbol{\Sigma}}_i
\end{aligned} \tag{C.6}
$$

Suppose the conditional moments derived above have the partition $\widetilde{\boldsymbol{\xi}}_{ic} = (\widetilde{\boldsymbol{\xi}}_{i1,c},\ldots,\widetilde{\boldsymbol{\xi}}_{id_x,c})^\top$ and $\widetilde{\boldsymbol{\xi}}_i = (\widetilde{\boldsymbol{\xi}}_{i1},\ldots,\widetilde{\boldsymbol{\xi}}_{id_x})^\top$, $\widetilde{\boldsymbol{\Sigma}}_{ic} = (\widetilde{\boldsymbol{\Sigma}}_{ic,vv'})_{v,v'=1}^{d_x} = (\widetilde{\boldsymbol{\xi}}_{iv,c}\widetilde{\boldsymbol{\xi}}_{iv',c}^\top + \boldsymbol{\Omega}_{ic,vv'})_{vv'=1}^{d_x}$ with $\boldsymbol{\Omega}_{ic} = (\boldsymbol{\Omega}_{ic,vv'})_{v,v'=1}^{d_x}$. Combine the calculations above, we have

$$
Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) = Q_1(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) + Q_2(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) + Q_3(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) + Q_4(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}),
$$

where

$$
\begin{aligned}
Q_1(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) &= \sum_{i=1}^n\sum_{c=1}^C -2\times\widetilde{\pi}_{ic}\times\log(\pi_{ic}), \\
Q_2(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) &= \sum_{i=1}^n\sum_{c=1}^C \widetilde{\pi}_{ic}\Bigg[m_{y,i}\log(\sigma_\epsilon^2) + \frac{1}{\sigma_\epsilon^2}\Bigg\{\sum_{v=1}^{d_x}\sum_{v'=1}^{d_x}\beta_{x,cv}\beta_{x,cv'}\,\mathrm{tr}(\boldsymbol{\Theta}_{\psi v}^\top\boldsymbol{B}_{iv}^{*\top}\boldsymbol{B}_{iv'}^*\boldsymbol{\Theta}_{\psi v'}\boldsymbol{\Omega}_{ic,v'v}) \\
&\quad +(\boldsymbol{Y}_i - \beta_{0,c}\mathbf{1} - \sum_{v=1}^{d_x}\beta_{x,cv}\boldsymbol{B}_{iv}^*\boldsymbol{\theta}_{\mu v} - \sum_{v=1}^{d_x}\beta_{x,cv}\boldsymbol{B}_{iv}^*\boldsymbol{\Theta}_{\psi v}\widetilde{\boldsymbol{\xi}}_{iv,c} - \boldsymbol{Z}_i\boldsymbol{\beta}_{z,c})^{\otimes 2}\Bigg\}\Bigg] \\
Q_3(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) &= \sum_{i=1}^n\sum_{v=1}^{d_x}\Bigg[\frac{1}{\sigma_v^2}\Bigg\{(\boldsymbol{W}_{iv} - \boldsymbol{B}_{iv}\boldsymbol{\theta}_{\mu v} - \boldsymbol{B}_{iv}\boldsymbol{\Theta}_{\psi v}\widetilde{\boldsymbol{\xi}}_{iv})^{\otimes 2} - \widetilde{\boldsymbol{\xi}}_{iv}^\top\boldsymbol{\Theta}_{\psi v}^\top\boldsymbol{B}_{iv}^\top\boldsymbol{B}_{iv}\boldsymbol{\Theta}_{\psi v}\widetilde{\boldsymbol{\xi}}_{iv} \\
&\quad + \mathrm{tr}(\boldsymbol{\Theta}_{\psi v}^\top\boldsymbol{B}_{iv}^\top\boldsymbol{B}_{iv}\boldsymbol{\Theta}_{\psi v}\widetilde{\boldsymbol{\Sigma}}_{i,vv})\Bigg\} + m_{xv,i}\log(\sigma_v^2)\Bigg],
\end{aligned}
$$

$$Q_4(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) \;\; = \;\; n\log(|\boldsymbol{\Sigma}|) + \sum_{i=1}^{n} \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\widetilde{\boldsymbol{\Sigma}}_i). \tag{C.7}$$

To update the parameters, the M-step involves executing a sequence of steps, each applied to one block at a time. Parameters can be updated by minimizing the Q function or using Newton-Raphson iteration.

C.1 Update $\boldsymbol{\gamma}$

Using the logistic regression model (2) in the main paper, Q_1 can be rewritten as

$$Q_1(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) = -2\sum_{i=1}^{n}\left[(\sum_{c=1}^{C-1}\widetilde{\pi}_{ic}\widetilde{\mathbb{Z}}_i^{\top}\boldsymbol{\gamma}_c) - \log\{1+\sum_{c'=1}^{C-1}\exp(\widetilde{\mathbb{Z}}_i^{\top}\boldsymbol{\gamma}_{c'})\}\right].$$

Then we use a Newton-Raphson algorithm to update our $\boldsymbol{\gamma}$. Given the value of $\boldsymbol{\gamma} = \widehat{\boldsymbol{\gamma}}^{(\zeta)}$ at the end of the ζth iteration, we update our $\widehat{\boldsymbol{\gamma}}$ as

$$\widehat{\boldsymbol{\gamma}}^{(\zeta+1)} = \widehat{\boldsymbol{\gamma}}^{(\zeta)} - \boldsymbol{H}^{-1}(\widehat{\boldsymbol{\gamma}}^{(\zeta)}) \bigtriangledown Q_1(\widehat{\boldsymbol{\gamma}}^{(\zeta)}),$$

where $\bigtriangledown Q_1(\widehat{\boldsymbol{\gamma}}^{(\zeta)}) = \{(\frac{\partial Q_1}{\partial \boldsymbol{\gamma}_1})^{\top}, (\frac{\partial Q_1}{\partial \boldsymbol{\gamma}_2})^{\top}, \ldots, (\frac{\partial Q_1}{\partial \boldsymbol{\gamma}_{C-1}})^{\top}\}$ with

$$\frac{\partial Q_1}{\partial \boldsymbol{\gamma}_c} = -2\sum_{i=1}^{n}\{\widetilde{\pi}_{ic}\widetilde{\mathbb{Z}}_i - \frac{\widetilde{\mathbb{Z}}_i\exp(\widetilde{\mathbb{Z}}_i^{\top}\boldsymbol{\gamma}_c)}{1+\sum_{c'=1}^{C-1}\exp(\widetilde{\mathbb{Z}}_i^{\top}\boldsymbol{\gamma}_{c'})}\}, \quad c = 1,\ldots,C-1, \tag{C.8}$$

and $\boldsymbol{H}$ is the Hessian matrix with diagonal and off-diagonal entries

$$\frac{\partial^2 Q_1}{\partial \boldsymbol{\gamma}_c \partial \boldsymbol{\gamma}_c^{\top}} = 2\sum_{i=1}^{n}\frac{\widetilde{\mathbb{Z}}_i\widetilde{\mathbb{Z}}_i^{\top}\exp(\widetilde{\mathbb{Z}}_i^{\top}\boldsymbol{\gamma}_c)\{1+\sum_{c'=1}^{C-1}\exp(\widetilde{\mathbb{Z}}_i^{\top}\boldsymbol{\gamma}_{c'})\} - \widetilde{\mathbb{Z}}_i\widetilde{\mathbb{Z}}_i^{\top}\exp(2\widetilde{\mathbb{Z}}_i^{\top}\boldsymbol{\gamma}_c)}{\{1+\sum_{c'=1}^{C-1}\exp(\widetilde{\mathbb{Z}}_i^{\top}\boldsymbol{\gamma}_{c'})\}^2},$$

$$\frac{\partial^2 Q_1}{\partial \boldsymbol{\gamma}_c \partial \boldsymbol{\gamma}_{c'}^\top} = -2\sum_{i=1}^{n} \frac{\widetilde{\mathbb{Z}}_i \widetilde{\mathbb{Z}}_i^\top \exp(\widetilde{\mathbb{Z}}_i^\top \boldsymbol{\gamma}_c + \widetilde{\mathbb{Z}}_i^\top \boldsymbol{\gamma}_{c'})}{\{1+\sum_{c'=1}^{C-1}\exp(\widetilde{\mathbb{Z}}_i^\top \boldsymbol{\gamma}_{c'})\}^2} \qquad \text{if } c \neq c'.$$

Since this Newton-Raphson iteration is embedded in our EM algorithm, we only iterate two steps to update $\widehat{\boldsymbol{\gamma}}$ within each EM iteration to save computation time.

C.2 Update $\boldsymbol{\beta}$

When updating $\boldsymbol{\beta}_c$, define $m_{y,i} \times d$ matrix $\mathbb{X}_{ic} = (\mathbf{1}_{m_{y,i}}, \boldsymbol{B}_i\widetilde{\boldsymbol{\Theta}}_\mu + \boldsymbol{B}_i\boldsymbol{\Theta}_\psi\boldsymbol{\Xi}_{ic}, \boldsymbol{Z}_i)$, where $d = 1+d_x+d_z$, $\boldsymbol{B}_i = (\boldsymbol{B}_{i1}, \ldots, \boldsymbol{B}_{id_x})$ is a $m_{y,i} \times q$ matrix, $\widetilde{\boldsymbol{\Theta}}_\mu = \text{diag}(\boldsymbol{\theta}_{\mu 1}, \ldots, \boldsymbol{\theta}_{\mu d_x})$ is a $q \times d_x$ matrix, and $\boldsymbol{\Xi}_{ic} = \text{diag}(\widetilde{\boldsymbol{\xi}}_{i1,c}, \ldots, \widetilde{\boldsymbol{\xi}}_{id_x,c})$ is a $p \times d_x$ matrix. A closed-form expression to update $\boldsymbol{\beta}_c$ is

$$\widehat{\boldsymbol{\beta}}_c = \{\sum_{i=1}^{n} \widetilde{\pi}_{ic}(\mathbb{X}_{ic}^\top \boldsymbol{V}_{ic}^{-1}\mathbb{X}_{ic} + \boldsymbol{A}_{ic})\}^{-1}(\sum_{i=1}^{n} \widetilde{\pi}_{ic}\mathbb{X}_{ic}^\top \boldsymbol{V}_{ic}^{-1}\boldsymbol{Y}_i) \tag{C.9}$$

where $\boldsymbol{A}_{ic} = \text{diag}(\mathbf{0}, [\text{tr}\{\boldsymbol{\Theta}_{\psi v}^\top \boldsymbol{B}_{iv}^\top \boldsymbol{V}_{ic}^{-1}\boldsymbol{B}_{iv'}\boldsymbol{\Theta}_{\psi v'}\boldsymbol{\Omega}_{ic,v'v}\}]_{v,v'=1}^{d_x}, \mathbf{0})$ is a $d \times d$ matrix.

C.3 Update the variance component: σ_ϵ^2 and σ_v^2 for $v = 1, \ldots, d_x$

Update the estimates of σ_ϵ^2 and σ_v^2 according to Q_2 and Q_3 separately. The updating formulae are

$$\begin{aligned}
\widehat{\sigma}_\epsilon^2 &= \frac{1}{\sum_{i=1}^n m_{y,i}} \sum_{i=1}^{n}\sum_{c=1}^{C} \widetilde{\pi}_{ic}\Big\{\boldsymbol{Y}_i^\top\boldsymbol{Y}_i - \boldsymbol{Y}_i^\top\mathbb{X}_{ic}\boldsymbol{\beta}_c - \boldsymbol{\beta}_c^\top\mathbb{X}_{ic}^\top\boldsymbol{Y}_i + \boldsymbol{\beta}_c^\top(\mathbb{X}_{ic}^\top\mathbb{X}_{ic} + \boldsymbol{A}_{ic})\boldsymbol{\beta}_c\Big\}, \\
\widehat{\sigma}_v^2 &= \frac{1}{m_{x,i}} \sum_{i=1}^{n}\Big[(\boldsymbol{W}_{iv} - \boldsymbol{B}_{iv}\boldsymbol{\theta}_{\mu v} - \boldsymbol{B}_{iv}\boldsymbol{\Theta}_{\psi v}\widetilde{\boldsymbol{\xi}}_{iv})^{\otimes 2} + \text{tr}\{\boldsymbol{\Theta}_{\psi v}^\top\boldsymbol{B}_{iv}^\top\boldsymbol{B}_{iv}\boldsymbol{\Theta}_{\psi v}(\widetilde{\boldsymbol{\Sigma}}_{i,vv} - \widetilde{\boldsymbol{\xi}}_{iv}\widetilde{\boldsymbol{\xi}}_{iv}^\top)\}\Big]
\end{aligned}$$

C.4 Update spline coefficients for the mean and eigenfunctions

We can update the spline coefficients according to $Q_2 + Q_3$ as

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}_{\mu v} &= \left\{ \sum_{i=1}^{n} \left[(\sum_{c=1}^{C} \frac{\widetilde{\pi}_{ic}}{\sigma_\epsilon^2} \beta_{x,cv}^2 \boldsymbol{B}_{iv}^{*\top} \boldsymbol{B}_{iv}^{*}) + \frac{1}{\sigma_v^2} \boldsymbol{B}_{iv}^{\top} \boldsymbol{B}_{iv} \right] \right\}^{-1} \left\{ \sum_{i=1}^{n} \left[\frac{1}{\sigma_v^2} (\boldsymbol{B}_{iv}^{\top} \boldsymbol{W}_{iv} - \boldsymbol{B}_{iv}^{\top} \boldsymbol{B}_{iv} \boldsymbol{\Theta}_{\psi v} \widetilde{\boldsymbol{\xi}}_{iv}) \right.\right. \\
&+ \left\{ \sum_{c=1}^{C} \frac{\widetilde{\pi}_{ic}}{\sigma_\epsilon^2} \beta_{x,cv} \boldsymbol{B}_{iv}^{*\top} (\boldsymbol{Y}_i - \beta_{0,c} \mathbf{1} - \sum_{v' \neq v; v'=1}^{d_x} \beta_{x,cv'} \boldsymbol{B}_{iv'}^{*} \boldsymbol{\theta}_{\mu v'} \right. \\
&\left.\left.\left. - \sum_{v=1}^{d_x} \beta_{x,cv} \boldsymbol{B}_{iv}^{*} \boldsymbol{\Theta}_{\psi v} \widetilde{\boldsymbol{\xi}}_{iv,c} - \boldsymbol{Z}_i \boldsymbol{\beta}_{z,c}) \right\} \right] \right\}. \qquad \text{(C.10)}
\end{aligned}$$

And let $\boldsymbol{\Theta}_{\psi v} = (\boldsymbol{\theta}_{\psi v,1}, \boldsymbol{\theta}_{\psi v,2}, \ldots, \boldsymbol{\theta}_{\psi v,p_v})$, where $\boldsymbol{\theta}_{\psi v,k}$ are the spline coefficients for the k^{th} eigenfunction for variable $\boldsymbol{X}_v$. We update the columns of $\boldsymbol{\Theta}_{\psi v}$ one at a time while holding all other columns fixed. Let $\boldsymbol{\Theta}_{\psi v}^{(\iota)}$ be its value at ι^{th} step, we update the k^{th} column by

$$\widehat{\boldsymbol{\theta}}_{\psi v,k}^{(\iota+1)} = \boldsymbol{\theta}_{\psi v,k}^{(\iota)} - [\boldsymbol{H}(\widehat{\boldsymbol{\theta}}_{\psi v,k}^{(\iota)})]^{-1} \bigtriangledown_{\boldsymbol{\theta}_{\psi v,k}} (Q_2 + Q_3)(\boldsymbol{\Theta}^{(\iota)}),$$

where $\boldsymbol{H}(\widehat{\boldsymbol{\theta}}_{\psi v,k}^{(\iota)})$ and $\bigtriangledown_{\boldsymbol{\theta}_{\psi v,k}}(Q_2+Q_3)(\boldsymbol{\Theta}^{(\iota)})$ are hessian and gradient of $\widehat{Q}_2+\widehat{Q}_3$ with respect to $\boldsymbol{\theta}_{\psi v,k}$. $\bigtriangledown_{\boldsymbol{\theta}_{\psi v,k}}(Q_2 + Q_3)(\boldsymbol{\Theta}^{(\iota)})$ is calculated from

$$\begin{aligned}
\bigtriangledown_{\boldsymbol{\theta}_{\psi v,k}}(Q_2 + Q_3)(\boldsymbol{\Theta}^{(\iota)}) &= -2 \sum_{i=1}^{n} \sum_{c=1}^{C} \left[\frac{\widetilde{\pi}_{ic}}{\sigma_\epsilon^2} \{ \beta_{x,cv} \widetilde{\xi}_{ivc}^{[k]} \boldsymbol{B}_{iv}^{*\top} (\widetilde{\boldsymbol{Y}}_{ic} - \sum_{v=1}^{d_x} \beta_{x,cv} \boldsymbol{B}_{iv}^{*} \boldsymbol{\Theta}_{\psi v} \widetilde{\boldsymbol{\xi}}_{iv,c}) \right. \\
&\quad \left. - \beta_{x,cv} \boldsymbol{B}_{iv}^{*\top} \sum_{v'=1}^{d_x} \beta_{x,cv'} \boldsymbol{B}_{iv'}^{*} \boldsymbol{\Theta}_{\psi v'} \boldsymbol{\Omega}_{ic,v'v}^{[;k]} \} \right] \\
&\quad -2 \sum_{i=1}^{n} \left[\frac{1}{\sigma_v^2} \{ \boldsymbol{B}_{iv}^{\top} \widetilde{\boldsymbol{\xi}}_{iv}^{[k]} (\widetilde{\boldsymbol{W}}_{iv} - \boldsymbol{B}_{iv} \boldsymbol{\Theta}_{\psi v} \widetilde{\boldsymbol{\xi}}_{iv}) \right. \\
&\quad \left. - \boldsymbol{B}_{iv}^{\top} \boldsymbol{B}_{iv} \boldsymbol{\Theta}_{\psi v} (\widetilde{\boldsymbol{\Sigma}}_{i,vv} - \widetilde{\boldsymbol{\xi}}_{iv} \widetilde{\boldsymbol{\xi}}_{iv}^{\top})^{[;k]}) \} \right], \qquad \text{(C.11)}
\end{aligned}$$

here $\boldsymbol{\Omega}_{ic,v'v}^{[;k]}$ denotes the k^{th} column of $\boldsymbol{\Omega}_{ic,v'v}$ for $v, v' = 1, \ldots, d_x$, $\widetilde{\xi}_{ivc}^{[k]}$ denotes the k^{th} element in the vector $\widetilde{\boldsymbol{\xi}}_{ivc}$.

Furthermore our hessian matrix $\boldsymbol{H}(\widehat{\boldsymbol{\theta}}_{\psi v,k}^{(\iota)})$ is

$$\begin{aligned}\boldsymbol{H}(\widehat{\boldsymbol{\theta}}_{\psi v,k}^{(\iota)}) &= 2\sum_{i=1}^{n}\sum_{c=1}^{C}\left\{\frac{\widetilde{\pi}_{ic}}{\sigma_\epsilon^2}(\beta_{x,cv}^2\widetilde{\omega}_{ic,vv,kk}\boldsymbol{B}_{iv}^{*\top}\boldsymbol{B}_{iv}^{*} + \beta_{x,cv}^2\boldsymbol{B}_{iv}^{*\top}\boldsymbol{B}_{iv}^{*}\widetilde{\xi}_{ivc}^{[k]^2})\right\} \\ &\quad +2\sum_{i=1}^{n}\left\{\frac{1}{\sigma_v^2}(\widetilde{\delta}_{i,vv,kk}\boldsymbol{B}_{iv}^{\top}\boldsymbol{B}_{iv} + \boldsymbol{B}_{iv}^{\top}\boldsymbol{B}_{iv}\widetilde{\boldsymbol{\xi}}_{iv}^{[k]^2})\right\} \qquad \text{(C.12)}\end{aligned}$$

with $\widetilde{\omega}_{ic,vv,kk}$ and $\widetilde{\delta}_{i,vv,kk}$ are the elements in k^{th} column and k^{th} row of $\boldsymbol{\Omega}_{ic,vv}$ and $(\widetilde{\boldsymbol{\Sigma}}_{i,vv} - \widetilde{\boldsymbol{\xi}}_{iv}\widetilde{\boldsymbol{\xi}}_{iv}^{\top})$, respectively.

Another way to update $\boldsymbol{\Theta}_{\psi v}$ is

$$\begin{aligned}\boldsymbol{\theta}_{\psi v}^{[;k]} &= \left[\sum_{i=1}^{n}\left\{\sum_{c=1}^{C}(\frac{\widetilde{\pi}_{ic}}{\sigma_\epsilon^2}\beta_{xc,v}^2\widetilde{\boldsymbol{\Sigma}}_{ic,vv}^{[k;k]}\boldsymbol{B}_{iv}^{*\top}\boldsymbol{B}_{iv}^{*}) + \frac{1}{\sigma_v^2}\widetilde{\boldsymbol{\Sigma}}_{i,vv}^{[k;k]}\boldsymbol{B}_{iv}^{\top}\boldsymbol{B}_{iv}\right\}\right]^{-1} \\ &\quad \left\{\sum_{i=1}^{n}\left(\sum_{c=1}^{C}\left[\frac{\widetilde{\pi}_{ic}}{\sigma_\epsilon^2}\boldsymbol{B}_{iv}^{*\top}\widetilde{\xi}_{ivc}^{[k]}\beta_{xc,v}\{\widetilde{\boldsymbol{Y}}_{ic} - (\sum_{v'=1;v'\neq v}^{d_x}\beta_{xc,v'}\boldsymbol{B}_{iv'}^{*}\boldsymbol{\Theta}_{\psi v'}\widetilde{\boldsymbol{\xi}}_{iv'c})\right.\right.\right. \\ &\quad -\beta_{xc,v}\boldsymbol{B}_{iv}^{*}(\sum_{j=1;j\neq k}^{p_v}\boldsymbol{\Theta}_{\psi v}^{[;j]}\widetilde{\xi}_{ivc}^{[j]})\} - \frac{\widetilde{\pi}_{ic}}{\sigma_\epsilon^2}\beta_{xc,v}\boldsymbol{B}_{iv}^{*\top}\{(\sum_{v'=1v'\neq v}^{d_x}\beta_{xc,v'}\boldsymbol{B}_{iv'}^{*}\boldsymbol{\Theta}_{\psi v'}\boldsymbol{\Omega}_{ic,v'v}^{[;k]}) \\ &\quad \left.\left.+\beta_{xc,v}\boldsymbol{B}_{iv}^{*}(\sum_{j=1;j\neq k}^{p_v}\boldsymbol{\Theta}_{\psi v}^{[;j]}\boldsymbol{\Omega}_{ic,vv}^{[j;k]})\}\right] + \frac{1}{\sigma_v^2}\boldsymbol{B}_{iv}^{\top}\widetilde{\boldsymbol{\xi}}_{iv}^{[k]}\{\widetilde{\boldsymbol{W}}_{iv} - \boldsymbol{B}_{iv}(\sum_{j=1;j\neq k}^{p_v}\boldsymbol{\Theta}_{\psi v}^{[;j]}\widetilde{\boldsymbol{\xi}}_{iv}^{[j]})\}\right. \\ &\quad \left.-\frac{1}{\sigma_v^2}\boldsymbol{B}_{iv}^{\top}\boldsymbol{B}_{iv}(\sum_{j=1;j\neq k}^{p_v}\boldsymbol{\Theta}_{\psi v}^{[;j]}\boldsymbol{\delta}_{i,vv}^{[j;k]})\right)\Bigg\}, \qquad \text{(C.13)}\end{aligned}$$

where $\widetilde{\boldsymbol{\Sigma}}_{ic,vv}^{[k;k]}$ represents the value in the k^{th} row and the k^{th} column in $\widetilde{\boldsymbol{\Sigma}}_{ic,vv}$, $\boldsymbol{\Theta}_{\psi v}^{[;k]}$ represents the k^{th} column of $\boldsymbol{\Theta}_{\psi v}$, $\widetilde{\xi}_{ivc}^{[k]}$ represents the k^{th} element in the $\widetilde{\boldsymbol{\xi}}_{ivc}$ vector, and $\boldsymbol{\delta}_{i,vv} = \widetilde{\boldsymbol{\Sigma}}_{i,vv} - \widetilde{\boldsymbol{\xi}}_{iv}\widetilde{\boldsymbol{\xi}}_{iv}^{\top}$.

Appendix D: Technical details of the EM step in Chapter 3

An expectation of our conditional log-likelihood function $Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)})$ can be derived from

$$
\begin{aligned}
Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) &= -2\mathrm{E}\Big[\{\log \pounds(\boldsymbol{\Theta};\boldsymbol{Y},\boldsymbol{W},\boldsymbol{\xi},\boldsymbol{L}|\boldsymbol{Z})\}|\boldsymbol{Y},\boldsymbol{W},\boldsymbol{Z};\boldsymbol{\Theta}^{(\kappa)}\Big] \\
&= \sum_{i=1}^{n}\sum_{c=1}^{C}\Bigg[\mathrm{E}(L_{ic}|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)})\Big\{-2\log(\pi_{ic})+\log(|\boldsymbol{V}_{ic}|)+\widetilde{\boldsymbol{Y}}_{ic}^{\top}\boldsymbol{V}_{ic}^{-1}\widetilde{\boldsymbol{Y}}_{ic} \\
&\quad -\widetilde{\boldsymbol{Y}}_{ic}^{\top}\boldsymbol{V}_{ic}^{-1}\widetilde{\boldsymbol{\Psi}}_{ic}\mathrm{E}(L_{ic}\boldsymbol{\xi}_i|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)})-\mathrm{E}(L_{ic}\boldsymbol{\xi}_i^{\top}|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)})\widetilde{\boldsymbol{\Psi}}_{ic}^{\top}\boldsymbol{V}_{ic}^{-1}\widetilde{\boldsymbol{Y}}_{ic} \\
&\quad +\mathrm{tr}\left\{\mathrm{E}(L_{ic}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^{\top}|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)})\widetilde{\boldsymbol{\Psi}}_{ic}^{\top}\boldsymbol{V}_{ic}^{-1}\widetilde{\boldsymbol{\Psi}}_{ic}\right\}\Bigg]+\sum_{i=1}^{n}\Big[\widetilde{\boldsymbol{W}}_i^{\top}\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i \\
&\quad -\widetilde{\boldsymbol{W}}_i^{\top}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\Psi}_i^{*}\mathrm{E}(\boldsymbol{\xi}_i|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)})-\mathrm{E}(\boldsymbol{\xi}_i^{\top}|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)})\boldsymbol{\Psi}_i^{*\top}\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i \\
&\quad +\mathrm{tr}\left\{\mathrm{E}(\boldsymbol{\xi}_i\boldsymbol{\xi}_i^{\top}|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i;\boldsymbol{\Theta}^{(\kappa)})(\boldsymbol{\Psi}_i^{*\top}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\Psi}_i^{*}+\boldsymbol{\Sigma}_{\xi}^{-1})\right\}\Big]
\end{aligned}
\tag{D.1}
$$

with $\widetilde{\boldsymbol{Y}}_{ic},\widetilde{\boldsymbol{\Psi}}_{ic},\widetilde{\boldsymbol{W}}_i,\boldsymbol{\Psi}_i^{*}$ defined in (B.1).

For the conditional probability mass function of L_{ic} given $\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i$, it is

$$
\begin{aligned}
P(L_{ic}=1|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i) &= \int f(L_{ic}=1,\boldsymbol{\xi}_i|\boldsymbol{Y}_i,\boldsymbol{W}_i,\boldsymbol{Z}_i)d\boldsymbol{\xi}_i \\
&= \int \frac{\pi_{ic}f(\boldsymbol{Y}_i|\boldsymbol{\xi}_i,\boldsymbol{Z}_i,L_{ic}=1)f(\boldsymbol{W}_i|\boldsymbol{\xi}_i)f(\boldsymbol{\xi}_i)}{f(\boldsymbol{Y}_i,\boldsymbol{W}_i|\boldsymbol{Z}_i)}d\boldsymbol{\xi}_i \\
&= \frac{\pi_{ic}|\boldsymbol{V}_{ic}|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{Y}}_{ic}^{\top}\boldsymbol{V}_{ic}^{-1}\widetilde{\boldsymbol{Y}}_{ic}+\widetilde{\boldsymbol{W}}_i^{\top}\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i)\}}{\sum_{c'=1}^{C}\pi_{ic'}(2\pi)^{\frac{p}{2}}|\boldsymbol{\Omega}_{ic'}|^{\frac{1}{2}}|\boldsymbol{V}_{ic'}|^{-\frac{1}{2}}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{Y}}_{ic'}^{\top}\boldsymbol{V}_{ic'}^{-1}\widetilde{\boldsymbol{Y}}_{ic'}+\widetilde{\boldsymbol{W}}_i^{\top}\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i-\widetilde{\boldsymbol{\xi}}_{ic'}^{\top}\boldsymbol{\Omega}_{ic'}^{-1}\widetilde{\boldsymbol{\xi}}_{ic'})\}} \\
&\quad \times\int\exp[-\frac{1}{2}\{\boldsymbol{\xi}_i^{\top}(\boldsymbol{\Psi}_i^{*\top}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\Psi}_i^{*}+\boldsymbol{\Sigma}_{\xi}^{-1}+\widetilde{\boldsymbol{\Psi}}_{ic}^{\top}\boldsymbol{V}_{ic}^{-1}\widetilde{\boldsymbol{\Psi}}_{ic})\boldsymbol{\xi}_i-(\widetilde{\boldsymbol{Y}}_{ic}^{\top}\boldsymbol{V}_{ic}^{-1}\widetilde{\boldsymbol{\Psi}}_{ic}+\widetilde{\boldsymbol{W}}_i^{\top}\boldsymbol{\Lambda}_i^{-1}\boldsymbol{\Psi}_i^{*})\boldsymbol{\xi}_i \\
&\quad -\boldsymbol{\xi}_i^{\top}(\widetilde{\boldsymbol{\Psi}}_{ic}^{\top}\boldsymbol{V}_{ic}^{-1}\widetilde{\boldsymbol{Y}}_{ic}+\boldsymbol{\Psi}_i^{*\top}\boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i)\}]d\boldsymbol{\xi}_i \\
&= \frac{\pi_{ic}|\boldsymbol{V}_{ic}|^{-\frac{1}{2}}|\boldsymbol{\Omega}_{ic}|^{\frac{1}{2}}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{Y}}_{ic}^{\top}\boldsymbol{V}_{ic}^{-1}\widetilde{\boldsymbol{Y}}_{ic}-\widetilde{\boldsymbol{\xi}}_{ic}^{\top}\boldsymbol{\Omega}_{ic}^{-1}\widetilde{\boldsymbol{\xi}}_{ic}))\}}{\sum_{c'=1}^{C}\pi_{ic'}|\boldsymbol{V}_{ic'}|^{-\frac{1}{2}}|\boldsymbol{\Omega}_{ic'}|^{\frac{1}{2}}\exp\{-\frac{1}{2}(\widetilde{\boldsymbol{Y}}_{ic'}^{\top}\boldsymbol{V}_{ic'}^{-1}\widetilde{\boldsymbol{Y}}_{ic'}-\widetilde{\boldsymbol{\xi}}_{ic'}^{\top}\boldsymbol{\Omega}_{ic'}^{-1}\widetilde{\boldsymbol{\xi}}_{ic'})\}} \\
&= \widetilde{\pi}_{ic}.
\end{aligned}
\tag{D.2}
$$

The conditional distribution of $\boldsymbol{\xi}_i$ given $\boldsymbol{Y}_i, \boldsymbol{W}_i$ and $\boldsymbol{Z}_i$ can be expressed as

$$
\begin{aligned}
f(\boldsymbol{\xi}_i|\boldsymbol{Y}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i) &= \sum_{c=1}^{C} f(L_{ic}=1, \boldsymbol{\xi}_i|\boldsymbol{Y}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i) \\
&= \sum_{c=1}^{C} \frac{f(L_{ic}=1, \boldsymbol{\xi}_i, \boldsymbol{Y}_i, \boldsymbol{W}_i|\boldsymbol{Z}_i)}{f(\boldsymbol{Y}_i, \boldsymbol{W}_i|\boldsymbol{Z}_i)} \\
&= \sum_{c=1}^{C} \frac{\pi_{ic}|\boldsymbol{V}_{ic}|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\widetilde{\boldsymbol{Y}}_{ic} - \widetilde{\boldsymbol{\Psi}}_{ic}\boldsymbol{\xi}_i)_{\boldsymbol{V}_{ic}}^{\otimes 2} - \frac{1}{2}(\widetilde{\boldsymbol{W}}_i - \boldsymbol{\Psi}_i^*\boldsymbol{\xi}_i)_{\boldsymbol{\Lambda}_i}^{\otimes 2} - \frac{1}{2}\boldsymbol{\xi}_i^\top \boldsymbol{\Sigma}_\xi^{-1}\boldsymbol{\xi}_i\}}{\sum_{c'=1}^{C} \pi_{ic'}(2\pi)^{\frac{p}{2}}|\boldsymbol{V}_{ic'}|^{-\frac{1}{2}}|\boldsymbol{\Omega}_{ic'}|^{\frac{1}{2}} \exp\{-\frac{1}{2}(\widetilde{\boldsymbol{Y}}_{ic'}^\top \boldsymbol{V}_{ic'}^{-1}\widetilde{\boldsymbol{Y}}_{ic'} + \widetilde{\boldsymbol{W}}_i^\top \boldsymbol{\Lambda}_i^{-1}\widetilde{\boldsymbol{W}}_i - \widetilde{\boldsymbol{\xi}}_{ic'}^\top \boldsymbol{\Omega}_{ic'}^{-1}\widetilde{\boldsymbol{\xi}}_{ic'})\}},
\end{aligned}
$$

here we define $\boldsymbol{A}_V^{\otimes 2} = \boldsymbol{A}^\top \boldsymbol{V}^{-1}\boldsymbol{A}$.

Next, the first two conditional moments of $L_{ic}\boldsymbol{\xi}_i$ given observed data are

$$
\begin{aligned}
\mathrm{E}(L_{ic}\boldsymbol{\xi}_i|\boldsymbol{Y}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i) &= \int \boldsymbol{\xi}_i \frac{\pi_{ic} f(\boldsymbol{Y}_i|\boldsymbol{\xi}_i, \boldsymbol{Z}_i, L_{ic}=1) f(\boldsymbol{W}_i|\boldsymbol{\xi}_i) f(\boldsymbol{\xi}_i)}{f(\boldsymbol{Y}_i, \boldsymbol{W}_i|\boldsymbol{Z}_i)} d\boldsymbol{\xi}_i \\
&= \frac{\pi_{ic}|\boldsymbol{\Omega}_{ic}|^{\frac{1}{2}}|\boldsymbol{V}_{ic}|^{-\frac{1}{2}}\widetilde{\boldsymbol{\xi}}_{ic} \exp(-\frac{1}{2}\widetilde{\boldsymbol{Y}}_{ic}^\top \boldsymbol{V}_{ic}^{-1}\widetilde{\boldsymbol{Y}}_{ic} + \frac{1}{2}\widetilde{\boldsymbol{\xi}}_{ic}^\top \boldsymbol{\Omega}_{ic}^{-1}\widetilde{\boldsymbol{\xi}}_{ic})}{\sum_{c'=1}^{C} \pi_{ic'}|\boldsymbol{\Omega}_{ic'}|^{\frac{1}{2}}|\boldsymbol{V}_{ic}|^{-\frac{1}{2}} \exp\{-\frac{1}{2}\widetilde{\boldsymbol{Y}}_{ic'}^\top \boldsymbol{V}_{ic'}^{-1}\widetilde{\boldsymbol{Y}}_{ic'} + \frac{1}{2}\widetilde{\boldsymbol{\xi}}_{ic'}^\top \boldsymbol{\Omega}_{ic'}^{-1}\widetilde{\boldsymbol{\xi}}_{ic'}\}} \\
&= \widetilde{\pi}_{ic}\widetilde{\boldsymbol{\xi}}_{ic}, \qquad \text{(D.3)} \\
\mathrm{E}(L_{ic}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^\top|\boldsymbol{Y}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i) &= \int \boldsymbol{\xi}_i\boldsymbol{\xi}_i^\top \frac{\pi_{ic} f(\boldsymbol{Y}_i|\boldsymbol{\xi}_i, \boldsymbol{Z}_i, L_{ic}=1) f(\boldsymbol{W}_i|\boldsymbol{\xi}_i) f(\boldsymbol{\xi}_i)}{f(\boldsymbol{Y}_i, \boldsymbol{W}_i|\boldsymbol{Z}_i)} d\boldsymbol{\xi}_i \\
&= \frac{\pi_{ic}|\boldsymbol{\Omega}_{ic}|^{\frac{1}{2}}|\boldsymbol{V}_{ic}|^{-\frac{1}{2}}(\widetilde{\boldsymbol{\xi}}_{ic}\widetilde{\boldsymbol{\xi}}_{ic}^\top + \boldsymbol{\Omega}_{ic}) \exp\{-\frac{1}{2}\widetilde{\boldsymbol{Y}}_{ic}^\top \boldsymbol{V}_{ic}^{-1}\widetilde{\boldsymbol{Y}}_{ic} + \frac{1}{2}\widetilde{\boldsymbol{\xi}}_{ic}^\top \boldsymbol{\Omega}_{ic}^{-1}\widetilde{\boldsymbol{\xi}}_{ic}\}}{\sum_{c'=1}^{C} \pi_{ic'}|\boldsymbol{\Omega}_{ic'}|^{\frac{1}{2}}|\boldsymbol{V}_{ic}|^{-\frac{1}{2}} \exp(-\frac{1}{2}\widetilde{\boldsymbol{Y}}_{ic'}^\top \boldsymbol{V}_{ic'}^{-1}\widetilde{\boldsymbol{Y}}_{ic'} + \frac{1}{2}\widetilde{\boldsymbol{\xi}}_{ic'}^\top \boldsymbol{\Omega}_{ic'}^{-1}\widetilde{\boldsymbol{\xi}}_{ic'})} \\
&= \widetilde{\pi}_{ic}(\widetilde{\boldsymbol{\xi}}_{ic}\widetilde{\boldsymbol{\xi}}_{ic}^\top + \boldsymbol{\Omega}_{ic}) = \widetilde{\pi}_{ic}\widetilde{\boldsymbol{\Sigma}}_{ic}. \qquad \text{(D.4)}
\end{aligned}
$$

Furthermore, the first two conditional moments of $[\boldsymbol{\xi}_i|\boldsymbol{Y}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i]$ are

$$
\begin{aligned}
\mathrm{E}(\boldsymbol{\xi}_i|\boldsymbol{Y}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i) &= \int \boldsymbol{\xi}_i f(\boldsymbol{\xi}_i|\boldsymbol{Y}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i) d\boldsymbol{\xi}_i \\
&= \sum_{c=1}^{C} \frac{\pi_{ic}|\boldsymbol{\Omega}_{ic}|^{\frac{1}{2}}|\boldsymbol{V}_{ic}|^{-\frac{1}{2}} \exp(-\frac{1}{2}\widetilde{\boldsymbol{Y}}_{ic}^\top \boldsymbol{V}_{ic}^{-1}\widetilde{\boldsymbol{Y}}_{ic} + \frac{1}{2}\widetilde{\boldsymbol{\xi}}_{ic}^\top \boldsymbol{\Omega}_{ic}^{-1}\widetilde{\boldsymbol{\xi}}_{ic})\widetilde{\boldsymbol{\xi}}_{ic}}{\sum_{c'=1}^{C} \pi_{ic'}|\boldsymbol{\Omega}_{ic'}|^{\frac{1}{2}}|\boldsymbol{V}_{ic'}|^{-\frac{1}{2}} \exp\{-\frac{1}{2}\widetilde{\boldsymbol{Y}}_{ic'}^\top \boldsymbol{V}_{ic'}^{-1}\widetilde{\boldsymbol{Y}}_{ic'} + \frac{1}{2}\widetilde{\boldsymbol{\xi}}_{ic'}^\top \boldsymbol{\Omega}_{ic'}^{-1}\widetilde{\boldsymbol{\xi}}_{ic'})\}}
\end{aligned}
$$

$$= \sum_{c=1}^{C} \widetilde{\pi}_{ic} \widetilde{\boldsymbol{\xi}}_{ic} = \widetilde{\boldsymbol{\xi}}_i, \tag{D.5}$$

$$\begin{aligned}
\mathrm{E}(\boldsymbol{\xi}_i \boldsymbol{\xi}_i^\top | \boldsymbol{Y}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i) &= \int \boldsymbol{\xi}_i \boldsymbol{\xi}_i^\top f(\boldsymbol{\xi}_i | \boldsymbol{Y}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i) d\boldsymbol{\xi}_i \\
&= \sum_{c=1}^{C} \int \boldsymbol{\xi}_i \boldsymbol{\xi}_i^\top f(L_{ic} = 1, \boldsymbol{\xi}_i | \boldsymbol{Y}_i, \boldsymbol{W}_i, \boldsymbol{Z}_i) d\boldsymbol{\xi}_i \\
&= \sum_{c=1}^{C} \frac{\pi_{ic} |\boldsymbol{\Omega}_{ic}|^{\frac{1}{2}} |\boldsymbol{V}_{ic}|^{-\frac{1}{2}} (\widetilde{\boldsymbol{\xi}}_{ic} \widetilde{\boldsymbol{\xi}}_{ic}^\top + \boldsymbol{\Omega}_{ic}) \exp(-\frac{1}{2} \widetilde{\boldsymbol{Y}}_{ic}^\top \boldsymbol{V}_{ic}^{-1} \widetilde{\boldsymbol{Y}}_{ic} + \frac{1}{2} \widetilde{\boldsymbol{\xi}}_{ic}^\top \boldsymbol{\Omega}_{ic}^{-1} \widetilde{\boldsymbol{\xi}}_{ic})}{\sum_{c'=1}^{C} \pi_{ic'} |\boldsymbol{\Omega}_{ic'}|^{\frac{1}{2}} |\boldsymbol{V}_{ic'}|^{-\frac{1}{2}} \exp\{-\frac{1}{2} \widetilde{\boldsymbol{Y}}_{ic'}^\top \boldsymbol{V}_{ic'}^{-1} \widetilde{\boldsymbol{Y}}_{ic'} + \frac{1}{2} \widetilde{\boldsymbol{\xi}}_{ic'}^\top \boldsymbol{\Omega}_{ic'}^{-1} \widetilde{\boldsymbol{\xi}}_{ic'})\}} \\
&= \sum_{c=1}^{C} \widetilde{\pi}_{ic} (\widetilde{\boldsymbol{\xi}}_{ic} \widetilde{\boldsymbol{\xi}}_{ic}^\top + \boldsymbol{\Omega}_{ic}) = \sum_{c=1}^{C} \widetilde{\pi}_{ic} \widetilde{\boldsymbol{\Sigma}}_{ic} = \widetilde{\boldsymbol{\Sigma}}_i.
\end{aligned} \tag{D.6}$$

Suppose the conditional moments derived above have the partition $\widetilde{\boldsymbol{\xi}}_{ic} = (\widetilde{\boldsymbol{\xi}}_{i1,c}, \ldots, \widetilde{\boldsymbol{\xi}}_{id_x,c})^\top$ and $\widetilde{\boldsymbol{\xi}}_i = (\widetilde{\boldsymbol{\xi}}_{i1}, \ldots, \widetilde{\boldsymbol{\xi}}_{id_x})^\top$, $\widetilde{\boldsymbol{\Sigma}}_{ic} = (\widetilde{\boldsymbol{\Sigma}}_{ic,vv'})_{v,v'=1}^{d_x} = (\widetilde{\boldsymbol{\xi}}_{iv,c} \widetilde{\boldsymbol{\xi}}_{iv',c}^\top + \boldsymbol{\Omega}_{ic,vv'})_{vv'=1}^{d_x}$ with $\boldsymbol{\Omega}_{ic} = (\boldsymbol{\Omega}_{ic,vv'})_{v,v'=1}^{d_x}$. Combine the calculations above, we have

$$Q(\boldsymbol{\Theta} | \boldsymbol{\Theta}^{(\kappa)}) = Q_1(\boldsymbol{\Theta} | \boldsymbol{\Theta}^{(\kappa)}) + Q_2(\boldsymbol{\Theta} | \boldsymbol{\Theta}^{(\kappa)}) + Q_3(\boldsymbol{\Theta} | \boldsymbol{\Theta}^{(\kappa)}) + Q_4(\boldsymbol{\Theta} | \boldsymbol{\Theta}^{(\kappa)}),$$

where

$$\begin{aligned}
Q_1(\boldsymbol{\Theta} | \boldsymbol{\Theta}^{(\kappa)}) &= \sum_{i=1}^{n} \sum_{c=1}^{C} -2 \times \widetilde{\pi}_{ic} \times \log(\pi_{ic}), \\
Q_2(\boldsymbol{\Theta} | \boldsymbol{\Theta}^{(\kappa)}) &= \sum_{i=1}^{n} \sum_{c=1}^{C} \Bigg(\widetilde{\pi}_{ic} \Bigg[(\boldsymbol{Y}_i - \beta_{0,c} \mathbf{1}_{m_{y,i}} - \sum_{v=1}^{d_x} \beta_{x,cv} \boldsymbol{B}_{iv} \boldsymbol{\theta}_{\mu v} - \sum_{v=1}^{d_x} \beta_{x,cv} \boldsymbol{B}_{iv} \boldsymbol{\Theta}_{\psi v} \widetilde{\boldsymbol{\xi}}_{iv,c} \\
&\qquad - \boldsymbol{Z}_i \boldsymbol{\beta}_{z,c})_{\boldsymbol{V}_{ic}}^{\otimes 2} + \sum_{v=1}^{d_x} \sum_{v'=1}^{d_x} \beta_{x,cv} \beta_{x,cv'} \operatorname{tr}\{\boldsymbol{\Theta}_{\psi v}^\top \boldsymbol{B}_{iv}^\top \boldsymbol{V}_{ic}^{-1} \boldsymbol{B}_{iv'} \boldsymbol{\Theta}_{\psi v'} \boldsymbol{\Omega}_{ic,v'v}\} + \log(|\boldsymbol{V}_{ic}|) \Bigg] \Bigg), \\
Q_3(\boldsymbol{\Theta} | \boldsymbol{\Theta}^{(\kappa)}) &= \sum_{i=1}^{n} \sum_{v=1}^{d_x} \Bigg[m_{x,iv} \log(\sigma_v^2) + \frac{1}{\sigma_v^2} \Bigg\{ (\boldsymbol{W}_{iv} - \boldsymbol{B}_{iv}^* \boldsymbol{\theta}_{\mu v} - \boldsymbol{B}_{iv}^* \boldsymbol{\Theta}_{\psi v} \widetilde{\boldsymbol{\xi}}_{iv})^{\otimes 2} \\
&\qquad - \widetilde{\boldsymbol{\xi}}_{iv}^\top \boldsymbol{\Theta}_{\psi v}^\top (\boldsymbol{B}_{iv}^*)^\top \boldsymbol{B}_{iv}^* \boldsymbol{\Theta}_{\psi v} \widetilde{\boldsymbol{\xi}}_{iv} + \operatorname{tr}(\boldsymbol{\Theta}_{\psi v}^\top (\boldsymbol{B}_{iv}^*)^\top \boldsymbol{B}_{iv}^* \boldsymbol{\Theta}_{\psi v} \widetilde{\boldsymbol{\Sigma}}_{i,vv}) \Bigg\} \Bigg],
\end{aligned}$$

$$Q_4(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) = n\log(|\boldsymbol{\Sigma}_\xi|) + \sum_{i=1}^{n} \mathrm{tr}(\boldsymbol{\Sigma}_\xi^{-1}\widetilde{\boldsymbol{\Sigma}}_i). \tag{D.7}$$

To update the parameters, the M-step involves executing a sequence of steps, each applied to one block at a time. Also, the updates for the linear fixed effect of the mixture component bear many similarities to those for the linear mixed effect of the mixture component.

D.1 Update $\boldsymbol{\gamma}$

Using the logistic regression model in the main paper, Q_1 can be rewritten as

$$Q_1(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) = -2\sum_{i=1}^{n}\left[(\sum_{c=1}^{C-1}\widetilde{\pi}_{ic}\widetilde{\mathbb{Z}}_i^\top\boldsymbol{\gamma}_c) - \log\{1+\sum_{c'=1}^{C-1}\exp(\widetilde{\mathbb{Z}}_i^\top\boldsymbol{\gamma}_{c'})\}\right].$$

Then we use a Newton-Raphson algorithm to update our $\boldsymbol{\gamma}$. Given the value of $\boldsymbol{\gamma} = \widehat{\boldsymbol{\gamma}}^{(\zeta)}$ at the end of the ζth iteration, we update our $\widehat{\boldsymbol{\gamma}}$ as

$$\widehat{\boldsymbol{\gamma}}^{(\zeta+1)} = \widehat{\boldsymbol{\gamma}}^{(\zeta)} - \boldsymbol{H}^{-1}(\widehat{\boldsymbol{\gamma}}^{(\zeta)}) \bigtriangledown Q_1(\widehat{\boldsymbol{\gamma}}^{(\zeta)}),$$

where $\bigtriangledown Q_1(\widehat{\boldsymbol{\gamma}}^{(\zeta)}) = \{(\frac{\partial Q_1}{\partial \boldsymbol{\gamma}_1})^\top, (\frac{\partial Q_1}{\partial \boldsymbol{\gamma}_2})^\top, \ldots, (\frac{\partial Q_1}{\partial \boldsymbol{\gamma}_{C-1}})^\top\}$ with

$$\frac{\partial Q_1}{\partial \boldsymbol{\gamma}_c} = -2\sum_{i=1}^{n}\{\widetilde{\pi}_{ic}\widetilde{\mathbb{Z}}_i - \frac{\widetilde{\mathbb{Z}}_i\exp(\widetilde{\mathbb{Z}}_i^\top\boldsymbol{\gamma}_c)}{1+\sum_{c'=1}^{C-1}\exp(\widetilde{\mathbb{Z}}_i^\top\boldsymbol{\gamma}_{c'})}\}, \quad c = 1,\ldots,C-1, \tag{D.8}$$

and $\boldsymbol{H}$ is the Hessian matrix with diagonal and off-diagonal entries as

$$\frac{\partial^2 Q_1}{\partial \boldsymbol{\gamma}_c \partial \boldsymbol{\gamma}_c^\top} = 2\sum_{i=1}^{n} \frac{\widetilde{\mathbb{Z}}_i \widetilde{\mathbb{Z}}_i^\top \exp(\widetilde{\mathbb{Z}}_i^\top \boldsymbol{\gamma}_c)\{1+\sum_{c'=1}^{C-1}\exp(\widetilde{\mathbb{Z}}_i^\top \boldsymbol{\gamma}_{c'})\} - \widetilde{\mathbb{Z}}_i \widetilde{\mathbb{Z}}_i^\top \exp(2\widetilde{\mathbb{Z}}_i^\top \boldsymbol{\gamma}_c)}{\{1+\sum_{c'=1}^{C-1}\exp(\widetilde{\mathbb{Z}}_i^\top \boldsymbol{\gamma}_{c'})\}^2},$$

$$\frac{\partial^2 Q_1}{\partial \boldsymbol{\gamma}_c \partial \boldsymbol{\gamma}_{c'}^\top} = -2\sum_{i=1}^{n} \frac{\widetilde{\mathbb{Z}}_i \widetilde{\mathbb{Z}}_i^\top \exp(\widetilde{\mathbb{Z}}_i^\top \boldsymbol{\gamma}_c + \widetilde{\mathbb{Z}}_i^\top \boldsymbol{\gamma}_{c'})}{\{1+\sum_{c'=1}^{C-1}\exp(\widetilde{\mathbb{Z}}_i^\top \boldsymbol{\gamma}_{c'})\}^2} \qquad \text{if } c \neq c'.$$

Since this Newton-Raphson iteration is embedded in our EM algorithm, we only iterate two steps to update $\widehat{\boldsymbol{\gamma}}$ within each EM iteration to save computation time.

D.2 Update $\boldsymbol{\beta}$

When updating $\boldsymbol{\beta}_c$, define $m_{y,i} \times d$ matrix $\mathbb{X}_{ic} = (\mathbf{1}_{m_{y,i}}, \boldsymbol{B}_i\widetilde{\boldsymbol{\Theta}}_\mu + \boldsymbol{B}_i\boldsymbol{\Theta}_\psi\boldsymbol{\Xi}_{ic}, \boldsymbol{Z}_i)$, where $d = 1+d_x+d_z$, $\boldsymbol{B}_i = (\boldsymbol{B}_{i1}, \ldots, \boldsymbol{B}_{id_x})$ is a $m_{y,i} \times q$ matrix, $\widetilde{\boldsymbol{\Theta}}_\mu = \text{diag}(\boldsymbol{\theta}_{\mu 1}, \ldots, \boldsymbol{\theta}_{\mu d_x})$ is a $q \times d_x$ matrix, and $\boldsymbol{\Xi}_{ic} = \text{diag}(\widetilde{\boldsymbol{\xi}}_{i1,c}, \ldots, \widetilde{\boldsymbol{\xi}}_{id_x,c})$ is a $p \times d_x$ matrix. A closed-form expression to update $\boldsymbol{\beta}_c$ is

$$\widehat{\boldsymbol{\beta}}_c = \{\sum_{i=1}^{n} \widetilde{\pi}_{ic}(\mathbb{X}_{ic}^\top \boldsymbol{V}_{ic}^{-1}\mathbb{X}_{ic} + \boldsymbol{A}_{ic})\}^{-1}(\sum_{i=1}^{n} \widetilde{\pi}_{ic}\mathbb{X}_{ic}^\top \boldsymbol{V}_{ic}^{-1}\boldsymbol{Y}_i) \tag{D.9}$$

where $\boldsymbol{A}_{ic} = \text{diag}(\mathbf{0}, [\text{tr}\{\boldsymbol{\Theta}_{\psi v}^\top \boldsymbol{B}_{iv}^\top \boldsymbol{V}_{ic}^{-1} \boldsymbol{B}_{iv'} \boldsymbol{\Theta}_{\psi v'} \boldsymbol{\Omega}_{ic,v'v}\}]_{v,v'=1}^{d_x}, \mathbf{0})$ is a $d \times d$ matrix.

D.3 Update the variance components: $\boldsymbol{V}_{ic}$ and σ_v^2 for $v = 1, \ldots, d_x$

Since the dimension $m_{y,i}$ of $\boldsymbol{V}_{ic}$ are subject-specific, there is no closed-form solution for the variance-covariance matrix of the linear mixed model. By straightforward algebra

$|\boldsymbol{V}_{ic}| = |\sigma_\epsilon^2 \boldsymbol{I}| \times (1 + \frac{m_{y,i}\sigma_{b,c}^2}{\sigma_\epsilon^2}), \boldsymbol{V}_{ic}^{-1} = \frac{1}{\sigma_\epsilon^2}\boldsymbol{I} - \frac{\sigma_{b,c}^2}{\sigma_\epsilon^4 + m_{y,i}\sigma_{b,c}^2\sigma_\epsilon^2}\mathbf{1}_{m_{y,i}}\mathbf{1}_{m_{y,i}}^\top$, we can rewrite Q_2 as

$$
\begin{aligned}
Q_2(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(\kappa)}) \;=\; & \sum_{i=1}^{n}\sum_{c=1}^{C}\Bigg[\widetilde{\pi}_{ic}\Big\{\log(1+\frac{m_{y,i}\sigma_{b,c}^2}{\sigma_\epsilon^2}) + \frac{1}{\sigma_\epsilon^2}(\boldsymbol{Y}_i - \mathbb{X}_{ic}\boldsymbol{\beta}_c)^\top(\boldsymbol{Y}_i - \mathbb{X}_{ic}\boldsymbol{\beta}_c) \\
& + \frac{1}{\sigma_\epsilon^2}\boldsymbol{\beta}_c^\top \boldsymbol{A}_{1,ic}\boldsymbol{\beta}_c - \frac{\sigma_{b,c}^2}{\sigma_\epsilon^4 + m_{y,i}\sigma_{b,c}^2\sigma_\epsilon^2}(\boldsymbol{Y}_i - \mathbb{X}_{ic}\boldsymbol{\beta}_c)^\top \mathbf{1}_{m_{y,i}}\mathbf{1}_{m_{y,i}}^\top(\boldsymbol{Y}_i - \mathbb{X}_{ic}\boldsymbol{\beta}_c) \\
& - \frac{\sigma_{b,c}^2}{\sigma_\epsilon^4 + m_{y,i}\sigma_{b,c}^2\sigma_\epsilon^2}\boldsymbol{\beta}_c^\top \boldsymbol{A}_{2,ic}\boldsymbol{\beta}_c + m_{y,i}\log(\sigma_\epsilon^2)\Big\}\Bigg],
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{A}_{1,ic} &= \mathrm{diag}\big[\mathbf{0}, \{\mathrm{tr}(\boldsymbol{\Theta}_{\psi v}^\top \boldsymbol{B}_{iv}^\top \boldsymbol{B}_{iv'}\boldsymbol{\Theta}_{\psi v'}\boldsymbol{\Omega}_{ic,v'v})\}_{v,v'=1}^{d_x}, \mathbf{0}\big] \quad \text{and} \\
\boldsymbol{A}_{2,ic} &= \mathrm{diag}[\mathbf{0}, \{\mathrm{tr}(\boldsymbol{\Theta}_{\psi v}^\top \boldsymbol{B}_{iv}^\top \mathbf{1}_{m_{y,i}}\mathbf{1}_{m_{y,i}}^\top \boldsymbol{B}_{iv'}\boldsymbol{\Theta}_{\psi v'}\boldsymbol{\Omega}_{ic,v'v})\}_{v,v'=1}^{d_x}, \mathbf{0}]
\end{aligned}
$$

are $d \times d$ matrices. We use a Newton-Raphson algorithm to update σ_ϵ^2: given the value of σ_ϵ^2 from the ζth step, denoted as $[\widehat{\sigma}_\epsilon^2]^{(\zeta)}$, we update σ_ϵ^2 as

$$
\big[\widehat{\sigma}_\epsilon^2\big]^{(\zeta+1)} = \big[\widehat{\sigma}_\epsilon^2\big]^{(\zeta)} - \Big[\frac{\partial^2 Q_2}{\partial(\sigma_\epsilon^2)^2}\Big|_{\sigma_\epsilon^2=[\widehat{\sigma}_\epsilon^2]^{(\zeta)}}\Big]^{-1}\Big[\frac{\partial Q_2}{\partial \sigma_\epsilon^2}\Big|_{\sigma_\epsilon^2=[\widehat{\sigma}_\epsilon^2]^{(\zeta)}}\Big]
$$

where

$$
\begin{aligned}
\frac{\partial Q_2}{\partial \sigma_\epsilon^2} \;=\; & \sum_{i=1}^{n}\sum_{c=1}^{C}\Bigg[\widetilde{\pi}_{ic}\Big\{\frac{2\sigma_{b,c}^2\sigma_\epsilon^2 + m_{y,i}\sigma_{b,c}^4}{\sigma_\epsilon^4(\sigma_\epsilon^2 + m_{y,i}\sigma_{b,c}^2)^2}(\boldsymbol{Y}_i - \mathbb{X}_{ic}\boldsymbol{\beta}_c)^\top \mathbf{1}_{m_{y,i}}\mathbf{1}_{m_{y,i}}^\top(\boldsymbol{Y}_i - \mathbb{X}_{ic}\boldsymbol{\beta}_c) \\
& + \frac{2\sigma_{b,c}^2\sigma_\epsilon^2 + m_{y,i}\sigma_{b,c}^4}{\sigma_\epsilon^4(\sigma_\epsilon^2 + m_{y,i}\sigma_{b,c}^2)^2}\boldsymbol{\beta}_c^\top \boldsymbol{A}_{2,ic}\boldsymbol{\beta}_c - \frac{1}{\sigma_\epsilon^4}\boldsymbol{\beta}_c^\top \boldsymbol{A}_{1,ic}\boldsymbol{\beta}_c - \frac{1}{\sigma_\epsilon^4}(\boldsymbol{Y}_i - \mathbb{X}_{ic}\boldsymbol{\beta}_c)^{\otimes 2} \\
& - \frac{m_{y,i}\sigma_{b,c}^2}{\sigma_\epsilon^4 + m_{y,i}\sigma_{b,c}^2\sigma_\epsilon^2} + \frac{m_{y,i}}{\sigma_\epsilon^2}\Big\}\Bigg],
\end{aligned}
\tag{D.10}
$$

$$\frac{\partial^2 Q_2}{\partial(\sigma_\epsilon^2)^2} = \sum_{i=1}^{n}\sum_{c=1}^{C}\Bigg(\widetilde{\pi}_{ic}\Bigg[\frac{2}{\sigma_\epsilon^6}(\boldsymbol{Y}_i-\mathbb{X}_{ic}\boldsymbol{\beta}_c)^{\otimes 2}+\frac{m_{y,i}\sigma_{b,c}^2(2\sigma_\epsilon^2+m_{y,i}\sigma_{b,c}^2)}{\sigma_\epsilon^4(\sigma_\epsilon^2+m_{y,i}\sigma_{b,c}^2)^2}-\frac{m_{y,i}}{\sigma_\epsilon^4}$$
$$+\frac{2}{\sigma_\epsilon^6}\boldsymbol{\beta}_c^\top\boldsymbol{A}_{1,ic}\boldsymbol{\beta}_c+\frac{2\sigma_{b,c}^2\{\sigma_\epsilon^2+m_{y,i}\sigma_{b,c}^2-(2\sigma_\epsilon^2+m_{y,i}\sigma_{b,c}^2)^2\}}{\sigma_\epsilon^6(\sigma_\epsilon^2+m_{y,i}\sigma_{b,c}^2)^3}(\boldsymbol{Y}_i-\mathbb{X}_{ic}\boldsymbol{\beta}_c)^\top$$
$$\times\mathbf{1}_{m_{y,i}}\mathbf{1}_{m_{y,i}}^\top(\boldsymbol{Y}_i-\mathbb{X}_{ic}\boldsymbol{\beta}_c)+\frac{2\sigma_{b,c}^2\{\sigma_\epsilon^2+m_{y,i}\sigma_{b,c}^2-(2\sigma_\epsilon^2+m_{y,i}\sigma_{b,c}^2)^2\}}{\sigma_\epsilon^6(\sigma_\epsilon^2+m_{y,i}\sigma_{b,c}^2)^3}\Bigg]\Bigg)$$
$$\times\boldsymbol{\beta}_c^\top\boldsymbol{A}_{2,ic}\boldsymbol{\beta}_c. \tag{D.11}$$

Then $\sigma_{b,c}^2$ and σ_v^2 can be updated as

$$\widehat{\sigma}_{b,c}^2 = \frac{1}{\sum_{i=1}^n\widetilde{\pi}_{ic}m_{y,i}^2}\sum_{i=1}^{n}\Bigg[\widetilde{\pi}_{ic}\Big\{\boldsymbol{\beta}_c^\top\boldsymbol{A}_{2,ic}\boldsymbol{\beta}_c+(\boldsymbol{Y}_i-\mathbb{X}_{ic}\boldsymbol{\beta}_c)^\top\boldsymbol{J}_{m_{y,i}}(\boldsymbol{Y}_i-\mathbb{X}_{ic}\boldsymbol{\beta}_c)-m_{y,i}\sigma_\epsilon^2\Big\}\Bigg],$$

for $c=1,\ldots,C$, and

$$\widehat{\sigma}_v^2 = \frac{1}{m_{x,i}}\sum_{i=1}^{n}\Bigg[(\boldsymbol{W}_{iv}-\boldsymbol{B}_{iv}^*\boldsymbol{\theta}_{\mu v}-\boldsymbol{B}_{iv}^*\boldsymbol{\Theta}_{\psi v}\widetilde{\boldsymbol{\xi}}_{iv})^{\otimes 2}+\mathrm{tr}\{\boldsymbol{\Theta}_{\psi v}^\top(\boldsymbol{B}_{iv}^*)^\top\boldsymbol{B}_{iv}^*\boldsymbol{\Theta}_{\psi v}(\widetilde{\boldsymbol{\Sigma}}_{i,vv}-\widetilde{\boldsymbol{\xi}}_{iv}\widetilde{\boldsymbol{\xi}}_{iv}^\top)\}\Bigg],$$

for $v=1,2,\ldots,d_x$.

D.4 Update spline coefficients for the mean and eigenfunctions

We first update the coefficients of the mean functions by

$$\widehat{\boldsymbol{\theta}}_{\mu v}=\Bigg(\sum_{i=1}^{n}\Bigg[\{\sum_{c=1}^{C}\widetilde{\pi}_{ic}\beta_{x,cv}^2\boldsymbol{B}_{iv}^\top\boldsymbol{V}_{ic}^{-1}\boldsymbol{B}_{iv}\}+\frac{1}{\sigma_v^2}(\boldsymbol{B}_{iv}^*)^\top\boldsymbol{B}_{iv}^*\Bigg]\Bigg)^{-1}\Bigg(\sum_{i=1}^{n}\Bigg[\frac{1}{\sigma_v^2}\{(\boldsymbol{B}_{iv}^*)^\top(\boldsymbol{W}_{iv}$$
$$-\boldsymbol{B}_{iv}^*\boldsymbol{\Theta}_{\psi v}\widetilde{\boldsymbol{\xi}}_{iv})\}+\{\sum_{c=1}^{C}\widetilde{\pi}_{ic}\beta_{x,cv}\boldsymbol{B}_{iv}^\top\boldsymbol{V}_{ic}^{-1}(\boldsymbol{Y}_i-\beta_{0,c}\mathbf{1}-\sum_{v'\neq v;v'=1}^{d_x}\beta_{x,cv'}\boldsymbol{B}_{iv'}\boldsymbol{\theta}_{\mu v'}$$
$$-\sum_{v=1}^{d_x}\beta_{x,cv}\boldsymbol{B}_{iv}\boldsymbol{\Theta}_{\psi v}\widetilde{\boldsymbol{\xi}}_{iv,c}-\boldsymbol{Z}_i\boldsymbol{\beta}_{z,c})\}\Bigg]\Bigg),\qquad v=1,\ldots,d_x. \tag{D.12}$$

When updating the coefficient for eigenfunctions, we will update the coefficient for one eigenfunction at a time until all the coefficients are updated. Specifically, denote the coefficient matrix for the eigenfunctions of $X_v(\cdot)$ as $\boldsymbol{\Theta}_{\psi v} = (\boldsymbol{\theta}_{\psi v,1}, \boldsymbol{\theta}_{\psi v,2}, \ldots, \boldsymbol{\theta}_{\psi v,p_v})$, where each column contains the spline coefficients of one eigenfunction. We update the columns of $\boldsymbol{\Theta}_{\psi v}$ one at a time using a Newton-Raphson step while holding all other columns fixed. Let $\boldsymbol{\Theta}_{\psi v}^{(\kappa)}$ be its value at κ^{th} step, we update the ιth column by

$$\widehat{\boldsymbol{\theta}}_{\psi v,\iota}^{(\kappa+1)} = \boldsymbol{\theta}_{\psi v,\iota}^{(\kappa)} - [\boldsymbol{H}(\widehat{\boldsymbol{\theta}}_{\psi v,\iota}^{(\kappa)})]^{-1} \bigtriangledown_{\boldsymbol{\theta}_{\psi v,\iota}} (Q_2 + Q_3)(\boldsymbol{\Theta}^{(\kappa)}),$$

where $\boldsymbol{H}(\widehat{\boldsymbol{\theta}}_{\psi v,\iota}^{(\kappa)})$ and $\bigtriangledown_{\boldsymbol{\theta}_{\psi v,\iota}}(Q_2+Q_3)(\boldsymbol{\Theta}^{(\kappa)})$ are hessian and gradient of $\widehat{Q}_2+\widehat{Q}_3$ with respect to $\boldsymbol{\theta}_{\psi v,\iota}$.

We have

$$\begin{aligned}
\bigtriangledown_{\boldsymbol{\theta}_{\psi v,\iota}}(Q_2 + Q_3)(\boldsymbol{\Theta}^{(\kappa)}) &= -2\sum_{i=1}^{n}\sum_{c=1}^{C}\widetilde{\pi}_{ic}\Big\{\beta_{x,cv}\widetilde{\xi}_{ivc}^{[\iota]}\boldsymbol{B}_{iv}^{\top}\boldsymbol{V}_{ic}^{-1}(\widetilde{\boldsymbol{Y}}_{ic} - \sum_{v=1}^{d_x}\beta_{x,cv}\boldsymbol{B}_{iv}\boldsymbol{\Theta}_{\psi v}\widetilde{\boldsymbol{\xi}}_{iv,c}) \\
&\quad -\beta_{x,cv}\boldsymbol{B}_{iv}^{\top}\boldsymbol{V}_{ic}^{-1}\sum_{v'=1}^{d_x}\beta_{x,cv'}\boldsymbol{B}_{iv'}\boldsymbol{\Theta}_{\psi v'}\boldsymbol{\Omega}_{ic,v'v}^{[;\iota]}\Big\} \\
&\quad -2\sum_{i=1}^{n}\Big[\frac{1}{\sigma_v^2}\{(\boldsymbol{B}_{iv}^{*})^{\top}\widetilde{\xi}_{iv}^{[\iota]}(\widetilde{\boldsymbol{W}}_{iv} - \boldsymbol{B}_{iv}^{*}\boldsymbol{\Theta}_{\psi v}\widetilde{\boldsymbol{\xi}}_{iv}) \\
&\quad -(\boldsymbol{B}_{iv}^{*})^{\top}\boldsymbol{B}_{iv}^{*}\boldsymbol{\Theta}_{\psi v}(\widetilde{\boldsymbol{\Sigma}}_{i,vv} - \widetilde{\boldsymbol{\xi}}_{iv}\widetilde{\boldsymbol{\xi}}_{iv}^{\top})^{[;\iota]})\}\Big], \qquad \text{(D.13)}
\end{aligned}$$

where $\boldsymbol{\Omega}_{ic,v'v}^{[;\iota]}$ denotes the ιth column of $\boldsymbol{\Omega}_{ic,v'v}$ for $v, v' = 1, \ldots, d_x$, $\widetilde{\xi}_{iv,c}^{[\iota]}$ denotes the ιth element in the vector $\widetilde{\boldsymbol{\xi}}_{iv,c}$. Furthermore,

$$\boldsymbol{H}(\widehat{\boldsymbol{\theta}}_{\psi v,\iota}^{(\kappa)}) = 2\sum_{i=1}^{n}\sum_{c=1}^{C}\Big[\widetilde{\pi}_{ic}\{\beta_{x,cv}^2\widetilde{\omega}_{ic,vv,\iota\iota}\boldsymbol{B}_{iv}^{\top}\boldsymbol{V}_{ic}^{-1}\boldsymbol{B}_{iv} + \beta_{x,cv}^2\boldsymbol{B}_{iv}^{\top}\boldsymbol{V}_{ic}^{-1}\boldsymbol{B}_{iv}\widetilde{\xi}_{iv,c}^{[\iota]^2}\}\Big]$$

$$+2\sum_{i=1}^{n}\left\{\frac{1}{\sigma_v^2}(\widetilde{\delta}_{i,vv,\iota\iota}(\boldsymbol{B}_{iv}^*)^\top\boldsymbol{B}_{iv}^* + (\boldsymbol{B}_{iv}^*)^\top\boldsymbol{B}_{iv}^*\widetilde{\boldsymbol{\xi}}_{iv}^{[\iota]^2})\right\}, \tag{D.14}$$

where $\widetilde{\omega}_{ic,vv,\iota\iota}$ and $\widetilde{\delta}_{i,vv,\iota\iota}$ are the elements in ιth column and ιth row of $\boldsymbol{\Omega}_{ic,vv}$ and $(\widetilde{\boldsymbol{\Sigma}}_{i,vv} - \widetilde{\boldsymbol{\xi}}_{iv}\widetilde{\boldsymbol{\xi}}_{iv}^\top)$, respectively.

To enforce the orthonormal constraint in $\boldsymbol{\Theta}_{\psi v}$ define $D_{\xi v} = \widetilde{\boldsymbol{\Theta}}_{\psi v}\widetilde{\boldsymbol{\Sigma}}_{vv}\widetilde{\boldsymbol{\Theta}}_{\psi v}^\top$, where $\widetilde{\boldsymbol{\Theta}}_{\psi v}$ is the estimators from Newton-Ralphson algorithm, and $\widetilde{\boldsymbol{\Sigma}}_{vv} = \frac{1}{n}\sum_{i=1}^{n}\widetilde{\boldsymbol{\Sigma}}_{i,vv}$. Then we perform an eigenvalue decomposition on $D_{\xi v}$ such that $D_{\xi v} = \widehat{\boldsymbol{\Theta}}_{\psi v}\widehat{\boldsymbol{\Sigma}}_{vv}\widehat{\boldsymbol{\Theta}}_{\psi v}^\top$. Here, $\widehat{\boldsymbol{\Sigma}}_{vv}$ is a diagonal matrix that consists of the leading eigenvalues of X_v, and $\widehat{\boldsymbol{\Theta}}_{\psi v}$ consists of the spline coefficients of the eigenfunctions which satisfy the orthonormal constraints. Thereupon we have the cross-covariance estimators as

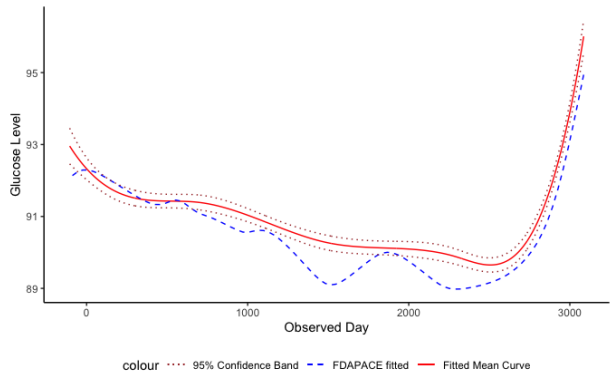
$$\widehat{\boldsymbol{\Sigma}}_{vv'} = \frac{1}{n}\sum_{i=1}^{n}\{\widehat{\boldsymbol{\Theta}}_{\psi v}^\top\widetilde{\boldsymbol{\Theta}}_{\psi v}\widetilde{\boldsymbol{\Sigma}}_{i,vv'}\widetilde{\boldsymbol{\Theta}}_{\psi v'}^\top\widehat{\boldsymbol{\Theta}}_{\psi v'}\}, \tag{D.15}$$

for any $v \neq v'$.

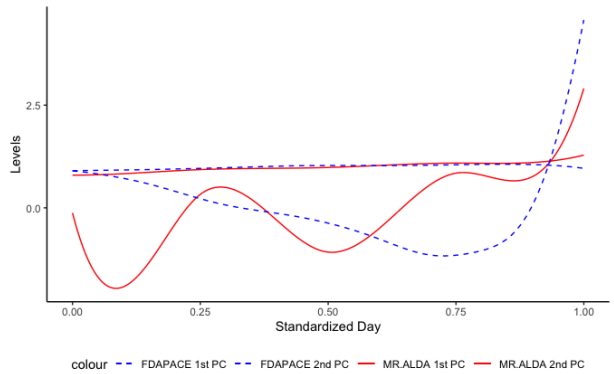
Appendix E: Additional results for real data analysis

We present the estimated mean and eigenfunctions for glucose, triglyceride, and systolic blood pressure in the real data analysis obtained from both the **fdapace** and **Emerald** models, as shown in Figure A.1, Figure A.2, and Figure A.3. In the mean curve plots, the fitted mean curve from **Emerald** is represented by a solid red line, accompanied by 95% confidence bands depicted as dotted brown lines. The confidence bands are computed based on the bootstrap samples outlined in Chapter 5. Additionally, the fitted mean curve from **fdapace** is illustrated as a dashed blue line. Through these plots, we can compare the mean and eigenfunctions between the **Emerald** and **fdapace** models. The inclusion of the confidence bands aids in assessing the uncertainty associated with the estimated mean curves.

These figures provide a visual representation of the estimated patterns and variations in the respective biomarkers, offering insights into their longitudinal behavior within the studied population. The mean curve for these three time-varying covariates appears to be relatively stable, indicating that over the course of 10 years, there is no significant overall change in the average levels of glucose, triglycerides, and systolic blood pressure within the entire population. Regarding the eigenfunctions, the first eigenfunction appears to be almost constant, suggesting that it captures the random intercept of the time-varying covariates. On the other hand, the second eigenfunction demonstrates an approximately linear trend. This implies that the second eigenfunction captures the random slope, indicating that there may be individual-specific changes in the biomarker levels over time.



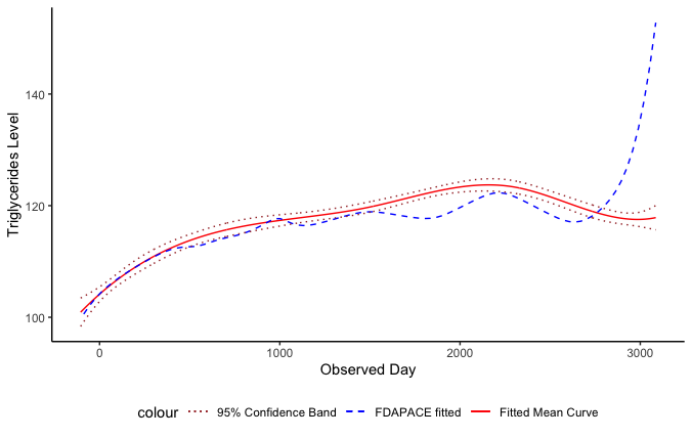
(a) Mean curve



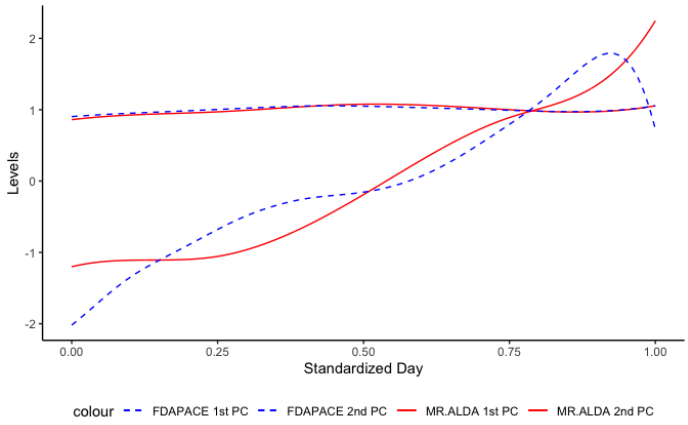
(b) Eigenfunction

Figure A.1: SWAN data: comparison of the fitted and eigenfunction for glucose. The mean plot shows the fitted mean functions (solid red), pointwise 95% bootstrap confidence intervals (dotted brown) as well as the fitted mean curve from **fdapace** (dashed blue). The eigenfunction plot shows the fitted eigenfunctions from **Emerald** (solid red) and from **fdapace** (dashed blue).

By examining these eigenfunctions, we can deduce that they effectively represent the random intercept and random slope characteristics of the three time-varying covariates.

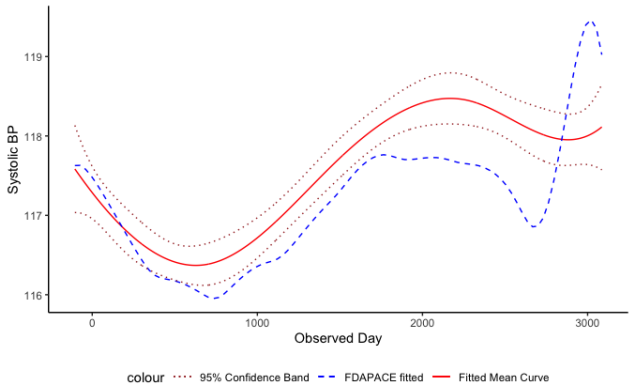


(a) Mean curve

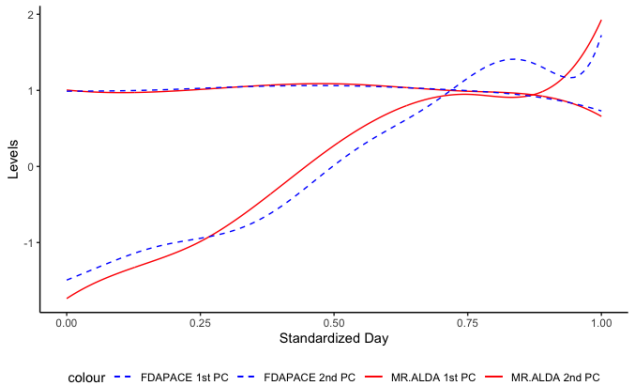


(b) Eigenfunction

Figure A.2: SWAN data: comparison of the fitted mean and eigenfunction for triglycerides. The mean plot shows the fitted eigenfunctions from **Emerald** (solid red) and from **fdapace** (dashed blue). The eigenfunction plot shows the fitted eigenfunctions from **Emerald** (solid red) and from **fdapace** (dashed blue).



(a) Mean curve



(b) Eigenfunction

Figure A.3: SWAN data: comparison of the fitted mean and eigenfunction for systolic BP. The mean plot shows the fitted eigenfunctions from **Emerald** (solid red) and from **fdapace** (dashed blue). The eigenfunction plot shows the fitted eigenfunctions from **Emerald** (solid red) and from **fdapace** (dashed blue).

This information is valuable in understanding the individual variations and trajectories of glucose, triglycerides, and systolic blood pressure within the studied population.

Appendix F: Additional simulation results with sample size equal to 500

We then increase the sample size to be $n = 500$ instead of $n = 200$, and report the bias, and standard deviations of model parameters, including logistic regression coefficients, mixture regression coefficients, and variance parameters. Furthermore, we compare our results with those of **fdapace** and **lcmm** to highlight our capabilities and demonstrate superior performance.

We follow the same instruction in Chapter 4 to simulate the data with 3 subgroups. We provide a summary of the mean and standard deviations (SD) for the Integrated Squared Error (ISE) of the functional estimators, including the mean and eigenfunctions for $X_1(t)$ and $X_2(t)$, respectively in Table A.1. In addition, we report the bias and SD of the scalar parameters, such as the eigenvalues, cross-covariance parameters, and error variance of the covariate processes, as well as the Relative Mean Squared Error (RMSE) of the Functional Principal Component (FPC) scores in Table A.1. Table A.1 and Figure A.4 provide further confirmation that increasing the sample size ($n = 500$) results in improved accuracy in the estimation of time-varying covariates. This is evident from the lower levels of bias and standard deviation observed in these evaluations. We also plot the fitted mean and eigenfunction for time-varying covariate $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ in Figure A.4 and A.5.

Furthermore, we have summarized the results for subgroup classification and effects. Table A.2 provides a summary of the bias and standard deviations of the model

Table A.1: Summary of FPCA estimators in the simulation study with $n = 500$

X_1	$\mu_1(t)$	$\psi_{11}(t)$	$\psi_{12}(t)$	ξ_{11}	ξ_{12}
Emerald	-0.002 (0.281)	0.005 (0.222)	-0.001 (0.221)	0.046(0.037)	0.057(0.044)
fdapace	-0.015 (0.047)	-0.022 (0.053)	0.039 (0.033)	0.258(0.378)	0.359(0.455)
X_2	$\mu_2(t)$	$\psi_{21}(t)$	$\psi_{22}(t)$	ξ_{21}	ξ_{22}
Emerald	0.003 (0.039)	-0.004 (0.045)	0.002 (0.011)	0.135(0.157)	0.151(0.225)
fdapace	0.099 (0.038)	0.163 (0.059)	-0.080 (0.061)	0.538(0.116)	0.764(0.144)

(a) Mean (SD) for the ISE of the functional estimators and RMSE of the FPCA scores

Eigenvalues	ω_{11}	ω_{12}	ω_{21}	ω_{22}
Emerald	0.042 (0.095)	0.072 (0.155)	-0.009 (0.102)	0.045 (0.079)
fdapace	-1.400 (0.155)	-0.849 (0.120)	-2.085 (0.062)	-1.450 (0.039)

(b) Bias (SD) for the eigenvalues

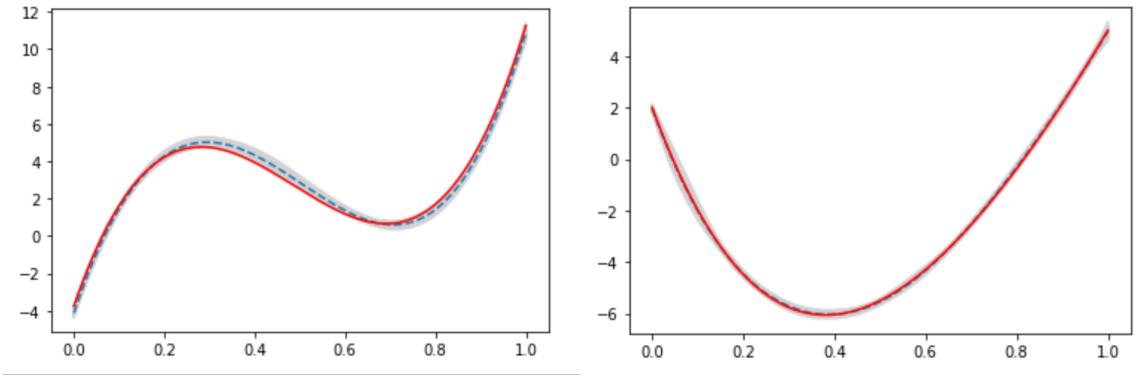
Variance	σ_1^2	σ_2^2
Emerald	0.024 (0.033)	0.014 (0.036)
fdapace	0.540 (0.260)	3.682 (0.137)

(c) Bias (SD) for the error term in $\boldsymbol{W}$

Cross-Covariance	$\omega_{12,11}$	$\omega_{12,11}$	$\omega_{12,21}$	$\omega_{12,22}$
Emerald	0.042 (0.152)	0.006 (0.181)	0.073 (0.195)	-0.019 (0.089)

(d) Bias (SD) of the cross-covariance parameters

parameters in the logistic regression coefficients. The bias and standard deviations of the model parameters in the mixture regression coefficients are summarized in Table A.3. Last, table A.4 presents the estimation of variance parameters obtained from both **lcmm** and **Emerald**. Upon careful analysis of the presented tables, it becomes evident that increasing the sample size has a significant positive impact on both the accuracy and stability of our



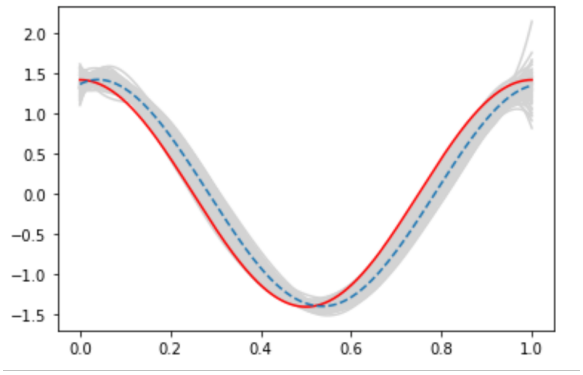
(a) Mean curve $\boldsymbol{X}_1$ (b) Mean curve of $\boldsymbol{X}_2$

Figure A.4: Fitted mean for time-varying covariates $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ in 3-subgroup simulation study with $n = 500$. The solid red line plots the true function and dashed blue line plots out the average of the fitted function and the light gray lines are 200 fitted lines.

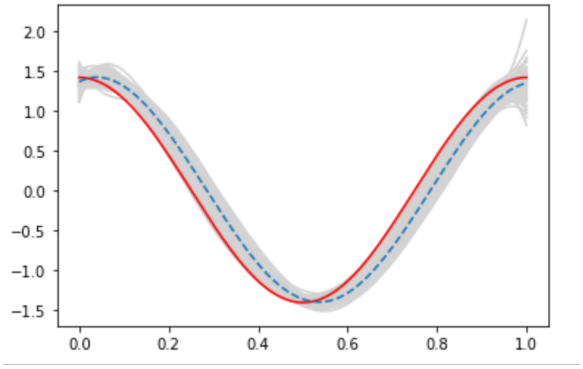
Table A.2: Bias (SD) for the logistic regression parameters for the subgroup labels in the 3-subgroup simulation study with $n = 500$

	lcmm$_O$	**lcmm**$_L$	**Emerald**
γ_{01}	0.018 (0.178)	0.061 (0.192)	0.012 (0.174)
γ_{z11}	-0.018 (0.252)	-0.016 (0.254)	-0.013 (0.206)
γ_{z21}	-0.026 (0.137)	-0.030 (0.144)	-0.027 (0.141)
γ_{02}	0.020 (0.174)	0.054 (0.172)	0.004 (0.152)
γ_{z12}	-0.083 (0.349)	-0.055 (0.349)	-0.007 (0.250)
γ_{z22}	-0.003 (0.164)	-0.003 (0.170)	-0.019 (0.134)

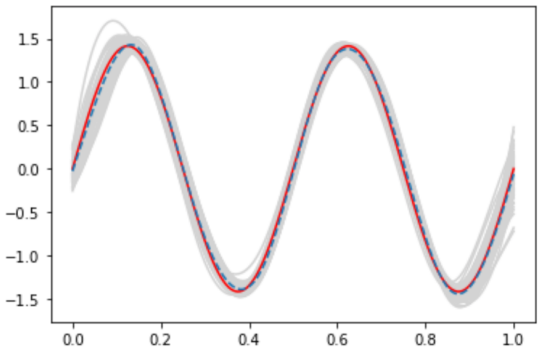
model. In other words, **Emerald** can lower the bias and standard deviation by increasing the sample size. As the number of observational data points increases, we observe improvements in terms of precision and consistency in both methods. However, it is still noteworthy that **Emerald** consistently outperforms **lcmm**$_L$, and in scenarios where time constraints are not a limiting factor, our model's performance rivals or even surpasses that of **lcmm**$_O$



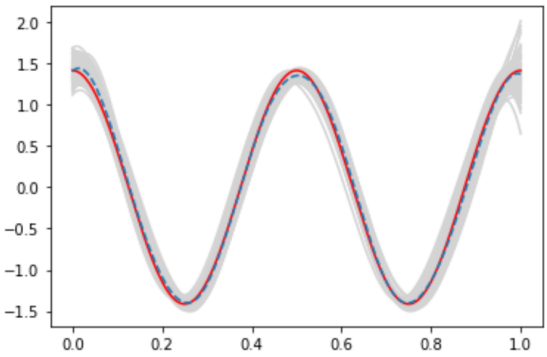
(a) First eigenfunction of$\boldsymbol{X}_1$



(b) Second eigenfunciton of $\boldsymbol{X}_1$



(c) First eigenfunction of $\boldsymbol{X}_2$



(d) Second eigenfunction of $\boldsymbol{X}_2$

Figure A.5: Fitted eigenfunction for time-varying covariates $\boldsymbol{X}_2$ in 3-subgroup simulation study with $n = 500$. The solid red line plots the true function and dashed blue line plots out the average of the fitted function and the light gray lines are 200 fitted lines.

without any time limitations. This is exemplified by the fact that the bias in the coefficients of the logistic regression model derived from **Emerald** is consistently smaller than that of $\textbf{lcmm}_O$. Furthermore, in this simulation study involving a sample size of $n = 500$, **Emerald** consistently achieved remarkable results in classification. It obtained an average

Table A.3: Bias (SD) for the mixture regression coefficients in the 3-subgroup simulation study with $n = 500$

	lcmm$_O$	**lcmm**$_L$	**Emerald**
$\beta_{0,1}$	-0.002 (0.183)	-0.066 (0.640)	-0.038 (0.196)
$\beta_{x,11}$	0.001 (0.012)	0.050 (0.477)	0.006 (0.027)
$\beta_{x,12}$	0.003 (0.014)	-0.001 (0.024)	-0.011 (0.028)
$\beta_{z,11}$	0.012 (0.230)	-0.016 (0.242)	-0.036 (0.212)
$\beta_{z,12}$	-0.005 (0.119)	0.011 (0.259)	0.001 (0.108)
$\beta_{0,2}$	-0.025 (0.175)	0.263 (1.039)	0.025 (0.188)
$\beta_{x,21}$	0.001 (0.018)	-0.126 (0.693)	-0.009 (0.029)
$\beta_{x,22}$	0.002 (0.018)	-0.089 (0.728)	0.006 (0.028)
$\beta_{z,21}$	0.062 (0.341)	0.002 (0.347)	0.054 (0.401)
$\beta_{z,22}$	-0.029 (0.161)	-0.048 (0.311)	-0.001 (0.141)
$\beta_{0,3}$	0.004 (0.172)	-0.106 (0.460)	-0.020 (0.228)
$\beta_{x,31}$	-0.002 (0.017)	0.007 (0.034)	-0.006 (0.031)
$\beta_{x,32}$	-0.002 (0.019)	0.164 (0.646)	-0.001 (0.030)
$\beta_{z,31}$	0.020 (0.199)	0.011 (0.482)	0.022 (0.225)
$\beta_{z,32}$	-0.003 (0.128)	-0.034 (0.189)	0.016 (0.113)

Table A.4: Bias(SD) for the variance parameters in the 3-subgroup simulation study with $n = 500$

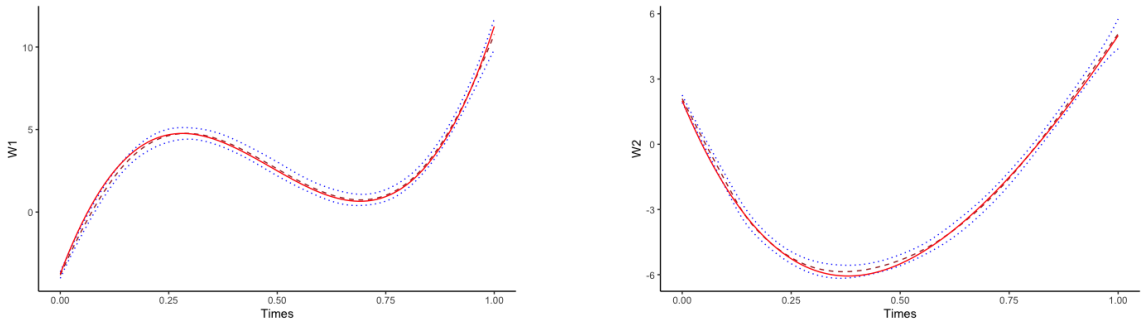
Method	σ^2_{b1}	σ^2_{b2}	σ^2_{b3}	σ^2_{ϵ}
lcmm$_O$	-0.062 (0.398)	-0.319 (0.214)	0.608 (0.398)	0.402 (0.434)
lcmm$_L$	-0.421 (0.882)	0.271 (0.018)	-0.331 (0.119)	2.886 (0.246)
Emerald	-0.076 (0.306)	-0.079 (0.316)	-0.097 (0.324)	0.468 (0.591)

Adjusted Rand Index (ARI) of 0.9982 with a narrow standard deviation of 0.024 across 200 runs. In comparison, the average ARIs for **lcmm**$_O$ and **lcmm**$_L$ were 0.9963 and 0.9474, respectively, with higher standard deviations of 0.037 and 0.017. These findings strongly

support the conclusion that **Emerald** outperforms **lcmm** in terms of its classification error rate.

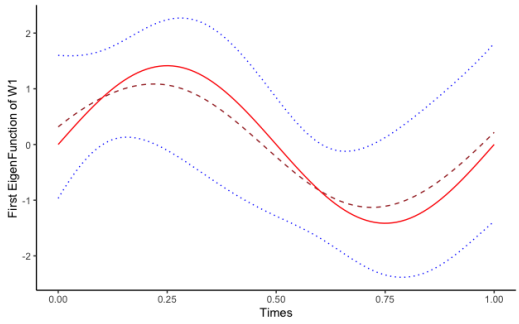
Web Appendix A Additional numerical results for 3-subgroup simulation study

Figures W1, W2, W3, and W4 showcase the estimated mean and eigenfunctions from the **fdapace** and our proposed method, EMERALD, in the simulation study described in Chapter 4. The study was conducted with a sample size of $n = 200$ and was repeated 200 times. These figures visually depict the outcomes obtained by both methods. The red solid curve represents the true function, while the brown dashed curve represents the mean of the estimator. The two blue dotted curves represent the pointwise 2.5% and 97.5% percentiles, serving as confidence bands. Notably, the confidence bands' standard errors were computed from the bootstrap samples, as outlined in Chapter 4.

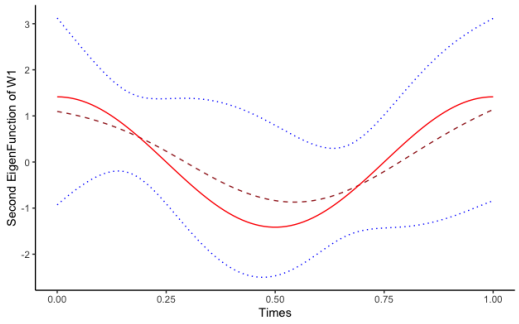


(a) μ_1(t) (b) μ_2(t)

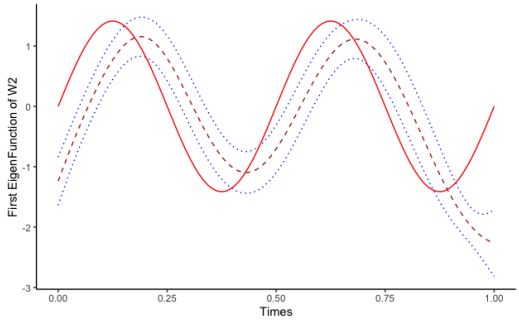
Figure W1: Simulation results: the mean along with 95% confidence bands, using **fdapace**. In each plot, the solid red line is the true function, the brown dashed line is the mean of the functional estimator, and the blue dotted lines are the 95% confidence bands.



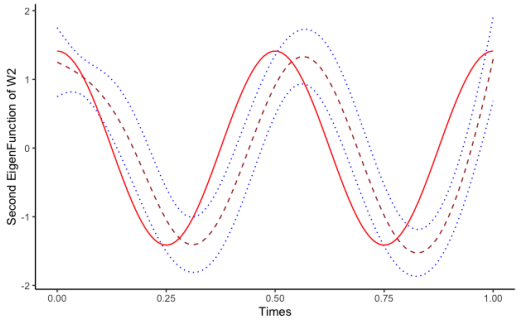
(a) ψ_{11} (t)



(b) ψ_{12} (t)



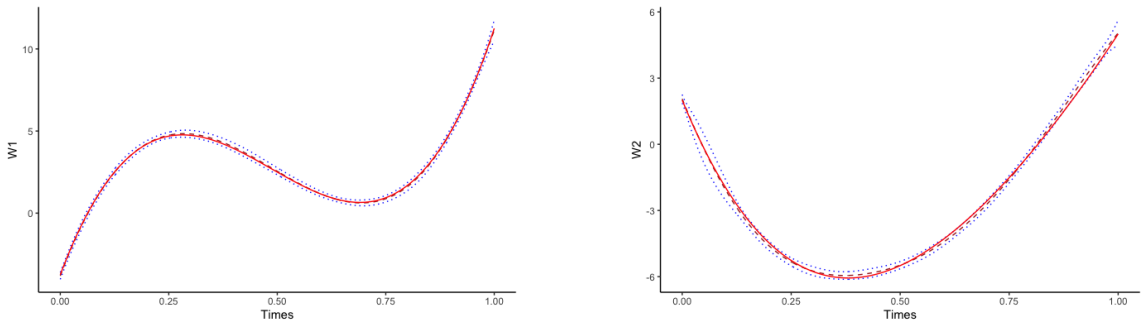
(c) ψ_{21} (t)



(d) ψ_{22} (t)

Figure W2: Simulation results: the eigenfunction along with 95% confidence bands, using **fdapace**. In each plot, the solid red line is the true function, the brown dashed line is the mean of the functional estimator, and the blue dotted lines are the 95% confidence bands.

The mean function plots show that both **fdapace** and our method are able to fit well with the true curve. However, the key differences between the two methods can be observed in the eigenfunction plots. These plots provide compelling evidence that our algorithm consistently produces accurate estimates of the mean and eigenfunctions, showcasing

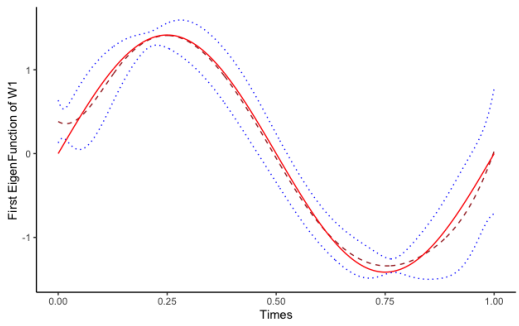


(a) μ_1 (t) (b) μ_2 (t)

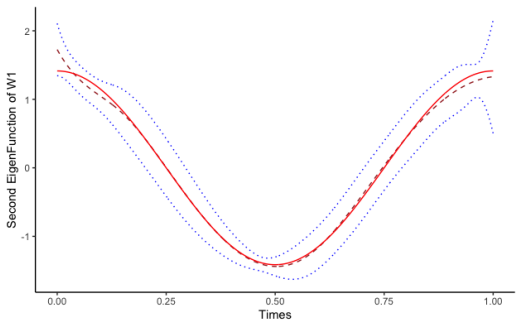
Figure W3: Simulation results: the mean curve along with their 95% confidence bands, using EMERALD. In each plot, the solid red line is the true function, the brown dashed line is the mean of the functional estimator, and the blue dotted lines are the 95% confidence bands.

its superior stability when compared to **fdapace**. The estimated curves closely resemble the true function, indicating the effectiveness of our proposed method in capturing the underlying patterns. Additionally, the narrower confidence bands associated with our method, in comparison to those of **fdapace**, suggest increased stability in the estimation process. These findings robustly validate the reliability of our algorithm in accurately estimating time-varying covariates.

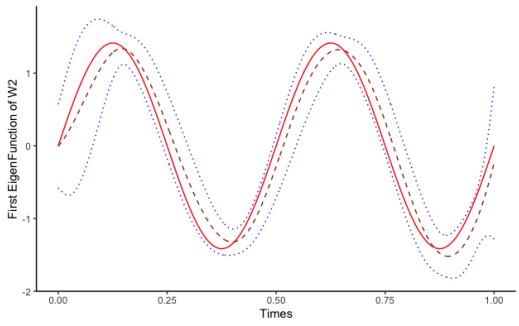
In conclusion, the comprehensive analysis of the figures above and the tables presented in Chapter 4 strongly supports the claim that our method outperforms **fdapace** in modeling time-varying covariates. By jointly modeling the trajectories, our approach enables the sharing of information among the covariates, resulting in accurate predictions of the overall function as well as capturing the underlying patterns and randomness (eigenfunctions, principal component scores) effectively. This advantage of joint modeling contributes



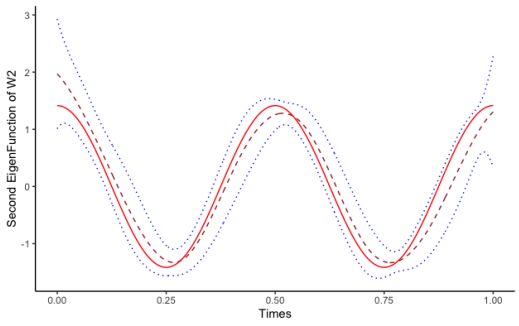
(a) ψ_{11} (t)



(b) ψ_{12} (t)



(c) ψ_{21} (t)



(d) ψ_{22} (t)

Figure W4: Simulation results: the eigenfunction along with their 95% confidence bands, using EMERALD. In each plot, the solid red line is the true function, the brown dashed line is the mean of the functional estimator, and the blue dotted lines are the 95% confidence bands.

to the superior performance of our method in accurately characterizing the dynamics of time-varying covariates.

Web Appendix B Additional simulation results for mixed effects under two subgroups

To generate the subgroup simulation setting in accordance with the main paper, we followed the same procedure. The covariate matrix $\boldsymbol{X}$ and random effects matrix $\boldsymbol{W}$ were generated using the specified method. The subgroup membership vector $\boldsymbol{L}_i$ was generated from a binomial distribution, utilizing the logistic regression model with coefficients $\boldsymbol{\gamma} = (0.7, -0.3, 1.5)^\top$. Given $\boldsymbol{L}_i$, the response vector $\boldsymbol{Y}_i$ was generated from the mixture regression model (1) with coefficients $\boldsymbol{\beta}_1 = (2.02, 1.23, 1.67, -1.72, -0.65)^\top$ and $\boldsymbol{\beta}_2 = (0.59, -2.60, -0.70, 1.85, 1.68)^\top$. The variance parameters $\sigma^2_{b,1}$ and $\sigma^2_{b,2}$ were set to 2 and 1.5, respectively, while σ^2_ϵ was set to 4.

Table W1 presents the bias and standard deviations for the coefficients of the logistic regression model in the two-subgroup setting. The table clearly indicates that our algorithm could predict $\boldsymbol{\gamma}$ with the smallest bias stability, even outperforming the true data in the **lcmm** package.

	$\textbf{lcmm}_O$	$\textbf{lcmm}_L$	EMERALD
γ_0	-0.005(0.260)	-0.030 (0.277)	-0.009(0.257)
γ_{z1}	-0.036(0.353)	-0.039(0.373)	0.028(0.322)
γ_{z2}	-0.024(0.266)	-0.015(0.279)	-0.027(0.247)

TableW1: Bias (SD) for the logistic regression parameters for the subgroup labels in the 2-subgroup simulation study

In this simulation scenario, EMERALD achieved an average Adjusted Rand Index (ARI) of 0.9932 with a standard deviation of 0.01 over 200 runs. Conversely, the average

ARIs for **lcmm**O and **lcmm**L were 0.9994 and 0.9380, respectively, with standard deviations of 0.0040 and 0.0341. These findings strongly support the superiority of EMERALD over **lcmm** in terms of classification error rate.

Table W2 presents an analysis of the bias and standard deviation of the coefficients of the mixture regression model obtained from different algorithms used in the simulation study, including our algorithm and **lcmm**. The results unequivocally indicate that our algorithm outperforms **lcmm** in terms of bias, even when employing the last observation carried forward method, which typically yields the highest bias. Table W3 provides a summary

	$\mathbf{lcmm}_O$	$\mathbf{lcmm}_L$	EMERALD
$\beta_{0,1}$	-0.014(0.171)	0.847(0.431)	-0.010(0.145)
$\beta_{x,11}$	< 0.001(0.005)	-0.147(0.048)	-0.004(0.015)
$\beta_{x,12}$	0.001(0.006)	0.235(0.058)	-0.002(0.016)
$\beta_{z,11}$	0.031(0.209)	0.067(0.497)	-0.011(0.196)
$\beta_{z,12}$	0.006(0.146)	-0.002(0.295)	0.015(0.121)
$\beta_{0,2}$	-0.026(0.127)	0.511(0.295)	0.035(0.214)
$\beta_{x,21}$	< 0.001(0.003)	-0.244(0.042)	-0.002(0.02)
$\beta_{x,22}$	-0.001(0.004)	-0.098(0.053)	0.021(0.019)
$\beta_{z,21}$	0.004(0.178)	-0.008(0.396)	0.049(0.236)
$\beta_{z,22}$	0.008(0.114)	0.027(0.246)	-0.004(0.16)

TableW2: Bias (SD) for the mixture regression coefficients in the 2-subgroup simulation study

of the average variances for the random intercept and measurement error, calculated over 200 runs in the simulation study. The estimated variances obtained from our algorithm exhibit a close correspondence to the true values. This suggests that our method accurately

captures the inherent variability in the data, highlighting its effectiveness and reliability. In

Method	σ^2_{b1}	σ^2_{b2}	σ^2_{ϵ}
$\mathbf{lcmm}_O$	-0.080 (0.370)	0.138 (0.263)	-0.026 (0.189)
$\mathbf{lcmm}_L$	-0.454 (0.971)	-0.936 (0.619)	0.804 (0.188)
EMERALD	-0.092 (0.148)	-0.174 (0.181)	0.472 (0.132)

TableW3: Bias(SD) for the variance parameters in the 2-subgroup simulation study

summary, the simulation studies conducted in both the 2-subgroup and 3-subgroup settings, as well as with sample sizes of $n = 200$ and $n = 500$, consistently demonstrate the superior performance of our proposed method. Our algorithm outperforms all existing packages, even when time limitations are not imposed. These results affirm the effectiveness and efficiency of our method in handling complex data scenarios and highlight its potential for practical applications.