Title
Semiparametric Modeling for Genome-Wide Association Studies and Repeated Measurements

Author
Huang, Zijian

Publication Date
2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Semiparametric Modeling for Genome-Wide Association Studies and Repeated Measurements

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

Doctor of Philosophy

in

Applied Statistics

by

Zijian Huang

August 2015

Dissertation Committee:

Dr. Shujie Ma , Chairperson
Dr. Xinping Cui
Dr. Gloria Gonzalez-Rivera

The Dissertation of Zijian Huang is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I would like to thank my advisor, Professor Shujie Ma, for her constant support and excellent guidance in developing this dissertation. She was always standing by me and gave me enormous amount of time, patience and assistance in guiding me towards my goals. Whenever I had doubts and questions, Professor Ma was ready to help me out and encouraged me to make further progress. Without her help and encouragement this dissertation would not have been written.

Also, I am indebted to other members of my dissertation committee, Professor Xinping Cui and Professor Gloria Gonzalez-Rivera. They gave me invaluable suggestions and inspirations at various stages of my studies at UCR. Whenever I needed, they were ready for helping me and spending their valuable time with me.

I would like to sincerely thank all professors who taught me during my graduate study at UCR and thank the Statistics Department for providing financial and professional assistance to me. I also want to thank Graduate Division for giving me Dean's Distinguished Fellowship and the Dissertation Year Program Fellowship. Thank for the permission to include Chapter 4 of my dissertation, which was originally published in Annals of Institute of Statistical Mathematics. In addition, I wish to thank my friends. They brought me countless joy during my stay in California.

Most importantly, I would like to thank my parents for all the tremendous support, endless care and love. They always stand by my side, encourage me to be strong during the hard time and motivate me to do my best. Without them, I would not be able to accomplish all my goals.

To my parents for all the support.

ABSTRACT OF THE DISSERTATION

Semiparametric Modeling for Genome-Wide Association Studies and Repeated Measurements

by

Zijian Huang

Doctor of Philosophy, Graduate Program in Applied Statistics
University of California, Riverside, August 2015
Dr. Shujie Ma , Chairperson

In recent years, advanced technologies have enabled people to collect complex data and the analysis of such data can be challenging. My dissertation focuses on developing new methodologies and computational algorithms in non- and semi- parametric regression models to analyze complex and large scaled data. Chapter 1 introduces commonly used semiparametric models and their properties. Chapter 2 reviews B-splines approximation to the nonparametric functions. Chapter 3 provides an overview of methodologies including generalized estimating equations and mixed models, which are used to analyze correlated data.

In chapter 4, we propose a flexible generalized semiparametric model for repeated measurements by combining generalized partially linear single index model with varying coefficient model. The proposed model is a useful analytic tool to explore dynamic patterns which naturally exist in longitudinal data and also to study possible nonlinear relationships between the response and covariates. We then employ the quadratic inference function and develop an estimation procedure to estimate unknown regression parameters and nonpara-

metric functions. To select variables and estimate parameters simultaneously, we further obtain penalized estimators. Moreover, we establish theoretical properties of the parametric and nonparametric estimators. Both simulations and an empirical example are presented to illustrate the use of the proposed model.

In chapter 5, we propose a semiparametric model in genome-wide association studies (GWAS). The use of linear mixed models (LMMs) in GWAS is now widely accepted because LMMs have been shown to be capable of correcting for several forms of confounding due to genetic relatedness of sampled data. On the other hand, gene and environment (G $\times$ E) interactions play a pivotal role in determining the risk of human diseases. Conventional parametric models such as LMMs may not reflect the underlying nonlinear G $\times$ E interactions, which will result in serious bias. Therefore, we propose a semiparametric mixed model to investigate important gene-disease associations in the context of possible nonlinear G $\times$ E interactions in GWAS. We further propose a profile maximum likelihood estimation procedure to estimate the parameters and nonparametric functions, and apply the restricted maximum likelihood estimation method to estimate the variance components. For these profile parameter and nonparametric function estimators, asymptotic consistency and normality are established. Moreover, the Rao-score-type test procedure is developed and a multiple testing process is employed to identify the important genetic factors. Both simulation studies and an empirical example are presented to illustrate the use of our proposed model and methods.

Contents

List of Figures

List of Tables

Chapter 1

Semiparametric Regression Models

1.1 Introduction

Regression analysis is a statistical process for estimating the relationships among variables. For example, the relationship between a response variable Y and an explanatory variable X for n data points $(X_i, Y_i), i = 1, ..., n$ can be modeled through a function m as

$$Y_i = m(X_i) + \epsilon_i,$$

where ϵ_i is the error term.

In parametric approaches, m is fully described by a finite set of parameters with some model assumptions, linearity being among the most convenient. Although their properties are well established, parametric models have many limitations in applications. Besides, misspecification of the data generation mechanism could lead to large bias in prediction. On the other hand, the nonparametric approach makes no assumption on the specification of the model and provides a versatile method to explore the features of function

m. However, this approach may not incorporate some prior information and the resulting estimator of m tends to incur larger variance. In addition, the standard nonparametric method is practically impotent in the presence of high-dimensional covariates.

Consequently, many different semiparametric regression methods have been proposed and developed. They offer more flexibility than standard parametric regressions and overcome the curse of dimensionality in nonparametric regressions. The most popular semiparametric models are single-index, partially linear, and varying coefficient models.

1.2 Single-Index Model

1.2.1 Introduction

Index models play an important role in econometrics. An index is a summary of different variables into one variable such as the price index, the growth index, and the cost-of-living index. By summarizing all the information contained in several variables into one "single index" term one can greatly reduce the dimensionality, and thereby, achieve greater estimation precision. Models based on such an index are known as single-index models. They relax some of the restrictive assumptions of familiar parametric models. Moreover, they are often easy to compute, and their results are easy to interpret.

Let Y be a scalar random variable and X be a $d \times 1$ random vector. In a single-index model, the conditional mean function $E(Y|X = x)$ has the form

$$E(Y|X) = m(X) = g\{v_\beta(X)\}, \tag{1.1}$$

where β is an unknown coefficients vector, $v_\beta(X)$ is an index function, and g is an arbitrary smooth function. Usually, the general index function $v_\beta(X)$ is replaced by the linear index $X^{\mathrm{T}}\beta$, where β is a $d \times 1$ constant vector. The inferential problem is to estimate β and g from observations of (Y, X).

Model (1.1) contains many widely used parametric models as special cases. If g is the identity function, Model (1.1) is a linear model. If g is the cumulative normal or logistic distribution function, Model (1.1) is a binary probit or logit model. When g is unknown, Model (1.1) provides a specification that is more flexible than parametric models and retains many of the desirable features of the parametric models.

1.2.2 Identification

Before estimation for β and g can be considered, restrictions must be imposed to ensure their identification. That is, β and g must be uniquely determined by the population distribution of (Y, X).

Suppose that $E(Y|X = x)$ satisfies Model (1.1) and X is a d-dimensional random variable. Then β and g are identified if the following conditions hold:

(a) g is differentiable and not constant on the support of $X^{\mathrm{T}}\beta$.

(b) The components of X are continuously distributed random variables.

(c) The support of X is not contained in any proper linear subspace of R^d.

(d) The first component of β: $\beta_1 = 1$ or the Euclidean norm of β: $\| \beta \| = 1$.

1.2.3 Estimation

When estimating a single-index model, one must take into account that the functional form of the link function is unknown. Thus, both the index and the link function have to be estimated. Let

$$\epsilon = m(X) - Y = E(Y|X) - Y$$

be the deviation of Y from its conditional expectation with respect to X. Considering a single-index model given as

$$Y = g(X^{\mathrm{T}}\beta) + \epsilon,$$

the goal is to find the estimators for β and $g(.)$. As β is inside the nonparametric link, the challenge is to find an appropriate estimator for β that reaches the $\sqrt{n}$-rate of convergence. Two essentially different approaches exist for this purpose: one is an iterative approximation of β by semiparametric least squares or pseudo maximum likelihood estimation, the other approach is estimating β directly through the average derivative of the regression function.

In both cases, the estimation procedures can be summarized as:

(a) estimate β by $\hat{\beta}$;

(b) compute index values $\hat{\eta} = X^{\mathrm{T}}\hat{\beta}$;

(c) estimate the link function $g(.)$ by using a univariate nonparametric method for the regression of Y on $\hat{\eta}$.

In summary, single-index models achieve dimension reduction and avoid the curse of dimensionality because the index $X^{\mathrm{T}}\beta$ aggregates the dimension of X. This dimension-reduction feature gives them a considerable advantage over nonparametric methods. Moreover, in terms of rate of convergence in probability, single-index models are as accurate as the parametric model for estimating β. They are also as accurate as the one dimensional nonparametric mean regression for estimating g. In applications where X is multidimensional, the single-index structure is plausible.

1.3 Generalized Partially Linear Model

1.3.1 Generalized linear model

Nelder and Wedderburn (1972) introduced the term generalized linear models (GLM). The essential feature of the GLM is that the expectation $\mu = E(Y|X)$ of Y is a monotone function G of the index $\eta = X^{\mathrm{T}}\beta$, which means

$$E(Y|X) = G(X^{\mathrm{T}}\beta) \longleftrightarrow \mu = G(\eta).$$

Function G is called the link function.

In the GLM framework, the distribution of Y is a member of the exponential family. A distribution is said to be a member of the exponential family if its probability or density function has the structure

$$f(y, \theta, \phi) = exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\}$$

with some specific functions $a(.)$, $b(.)$ and $c(.)$. These functions differ for the distinct Y distributions. The primary interest is to estimate the parameter θ. McCullagh and Nelder (1989) called θ the canonical parameter and ϕ the nuisance parameter.

1.3.2 Quasi-likelihood estimation for GLM

Assuming the distribution of Y is a member of the exponential family, it can be derived that

$$\begin{aligned} E(Y) &= \mu = b'(\theta), \\ \text{Var}(Y) &= V(\mu)a(\phi) = b''(\theta)a(\phi). \end{aligned}$$

The expectation of Y only depends on θ, whereas the variance of Y depends on θ and ϕ.

Consider independent observations $Y_1, ..., Y_n$ with $\mu_i = G(x_i'\beta)$ and $\text{Var}(Y_i) = a(\phi)V(\mu_i)$. Define

$$U_j = \sum_{i=1}^{n} \left[\frac{y_i - \mu_i}{a(\phi)V(\mu_i)} \frac{\partial \mu_i}{\partial \beta_j} \right]$$

and

$$Q(\beta; y) = \sum_{i=1}^{n} \int_{y_i}^{\mu_i} \frac{y_i - t}{a(\phi)V(t)} dt.$$

U_j is called the quasi-score function and Q is the quasi likelihood. The equations $U_j = 0$, $j = 1, ..., p$ are quasi-likelihood estimation equations. In matrix form, there is

$$U(\beta; y) = D'V^{-1}(y - \mu)/a(\phi),$$

where D is the $n \times p$ matrix with $(i,j)^{th}$ entry $\partial\mu_i/\partial\beta_j$, V is the $n \times n$ diagonal matrix with i^{th} diagonal entry $V(\mu_i)$, $y = (y_1, ..., y_n)$, and $\mu = (\mu_1, ..., \mu_n)$. One can find the value $\hat{\beta}_{QL}$ that maximizes $Q(\beta; y)$ by setting $U(\hat{\beta}_{QL}; y) = 0$ and solving for $\hat{\beta}_{QL}$.

The quasi-likelihood technique is used for estimating regression coefficients without fully specifying the distribution of the observed data. As a result, it provides a more flexible approach than the maximum-likelihood approach.

1.3.3 Generalized partially linear model

A partially linear model (PLM) consists of two additive components: a linear and a nonparametric part as

$$E(Y|U,T) = U^{\mathrm{T}}\beta + m(T),$$

where $\beta = (\beta_1, ..., \beta_p)^{\mathrm{T}}$ is a finite dimensional parameter and $m(.)$ is a smooth function. The vector U typically covers continuous variables that are known to influence the index in a linear way, as well as categorical variables. The vector T contains continuous explanatory variables that are to be modeled in a nonparametric way. This model can be extended to the generalized partially linear model (GPLM) as

$$E(Y|U,T) = G\{U^{\mathrm{T}}\beta + m(T)\},$$

where $G(.)$ is a known link function.

1.3.4 Estimation algorithm for PLM and GPLM

Considering PLM, the goal is to find β and $m(.)$ in the following structural equation

$$Y = U^{\mathrm{T}}\beta + m(T) + \epsilon,$$

where ϵ denotes the error term with zero mean and finite variance. Taking expectations conditioned on T, i.e.

$$E(Y|T) = E(U^{\mathrm{T}}\beta|T) + E\{m(T)|T\} + E(\epsilon|T),$$

one can obtain

$$Y - E(Y|T) = \{U - E(U|T)\}^{\mathrm{T}}\beta + \epsilon - E(\epsilon|T). \tag{1.2}$$

There are two alternative approaches to estimate β and $m(.)$. The first approach is a backfitting approach. Subtract $U^T\beta$ from Y to get

$$E\{Y - U^{\mathrm{T}}\beta|T\} = m(T). \tag{1.3}$$

Estimate the parametric coefficient β by least squares regression of Y on U. Plugging $U^{\mathrm{T}}\hat{\beta}$ into (1.3) yields a classic nonparametric regression problem, so non-parametric estimators may be employed to estimate the values of $\hat{m}(T)$. The alternative approach is based on (1.2). $Y - E(Y|T)$ and $U - E(U|T)$ in equation (1.2) can be replaced by $Y - \hat{Y}$ and $U - \hat{U}$, where $\hat{Y}$ and $\hat{U}$ are the empirical counterparts of $E(Y|T)$ and $E(U|T)$. Then the standard linear regression can be applied to estimate β. Consequently, $m(.)$ can be estimated by the nonparametric regression of $Y - U^{\mathrm{T}}\hat{\beta}$ on T. Under regularity conditions, $\hat{\beta}$ can be shown to be $\sqrt{n}$-consistent for β and asymptotically normal, and there exists a consistent estimator of its limiting covariance matrix. In addition, $m(.)$ can be estimated with the usual univariate rate of convergence.

In order to estimate the GPLM, consider the same distributional assumptions for Y as in the GLM. As a result, there is an option between two cases: (a) the distribution of Y belongs to the exponential family, or (b) the first two (conditional) moments of Y are specified in order to use the quasi-likelihood function. To summarize, the estimation of the GPLM will be based on

$$E(Y|U,T) = \mu = G(\eta) = G\{U^{\mathrm{T}}\beta + m(T)\},$$

$$Var(Y|U,T) = a(\phi)V(\mu),$$

where the nuisance parameter ϕ is the dispersion parameter. The profile-likelihood method can be used to estimate β and $m(.)$. This method starts from keeping β fixed and estimates a least favorable nonparametric function $m_\beta(.)$ in dependence of the fixed β. The resulting estimate for $m_\beta(.)$ is then used to construct the profile likelihood for β. Consequently, the resulting $\hat{\beta}$ is estimated at $\sqrt{n}$ rate, it has an asymptotic normal distribution and is asymptotically efficient. The nonparametric function $m(.)$ can be estimated consistently by $\hat{m}(.) = \hat{m}_{\hat{\beta}}(.)$.

1.4 Varying Coefficient Model

Varying coefficient models are important tools to explore the dynamic pattern in many scientific areas, such as economics, finance, etc. They are natural extensions of classical parametric models with good interpretability and are becoming more and more popular in data analysis. Due to their flexibility and interpretability, in the past ten years, the varying coefficient models have experienced deep and exciting developments on methodological, theoretical, and applied sides.

Considering multivariate predictor variables containing a scalar U and a vector $X = (x_1, .., x_p)^{\mathrm{T}}$, the varying coefficient models have the form as

$$E(y|U, X) = m(U, X) = X^{\mathrm{T}} a(U),$$

where $a(U) = (a_1(U), .., a_p(U))^{\mathrm{T}}$ is the unknown functional coefficient. Therefore, they allow the coefficients to vary smoothly over the group stratified by U, ultimately permitting nonlinear interactions between U and X. To estimate a(.), one can use kernel-local polynomial smoothing, polynomial spline or smoothing spline. In my research, I adopt a B-spline approximation technique.

The varying coefficient models can be extended to the exponential family of conditional distributions. Via the canonical link function $G(.)$, the generalized varying coefficient model is

$$E(y|U,X) = G(m(U,X)) = G\{X^{\mathrm{T}} a(U)\}.$$

Chapter 2

B-splines Approximation

2.1 Introduction

The term B-spline is an abbreviation for basis spline. It is a spline function that has minimal support with respect to a given degree, smoothness, and domain partition. Any spline function of a given degree can be expressed as a linear combination of B-splines of that degree. Consequently, a spline function is a piecewise polynomial function. The places where the pieces meet are known as knots. The key property of spline functions is that they are continuous at the knots.

B-splines provide a better curve fit than other interpolation methods. They possess a variation-diminishing property, which means that increasing the order of a B-spline function does not create oscillation in the entire curve. Moreover, they have local support such that a portion of the B-spline curve may be modified without affecting the shape of the whole curve. They also maintain the smoothness and continuity of higher-order derivatives.

B-splines approximate the value of a function using control point values. Consider the following B-spline approximation:

$$P(t) = \sum_{i=1}^{n+1} N_{i,k}(t)P_i, \quad \text{where} \quad t_{min} \leq t \leq t_{max}.$$

There are $n+1$ control points: $P_1, P_2, ..., P_{n+1}$. The $N_{i,k}(t)$ basis functions are of order k (degree $k-1$). k must be at least 2 (linear) and can be no more than $n+1$ (the number of control points). A knot vector $(t_1, t_2, ..., t_{k+(n+1)})$ must be specified. It is necessary that $t_i \leq t_{i+1}$. $N_{i,k}(t)$ depends only on the value of k and the values in the knot vector. $N_{i,k}(t)$ is defined recursively as:

$$N_{i,1}(t) = \begin{cases} 1, & t_i \leq t < t_{i+1} \\ 0, & \text{otherwise} \end{cases}$$

$$N_{i,k}(t) = \frac{t - t_i}{t_{i+k-1} - t_i} N_{i,k-1}(t) + \frac{t_{i+k} - t}{t_{i+k} - t_{i+1}} N_{i+1,k-1}(t).$$

There are several things that one should understand regarding these equations. Each $N_{i,k}(t)$ depends only on the $k+1$ knot values from t_i to t_{i+k}. $N_{i,k}(t) = 0$ for $t < t_i$ or $t \geq t_{i+k}$, so P_i only influences the curve for $t_i \leq t < t_{i+k}$.

2.2 Knot Vector

The knot vector subdivides a domain into sub-regions or knot intervals. There are uniform, non-uniform, and open knot vectors.

For uniform knot vectors, $t_{i+1} - t_i = \text{constant}, \forall i$. For non-uniform knot vectors, the only constraint is $t_i \leq t_{i+1}, \forall i$. For open knot vectors, there are k equal knot values at each end:

$$t_i = t_1, \quad i \leq k$$

$$t_i \leq t_{i+1}, \quad k < i < n+2$$

$$t_i = t_{k+(n+1)}, \quad i \geq n+2.$$

Repetition of a knot value has the effect of drawing the curve closer to a specific control point. If a knot value is repeated $k-1$ times, the B-spline curve will pass through the associated control point. Therefore, in the case of an open B-spline (constructed from an open knot vector), the first and last control points will be interpolated. The shapes of the basis functions are determined entirely by the relative spacing between the knots. Scaling ($\tilde{t_i} = \alpha t_i, \forall i$) or translating ($\tilde{t_i} = t_i + \delta t, \forall i$) the knot vector has no effect on the shapes of the basis functions.

2.3 B-spline Basis Functions: Important Properties

Assume that a B-spline curve $C(u)$ of degree p is defined by $n+1$ control points and a knot vector $U = (u_0, u_1, ..., u_m)$ with the first $p+1$ and last $p+1$ knots being "clamped" (i.e., $u_0 = u_1 = ... = u_p$ and $u_{m-p} = u_{m-p+1} = ... = u_m$). Spline basis functions $N_{0,p}(u), N_{1,p}(u), ..., N_{n,p}(u)$ are defined as

$$N_{i,0}(u) = \begin{cases} 1, & u_i \leq u < u_{i+1} \\ 0, & \text{otherwise} \end{cases}$$

$$N_{i,p}(u) = \frac{u - u_i}{u_{i+p} - u_i} N_{i,p-1}(u) + \frac{u_{i+p+1} - u}{u_{i+p+1} - u_{i+1}} N_{i+1,p-1}(u).$$

This set of basis functions has the following properties:

1. $N_{i,p}(u)$ is a polynomial function in u of degree p.

2. Nonnegativity: for all i, p and u, $N_{i,p}(u)$ is non-negative.

3. Local Support: $N_{i,p}(u)$ is a non-zero polynomial on $[u_i, u_{i+p+1})$. If u is outside the interval $[u_i, u_{i+p+1})$, $N_{i,p}(u) = 0$.

4. On any span $[u_i, u_{i+1})$, at most $p + 1$ basis functions of degree p are non-zero, namely: $N_{i-p,p}(u), N_{i-p+2,p}(u), ..., N_{i,p}(u)$.

5. Partition of Unity: The sum of all non-zero basis functions of degree p on span $[u_i, u_{i+1})$ is 1.

6. If the number of the knots is $m + 1$, the degree of basis function is p, and the number of degree p basis functions is $n + 1$, then the equation $m = n + p + 1$ holds.

7. Basis function $N_{i,p}(u)$ is a composite curve of degree p polynomials with joining points at knots in $[u_i, u_{i+p+1})$.

8. At a knot of multiplicity k, basis function $N_{i,p}(u)$ is C^{p-k} continuous.

2.4 B-spline Curves: Important Properties

1. The B-spline curve $C(u)$ is a piecewise curve with each component a curve of degree p.

2. $C(u)$ passes through the two end control points P_0 and P_n.

3. Strong Convex Hull Property: The B-spline curve is contained in the convex hull of its control points.

4. Local Modification Scheme: Changing the position of control point P_i only affects the curve $C(u)$ on interval $[u_i, u_{i+p+1})$. Therefore, the curve can be modified locally

without changing the shape in a global way. Moreover, if fine-tuning curve shape is required, one can insert more knots so that the affected area could be restricted to a very narrow region.

5. $C(u)$ is C^{p-k} continuous at a knot of multiplicity k.

Chapter 3

GEE and Mixed Effects Model for Correlated Data

3.1 Introduction

The data observed are likely to be correlated in many cases. The most common instances are repeated observations over time, either in the form of panel studies or time-series of cross-sections. The widely used approaches to analyze correlated data include mixed effects regression models (MRM) and generalized estimating equation (GEE) models.

The primary distinction between these two approaches is that MRM are full-likelihood methods and GEE models are based on quasi-likelihood estimation. The advantages of the GEE models are that (a) they are computationally easier than full-likelihood methods, and (b) they can be generalized easily to a wide variety of outcome measures with quite different distributional forms.

3.2 Generalized Estimating Equation Models

3.2.1 Introduction

Generalized linear models (GLM) have received widespread use in cross-sectional analyses as we saw in Chapter 1. Liang and Zeger (1986) extended the GLM approach to correlated data in the context of repeated observations over time.

Consider observation on a dependent variable Y_{it} and k covariates X_{it}, where i indexes the N units of analysis (cases or clusters), $i = 1, 2, ..., N$ and t indexes the T time points (or repeated measurements), $t = 1, 2, ..., T$. Let $Y_i = (Y_{i1}, Y_{i2}, ..., Y_{iT})$ denote the corresponding column vector of observations on the outcome variable, and X_i indicate the $T \times k$ matrix of covariates of observation i. Define a function h to specify the relationship between Y_i and X_i: $E(Y_i) = \mu_i = h(X_i\beta)$, where β is a $k \times 1$ vector of parameters. The quasi-likelihood estimate of β is the solution to a set of k quasi-score differential equations:

$$U_k(\beta) = \sum_i^N D_i' V_i^{-1}(Y_i - \mu_i) = 0, \tag{3.1}$$

where $D_i = \partial\mu_i/\partial\beta$, and V_i is the variance of Y_i.

In the cross-sectional case (i.e., $T = 1$), V_i can be specified as a function g of the mean (i.e.,$V_i = \frac{g(\mu_i)}{\phi}$). For cases where $T > 1$, some provision must be made for the dependence across t. Liang and Zeger's (1986) solution was to specify a $T \times T$ matrix $R_i(\alpha)$ of the working correlations across t for a given Y_i. While $R_i(\alpha)$ can vary across cases, it is assumed to be fully specified by the vector of unknown parameters α. This correlation matrix then enters the variance term of equation (3.1):

$$V_i = \frac{(A_i)^{1/2} R_i(\alpha)(A_i)^{1/2}}{\phi}, \tag{3.2}$$

where A_i are $T \times T$ diagonal matrices with $g(\mu_{it})$ as the tth diagonal element. Substitution of (3.2) into (3.1) yields the GEE estimator.

Since V_i is the function of both α and β, estimation is typically accomplished by an iterative procedure. Liang and Zeger (1986) noted that (3.1) can be expressed as a function of β alone by substituting $\sqrt{N}$-consistent estimates of α and ϕ into (3.1). One can solve for β using the fisher score and calculate standardized residuals to consistently estimate α and ϕ. These two steps are iterated until the estimates reach convergence.

The GEE model has a number of attractive properties for applied researchers. Assuming that the model for μ is correctly specified, GEE estimates of β ($\hat{\beta}_{GEE}$) will be consistent in N. Moreover, $\sqrt{N}(\hat{\beta}_{GEE} - \beta)$ is asymptotically multivariate normal, and the covariance matrix of the estimates can be consistently estimated. Most importantly, the asymptotic consistency of $\hat{\beta}_{GEE}$ holds even when the working correlation structure is misspecified.

An additional advantage of the GEE approach is the broad range of options available for specifying the within-cluster correlation structure. Fitzmaurice, Laird and Rotnitzky (1993) discussed four common specifications of the working correlation matrix $R_i(\alpha)$:

1. $R_i(\alpha) = I$, a $T \times T$ identity matrix. This working independence assumption is equivalent to assuming no intra-cluster correlation.

2. $R_i(\alpha) = \alpha$ for off diagonal elements, an exchangeable correlation structure. Values of Y_i are assumed to covary equally across all observations within a cluster. In this specification, α is a scalar, which is estimated by the model.

3. $R_i(\alpha) = \alpha^{|t-s|}$ for element in row s and column t, an autoregressive specification.

As is typically the case in AR models, $|\alpha| \leq 1.0$. Higher-order autoregressive specifications are also available.

4. $R_i(\alpha) = \alpha_{st}$ for element in row s and column t, an unstructured correlation structure. In this context, α is a $T \times T$ matrix containing the $\frac{T(T-1)}{2}$ unique pairwise correlations for all possible combinations of time points.

In addition to these, a number of other specifications of the working correlation matrix are possible, including stationary and nonstationary models for varying orders. Alternatively, the researcher may specify $R_i(\alpha)$ explicitly; this option is valuable for testing the robustness of estimates to the correlation specification.

3.2.2 Quadratic inference function

GEEs enable one to estimate regression parameters consistently in longitudinal data analysis even when the correlation structure is misspecified. However, under such misspecification, the estimators of the regression parameters can be inefficient. The method of quadratic inference functions (QIF) does not involve direct estimation of the correlation parameters, and thus remains optimal even if the working correlation structure is misspecified. The idea is to represent the inverse of the working correlation matrix by a linear combination of basis matrices. Under misspecified working assumptions, their estimators are more efficient than the ones from GEEs.

In QIF, R^{-1} is modeled by the class of matrices

$$\sum_{i=1}^{m} a_i M_i, \tag{3.3}$$

where $M_1, ..., M_m$ are known matrices, and $a_1, ..., a_m$ are unknown constants. This is a sufficiently rich class that accommodates the most commonly used correlation structures.

For example, suppose $R(\alpha)$ has 1 on the diagonal and α everywhere off the diagonal. Then R^{-1} can be written as $a_0M_0 + a_1M_1$, where M_0 is the identity matrix, and M_1 has 0 on the diagonal and 1 off the diagonal. Here $a_0 = -(n-2)\alpha + 1/k_1$ and $a_1 = \alpha/k_1$, where $k_1 = (n-1)\alpha^2 - (n-2)\alpha - 1$ and n is the dimension of R. Note that this is not a unique linear representation of R^{-1}; M_1 can also be the rank-1 matrix with 1 everywhere.

Substituting (3.3) into (3.1) gives the following class of estimating functions:

$$\sum_{i=1}^{N} \dot{\mu}_i' A_i^{-1/2}(a_1M_1 + ... + a_mM_m)A_i^{-1/2}(y_i - \mu_i), \tag{3.4}$$

where $\dot{\mu}_i$ is the derivative of μ_i with respect to regression parameters β, and A_i is the diagonal marginal covariance matrix for the ith cluster. QIF approach proceeds as follows. Based on the form of the quasi-score, one can define the 'extended score' g_N to be

$$g_N(\beta) = \frac{1}{N}\sum_{i=1}^{N} g_i(\beta) = \frac{1}{N}\begin{bmatrix} \sum_{i=1}^{N}(\dot{\mu})'A_i^{-\frac{1}{2}}M_1A_i^{-\frac{1}{2}}(y_i - \mu_i) \\ \vdots \\ \sum_{i=1}^{N}(\dot{\mu})'A_i^{-\frac{1}{2}}M_mA_i^{-\frac{1}{2}}(y_i - \mu_i) \end{bmatrix}. \tag{3.5}$$

Therefore, the estimating function (3.4) is a linear combination of elements of the extended score vector (3.5).

Based on the extended scores g_N, quadratic inference function is defined as

$$Q_N(\beta) = g_N' C_N^{-1} g_N,$$

where $C_N = (1/N^2)\sum_{i=1}^{N} g_i(\beta)g_i'(\beta)$. The quadratic inference function estimator $\hat{\beta}$ is then defined to be

$$\hat{\beta} = \arg\min_{\beta} Q_N(\beta).$$

The corresponding estimating equation for β is

$$\dot{Q}_N(\beta) = 2\dot{g}'_N C_N^{-1} g_N - g'_N C_N^{-1} \dot{C}_N C_N^{-1} g_N = 0, \tag{3.6}$$

where $\dot{C}_N$ is a three-dimensional array $(\partial C_N/\partial\beta_1, ..., \partial C_N/\partial\beta_q)$, $\dot{g}_N$ is an $mq \times q$ matrix $\{\partial g_N/\partial\beta\}$, and $g'_N C_N^{-1} \dot{C}_N C_N^{-1} g_N$ is a $q \times 1$ vector

$$\{g'_N C_N^{-1} (\partial C_N/\partial\beta_i) C_N^{-1} g_N : i = 1, ..., q\}.$$

To solve equation (3.6), Newton-Raphson algorithm is implemented, which requires the second derivative of Q_N in β:

$$\ddot{Q}_N(\beta) = 2\dot{g}'_N C_N^{-1} \dot{g}_N + R_N,$$

where

$$R_N = 2\ddot{g}' C_N^{-1} g - 4\dot{g}'_N C_N^{-1} \dot{C}_N^{-1} g_N + 2g'_N C_N^{-1} C_N^{-1} \dot{C}_N C_N^{-1} g_N - g'_N C_N^{-1} \ddot{C}_N C_N^{-1} g_N.$$

Here $\ddot{C}_N$ is a four-dimensional array $\{\partial^2 C_N/\partial\beta_i\partial\beta_j : i, j = 1, ..., q\}$, and $g'_N C_N^{-1} \ddot{C}_N C'_N g_N$ is a $q \times q$ matrix $\{g'_N C_N^{-1} (\partial^2 C_N/\partial\beta_i\partial\beta_j) C_N^{-1} g_N : i, j = 1, ..., q\}$. Asymptotically $\ddot{Q}_N(\beta)$ can be approximately by $2\dot{g}'_N C_N^{-1} \dot{g}_N$ since R_N is $o_p(1)$. The Newton-Raphson method iterates the following relationship to convergence:

$$\hat{\beta}^{(j+1)} = \hat{\beta}^{(j)} - \ddot{Q}_N^{-1}(\hat{\beta}^{(j)}) \dot{Q}_N(\hat{\beta}^{(j)}).$$

The optimality of the QIF estimator is easily established.

3.3 Mixed Effects Model

3.3.1 Introduction

The term mixed model refers to the use of both fixed and random effects in the same analysis. It is particularly useful in settings where repeated measurements are made on the same statistical units (longitudinal study), or where measurements are made on clusters of related statistical units. Fixed effects have levels that are of primary interest and would be used again if the experiment is repeated. The levels of random effects are assumed to be randomly selected from an infinite population of possible levels.

3.3.2 Linear mixed model in genome-wide association study

A genome-wide association study (GWAS) is a powerful tool to investigate the associations between genes and a disease. Traditional approaches in GWAS assume that individuals sharing the same population background are unrelated. However, this independence assumption is always violated in the real world due to the genetic relatedness in study samples, which puts conventional statistical tests at the risk of spurious associations. Consequently, mixed models are widely used in GWAS as promising statistical methods to account for the hidden relatedness resulting from genealogy. Consider the following standard linear mixed model used in GWAS

$$\begin{aligned} y &= W\alpha + x\beta + Zu + \epsilon, \qquad (3.7) \\ u &\sim \mathrm{MVN}_m(0, \lambda\tau^{-1}\mathbf{K}), \\ \epsilon &\sim \mathrm{MVN}_n(0, \tau^{-1}\mathbf{I}_n), \end{aligned}$$

where n is the number of individuals, and m is the number of groups. y is an $n \times 1$ vector of quantitative traits, $W = (w_1, w_2, ..., w_c)$ is an $n \times c$ matrix of covariates (fixed effects) including a column vector of 1, α is a $c \times 1$ vector of corresponding coefficients including the intercept, x is an $n \times 1$ vector of marker genotypes, β is the effect size of the marker, Z is an $n \times m$ loading matrix, u is an $m \times 1$ vector of random effects, and ϵ is an $n \times 1$ vector of errors. τ^{-1} is the variance of the residual errors, λ is the ratio between the two variance components, $\mathbf{K}$ is a known $m \times m$ relatedness matrix, $\mathbf{I}_n$ is the $n \times n$ identity matrix, and MVN denotes multivariate normal distribution.

3.3.3 Restricted maximum likelihood estimation

Restricted maximum likelihood estimation (REML) is a way to estimate the variance components. In contrast to maximum likelihood estimation (MLE), REML can produce unbiased estimates of variance and covariance parameters. After estimation of random effects parameters, one can compute generalized least squares estimates of the fixed effects parameters.

The log-likelihood and log-restricted likelihood functions for the standard linear mixed model (3.7) are

$$l(\lambda, \tau, \alpha, \beta) = \frac{n}{2}\log(\tau) - \frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{H}| - \frac{1}{2}\tau(y - W\alpha - x\beta)^{\mathrm{T}}\mathbf{H}^{-1}(y - W\alpha - x\beta)$$

and

$$\begin{aligned} l_r(\lambda, \tau) \quad &= \quad \frac{n-c-1}{2}\log(\tau) - \frac{n-c-1}{2}\log(2\pi) + \frac{1}{2}\log|(W,x)^{\mathrm{T}}(W,x)| \\ &\quad -\frac{1}{2}\log|\mathbf{H}| - \frac{1}{2}\log|(W,x)^{\mathrm{T}}\mathbf{H}^{-1}(W,x)| - \frac{1}{2}\tau y^{\mathrm{T}}\mathbf{P}_x y, \end{aligned}$$

where

$$\mathbf{G} = Z\mathbf{K}Z^T, \mathbf{H} = \lambda\mathbf{G} + \mathbf{I}_n,$$

$$\mathbf{P}_x = \mathbf{H}^{-1} - \mathbf{H}^{-1}(W,x)((W,x)^{\mathrm{T}}\mathbf{H}^{-1}(W,x))^{-1}(W,x)^{\mathrm{T}}\mathbf{H}^{-1}.$$

If λ is known, the log-likelihood is maximized at:

$$\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix} = ((W,x)^{\mathrm{T}}\mathbf{H}^{-1}(W,x))^{-1}(W,x)^{\mathrm{T}}\mathbf{H}^{-1}y,$$

$$\hat{\tau} = \frac{n}{(y - W\hat{\alpha} - x\hat{\beta})^{\mathrm{T}}\mathbf{H}^{-1}(y - W\hat{\alpha} - x\hat{\beta})} = \frac{n}{y^{\mathrm{T}}\mathbf{P}_x y}.$$

Similarly, the log-restricted likelihood is maximized at

$$\hat{\tau} = \frac{n-c-1}{y^{\mathrm{T}}\mathbf{P}_x y}.$$

Therefore, finding MLEs and REML estimates is equivalent to optimizing the following target functions with respect to λ:

$$l(\lambda) = \frac{n}{2}\log(\frac{n}{2\pi}) - \frac{n}{2} - \frac{1}{2}\log|\mathbf{H}| - \frac{n}{2}\log(y^{\mathrm{T}}\mathbf{P}_x y),$$

$$\begin{aligned} l_r(\lambda) \quad = \quad & \frac{n-c-1}{2}\log(\frac{n-c-1}{2\pi}) - \frac{n-c-1}{2} + \frac{1}{2}\log|(W,x)^{\mathrm{T}}(W,x)| \\ & -\frac{1}{2}\log|\mathbf{H}| - \frac{1}{2}\log|(W,x)^{\mathrm{T}}\mathbf{H}^{-1}(W,x)| - \frac{n-c-1}{2}\log(y^{\mathrm{T}}\mathbf{P}_x y). \end{aligned}$$

After $\hat{\lambda}$ is obtained, one can then get $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\tau}$.

In Chapter 5, we propose a semiparametric mixed model to study the gene-disease associations in the context of possible nonlinear gene and environment interactions in GWAS. REML is then employed to estimate the variance components.

Chapter 4

Parameter Estimation for A Generalized Semiparametric Model with Repeated Measurements

4.1 Introduction

In the last three decades, advanced computing and telecommunication technologies have enabled researchers to collect data effectively and accurately. Hence, it is not surprising that the collected data can be complex and the analysis of such data is challenging. For example, in the regression context, the response variable can be discrete with repeated measurements, the relationship between the mean of the response variable and covariates can be non-linear, and the coefficients of explanatory variables can be dynamic. This motivates us to propose a model that can simultaneously account for these characteristics.

To take into account discrete responses and nonlinearity, Carroll (1997) proposed generalized partially linear single-index models (GPLSIM). These models encompass several important models, e.g., single-index models (Brillinger (1983); Horowitz (1998), Cui (2011)), generalized linear models (McCullagh and Nelder (1989)), partially linear models (Speckman (1998); Hardle (2000)), generalized partially linear models (Boente (2006)), and partially linear single-index models (Yu and Ruppert (2002); Xia and Hardle (2006); Ma and Zhu (2013)). The above references mainly focus on parameter estimation. Recently, researchers have employed penalized procedures (e.g., LASSO by Tibshirani (1996); SCAD by Fan and Li (2001)) to simultaneously select variables and estimate parameters for those models (e.g., Xie and Huang (2009); Liang (2010); Zhang (2010); Zeng (2012)).

Although the GPLSIM has played an important role in data analysis, it does not allow regression coefficients to be dynamic. To this end, Cleveland (1991) and Hastie (1993) proposed varying coefficient models, which have been applied in diverse fields, such as biological science, economics, finance, medicine, and social science. Further extensions to broad models are developed; see, for example, generalized varying coefficient models (Cai (2000)), semi-varying coefficient models (Zhang (2002)), survival models (Fan (2006)) and the newly proposed varying index coefficient model (Ma and Song (2014)). It is also noteworthy that an analog to the varying coefficient structure has been studied in the field of time series (e.g., see Chen and Tsay (1993); Cai (2000)). An excellent review paper on varying coefficient models can be found in Fan and Zhang (2008).

To better understand the performance of a response variable for each individual subject, a number of GPLSIMs as well as varying coefficient models have been extended to take into account repeated measurements (or longitudinal data or clustered data). Accordingly, various parameter estimation and model selection procedures are proposed (e.g., see Lin and Ying (2001); Davis, C. S. (2002); Diggle (2002); Huang (2002); Fan and Huang (2005); Lin and Carroll (2006); Ma (2012); Xu and Zhu (2012)). To obtain parameter estimation in repeated measurements, one needs to incorporate the correlation structure. Among available approaches, Qu and Li (2006) employed quadratic inference function (QIF) in Qu (2000) to directly incorporate correlations into their varying coefficient models without estimating nuisance parameters associated with correlations. Recently, Zhou and Qu (2012) adopted the QIF approach to obtain estimation and selection of correlation structure.

In this chapter, we introduce a generalized semiparametric model for repeated measurements by combining the GPLSIM with varying coefficient models. The proposed model is a useful analytic tool to investigate dynamic patterns of slope functions with some covariate such as time which naturally exist in longitudinal data as well as to capture possible nonlinear relationships between the response and covariates. Moreover, it contains many existing known parametric and nonparametric models as special cases, and thus it can be used for different types of data. Since each of GPLSIM and varying coefficient models has its own special feature, it is not surprising that obtaining parameter estimators and their theoretical properties become more challenging. For the sake of estimation, we first approximate the nonparametric function and coefficient functions by their corresponding linear combinations of spline basis functions. We then propose a profile-QIF procedure to obtain parameter estimates. It is worth noting that the profile procedure induces a single objective function of the parameters, which allows us to consider the penalization method for variable estimations and selections. The resulting penalized estimators of the nonzero coefficients are asymptotically normal and have the oracle property.

The rest of this chapter is organized as follows. Section 2 introduces the model structure and notation. Section 3 presents the estimation procedure, and demonstrates the consistency and asymptotic normality of parametric estimators as well as the consistency of nonparametric estimators. Section 4 proposes penalized estimators and shows their oracle properties. Simulation studies and an empirical example are presented in Section 5. We conclude this approach with discussions in Section 6, and technical proofs are relegated in the Appendix A.

4.2 A Generalized Semiparametric Model

To introduce the generalized semiparametric model by unifying partially linear single-index and varying coefficient models with repeated measurements, the j-th repeated observation for the i-th subject (or experimental unit) is denoted as $(Y_{ij}, X_{ij}, Z_{ij}, T_{ij})$ for $1 \le i \le n$ and $1 \le j \le m_i$, where Y_{ij} is the response variable and it is independent of other subjects, $X_{ij} = (X_{ij,1}, \ldots, X_{ij,p})^{\mathrm{T}}$ and $Z_{ij} = (Z_{ij,1}, \ldots, Z_{ij,d})^{\mathrm{T}}$ are p-dimensional and d -dimensional vectors of covariates, respectively, and T_{ij} represents a single predictor. Let $\mathbb{C}_{ij} = \left(X_{ij}^{\mathrm{T}}, Z_{ij}^{\mathrm{T}}, T_{ij}\right)^{\mathrm{T}}$ be the collection of covariates for the j-th observation of the i-th subject. We then consider the marginal model and assume that $E\left(Y_{ij} \left|\mathbb{C}_{ij}\right.\right) = \mu_{ij}$, where the marginal mean μ_{ij} depends on $\mathbb{C}_{ij}$ through a known monotonic and differentiable link function ϑ. This leads to the predictor function

$$\eta_{ij} = \vartheta(\mu_{ij}) = g\left(X_{ij}^{\mathrm{T}}\boldsymbol{\beta}\right) + \sum\nolimits_{l=1}^{d_1} \alpha_l\left(T_{ij}\right) Z_{ij,l} + \sum\nolimits_{l=d_1+1}^{d} \alpha_l Z_{ij,l},\ j = 1, \ldots, m_i,\ i = 1, ..., n, \tag{4.1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ is a p-dimensional index parameter, $\alpha_l(\cdot)$, $l = 1, ..., d_1$, are unknown smooth functions, α_l, $l = d_1 + 1, ..., d$ are coefficients, and $g(\cdot)$ is an unknown differentiable function of $U_{ij}(\boldsymbol{\beta}) = X_{ij}^{\mathrm{T}}\boldsymbol{\beta}$. For identifiability, we assume that $\boldsymbol{\beta}$ belongs to the parameter space:

$$\Theta = \left\{\boldsymbol{\beta} : \|\boldsymbol{\beta}\| = 1, \beta_1 > 0, \boldsymbol{\beta} \in R^p\right\}, \tag{4.2}$$

where $\|\cdot\|$ denotes the Euclidean norm of any vector $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_s)^{\mathrm{T}} \in R^s$ such that $\|\boldsymbol{\zeta}\| = (|\zeta_1|^2 + \cdots + |\zeta_s|^2)^{1/2}$.

Model (4.1) contains many existing models as special cases. When $\alpha_l(T_{ij}) = \alpha_l$ for $1 \leq l \leq d_1$, α_l are unknown constants, and $m_i = 1$ for $i = 1, \cdots, n$, it leads to a generalized partially linear single-index model (Carroll (1997)); when $g\left(X_{ij}^{\mathrm{T}}\boldsymbol{\beta}\right) = 0$, it yields a semiparametric varying-coefficient partially linear model (Fan and Huang (2005)); when $p = 1$ and $\alpha_l(T_{ij}) = \alpha_l$ for $1 \leq l \leq d_1$, it results to a generalized partially linear model (Hardle (2000)); when $g\left(X_{ij}^{\mathrm{T}}\boldsymbol{\beta}\right) = 0$ and $\alpha_l = 0$ for $l = d_1 + 1, \cdots, d$, it gives a generalized varying coefficient model (Hastie (1993) and Cai (2000)). It is worth noting that model (4.1) is different from the varying index coefficient model proposed by Ma and Song (2014), since the latter aims to assess nonlinear interaction effects of index variables with other covariates on the response in the cross-sectional data setting.

4.3 Parameter Estimates

4.3.1 The approximation of predictor function

In this subsection, we approximate the unknown functions $g(\cdot)$ and $\alpha_l(\cdot)$ in (4.1) by B-splines described as follows. Based on the given knots, we define the sets of q-th order B-spline basis functions

$$B_1(u) = \{B_{1,J}(u) : 1 \leq J \leq N+q\}^{\mathrm{T}}$$

and

$$B_2(t) = \{B_{2,J}(t) : 1 \leq J \leq N+q\}^{\mathrm{T}}$$

(see de Boor (2001)), where N is the number of interior knots with the distance between neighboring knots satisfying the conditions given in Zhou (1998). Then, the unknown function g in (4.1) can be approximated by a linear combination of the B-spline functions such that $g(u) \approx \sum_{J=1}^{N+q} \gamma_{J,0} B_{1,J}(u) = B_1(u)^{\mathrm{T}} \boldsymbol{\gamma}_0$ with a set of coefficients $\boldsymbol{\gamma}_0 = (\gamma_{1,0}, \ldots, \gamma_{N+q,0})^{\mathrm{T}}$. Analogously, $\alpha_l(t)$ in (4.1) can be approximated by $\alpha_l(t) \approx \sum_{J=1}^{N+q} \gamma_{J,l} B_{2,J}(t) = B_2(t)^{\mathrm{T}} \boldsymbol{\gamma}_l$, where $\boldsymbol{\gamma}_l = (\gamma_{1,l}, \ldots, \gamma_{N+q,l})^{\mathrm{T}}$. Accordingly, we obtain an approximation of the predictor function η_{ij}, which is

$$\widetilde{\eta}_{ij} = \sum_{J=1}^{N+q} \gamma_{J,0} B_{1,J}\left(X_{ij}^{\mathrm{T}} \boldsymbol{\beta}\right) + \sum_{l=1}^{d_1} \sum_{J=1}^{N+q} \gamma_{J,l} B_{2,J}(T_{ij}) Z_{ij,l} + \sum_{l=d_1+1}^{d} \alpha_l Z_{ij,l}. \quad (4.3)$$

From the above equation, we propose a two-step estimation procedure to estimate parametric and nonparametric components in the following two subsections, respectively.

4.3.2 The profile QIF estimators of parametric vectors

Let $Y_i = (Y_{i1}, \ldots, Y_{im_i})^{\mathrm{T}}$ and $\mu_i(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\alpha}) = (\mu_{i1}(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\alpha}), \ldots, \mu_{im_i}(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\alpha}))^{\mathrm{T}}$. Denote $\widetilde{\mu}_i(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\alpha}) = (\widetilde{\mu}_{i1}(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\alpha}), \ldots, \widetilde{\mu}_{im_i}(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\alpha}))^{\mathrm{T}}$, where $\widetilde{\mu}_{ij}(\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \vartheta^{-1}(\widetilde{\eta}_{ij})$, $\boldsymbol{\alpha} = (\alpha_{d_1+1}, \ldots, \alpha_d)^{\mathrm{T}}$ is a $d_2 \times 1$ vector, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_0^{\mathrm{T}}, \ldots, \boldsymbol{\gamma}_{d_1}^{\mathrm{T}})^{\mathrm{T}}$ is a $(1+d_1) J_n \times 1$ vector, and $J_n = N + q$. For given $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, the quasi-likelihood estimator of $\boldsymbol{\gamma}$ is the solution of the following estimating equations,

$$\sum\nolimits_{i=1}^{n} \widetilde{\mu}_i'(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\alpha}) \mathbf{V}_i^{-1} (Y_i - \widetilde{\mu}_i(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\alpha})) = 0, \tag{4.4}$$

where $\widetilde{\mu}_i'(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\alpha}) = \left[\left(\widetilde{\mu}_{i1}', \ldots, \widetilde{\mu}_{im_i}'\right)\right]_{(1+d_1)J_n \times m_i}$, $\widetilde{\mu}_{ij}' = \partial \widetilde{\mu}_{ij} / \partial \boldsymbol{\gamma}$ for $j = 1, \cdots, m_i$, and $\mathbf{V}_i$ is the $m_i \times m_i$ covariance matrix of Y_i. Since $\mathbf{V}_i$ is often unknown in practice, we adopt the approach in Liang and Zeger (1986) and assume that $\mathbf{V}_i = \mathbf{A}_i^{1/2} \mathbf{R}_i(\varsigma) \mathbf{A}_i^{1/2} / \phi$, where $\mathbf{R}_i(\varsigma)$ is the working correlation matrix of Y_i, ς is a vector of nuisance parameters, and $\mathbf{A}_i$ is an $m_i \times m_i$ diagonal matrix with the marginal variance of Y_{ij} as its j-th diagonal element. However, the working correlation structure may be misspecified. Hence, we further apply the quadratic inference function (QIF) in Qu (2000) to efficiently incorporate the within-cluster correlation structure. For the sake of simplicity, we assume that cluster sizes are equal, i.e., $m_i = m < \infty$, and let $\mathbf{R}$ be a common working correlation matrix. When the cluster sizes are unequal, our estimation procedure given below can be modified via the same technique proposed by Xue (2010). Following the QIF approach, the inverse of $\mathbf{R}$ can be approximated by a linear combination of κ basis matrices, i.e.,

$$\mathbf{R}^{-1} \approx a_1 \mathbf{M}_1 + \cdots + a_\kappa \mathbf{M}_\kappa, \tag{4.5}$$

where $\mathbf{M}_1 = \mathbf{I}$ (the identity matrix) and $\mathbf{M}_k$ are known symmetric basis matrices for $1 \le k \le \kappa$.

We next construct the QIF to obtain parameter estimators. To this end, consider the estimating function of $\boldsymbol{\gamma}$, for the given $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$:

$$\widetilde{\phi}_n(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\alpha}) = n^{-1}\sum\nolimits_{i=1}^{n} \widetilde{\phi}_{in}(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\alpha}) = \left\{\widetilde{\phi}_{n,1}(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\alpha})^{\mathrm{T}},\ldots,\widetilde{\phi}_{n,\kappa}(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\alpha})^{\mathrm{T}}\right\}^{\mathrm{T}}$$

$$= n^{-1}\left\{\begin{array}{c} \sum_{i=1}^{n}\mathbf{Q}_i(\boldsymbol{\beta})^{\mathrm{T}}\widetilde{\Delta}_i(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\alpha})\mathbf{A}_i^{-1/2}\mathbf{M}_1\mathbf{A}_i^{-1/2}(Y_i-\widetilde{\mu}_i(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\alpha})) \\ \vdots \\ \sum_{i=1}^{n}\mathbf{Q}_i(\boldsymbol{\beta})^{\mathrm{T}}\widetilde{\Delta}_i(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\alpha})\mathbf{A}_i^{-1/2}\mathbf{M}_\kappa\mathbf{A}_i^{-1/2}(Y_i-\widetilde{\mu}_i(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\alpha})) \end{array}\right\}_{\kappa J_n(1+d_1)\times 1}, \tag{4.6}$$

where $\widetilde{\Delta}_i(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\alpha}) = \mathrm{diag}(\widetilde{\nu}_{i1}(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\alpha}),\ldots,\widetilde{\nu}_{im}(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\alpha}))$, $\widetilde{\nu}_{ij}(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\alpha}) = \partial\widetilde{\mu}_{ij}(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\alpha})/\partial\widetilde{\eta}_{ij}$, $\mathbf{Q}_i(\boldsymbol{\beta}) = \left(Q_{i1}(\boldsymbol{\beta})^{\mathrm{T}},\ \ldots, Q_{im_i}(\boldsymbol{\beta})^{\mathrm{T}}\right)^{\mathrm{T}}$, and

$$Q_{ij}(\boldsymbol{\beta}) = \left[B_1(U_{ij}(\boldsymbol{\beta}))^{\mathrm{T}}, \left\{B_2(T_{ij})^{\mathrm{T}} Z_{ij,l} : 1 \le l \le d_1\right\}\right]^{\mathrm{T}}_{J_n(1+d_1)\times 1}.$$

Then, define the QIF to be

$$\widetilde{Q}_n(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\alpha}) = \widetilde{\phi}_n(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\alpha})^{\mathrm{T}}\widetilde{C}_n(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\alpha})^{-1}\widetilde{\phi}_n(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\alpha}), \tag{4.7}$$

where $\widetilde{C}_n(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\alpha}) = n^{-2}\sum_{i=1}^{n}\widetilde{\phi}_{in}(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\alpha})\widetilde{\phi}_{in}(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\alpha})^{\mathrm{T}}$. Accordingly, the QIF estimator of $\boldsymbol{\gamma}$ is

$$\begin{aligned}\widetilde{\gamma}(\boldsymbol{\beta},\boldsymbol{\alpha}) &= \left[\left\{\widetilde{\gamma}_0(\boldsymbol{\beta},\boldsymbol{\alpha})^{\mathrm{T}},\ldots,\widetilde{\gamma}_{d_1}(\boldsymbol{\beta},\boldsymbol{\alpha})^{\mathrm{T}}\right\}^{\mathrm{T}}\right] \\ &= \arg\min_{\boldsymbol{\gamma}\in\mathbf{R}^{(1+d_1)J_n}}\widetilde{Q}_n(\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\alpha}).\end{aligned}$$

As a result, the QIF estimators of $g(\cdot)$, $g'(\cdot)$ (the first derivative of g), and $\alpha_l(\cdot)$ are

$$\begin{aligned}\widetilde{g}(u;\boldsymbol{\beta},\boldsymbol{\alpha}) &= B_1(u)^{\mathrm{T}}\widetilde{\boldsymbol{\gamma}}_0(\boldsymbol{\beta},\boldsymbol{\alpha})\\ \widetilde{\alpha}_l(t;\ \boldsymbol{\beta},\boldsymbol{\alpha}) &= B_2(t)^{\mathrm{T}}\ \widetilde{\boldsymbol{\gamma}}_l(\boldsymbol{\beta},\boldsymbol{\alpha})\\ \widetilde{g}'(u;\boldsymbol{\beta},\boldsymbol{\alpha}) &= B_1'(u)^{\mathrm{T}}\widetilde{\boldsymbol{\gamma}}_0(\boldsymbol{\beta},\boldsymbol{\alpha})\end{aligned}$$

where $B_1'(u)$ is the first derivative of $B_1(u)$. By replacing $g(\cdot)$ and $\alpha_l(\cdot)$ with $\widetilde{g}(\cdot;\boldsymbol{\beta},\boldsymbol{\alpha})$ and $\widetilde{\alpha}_l(\cdot;\boldsymbol{\beta},\boldsymbol{\alpha})$ in (4.3), we obtain

$$\widehat{\eta}_{ij}(\boldsymbol{\beta},\boldsymbol{\alpha}) = \widetilde{g}\left(X_{ij}^{\mathrm{T}}\boldsymbol{\beta};\boldsymbol{\beta},\boldsymbol{\alpha}\right) + \sum\nolimits_{l=1}^{d_1}\widetilde{\alpha}_l(T_{ij};\ \boldsymbol{\beta},\boldsymbol{\alpha})Z_{ij,l} + \sum\nolimits_{l=d_1+1}^{d}\alpha_l Z_{ij,l}. \tag{4.8}$$

Before estimating $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, we use the assumptions of $||\boldsymbol{\beta}|| = 1$ and $\beta_1 > 0$ in (4.2) to reform the space of $\boldsymbol{\beta}$ given below, which ensures identifiability.

$$\left\{\left(\left(1-\sum\nolimits_{s=2}^{p}\beta_s^2\right)^{1/2},\beta_2,\ldots,\beta_p\right)^{\mathrm{T}} : \sum\nolimits_{s=2}^{p}\beta_s^2 < 1\right\}.$$

Denote $\widehat{\eta}_i(\boldsymbol{\beta},\boldsymbol{\alpha}) = \{\widehat{\eta}_{i1}(\boldsymbol{\beta},\boldsymbol{\alpha}),\ldots,\widehat{\eta}_{im}(\boldsymbol{\beta},\boldsymbol{\alpha})\}^{\mathrm{T}}$ and its gradient with respect to $\left(\boldsymbol{\beta}_{-1}^{\mathrm{T}},\ \boldsymbol{\alpha}^{\mathrm{T}}\right)^{\mathrm{T}}$ by $\widehat{\mathbf{D}}_i(\boldsymbol{\beta},\boldsymbol{\alpha}) = \nabla\widehat{\eta}_i(\ \boldsymbol{\beta},\boldsymbol{\alpha}) = \left\{\frac{\partial\widehat{\eta}_{i1}(\boldsymbol{\beta},\boldsymbol{\alpha})}{\partial\left(\boldsymbol{\beta}_{-1}^{\mathrm{T}},\boldsymbol{\alpha}^{\mathrm{T}}\right)^{\mathrm{T}}},\ldots,\frac{\partial\widehat{\eta}_{im}(\boldsymbol{\beta},\boldsymbol{\alpha})}{\partial\left(\boldsymbol{\beta}_{-1}^{\mathrm{T}},\boldsymbol{\alpha}^{\mathrm{T}}\right)^{\mathrm{T}}}\right\}_{m\times(p-1+d_2)}^{\mathrm{T}}$, where $\boldsymbol{\beta}_{-1} = (\beta_2,\ldots,\beta_p)^{\mathrm{T}}$. Consider the profiled estimating function of $(\boldsymbol{\beta},\boldsymbol{\alpha})$,

$$\begin{aligned}&\psi_n^*(\boldsymbol{\beta},\boldsymbol{\alpha})\\ =\ & n^{-1}\sum\nolimits_{i=1}^{n}\psi_{in}(\boldsymbol{\beta},\boldsymbol{\alpha}) = \left\{\psi_{n,1}(\boldsymbol{\beta},\boldsymbol{\alpha})^{\mathrm{T}},\ldots,\psi_{n,\kappa}(\boldsymbol{\beta},\boldsymbol{\alpha})^{\mathrm{T}}\right\}^{\mathrm{T}}\\ =\ & n^{-1}\left\{\begin{array}{c}\sum_{i=1}^{n}\widehat{\mathbf{D}}_i^{\mathrm{T}}(\boldsymbol{\beta},\boldsymbol{\alpha})\Delta_i(\boldsymbol{\beta},\boldsymbol{\alpha})\mathbf{A}_i^{-1/2}\mathbf{M}_1\mathbf{A}_i^{-1/2}(Y_i-\widehat{\mu}_i(\boldsymbol{\beta},\boldsymbol{\alpha}))\\ \vdots\\ \sum_{i=1}^{n}\widehat{\mathbf{D}}_i^{\mathrm{T}}(\boldsymbol{\beta},\boldsymbol{\alpha})\Delta_i(\boldsymbol{\beta},\boldsymbol{\alpha})\mathbf{A}_i^{-1/2}\mathbf{M}_\kappa\mathbf{A}_i^{-1/2}(Y_i-\widehat{\mu}_i(\boldsymbol{\beta},\boldsymbol{\alpha}))\end{array}\right\}_{\kappa(p-1+d_2)\times 1},\end{aligned}$$

where $\widehat{\mu}_i(\boldsymbol{\beta},\boldsymbol{\alpha}) = \{\widehat{\mu}_{i1}(\boldsymbol{\beta},\boldsymbol{\alpha}),\ldots,\widehat{\mu}_{im}(\boldsymbol{\beta},\boldsymbol{\alpha})\}^{\mathrm{T}}$, $\widehat{\mu}_{ij}(\boldsymbol{\beta},\boldsymbol{\alpha}) = \vartheta^{-1}\{\widehat{\eta}_{ij}(\boldsymbol{\beta},\boldsymbol{\alpha})\}$, $\Delta_i(\boldsymbol{\beta},\boldsymbol{\alpha}) = \mathrm{Diag}(\widehat{\nu}_{ij}(\boldsymbol{\beta},\boldsymbol{\alpha}))$, and $\widehat{\nu}_{ij}(\boldsymbol{\beta},\boldsymbol{\alpha}) = \partial\widehat{\mu}_{ij}(\boldsymbol{\beta},\boldsymbol{\alpha})/\partial\widehat{\eta}_{ij}$. Then, the profiled QIF estimator of

$\left(\boldsymbol{\beta}_{-1}^{\mathrm{T}}, \boldsymbol{\alpha}^{\mathrm{T}}\right)^{\mathrm{T}}$ is

$$\left(\widehat{\boldsymbol{\beta}}_{-1}^{\mathrm{T}}, \widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\right)^{\mathrm{T}} = \arg \min_{\left(\boldsymbol{\beta}_{-1}, \boldsymbol{\alpha}\right)} Q_n^*\left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right),$$

where

$$Q_n^*\left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right) = \psi_n^*\left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)^{\mathrm{T}} \Psi_n^*\left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)^{-1} \psi_n^*\left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right), \tag{4.9}$$

and $\Psi_n^*\left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right) = n^{-2} \sum_{i=1}^n \psi_n^*\left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)^{\mathrm{T}} \psi_n^*\left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)$. Using the fact that $\beta_1 = \sqrt{1 - \left\|\boldsymbol{\beta}_{-1}\right\|^2}$, we also obtain the estimator $\widehat{\beta}_1$. The detailed procedure for computing $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\alpha}}$ is given in Section 5.5.

To study the asymptotic properties of the parametric estimators, we need to introduce a few quantities evaluated at the true parameter values. To this end, let $\boldsymbol{\beta}^0 = \left(\beta_1^0, \boldsymbol{\beta}_{-1}^{0\mathrm{T}}\right)^{\mathrm{T}}$ and $\boldsymbol{\alpha}^0$ be the true parameter vectors, $\boldsymbol{\beta}_{-1}^0 = \left(\beta_2^0, \ldots, \beta_p^0\right)^{\mathrm{T}}$, and $\mathbf{J}^0 = \frac{\partial \boldsymbol{\beta}^0}{\partial \boldsymbol{\beta}^{(1)}}$ given below be the Jacobian matrix of size $p \times (p-1)$.

$$\mathbf{J}^0 = \begin{pmatrix} -\boldsymbol{\beta}_{-1}^{0\mathrm{T}} / \sqrt{1 - \left\|\boldsymbol{\beta}_{-1}^0\right\|^2} \\ \mathbf{I}_{p-1} \end{pmatrix}_{p \times (p-1)}.$$

For $1 \leq s \leq p$ and $0 \leq l \leq d_1$, let $\boldsymbol{\xi}_{s,l} = \left(\xi_{s,1,l}, \ldots, \xi_{s,N+q,l}\right)^{\mathrm{T}}$ be a $J_n \times 1$ vector of parameters. Let $\boldsymbol{\xi}_s = \left\{\left(\boldsymbol{\xi}_{s,0}^{\mathrm{T}}, \ldots, \boldsymbol{\xi}_{s,d_1}^{\mathrm{T}}\right)^{\mathrm{T}}\right\}_{(1+d_1)J_n \times 1}$. For $1 \leq s \leq p$, we further define

$$\omega_{n,s}\left(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, \boldsymbol{\xi}_s\right) = n^{-1} \sum_{i=1}^n \begin{bmatrix} \mathbf{Q}_i\left(\boldsymbol{\beta}^0\right)^{\mathrm{T}} \Delta_i \Lambda_1 \Delta_i \left\{X_{ij,s} - \mathbf{Q}_i\left(\boldsymbol{\beta}^0\right)^{\mathrm{T}} \boldsymbol{\xi}_s\right\} \\ \vdots \\ \mathbf{Q}_i\left(\boldsymbol{\beta}^0\right)^{\mathrm{T}} \Delta_i \Lambda_\kappa \Delta_i \left\{X_{ij,s} - \mathbf{Q}_i\left(\boldsymbol{\beta}^0\right)^{\mathrm{T}} \boldsymbol{\xi}_s\right\} \end{bmatrix}_{\kappa J_n (1+d_1) \times 1} \quad \text{and}$$

$$\Xi_n\left(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right) = \frac{1}{n^2} \sum_{i=1}^n \begin{bmatrix} \mathbf{Q}_i\left(\boldsymbol{\beta}^0\right)^{\mathrm{T}} \Delta_i \Gamma_{1,1} \Delta_i \mathbf{Q}_i\left(\boldsymbol{\beta}^0\right) & \cdots & \mathbf{Q}_i\left(\boldsymbol{\beta}^0\right)^{\mathrm{T}} \Delta_i \Gamma_{1,\kappa} \Delta_i \mathbf{Q}_i\left(\boldsymbol{\beta}^0\right) \\ \vdots & \ddots & \vdots \\ \mathbf{Q}_i\left(\boldsymbol{\beta}^0\right)^{\mathrm{T}} \Delta_i \Gamma_{\kappa,1} \Delta_i \mathbf{Q}_i\left(\boldsymbol{\beta}^0\right) & \cdots & \mathbf{Q}_i\left(\boldsymbol{\beta}^0\right)^{\mathrm{T}} \Delta_i \Gamma_{\kappa,\kappa} \Delta_i \mathbf{Q}_i\left(\boldsymbol{\beta}^0\right) \end{bmatrix}, \tag{4.10}$$

with dimension $\kappa J_n(1+d_1)\times\kappa J_n(1+d_1)$, where $\Lambda_k=\mathbf{A}_i^{-1/2}\mathbf{M}_k\mathbf{A}_i^{-1/2}$, $\Gamma_{k,k'}=\Lambda_k\mathbf{V}_i\Lambda_{k'}$ for $1\le k,k'\le\kappa$, and Δ_i, $\mathbf{V}_i$, and Λ_k are evaluated at $(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0)$. Then, we obtain the estimate of $\boldsymbol{\xi}_s$,

$$\widehat{\boldsymbol{\xi}}_s=\arg\min_{\boldsymbol{\xi}_s\in R^{(1+d_1)J_n}}\left\{\omega_{n,s}\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0,\boldsymbol{\xi}_s\right)^{\mathrm{T}}\Xi_n\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)^{-1}\omega_{n,s}\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0,\boldsymbol{\xi}_s\right)\right\}. \qquad (4.11)$$

In addition, replace $X_{ij,s}$ and $\boldsymbol{\xi}_s$ in $\omega_{n,s}\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0,\boldsymbol{\xi}_s\right)$ by $Z_{ij,l}$ and $\boldsymbol{\tau}_l$, respectively, which yields $\omega_{n,l}\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0,\boldsymbol{\tau}_l\right)$ for $d_1+1\le l\le d$. Adapting (4.11), we obtain the estimate $\widehat{\boldsymbol{\tau}}_l$.

Define $\widehat{X}_{ij,s}=X_{ij,s}-Q_{ij}\left(\boldsymbol{\beta}^0\right)^{\mathrm{T}}\widehat{\boldsymbol{\xi}}_s$, $\widehat{X}_{ij}=\left(\widehat{X}_{ij,1},\ldots,\widehat{X}_{ij,p}\right)^{\mathrm{T}}$, $\widehat{Z}_{ij,l}=Z_{ij,l}-Q_{ij}\left(\boldsymbol{\beta}^0\right)^{\mathrm{T}}\widehat{\boldsymbol{\tau}}_l$, and $\widehat{Z}_{ij}^{(2)}=\left(\widehat{Z}_{ij,d_1+1},\ldots,\widehat{Z}_{ij,d}\right)^{\mathrm{T}}$. In Lemma 10 of Appendix, we demonstrate that

$$\frac{\partial\widehat{\eta}_{ij}\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)}{\partial\left(\boldsymbol{\beta}_{-1}^{\mathrm{T}},\ \boldsymbol{\alpha}^{\mathrm{T}}\right)^{\mathrm{T}}}=\left\{\widetilde{g}'\left(X_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0;\boldsymbol{\beta}^0,\ \boldsymbol{\alpha}^0\right)\widehat{X}_{ij}^{\mathrm{T}}\mathbf{J}^0,\widehat{Z}_{ij}^{(2)\mathrm{T}}\right\}^{\mathrm{T}}\{1+o_p(1)\}.$$

Accordingly,

$$\begin{aligned}&\widehat{\mathbf{D}}_i\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)\\ =\ &\left[\left\{\widetilde{g}'\left(X_{i1}^{\mathrm{T}}\boldsymbol{\beta}^0;\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)\widehat{X}_{i1},\ldots,\widetilde{g}'\left(X_{im}^{\mathrm{T}}\boldsymbol{\beta}^0;\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)\widehat{X}_{im}\right\}^{\mathrm{T}}\mathbf{J}^0,\left(\widehat{Z}_{i1}^{(2)},\ldots,\widehat{Z}_{im}^{(2)}\right)^{\mathrm{T}}\right]\\ &\times\{1+o_p(1)\}\end{aligned}$$

Define $D_{ij}\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)=\left\{g'\left(X_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0\right)\widehat{X}_{ij}^{\mathrm{T}}\mathbf{J}^{\mathbf{0}},\widehat{Z}_{ij}^{(2)\mathrm{T}}\right\}^{\mathrm{T}}$,

$$\dot{\psi}_n\left(\boldsymbol{\beta}^{\mathbf{0}},\boldsymbol{\alpha}^{\mathbf{0}}\right)=E\left\{\begin{array}{c}\mathbf{D}_i^{\mathrm{T}}\Delta_i\left(\boldsymbol{\beta}^{\mathbf{0}},\boldsymbol{\alpha}^{\mathbf{0}}\right)\Lambda_1\Delta_i\left(\boldsymbol{\beta}^{\mathbf{0}},\boldsymbol{\alpha}^{\mathbf{0}}\right)\mathbf{D}_i\\ \vdots\\ \mathbf{D}_i^{\mathrm{T}}\Delta_i\left(\boldsymbol{\beta}^{\mathbf{0}},\boldsymbol{\alpha}^{\mathbf{0}}\right)\Lambda_\kappa\Delta_i\left(\boldsymbol{\beta}^{\mathbf{0}},\boldsymbol{\alpha}^{\mathbf{0}}\right)\mathbf{D}_i\end{array}\right\}\text{ and}$$

$$
\begin{aligned}
& \Psi_n\left(\boldsymbol{\beta}^{\mathbf{0}}, \boldsymbol{\alpha}^{\mathbf{0}}\right) \\
= \ & E\left\{\begin{array}{ccc}
\mathbf{D}_i^{\mathrm{T}} \Delta_i\left(\boldsymbol{\beta}^{\mathbf{0}}, \boldsymbol{\alpha}^{\mathbf{0}}\right) \Gamma_{1,1} \Delta_i\left(\boldsymbol{\beta}^{\mathbf{0}}, \boldsymbol{\alpha}^{\mathbf{0}}\right) \mathbf{D}_i & \cdots & \mathbf{D}_i^{\mathrm{T}} \Delta_i\left(\boldsymbol{\beta}^{\mathbf{0}}, \boldsymbol{\alpha}^{\mathbf{0}}\right) \Gamma_{1,\kappa} \Delta_i\left(\boldsymbol{\beta}^{\mathbf{0}}, \boldsymbol{\alpha}^{\mathbf{0}}\right) \mathbf{D}_i \\
\vdots & \ddots & \vdots \\
\mathbf{D}_i^{\mathrm{T}} \Delta_i\left(\boldsymbol{\beta}^{\mathbf{0}}, \boldsymbol{\alpha}^{\mathbf{0}}\right) \Gamma_{\kappa,1} \Delta_i\left(\boldsymbol{\beta}^{\mathbf{0}}, \boldsymbol{\alpha}^{\mathbf{0}}\right) \mathbf{D}_i & \cdots & \mathbf{D}_i^{\mathrm{T}} \Delta_i\left(\boldsymbol{\beta}^{\mathbf{0}}, \boldsymbol{\alpha}^{\mathbf{0}}\right) \Gamma_{\kappa,\kappa} \Delta_i\left(\boldsymbol{\beta}^{\mathbf{0}}, \boldsymbol{\alpha}^{\mathbf{0}}\right) \mathbf{D}_i
\end{array}\right\},
\end{aligned}
$$

where $\mathbf{D}_i = \mathbf{D}_i\left(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right) = \left(D_{i1}\left(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right), \ldots, D_{im}\left(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)\right)^{\mathrm{T}}$. Then, the asymptotic properties of parametric estimators are given below.

Theorem 1 *Assume that Conditions (C1)-(C5) in the Appendix hold, $N^4 n^{-1} = o(1)$, and $N^{-4r+2} n = o(1)$ with $r > 3/2$ in Condition (C2). Then, we have*

$$
\left\|\left(\widehat{\boldsymbol{\beta}}_{-1}^{T}, \ \widehat{\boldsymbol{\alpha}}^{T}\right)^{T} - \left(\boldsymbol{\beta}_{-1}^{0T}, \boldsymbol{\alpha}^{0T}\right)^{T}\right\| = o_p(1),
$$

and, as $n \to \infty$,

$$
\sqrt{n}\left(\Sigma_n^{(1)}\right)^{1/2}\left(\left(\widehat{\boldsymbol{\beta}}_{-1}^{T}, \widehat{\boldsymbol{\alpha}}^{T}\right)^{T} - \left(\boldsymbol{\beta}_{-1}^{0T}, \boldsymbol{\alpha}^{0T}\right)^{T}\right) \to N\left(0, \mathbf{I}_{p-1+d_2}\right),
$$

where $\Sigma_n^{(1)} = \dot{\psi}_n\left(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)^T \Psi_n\left(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right) \dot{\psi}_n\left(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)$ and $\mathbf{I}_a$ denotes the identity matrix with dimension $a \times a$.

Let $\boldsymbol{\Upsilon} = \begin{pmatrix} \mathbf{J}_{p\times(p-1)} & \mathbf{0}_{p\times d_2} \\ \mathbf{0}_{d_2\times(p-1)} & \mathbf{I}_{d_2\times d_2} \end{pmatrix}$. The above theorem, together with the multivariate delta-method, establishes the asymptotic normality of parametric estimators,

$$
\sqrt{n}\Sigma_n^{1/2}\left(\left(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}, \widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\right)^{\mathrm{T}} - \left(\boldsymbol{\beta}^{0\mathrm{T}}, \boldsymbol{\alpha}^{0\mathrm{T}}\right)^{\mathrm{T}}\right) \to N\left(0, \mathbf{I}_{p+d_2}\right),
$$

as $n \to \infty$, where $\Sigma_n = \boldsymbol{\Upsilon}\Sigma_n^{(1)}\boldsymbol{\Upsilon}^{\mathrm{T}}$. It is also worth noting that the resulting estimators are not semiparametric efficient since we assume that the true correlation structure is unknown and the working correlation may be misspecified.

4.3.3 The QIF estimator of nonparametric functions

After obtaining the parametric estimators $\left(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}, \widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\right)^{\mathrm{T}}$, we replace $\left(\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\alpha}^{\mathrm{T}}\right)^{\mathrm{T}}$ by $\left(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}, \widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\right)^{\mathrm{T}}$ in $\widetilde{Q}_n\left(\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\beta}\right)$ of (4.7). This allows us to find the estimator

$$\widehat{\boldsymbol{\gamma}} = \left(\widehat{\boldsymbol{\gamma}}_0^{\mathrm{T}}, \ldots, \widehat{\boldsymbol{\gamma}}_{d_1}^{\mathrm{T}} \right)^{\mathrm{T}} = \arg \min_{\boldsymbol{\gamma} \in \mathbf{R}^{(1+d_1)J_n}} \widetilde{Q}_n \left(\boldsymbol{\gamma}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}} \right).$$

Accordingly, the estimators of nonparametric functions $g\left(\cdot\right)$ and $\alpha_l\left(\cdot\right)$ are $\widehat{g}\left(u\right) = B_1\left(u\right)^{\mathrm{T}} \widehat{\boldsymbol{\gamma}}_0$ and $\widehat{\alpha}_l\left(t\right) = B_2\left(t\right)^{\mathrm{T}} \widehat{\boldsymbol{\gamma}}_l$, respectively. Next we present the L_2 convergence rates of $\widehat{g}$ and $\widehat{\alpha}_l$. With a slight abuse of notation in using $\|\cdot\|$, let $\|\phi\| = \left\{\int_{\mathcal{S}} \phi\left(t\right)^2 dt\right\}^{1/2}$ be the L_2 norm of any square integrable real-valued function $\phi\left(t\right)$ on its support $\mathcal{S}$.

Theorem 2 *Assume that* $N^4 n^{-1} = o\left(1\right)$ *and* $N^{-2-2r} n = o\left(1\right)$ *with* $r > 3/2$ *in Condition (C2). Then, under Conditions (C1)-(C5), we have* $\|\widehat{g}\left(\cdot\right) - g\left(\cdot\right)\| = O_p\left(\sqrt{N/n} + N^{-r}\right)$ *and* $\|\widehat{\alpha}_l\left(\cdot\right) - \alpha_l\left(\cdot\right)\| = O_p\left(\sqrt{N/n} + N^{-r}\right)$.

The optimal order requirements in the above theorem are achieved when the number of interior knots N is chosen to be $N \asymp n^{1/(2r+1)}$. As a result, the estimators $\widehat{g}$ and $\widehat{\alpha}_l$ of the nonparametric functions g and α_l have the optimal convergence rate $O_p\left(N^{-r/(2r+1)}\right)$.

4.4 Penalized-QIF Estimation

4.4.1 Penalized estimators

In data analysis, the true model is often unknown. Hence, researchers have employed the penalized approach to simultaneously select relevant variables and estimate unknown parameters for partially linear single-index models (see, e.g., Xie and Huang

(2009); Liang (2010)) and varying coefficients models (see, e.g., Li and Liang (2008)). This motivates us to propose a penalized QIF method for the proposed generalized semi-parametric model. Without loss of generality, we assume that the correct model in (4.1) has the true regression coefficients $\boldsymbol{\beta}^0 = \left(\boldsymbol{\beta}^{0\mathrm{T}}_{(1)}, \boldsymbol{\beta}^{0\mathrm{T}}_{(2)}\right)^{\mathrm{T}}$ and $\boldsymbol{\alpha}^0 = \left(\boldsymbol{\alpha}^{0\mathrm{T}}_{(1)}, \boldsymbol{\alpha}^{0\mathrm{T}}_{(2)}\right)^{\mathrm{T}}$, where $\boldsymbol{\beta}^0_{(1)} = \left[\beta^0_1, \left\{\left(\boldsymbol{\beta}^0_{(1),-1}\right)_{(p_1-1)\times 1}\right\}^{\mathrm{T}}\right]^{\mathrm{T}}$ is the $p_1 \times 1$ vector of non-zeros, $\boldsymbol{\beta}^0_{(2)}$ is the $(p-p_1)\times 1$ vector of zeros, $\boldsymbol{\alpha}^0_{(1)}$ is the $d_{20} \times 1$ vector of non-zeros and $\boldsymbol{\alpha}^0_{(2)}$ is the $(d_2 - d_{20}) \times 1$ vector of zeros. Their corresponding covariates are given as

$$\begin{aligned} X_{ij} &= \left[\left\{\left(X^{(1)}_{ij}\right)_{p_1\times 1}\right\}^{\mathrm{T}}, \left\{\left(X^{(2)}_{ij}\right)_{(p-p_1)\times 1}\right\}^{\mathrm{T}}\right]^{\mathrm{T}}, \\ Z^{(2)}_{ij} &= \left[\left\{\left(Z^{(21)}_{ij}\right)_{d_{20}\times 1}\right\}^{\mathrm{T}}, \left\{\left(Z^{(22)}_{ij}\right)_{(d_2-d_{20})\times 1}\right\}^{\mathrm{T}}\right]^{\mathrm{T}} \end{aligned}$$

To find the penalized parametric estimators, we propose the penalized-QIF,

$$\mathcal{L}^*_n(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \frac{1}{2} Q^*_n(\boldsymbol{\beta}, \boldsymbol{\alpha}) + n \sum\nolimits_{s=2}^{p} p_{\lambda_{n1}}(|\beta_s|) + n \sum\nolimits_{l=d_1+1}^{d} p_{\lambda_{n2}}(|\alpha_l|), \tag{4.12}$$

where $Q^*_n(\boldsymbol{\beta}, \boldsymbol{\alpha})$ is the unpenalized objective function defined in (4.9) and $p_{\lambda_n}(\cdot)$ is a penalty function with a regularization parameter λ_n. There are various penalty functions available in the literature, such as the L_1 and L_2 penalties, which yield the LASSO-type in Tibshirani (1996) and ridge-type estimators in Goldstein and Smith (1974), respectively. Here, we consider the smoothly clipped absolute deviation (SCAD) penalty proposed by Fan and Li (2001), whose first derivative is defined as

$$p'_\lambda(\theta) = \lambda \left\{ I(\theta \le \lambda) + \frac{(a\lambda - \theta)_+}{(a-1)\lambda} I(\theta > \lambda) \right\},$$

where $p_\lambda(0) = 0$, $a = 3.7$, and $(t)_+ = tI(t > 0)$. By minimizing $\mathcal{L}^*_n(\boldsymbol{\beta}, \boldsymbol{\alpha})$, we obtain the penalized-QIF estimators $\widehat{\boldsymbol{\beta}}^{\mathrm{PQIF}}_{-1} = \left(\left(\widehat{\boldsymbol{\beta}}^{\mathrm{PQIF}}_{(1),-1}\right)^{\mathrm{T}}, \left(\widehat{\boldsymbol{\beta}}^{\mathrm{PQIF}}_{(2)}\right)^{\mathrm{T}}\right)^{\mathrm{T}}$ of $\boldsymbol{\beta}_{-1} = \left(\left(\boldsymbol{\beta}_{(1),-1}\right)^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}}_{(2)}\right)^{\mathrm{T}}$

and $\widehat{\boldsymbol{\alpha}}^{\mathrm{PQIF}} = \left(\left(\widehat{\boldsymbol{\alpha}}_{(1)}^{\mathrm{PQIF}}\right)^{\mathrm{T}}, \left(\widehat{\boldsymbol{\alpha}}_{(2)}^{\mathrm{PQIF}}\right)^{\mathrm{T}}\right)^{\mathrm{T}}$ of $\boldsymbol{\alpha} = \left(\ \boldsymbol{\alpha}_{(1)}^{\mathrm{T}}, \boldsymbol{\alpha}_{(2)}^{\mathrm{T}}\right)^{\mathrm{T}}$.

To study asymptotic properties of penalized estimators, we follow the same approach for obtaining $\widehat{X}_{ij}$ and $\widehat{Z}_{ij}^{(2)}$ in Section 5.2 to get $\widehat{X}_{ij}^{(1)}$ and $\widehat{Z}_{ij}^{(21)}$. Let $\mathbf{D}_{1i}\left(\boldsymbol{\beta}_{(1)}^{0}\right) = \left(D_{1,i1}\left(\boldsymbol{\beta}_{(1)}^{0}\right), \ldots, D_{1,im}\left(\boldsymbol{\beta}_{(1)}^{0}\right)\right)^{\mathrm{T}}$ and $\Delta_{1i}\left(\boldsymbol{\beta}_{(1)}^{0},\ \boldsymbol{\alpha}_{(1)}^{0}\right) = \mathrm{Diag}(\widehat{\nu}_{1,ij}(\boldsymbol{\beta}_{(1)}^{0},\ \boldsymbol{\alpha}_{(1)}^{0}))$, where

$$
\begin{aligned}
D_{1,ij}\left(\boldsymbol{\beta}_{(1)}^{0}\right) &= \left\{g'\left(\boldsymbol{\beta}_{(1)}^{0\mathrm{T}} X_{ij}^{(1)}\right) \widehat{X}_{ij}^{(1)\mathrm{T}} \mathbf{J}_{1}^{0}, \widehat{Z}_{ij}^{(21)\mathrm{T}}\right\}^{\mathrm{T}}, \\
\widehat{\mu}_{1,ij}\left(\boldsymbol{\beta}_{(1)}^{0},\ \boldsymbol{\alpha}_{(1)}^{0}\right) &= \vartheta^{-1}\left\{\widehat{\eta}_{1,ij}\left(\boldsymbol{\beta}_{(1)}^{0},\ \boldsymbol{\alpha}_{(1)}^{0}\right)\right\}, \\
\widehat{\nu}_{1,ij}(\boldsymbol{\beta}_{(1)}^{0},\ \boldsymbol{\alpha}_{(1)}^{0}) &= \partial \widehat{\mu}_{1,ij}(\boldsymbol{\beta}_{(1)}^{0},\ \boldsymbol{\alpha}_{(1)}^{0}) / \partial \widehat{\eta}_{1,ij}(\boldsymbol{\beta}_{(1)}^{0},\ \boldsymbol{\alpha}_{(1)}^{0}), \\
\widehat{\eta}_{1,ij}(\boldsymbol{\beta}_{(1)}^{0},\ \boldsymbol{\alpha}_{(1)}^{0}) &= \widetilde{g}\left(\ \boldsymbol{\beta}_{(1)}^{0\mathrm{T}} X_{ij}^{(1)}\right) + \sum\nolimits_{l=1}^{d_1} \widetilde{\alpha}_{l}\left(T_{ij}\right) Z_{ij,l} + \ \boldsymbol{\alpha}_{(1)}^{0\mathrm{T}} Z_{ij}^{(21)},
\end{aligned}
$$

and

$$
\mathbf{J}_{1}^{0} = \begin{pmatrix} -\boldsymbol{\beta}_{(1),-1}^{0\mathrm{T}} / \sqrt{1 - \left\|\boldsymbol{\beta}_{(1),-1}^{0}\right\|^{2}} \\ \mathbf{I}_{d_{10}-1} \end{pmatrix}.
$$

In addition, let $\dot{\psi}_{n1}\left(\boldsymbol{\beta}_{(1)}^{0}, \boldsymbol{\alpha}_{(1)}^{0}\right)$ and $\Psi_{n1}\left(\boldsymbol{\beta}_{(1)}^{0}, \boldsymbol{\alpha}_{(1)}^{0}\right)$ be defined in the same manner as $\dot{\psi}_{n}\left(\ \boldsymbol{\beta}^{\mathbf{0}}, \boldsymbol{\alpha}^{\mathbf{0}}\right)$ and $\Psi_{n}\left(\boldsymbol{\beta}^{\mathbf{0}}, \boldsymbol{\alpha}^{\mathbf{0}}\right)$ in Section 5.2 by replacing their $\mathbf{D}_{i}\left(\boldsymbol{\beta}^{\mathbf{0}}\right)$ and $\Delta_{i}\left(\ \boldsymbol{\beta}^{\mathbf{0}}, \boldsymbol{\alpha}^{\mathbf{0}}\right)$ with $\mathbf{D}_{1i}\left(\ \boldsymbol{\beta}_{(1)}^{0}\right)$ and $\Delta_{1i}\left(\boldsymbol{\beta}_{(1)}^{0},\ \boldsymbol{\alpha}_{(1)}^{0}\right)$, respectively. Then, we establish the following oracle properties of the penalized estimators.

Theorem 3 *Assume that $N^{4} n^{-1} = o\left(1\right)$, $N^{-4r+2} n = o\left(1\right)$ with $r > 3/2$ in Condition (C2), and the tuning parameters satisfy $\lambda_{n1} \to 0$, $\lambda_{n2} \to 0$, $n^{1/2}\lambda_{n1} \to \infty$ and $n^{1/2}\lambda_{n2} \to \infty$. Then, under Conditions (C1)-(C5), the penalized estimators satisfy:*
(1) (sparsity) $P\left(\left\{\left(\widehat{\boldsymbol{\beta}}_{(2)}^{PQIF}\right)^{T}, \left(\widehat{\boldsymbol{\alpha}}_{(2)}^{PQIF}\right)^{T}\right\}^{T} = \mathbf{0}\right) \to 1$; and (2) (asymptotic normality)

$$
\sqrt{n}\left(\Sigma_{n1}^{(1)}\right)^{1/2}\left\{\left\{\left(\widehat{\boldsymbol{\beta}}_{(1),-1}^{PQIF}\right)^{T}, \left(\widehat{\boldsymbol{\alpha}}_{(1)}^{PQIF}\right)^{T}\right\}^{T} - \left(\boldsymbol{\beta}_{(1),-1}^{0T}, \boldsymbol{\alpha}_{(1)}^{0T}\right)^{T}\right\} \to N\left(0, \mathbf{I}_{(p_{1}+d_{20}-1)}\right),
$$

where $\Sigma_{n1}^{(1)} = \dot{\psi}_{n1}\left(\boldsymbol{\beta}_{(1)}^{0}, \boldsymbol{\alpha}_{(1)}^{0}\right)^{T} \Psi_{n1}\left(\boldsymbol{\beta}_{(1)}^{0}, \boldsymbol{\alpha}_{(1)}^{0}\right) \ \dot{\psi}_{n1}\left(\boldsymbol{\beta}_{(1)}^{0}, \boldsymbol{\alpha}_{(1)}^{0}\right)$.

Let $\boldsymbol{\Upsilon}_1 = \begin{pmatrix} \mathbf{J}^{\mathbf{0}}_{\mathbf{1}_{p_1 \times (p_1-1)}} & \mathbf{0}_{p_1 \times d_{20}} \\ \mathbf{0}_{d_{20} \times (p_1-1)} & \mathbf{I}_{d_{20} \times d_{20}} \end{pmatrix}$. The above theorem, together with the multivariate delta-method, leads to the asymptotic normality of penalized parametric estimators, $\sqrt{n}\Sigma_{n1}^{1/2}\left(\left\{\left(\widehat{\boldsymbol{\beta}}_{(1)}^{\mathrm{PQIF}}\right)^{\mathrm{T}}, \left(\widehat{\boldsymbol{\alpha}}_{(1)}^{\mathrm{PQIF}}\right)^{\mathrm{T}}\right\}^{\mathrm{T}} - \left(\boldsymbol{\beta}_{(1)}^{0\mathrm{T}}, \boldsymbol{\alpha}_{(1)}^{0\mathrm{T}}\right)^{\mathrm{T}}\right) \to N\left(0, \mathbf{I}_{p_1+d_{20}}\right)$, as $n \to \infty$, where $\Sigma_{n1} = \boldsymbol{\Upsilon}_1 \Sigma_{n1}^{(1)} \boldsymbol{\Upsilon}_1^{\mathrm{T}}$.

We next study the asymptotic properties of the penalized nonparametric estimators. To this end, assume that $\alpha_l(\cdot) \equiv 0$ for $(d_{10}+1) \le l \le d_1$ in the true model. By the density assumption of T_{ij} in Condition (C1) of the Appendix, we obtain that $\widetilde{\alpha}_l(\cdot) = 0$ if and only if $E\left\{\widetilde{\alpha}_l(T_{ij})^2\right\} = 0$. In addition, $\alpha_l(t) \approx \widetilde{\alpha}_l(t) = B_2(t)^{\mathrm{T}} \boldsymbol{\gamma}_l$. This motivates us to consider the empirical L_2 norm as a metric, that is, $\|\widetilde{\alpha}_l\| = \|\boldsymbol{\gamma}_l\|_{\mathbf{W}_n} = \left(\boldsymbol{\gamma}_l^{\mathrm{T}} \mathbf{W}_n \boldsymbol{\gamma}_l\right)^{1/2}$, where $\mathbf{W}_n = n_{\mathrm{T}}^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} B_2(T_{ij}) B_2(T_{ij})^{\mathrm{T}}$ and $n_{\mathrm{T}} = nm$. Using this metric and replacing $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ in $Q_n(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\alpha})$ by its $\sqrt{n}$ consistent estimator $(\widehat{\boldsymbol{\beta}}^*, \widehat{\boldsymbol{\alpha}}^*)$ (e.g., $(\widehat{\beta}^{\mathrm{PQIF}}, \widehat{\alpha}^{\mathrm{PQIF}})$), we adopt Wang (2007) group-penalized approach and propose the following penalized-QIF for spline coefficients,

$$\mathcal{L}_n(\boldsymbol{\gamma}) = \frac{1}{2} Q_n\left(\boldsymbol{\gamma}, \widehat{\boldsymbol{\beta}}^*, \ \widehat{\boldsymbol{\alpha}}^*\right) + n \sum_{l=1}^{d_1} p_{\lambda_{n3}}\left(\|\boldsymbol{\gamma}_l\|_{\mathbf{W}_n}\right). \tag{4.13}$$

The resulting penalized estimator of $\boldsymbol{\gamma}$ is

$$\widehat{\boldsymbol{\gamma}}^{\mathrm{PQIF}} = \left\{\left(\ \widehat{\boldsymbol{\gamma}}_l^{\mathrm{PQIF}}\right)^{\mathrm{T}}, 0 \le l \le d_1\right\}^{\mathrm{T}} = \arg\min_{\boldsymbol{\gamma}} \left(\mathcal{L}_n(\ \boldsymbol{\gamma})\right).$$

Subsequently, we obtain the estimators of $g(u)$ and $\alpha_l(t)$, $\widehat{g}^{\mathrm{PQIF}}(u) = B_1(u)^{\mathrm{T}} \widehat{\boldsymbol{\gamma}}_0^{\mathrm{PQIF}}$ and $\widehat{\alpha}_l^{\mathrm{PQIF}}(t) = B_2(t)^{\mathrm{T}} \ \widehat{\boldsymbol{\gamma}}_l^{\mathrm{PQIF}}$. Then, we demonstrate the following asymptotic properties of nonparametric estimators.

Theorem 4 *Assume that* $\lambda_{n3} \to 0$ *and* $\lambda_{n3} n^{r/(2r+1)} \to \infty$ *with* $r > 3/2$ *in Condition (C2). Then, under Conditions (C1)-(C5),* $\widehat{\boldsymbol{\gamma}}^{PQIF}$ *satisfies (1) (sparsity)* $P(\widehat{\gamma}_l^{PQIF} = 0) \to 1$ *for* $d_{10} + 1 \leq l \leq d_1$*; and (2) (*$L_2$ *rate of convergence)* $\left\|\widehat{g}^{PQIF}(\cdot) - g(\cdot)\right\| = O_p\left(N^{-r/(2r+1)}\right)$ *and* $\left\|\widehat{\alpha}_l^{PQIF}(\cdot) - \alpha_l(\cdot)\right\| = O_p\left(N^{-r/(2r+1)}\right)$ *for* $1 \leq l \leq d_{10}$*, where* $N \asymp n^{1/(2r+1)}$.

Theorem 4 indicates that, under some regularity conditions, the penalized-QIF estimator has the same optimal convergence rate as the unpenalized estimator. In addition, the penalized procedure is able to correctly select relevant B-spline coefficients with probability approaching 1.

4.4.2 Estimation algorithm

The algorithm for obtaining unpenalized estimators is a special case of the procedure to calculate penalized estimators. Hence, we only focus on the penalized estimates. To this end, we consider three possible scenarios: (i.) $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are penalized, while $\boldsymbol{\gamma}$ is unpenalized; (ii.) $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are unpenalized, but $\boldsymbol{\gamma}$ is penalized; (iii.) $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\gamma}$ are penalized. In the first scenario, let $\left(\widehat{\boldsymbol{\beta}}^i, \widehat{\boldsymbol{\alpha}}^i\right)$ and $\widehat{\boldsymbol{\gamma}}^i$ be the i-th iterative estimators of $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ and $\boldsymbol{\gamma}$, respectively. For given $\left(\widehat{\boldsymbol{\beta}}^i, \widehat{\boldsymbol{\alpha}}^i\right)$, we employ (4.7) to obtain the estimator $\widehat{\boldsymbol{\gamma}}^{i+1}$ of $\boldsymbol{\gamma}$ at the $(i+1)^{\text{th}}$ step. That is,

$$\widehat{\boldsymbol{\gamma}}^{i+1} = \widehat{\boldsymbol{\gamma}}^i - \ddot{\widetilde{Q}}_n\left(\widehat{\boldsymbol{\gamma}}^i, \widehat{\boldsymbol{\beta}}^i, \widehat{\boldsymbol{\alpha}}^i\right)^{-1} \dot{\widetilde{Q}}_n\left(\widehat{\boldsymbol{\gamma}}^i, \widehat{\boldsymbol{\beta}}^i, \widehat{\boldsymbol{\alpha}}^i\right), \tag{4.14}$$

where $\dot{\widetilde{Q}}_n(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\alpha}) = \partial \widetilde{Q}_n(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\alpha}) / \partial \boldsymbol{\gamma}$ and $\ddot{\widetilde{Q}}_n(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\alpha}) = \partial^2 \widetilde{Q}_n(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\alpha}) / \partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^{\mathrm{T}}$.

Based on $\widehat{\boldsymbol{\gamma}}^{i+1}$, we next obtain the $(i+1)$-th iterative estimators of $(\boldsymbol{\beta}, \boldsymbol{\alpha})$. To this end, we use the fact that $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ is a function of γ and then denote $Q_n^*(\boldsymbol{\beta}, \boldsymbol{\alpha})$ in (4.9) as $Q_n^*\left(\widehat{\boldsymbol{\beta}}^i,\widehat{\boldsymbol{\alpha}}^i,\widehat{\boldsymbol{\gamma}}^{i+1}\right)$ and its associated component $\widehat{\mathbf{D}}_i\left(\boldsymbol{\beta},\boldsymbol{\alpha}\right)$ results to

$$
\begin{aligned}
&\widehat{\mathbf{D}}_i\left(\widehat{\boldsymbol{\beta}}^i,\widehat{\boldsymbol{\alpha}}^i,\widehat{\boldsymbol{\gamma}}^{i+1}\right)\\
= &\left[\left\{B_1'\left(X_{i1}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}^i\right)^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_0^{i+1}\widehat{X}_{i1},\ldots,B_1'\left(X_{im}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}^i\right)^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_0^{i+1}\widehat{X}_{im}\right\}^{\mathrm{T}}\mathbf{J}^{\mathbf{0}},\left(\widehat{Z}_{i1}^{(2)},\ldots,\widehat{Z}_{im}^{(2)}\right)^{\mathrm{T}}\right].
\end{aligned}
$$

For the sake of simplicity, let $\theta=\left(\boldsymbol{\beta}_{-1}^{\mathrm{T}},\boldsymbol{\alpha}^{\mathrm{T}}\right)^{\mathrm{T}}$, and denote $\dot{Q}_n^*\left(\boldsymbol{\beta},\boldsymbol{\alpha},\widehat{\gamma}\right)=\partial Q_n^*\left(\boldsymbol{\beta},\boldsymbol{\alpha},\widehat{\gamma}\right)/\partial\boldsymbol{\theta}$ and $\ddot{Q}_n^*\left(\boldsymbol{\beta},\boldsymbol{\alpha},\widehat{\gamma}\right)=\partial Q_n^*\left(\boldsymbol{\beta},\boldsymbol{\alpha},\widehat{\gamma}\right)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}$.

To obtain the penalized estimate of $\boldsymbol{\theta}$, we adopt the approach of Fan and Li (2001) and obtain the locally quadratic approximation of $2\mathcal{L}_n^*\left(\boldsymbol{\beta}^{i+1},\boldsymbol{\alpha}^{i+1},\widehat{\boldsymbol{\gamma}}^{i+1}\right)$ in (4.12) as follows:

$$
\begin{aligned}
&Q_n^*\left(\widehat{\boldsymbol{\beta}}^i,\widehat{\boldsymbol{\alpha}}^i,\widehat{\boldsymbol{\gamma}}^{i+1}\right)+\dot{Q}_n^*\left(\widehat{\boldsymbol{\beta}}^i,\widehat{\boldsymbol{\alpha}}^i,\widehat{\boldsymbol{\gamma}}^{i+1}\right)^{\mathrm{T}}\left(\boldsymbol{\theta}^{i+1}-\widehat{\boldsymbol{\theta}}^i\right)\\
&+2^{-1}\left(\boldsymbol{\theta}^{i+1}-\widehat{\boldsymbol{\theta}}^i\right)^{\mathrm{T}}\ddot{Q}_n^*\left(\widehat{\boldsymbol{\beta}}^i,\widehat{\boldsymbol{\alpha}}^i,\widehat{\boldsymbol{\gamma}}^{i+1}\right)\left(\boldsymbol{\theta}^{i+1}-\widehat{\boldsymbol{\theta}}^i\right)^{\mathrm{T}}\\
&+2n\sum\nolimits_{s=2}^{p}p_{\lambda_{n1}}\left(\left|\widehat{\beta}_s^i\right|\right)+n\left(\boldsymbol{\beta}_{-1}^{i+1}\right)^{\mathrm{T}}\Phi_{\lambda_{n1}}\left(\widehat{\boldsymbol{\beta}}_{-1}^i\right)\boldsymbol{\beta}_{-1}^{i+1}-n\left(\widehat{\boldsymbol{\beta}}_{-1}^i\right)^{\mathrm{T}}\Phi_{\lambda_{n1}}\left(\widehat{\boldsymbol{\beta}}_{-1}^i\right)\widehat{\boldsymbol{\beta}}_{-1}^i\\
&+2n\sum\nolimits_{l=d_1+1}^{d}p_{\lambda_{n2}}\left(\left|\widehat{\alpha}_l^i\right|\right)+n\left(\boldsymbol{\alpha}^{i+1}\right)^{\mathrm{T}}\Phi_{\lambda_{n2}}\left(\widehat{\boldsymbol{\alpha}}^i\right)\boldsymbol{\alpha}^{i+1}-n\left(\widehat{\boldsymbol{\alpha}}^i\right)^{\mathrm{T}}\Phi_{\lambda_{n2}}\left(\widehat{\boldsymbol{\alpha}}^i\right)\widehat{\boldsymbol{\alpha}}^i,
\end{aligned}
$$

where

$$
\begin{aligned}
\Phi_{\lambda_{n1}}\left(\boldsymbol{\beta}_{-1}\right) &= \mathrm{diag}\left\{p_{\lambda_{n1}}'\left(|\beta_2|\right)/|\beta_2|,\ldots,p_{\lambda_{n1}}'\left(|\beta_p|\right)/|\beta_p|\right\},\\
\Phi_{\lambda_{n2}}\left(\boldsymbol{\alpha}\right) &= \mathrm{diag}\left\{p_{\lambda_{n2}}'\left(|\alpha_{d_1+1}|\right)/|\alpha_{d_1+1}|,\ldots,p_{\lambda_{n2}}'\left(|\alpha_d|\right)/|\alpha_d|\right\}.
\end{aligned}
$$

Minimizing the above function with respect to θ^{i+1}, we obtain that

$$
\widehat{\boldsymbol{\theta}}^{i+1}=\widehat{\boldsymbol{\theta}}^i-\left\{\ddot{Q}_n^*\left(\widehat{\boldsymbol{\beta}}^i,\widehat{\boldsymbol{\alpha}}^i,\widehat{\boldsymbol{\gamma}}^{i+1}\right)+2n\Phi\left(\widehat{\boldsymbol{\theta}}^i\right)\right\}^{-1}\left\{\dot{Q}_n^*\left(\widehat{\boldsymbol{\beta}}^i,\widehat{\boldsymbol{\alpha}}^i,\widehat{\boldsymbol{\gamma}}^{i+1}\right)+2n\Phi\left(\widehat{\boldsymbol{\theta}}^i\right)\widehat{\boldsymbol{\theta}}^i\right\}, \quad (4.15)
$$

where $\Phi\left(\boldsymbol{\theta}\right)=\begin{pmatrix}\Phi_{\lambda_{n1}}\left(\boldsymbol{\beta}^{(1)}\right) & \mathbf{0}_{(p-1)\times d_2}\\ \mathbf{0}_{d_2\times(p-1)} & \Phi_{\lambda_{n2}}\left(\boldsymbol{\alpha}\right)\end{pmatrix}$. Subsequently, $\widehat{\beta}_1^{i+1}=\left(1-\left\|\widehat{\boldsymbol{\beta}}_{-1}^{i+1}\right\|^2\right)^{1/2}$. If the i-th iterative penalized estimate $\widehat{\beta}_s^i$ is close to zero (i.e., $|\widehat{\beta}_s^i|<\epsilon_1^*$ for a small threshold value ϵ_1^*), we set $\widehat{\beta}_s^{i+1} = 0$. The iteration is stopped at the $(i+1)^{\text{th}}$ step if $\left\|\widehat{\boldsymbol{\theta}}^{i+1} - \widehat{\boldsymbol{\theta}}^{i}\right\| < \delta_1^*$ and $\left\|\widehat{\boldsymbol{\gamma}}^{i+1} - \widehat{\boldsymbol{\gamma}}^{i}\right\| < \delta_1^*$ for a small threshold value δ_1^*. Accordingly, the penalized estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are $\widehat{\boldsymbol{\beta}}^{\text{PQIF}} = \widehat{\boldsymbol{\beta}}^{i+1}$ and $\widehat{\boldsymbol{\alpha}}^{\text{PQIF}} = \widehat{\boldsymbol{\alpha}}^{i+1}$. It is noteworthy that unpenalized QIF estimators $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\alpha}}$, and $\widehat{\boldsymbol{\gamma}}$ can be obtained iteratively from equations (4.14) and (4.15) by setting $\Phi(\widehat{\boldsymbol{\theta}}^{i}) = 0$ in (4.15).

In the second scenario, we can show that the unpenalized QIF estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\alpha}}$ are $\sqrt{n}$ - consistent. Hence, we use them to replace $\widehat{\boldsymbol{\beta}}^*$ and $\widehat{\boldsymbol{\alpha}}^*$ in equation (4.13), and then employ the same techniques as those used for obtaining equation (4.15) to yield the penalized estimator $\widehat{\boldsymbol{\gamma}}^{\text{PQIF},i+1}$ at the $(i+1)^{\text{th}}$ step given below.

$$\begin{aligned}\widehat{\boldsymbol{\gamma}}^{\text{PQIF},i+1} \quad = \quad & \widehat{\boldsymbol{\gamma}}^{\text{PQIF},i} - \left\{\ddot{Q}_n\left(\widehat{\boldsymbol{\gamma}}^{\text{PQIF},i},\widehat{\boldsymbol{\beta}}^{\text{QIF}},\widehat{\boldsymbol{\alpha}}^{\text{QIF}}\right) + 2n\Phi_{\lambda_{n3}}\left(\widehat{\boldsymbol{\gamma}}^{\text{PQIF},i}\right)\right\}^{-1} \times \\ & \left\{\dot{Q}_n\left(\widehat{\boldsymbol{\gamma}}^{\text{PQIF},i},\widehat{\boldsymbol{\beta}}^{\text{QIF}},\widehat{\boldsymbol{\alpha}}^{\text{QIF}}\right) + 2n\Phi_{\lambda_{n3}}\left(\widehat{\boldsymbol{\gamma}}^{\text{PQIF},i}\right)\widehat{\boldsymbol{\gamma}}^{\text{PQIF},i}\right\}, \qquad (4.16)\end{aligned}$$

where

$$\Phi_{\lambda_{n_3}}(\boldsymbol{\gamma}) = \text{diag}\left\{p'_{\lambda_{n3}}\left(\|\boldsymbol{\gamma}_1\|_{\mathbf{W}n}\right)/\|\boldsymbol{\gamma}_1\|_{\mathbf{W}n}, \ldots, p'_{\lambda_{n3}}\left(\|\boldsymbol{\gamma}_p\|_{\mathbf{W}n}\right)/\|\boldsymbol{\gamma}_p\|_{\mathbf{W}n}\right\}.$$

If the i-th iterative penalized estimator $\widehat{\boldsymbol{\gamma}}^{\text{PQIF},i}$ is close to zero (i.e., $\left\|\widehat{\boldsymbol{\gamma}}^{\text{PQIF},i}\right\|_{\mathbf{W}n} < \epsilon_2^*$ for a small threshold value ϵ_2^*), we set $\widehat{\boldsymbol{\gamma}}^{\text{PQIF},i+1} = \mathbf{0}$. The iteration stops when

$$\left\|\widehat{\boldsymbol{\gamma}}^{\text{PQIF},i+1} - \widehat{\boldsymbol{\gamma}}^{\text{PQIF},i}\right\| < \delta_2^*$$

for a small threshold value δ_2^*, which leads to $\widehat{\boldsymbol{\gamma}}^{\text{PQIF}} = \widehat{\boldsymbol{\gamma}}^{\text{PQIF},i+1}$.

In the third scenario, we are able to demonstrate that the penalized estimators, $\widehat{\boldsymbol{\beta}}^{\text{PQIF}}$ and $\widehat{\boldsymbol{\alpha}}^{\text{PQIF}}$, obtained from the first scenario are consistent. Thus, we substitute $\widehat{\boldsymbol{\beta}}^*$

and $\widehat{\boldsymbol{\alpha}}^*$ in equation (4.13) with these estimators. Afterwards, we adopt the same procedure as given in equation (4.16) by replacing its $\widehat{\boldsymbol{\beta}}^{\mathrm{QIF}}$ and $\widehat{\boldsymbol{\alpha}}^{\mathrm{QIF}}$ with $\widehat{\boldsymbol{\beta}}^{\mathrm{PQIF}}$ and $\widehat{\boldsymbol{\alpha}}^{\mathrm{PQIF}}$, respectively, to obtain $\widehat{\boldsymbol{\gamma}}^{\mathrm{PQIF}}$.

To facilitate computations, we recommend using the unpenalized estimators as initial estimators in the iterative equations, (4.14), (4.15), and (4.16). It is worth noting that the tuning parameters are unknown in those equations, and we adapt Wang (2007)'s BIC criterion to choose the tuning parameters λ_{n1}, λ_{n2} and λ_{n3} in the penalized-QIF procedure. They are

$$\begin{aligned}\mathrm{BIC}\left(\lambda_{n1},\lambda_{n2}\right) &= \mathcal{L}_n^*\left(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\alpha}}\right)+\log\left(n\right)\times\left(\widehat{p}_1-1+\widehat{d}_{21}\right) \quad \text{and} \\ \mathrm{BIC}\left(\lambda_{n3}\right) &= \mathcal{L}_n\left(\widehat{\boldsymbol{\gamma}}\right)+\log\left(n\right)\times\left\{J_n\left(1+\widehat{d}_{10}\right)\right\},\end{aligned}$$

where $\widehat{d}_{21}$ and $\widehat{p}_1$ are the number of nonzero components in $\widehat{\boldsymbol{\alpha}}^{\mathrm{PQIF}}$ and $\widehat{\boldsymbol{\beta}}^{\mathrm{PQIF}}$, and $\widehat{d}_{10}$ is the number of nonzero estimated functions $\widehat{\alpha}_l^{\mathrm{PQIF}}\left(\cdot\right)$. Accordingly,

$$\left(\widehat{\lambda}_{n1},\widehat{\lambda}_{n2}\right)=\arg\min_{\left(\lambda_{n1},\lambda_{n2}\right)} BIC\left(\lambda_{n1},\lambda_{n2}\right)$$

and $\widehat{\lambda}_{n3}=\arg\min_{\lambda_{n3}}\mathrm{BIC}(\lambda_{n3})$. In our numerical studies given below, we use cubic splines with $q=4$ to estimate nonparametric functions. In addition, the number of interior knots is set at $N=\left[n^{1/(2q+1)}\right]+1$, which is of the optimal order and $[a]$ denotes the greatest integer less than or equal to a. In the empirical implementations, we use the minimal and maximal values of $X_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}$ and T_{ij} as the two boundary points to generate B-spline basis functions $B_{1,J}\left(u\right)$ and $B_{2,J}\left(t\right)$, respectively.

4.5 Numerical Examples

4.5.1 Simulation studies

In this subsection, we conduct two Monte Carlo studies to evaluate the finite sample performance of the proposed estimators. The first two example focuse on scenario (i) $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are penalized, while $\boldsymbol{\gamma}$ used for computing nonparametric functions is unpenalized. In contrast, the third example addresses scenario (ii) $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are unpenalized, but $\boldsymbol{\gamma}$ is penalized.

Example 1. Within each cluster, the correlated binary responses Y_{ij} are generated from a marginal logit model,

$$\text{logit}P\left(Y_{ij}=1\,|X_{ij},Z_{ij},T_{ij}\right)=g\left(X_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{0}\right)+\sum\nolimits_{l=1}^{2}\alpha_{l}\left(T_{ij}\right)Z_{ij,l}+\sum\nolimits_{l=3}^{6}\alpha_{l}^{0}Z_{ij,l}, \tag{4.17}$$

where $g\left(U\right)=0.5\cos(2\pi U)$, $\alpha_{1}\left(T\right)=0.1\cos(2\pi T)$, $\alpha_{2}\left(T\right)=0.1\sin(2\pi T)$, $\boldsymbol{\beta}^{0}=\frac{1}{\sqrt{22}}\left(3,3,2,0,0,0,0\right)^{\mathrm{T}}$, $\boldsymbol{\alpha}^{0}=\left(\alpha_{3},\alpha_{4,}\alpha_{5},\alpha_{6}\right)^{\mathrm{T}}=\left(-0.5,0,0,0.4\right)^{\mathrm{T}}$, $j=1,\ldots,5$, $i=1,\ldots,n$, and $n=200$ and 500. We then use the algorithm in [15] to generate correlated binary responses with an exchangeable correlation structure and the correlation parameter is 0.3 within each cluster. Furthermore, covariates $X_{ij}=\left(X_{ij,1}\ldots,X_{ij,7}\right)^{\mathrm{T}}$ are independently generated from uniform$[0,1]$, T_{ij} are randomly simulated from uniform$[0,1]$, and $\left(Z_{ij,1}\ldots,Z_{ij,6}\right)^{\mathrm{T}}$ are independently generated from $N\left(0,0.5^{2}\right)$. To assess the robustness of covariance setting, we consider three different working correlation structures: independent (IND), exchangeable (EX), and AR(1), although the data are simulated from the exchangeable setting.

To examine the selection performance of parametric components, we conduct 200 realizations and report the proportions of parameters correctly fitted (C), overfitted (O), and underfitted (U) as well as the average of true positives (TP), i.e., the average number of covariates being correctly selected from all possible candidates, and the average number of false positives (FP), i.e., the average number of covariates being incorrectly selected from all possible candidates. To evaluate the estimation accuracy, we compare the SCAD-penalized QIF (PQIF) estimate with the ORACLE estimate obtained by assuming that we know the zero components in $\boldsymbol{\beta}^0$ and $\boldsymbol{\alpha}^0$. The assessment measure is the median of squared errors (MSE) defined as the median of $\left\|\widehat{\boldsymbol{\beta}}_{(k)}^{\mathrm{PQIF}} - \boldsymbol{\beta}^0\right\|^2$ and the median of $\left\|\widehat{\boldsymbol{\alpha}}_{(k)}^{\mathrm{PQIF}} - \boldsymbol{\alpha}^0\right\|^2$ in 200 realizations, where $\widehat{\boldsymbol{\beta}}_{(k)}^{\mathrm{PQIF}}$ and $\widehat{\boldsymbol{\alpha}}_{(k)}^{\mathrm{PQIF}}$ are the PQIF estimates of $\boldsymbol{\beta}^0$ and $\boldsymbol{\alpha}^0$ calculated in the k^{th} realization.

Table 4.1: Variable selection and estimation results for $\boldsymbol{\beta}^0$ with the exchangeable (EX), AR(1), and independent (IND) working correlation structures in Example 1. The symbols C, O, and U denote the proportion of correct fitting, overfitting, and underfitting, respectively. The TP and FP denote the average number of true positives and false positives, respectively. The columns PQIF and ORACLE report the median of squared errors (MSEs) of the penalized and oracle estimates.

		Variable selection and parameter estimation						
n		C	O	U	TP	FP	PQIF	ORACLE
	EX	0.800	0.120	0.080	2.920	0.190	0.0209	0.0184
200	AR(1)	0.745	0.140	0.115	2.885	0.190	0.0231	0.0184
	IND	0.720	0.175	0.105	2.895	0.265	0.0302	0.0182
	EX	1.000	0.000	0.000	3.000	0.000	0.0062	0.0062
500	AR(1)	1.000	0.000	0.000	3.000	0.000	0.0062	0.0062
	IND	0.995	0.005	0.000	3.000	0.005	0.0063	0.0063

Tables 4.1 and 4.2 report variable selection and estimation results for $\boldsymbol{\beta}^0$ and $\boldsymbol{\alpha}^0$, respectively. Both tables show that the proportions of correctly fitted models increase and the proportions of overfitted and underfitted models decrease when the sample size

Table 4.2: Variable selection and estimation results for α^0 with the exchangeable (EX), AR(1), and independent (IND) working correlation structures in Example 1. The symbols C, O, and U denote the proportion of correct fitting, overfitting, and underfitting, respectively. The TP and FP denote the average numbers of true positives and false positives, respectively. The columns PQIF and ORACLE report the median of squared errors (MSEs) of the penalized and oracle estimates.

		Variable selection and parameter estimate						
n		C	O	U	TP	FP	PQIF	ORACLE
	EX	0.805	0.055	0.140	1.895	0.095	0.0285	0.0228
200	AR(1)	0.795	0.050	0.155	1.895	0.105	0.0291	0.0224
	IND	0.765	0.090	0.145	1.920	0.160	0.0274	0.0224
	EX	0.980	0.014	0.006	1.990	0.014	0.0103	0.0103
500	AR(1)	0.970	0.020	0.010	1.995	0.025	0.0107	0.0101
	IND	0.955	0.020	0.025	1.985	0.030	0.0108	0.0105

becomes larger. In addition, the number of true positives is closer to the correct number of nonzero parameters and the number of false positives decreases to zero, as the sample size increases. Moreover, the difference between the PQIF and ORACLE estimates measured by MSE becomes negligible as the sample size increases. The above findings support the theoretical results. It is noteworthy that the three working correlation structures yield similar performance, although EX is the correct structure. This indicates that the PQIF estimators are robust even though the working correlation is mis-specified.

To evaluate the performance of the estimates of the nonparametric functions, we next define the integrated squared error (ISE) of the estimated functions $\widehat{g}$, $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$, given as

$$\begin{aligned} \mathrm{ISE}\left(\widehat{g}\right) &= (nm)^{-1}\sum\nolimits_{i=1}^{n}\sum\nolimits_{j=1}^{m}\left\{\widehat{g}\left(X_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}^{\mathrm{PQIF}}\right)-g\left(X_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{0}\right)\right\}^{2}, \\ \mathrm{ISE}\left(\widehat{\alpha}_l\right) &= (nm)^{-1}\sum\nolimits_{i=1}^{n}\sum\nolimits_{j=1}^{m}\left\{\widehat{\alpha}_l\left(T_{ij}\right)-\alpha_l\left(T_{ij}\right)\right\}^{2}, l=1,2. \end{aligned}$$

When $n = 200$, the averages of the ISEs for $\widehat{g}$, $\widehat{\alpha}_1$, and $\widehat{\alpha}_2$ across 200 realizations are

0.100, 0.207 and 0.215, respectively. As the sample size increases to 500, the corresponding averages of the ISEs decrease to 0.02, 0.048 and 0.048, which corroborates the theoretical results in Theorem 2.

Example 2. This example addresses the case where the covariates are correlated and some are discrete. To this end, we generate the response observations using model (4.17) with the same true parameters, nonparametric functions, and distribution of variable T_{ij} as those given in Example 1. In addition, the covariates $(Z_{ij,1} \ldots, Z_{ij,7})^{\mathrm{T}}$ are simulated from a multivariate normal distribution with mean zero, marginal variance 0.5^2, and AR(1) correlation matrix with autocorrelation coefficient 0.3, while the covariate $Z_{ij,8}$ is generated from Bernoulli(0.5). Moreover, the covariates $(X_{ij,1} \ldots, X_{ij,7})^{\mathrm{T}}$ are simulated from the same distribution as that of $(Z_{ij,1} \ldots, Z_{ij,7})^{\mathrm{T}}$. To assess the robustness of estimates against the working correlation, we consider three different working correlation structures: independent (IND), exchangeable (EX), and AR(1), whereas the data are simulated from the exchangeable setting.

Table 4.3: Variable selection and estimation results for $\boldsymbol{\beta}^0$ and $\boldsymbol{\alpha}^0$ with the exchangeable (EX), AR(1), and independent (IND) working correlation structures in Example 2. The symbols C, O, and U denote the proportion of correct fitting, overfitting, and underfitting, respectively. The TP and FP denote the average number of true positives and false positives, respectively. The columns PQIF and ORACLE report the median of squared errors (MSEs) of the penalized and oracle estimates.

		Variable selection and parameter estimation						
		C	O	U	TP	FP	PQIF	ORACLE
	EX	0.975	0.002	0.023	2.973	0.009	0.0108	0.0104
$\boldsymbol{\beta}^0$	AR(1)	0.970	0.005	0.025	2.973	0.015	0.0136	0.0128
	IND	0.960	0.015	0.025	2.973	0.045	0.0170	0.0155
	EX	0.910	0.080	0.010	1.991	0.109	0.0132	0.0126
$\boldsymbol{\alpha}^0$	AR(1)	0.905	0.085	0.010	1.986	0.120	0.0145	0.0137
	IND	0.900	0.085	0.015	1.982	0.127	0.0168	0.0141

Tables 4.3 presents variable selection and estimation results for $\boldsymbol{\beta}^0$ and $\boldsymbol{\alpha}^0$ with $n = 500$ in 200 realizations. They indicate that the proportions of correctly fitted models are closer to one and the proportions of overfitted and underfitted models are closer to zero. In addition, the number of true positives is closer to the correct number of nonzero parameters and the number of false positives is small. Moreover, the MSE values of the PQIF and ORACLE estimates are similar, which confirms our theoretical results.

Example 3. In this example, within each cluster, the correlated binary responses Y_{ij} are generated from a marginal logit model,

$$\text{logit}P\left(Y_{ij} = 1 \,|X_{ij}, Z_{ij}, T_{ij}\right) = g\left(X_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0\right) + \sum\nolimits_{l=1}^{5} \alpha_l\left(T_{ij}\right) Z_{ij,l} + \sum\nolimits_{l=6}^{7} \alpha_l^0 Z_{ij,l},$$

where $g\left(U\right) = 0.5\cos(2\pi U)$, $\alpha_1\left(T\right) = 0.7\cos(2\pi T)$, $\alpha_2\left(T\right) = 0.7\sin(2\pi T)$, $\alpha_l\left(\cdot\right) = 0$ for $3 \le l \le 5$, $\boldsymbol{\beta}^0 = \frac{1}{\sqrt{22}}\left(3, 3, 2\right)^{\mathrm{T}}$, $\alpha^0 = \left(\alpha_6, \alpha_7\right)^{\mathrm{T}} = \left(-0.5, 0.4\right)^{\mathrm{T}}$, $j = 1, \ldots, 5$, $i = 1, \ldots, n$, and $n = 200$ and 500. In addition, the binary responses are generated from an exchangeable correlation structure with the correlation parameter 0.15. Moreover, covariates X_{ij}, T_{ij} and Z_{ij} are independently simulated from the same distributions as given in Example 1.

Table 4.4: Variable selection and estimation results for the varying coefficient functions $\alpha_l(T)$ with the exchangeable (EX), AR(1), and independent (IND) working correlation structures in Example 3. The symbols C, O, and U denote the proportion of correct-fitting, over-fitting, and under-fitting, respectively. The TP and FP denote the average of true positives and the average of false positives.

		Variable selection and estimation				
n		C	O	U	TP	FP
	EX	0.525	0.280	0.195	1.809	0.418
200	AR(1)	0.510	0.290	0.200	1.785	0.425
	IND	0.495	0.290	0.215	1.755	0.427
	EX	0.955	0.010	0.035	1.955	0.010
500	AR(1)	0.945	0.000	0.055	1.945	0.000
	IND	0.955	0.015	0.030	1.955	0.015

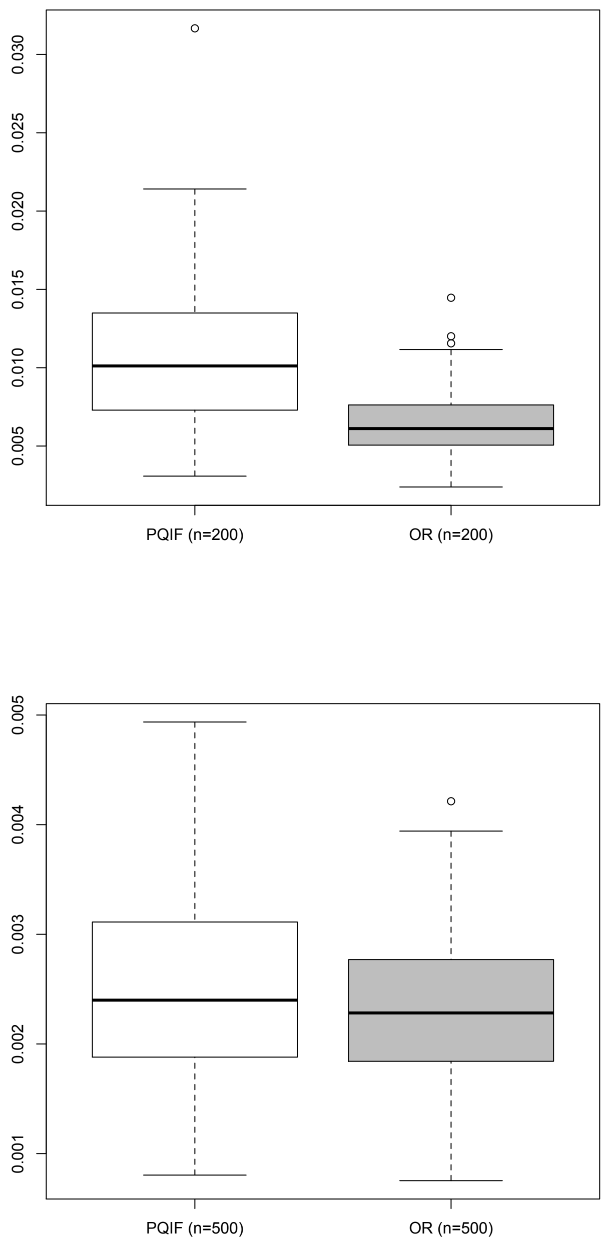
To assess the selection performance for varying coefficient components, we conduct 200 realizations. Table 4.4 reports the selection and estimation results for the varying coefficients with covariates $(Z_{ij,1}, \ldots, Z_{ij,5})^{\mathrm{T}}$. It shows that the proportions of correct fittings are close to 1 (above 95%) for all the three correlation structures for $n = 500$, while they are relatively low for $n = 200$. The high proportion of correct fitting in the large sample size corroborates the model selection consistency established in Theorem 4. In addition, the number of true positives gets closer to 2, and the number of false positives decreases to zero, as the sample size n increases.

Table 4.5: The average MSEs of the parameter estimates for $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^{\mathrm{T}}$ and $\alpha = (\alpha_6, \alpha_7)^{\mathrm{T}}$. and the empirical coverage probabilities (CP) of the 95% confidence intervals for parameters $(\beta_1, \beta_2, \beta_3)$ and (α_6, α_7) based on 200 realizations in Example 3.

		MSE		CP				
n		$\boldsymbol{\beta}$	$\boldsymbol{\alpha}$	β_1	β_2	β_3	α_6	α_7
	EX	0.0168	0.0267	0.855	0.865	0.835	0.915	0.935
200	AR(1)	0.0170	0.0269	0.855	0.825	0.865	0.915	0.925
	IND	0.0171	0.0271	0.865	0.865	0.860	0.910	0.940
	EX	0.0047	0.0073	0.955	0.935	0.925	0.920	0.940
500	AR(1)	0.0050	0.0078	0.955	0.920	0.920	0.915	0.945
	IND	0.0051	0.0078	0.960	0.935	0.925	0.920	0.950

In addition to varying coefficient components, we next study the performance of parametric components. Table 4.5 shows the MSEs of the parameter estimates and the empirical coverage probabilities of the 95% confidence intervals for the parametric components. All three working correlation structures result in similar average MSE values for both parameter estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. Furthermore, the MSE values decrease as n increases, which confirms the consistency property of the parameter estimates. Moreover, the empirical coverage probabilities get closer to the nominal coverage level of 95% as n increases, which corroborates the asymptotic normality of the parameter estimators. Next, we assess the overall model fitting. To this end, we define the model error (ME) as the average of the squared difference of the estimated and true conditional means of Y_{ij}. Figure 4.1 depicts the boxplots of the model errors by comparing the PQIF and oracle (OR) estimates, where OR is computed by assuming the true model is known *a priori*. It is not surprising that the model errors of the oracle estimates are smaller than those of the PQIF estimates. As the sample size gets large, however, the model errors of PQIF and OR are very similar. It is also noteworthy that the model errors are small even though $n = 200$, which demonstrates the accuracy of PQIF estimates.

Figure 4.1: Boxplots of the model errors calculated from the PQIF and oracle (OR) estimates with the EX, AR(1) and IND working correlation structures for $n = 200$ (top panel) and $n = 500$ (bottom panel).



Remark. To study the performance of the proposed estimation and selection methods in scenario (iii), we generate data from the following model:

$$\text{logit}P\left(Y_{ij}=1\,|X_{ij},Z_{ij},T_{ij}\right)=g\left(X_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{0}\right)+\sum\nolimits_{l=1}^{5}\alpha_{l}\left(T_{ij}\right)Z_{ij,l}+\sum\nolimits_{l=6}^{9}\alpha_{l}^{0}Z_{ij,l},$$

where $\boldsymbol{\beta}^0=\frac{1}{\sqrt{22}}\left(3,3,2,0,0,0,0\right)^{\mathrm{T}}$, $\boldsymbol{\alpha}^0=\left(\alpha_6,\alpha_7,\alpha_8,\alpha_9\right)^{\mathrm{T}}=\left(-0.5,0,0,0.4\right)^{\mathrm{T}}$, and $\alpha_l\left(T\right)$ are defined as Example 2 for $1\le l\le 5$. In addition, covariates X_{ij}, T_{ij} and Z_{ij} are independently simulated from the same distributions as given in Example 2, and Y_{ij} have the same correlation structure as given in Example 2. Since Monte Carlo results show similar findings as those in Examples 1 and 2, we do not present them here.

4.5.2 Empirical example

Following Klein (1984), we consider a data set from the Wisconsin epidemiologic study of diabetic retinopathy (WESDR). The aim of this study is to investigate the risk factors for diabetic retinopathy. The response is a binary variable indicating the presence of diabetic retinopathy in each of two eyes from 720 individuals in the study. In addition, the data set contains 13 risk factors including: *eye refractive error*, *eye intraocular pressure*, *age at diabetes diagnosis (years)*, *duration of diabetes (years)*, *glycosylated hemoglobin level*, *systolic blood pressure*, *diastolic blood pressure*, *body mass index*, *pulse rate (beats/30 seconds)*, *sex (male=1, female=2)*, *proteinuria (absent=0, present=1)*, *doses of insulin per day taken by the patient*, and *type of county of residence (urban=1, rural=2).*

Based on a preliminary fitting of the data to a logistic linear regression model, we found that there exist significant interaction effects between the logarithm of diabetes' duration, respectively, with *glycosylated hemoglobin level*, *systolic blood pressure*, and *dias-*

tolic blood pressure, where the logarithmic transformation of diabetes duration is used to amend its right skewness. This motivates us to consider $Z_{ij,1}$ = *glycosylated hemoglobin level*, $Z_{ij,2}$ = *systolic blood pressure*, and $Z_{ij,3}$ = *diastolic blood pressure* as the covariates associated with their corresponding varying coefficients $\alpha_l(T_{ij})$ $(l = 1, 2, 3)$, where T_{ij} = *logarithm of diabetes duration.* We then assign the rest of the continuous variables to be index covariates such that $X_{ij,1}$ = *age at diagnosis of diabetes* , $X_{ij,2}$ = *body mass index* , $X_{ij,3}$ = *eye refractive error*, $X_{ij,4}$ = *eye intraocular pressure*, and $X_{ij,5}$ = *pulse rate.* The remaining categorical variables, $Z_{ij,4}$ = *sex*, $Z_{ij,5}$ = *proteinuria*, $Z_{ij,6}$ = *doses of insulin*, and $Z_{ij,7}$ = *type of county of residence*, are used as the covariates in the linear part with constant coefficients. As a result, we fit the data with the following equation,

$$\eta_{ij} = \text{logit}(\mu_{ij}) = g\left(X_{ij,1}\beta_1 + \cdots + X_{ij,5}\beta_5\right) + \sum\nolimits_{l=1}^{3} \alpha_l\left(T_{ij}\right) Z_{ij,l} + \sum\nolimits_{l=4}^{7} \alpha_l Z_{ij,l}, \quad (4.18)$$

where $j = 1, 2$, $i = 1, ..., 720$. It is worth noting that we only consider IND and EX correlation structures since there are two repeated measurements for each subject and the results are the same for EX and AR(1) structures. In addition, all continuous variables are centered and standardized for parameter estimation.

By applying the penalized-QIF method in Section 4.4.1, two index variables ($X_{ij,1}$ = *age at diabetes diagnosis* and $X_{ij,2}$ = *body mass index*) and one linear variable ($Z_{ij,5}$ = *proteinuria*) are selected under the IND and EX working correlation structures. Table 4.6 presents the parameter estimates (EST) and their standard errors (SE) for the selected variables. The resulting Wald test statistics show that these variables are significant at the 5% level. Furthermore, the estimated coefficient of *proteinuri* (0.307 in IND and 0.311 in EX) indicates that the presence of diabetic retinopathy is approximately $\exp(0.31) = 1.35$

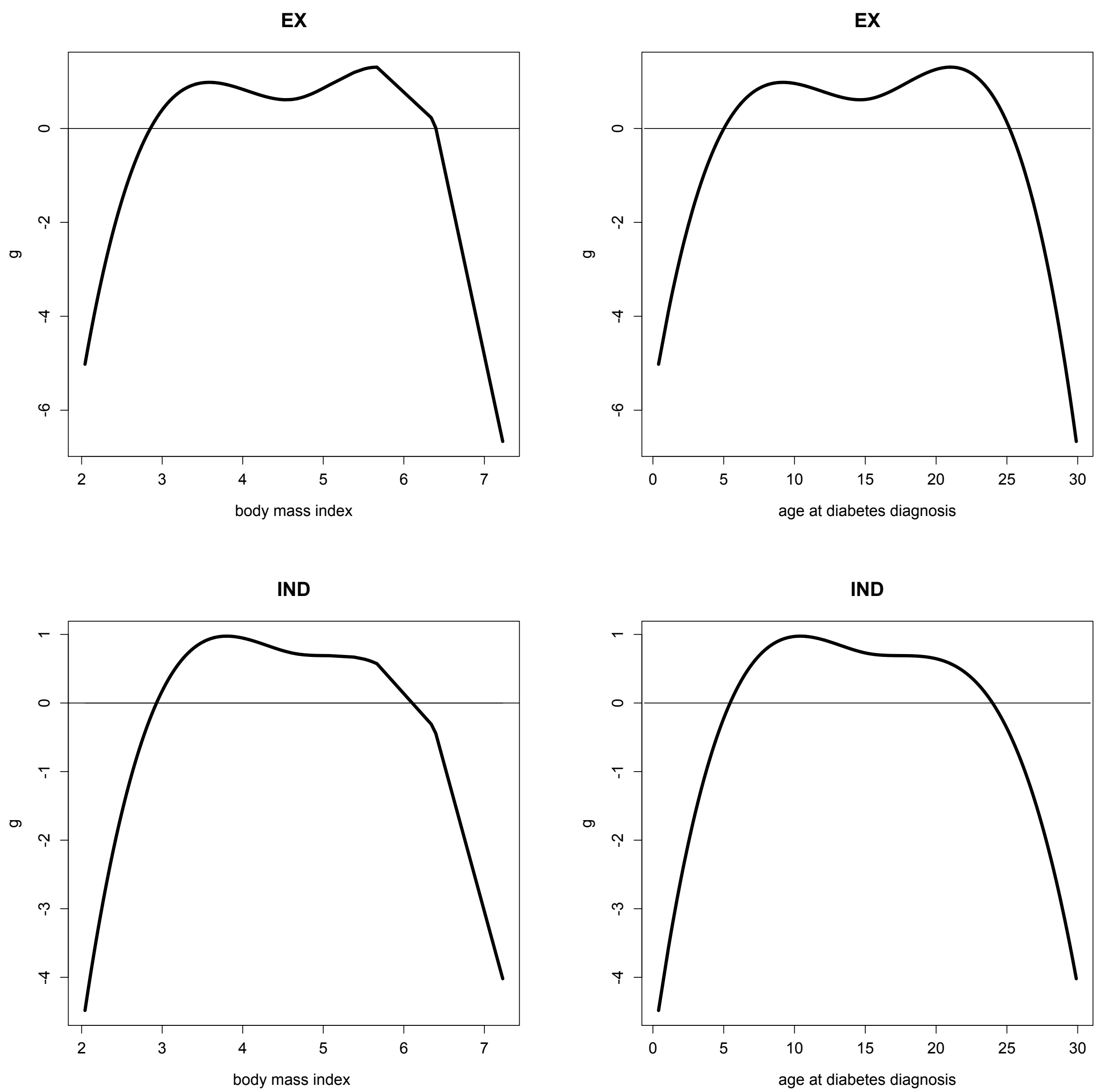
times as frequent among proteinuri than among no-proteinuri, after adjusting for the other variables in the model.

Table 4.6: The PQIF estimates (EST) and their associated standard errors (SE) of regression coefficients for the selected variables, respectively, under the IND and EX working correlation structures for the Wisconsin epidemiologic study.

		β_1	β_2	α_5
IND	EST	0.303	0.953	0.307
	SD	0.132	0.042	0.099
EX	EST	0.447	0.894	0.311
	SD	0.151	0.075	0.103

Next, we plot the estimated index functions $\widehat{g}(\cdot)$ against the variables of *age at diabetes diagnosis* and *body mass index*, respectively, by setting the rest of their corresponding index components to zero. Figure 4.2 depicts the estimated functions $\widehat{g}(\cdot)$ under the IND and EX working correlation structures. The function $\widehat{g}(\cdot)$ displays a quadratic pattern over the *body mass index*, which is consistent with the findings of Barnhart (1998) and Lian (2013). For example, under the EX structure, the value of $\widehat{g}(\cdot)$ above 0 indicates that the presence of diabetic retinopathy is higher when *body mass index* lies between 2.854 and 6.397 than in the tail regions (i.e., 2.042 to 2.854 and 6.397 to 7.228). It is interesting to note that $\widehat{g}(\cdot)$ also exhibits a quadratic pattern across the variable of *age at diabetes diagnosis*, and the value of $\widehat{g}(\cdot)$ above 0 shows that the presence of diabetic retinopathy is higher when age ranges between 5.1 and 25.1 than in the tail regions (i.e., 0.4 to 5.1 and 25.1 to 29.9). Accordingly, it is not surprising that the plot of $\widehat{g}(\cdot)$ versus the index exhibits a quadratic shape. In sum, the diabetic retinopathy risk is highest in this study among people with middle values for body mass and middle values for age at diagnosis of diabetes.

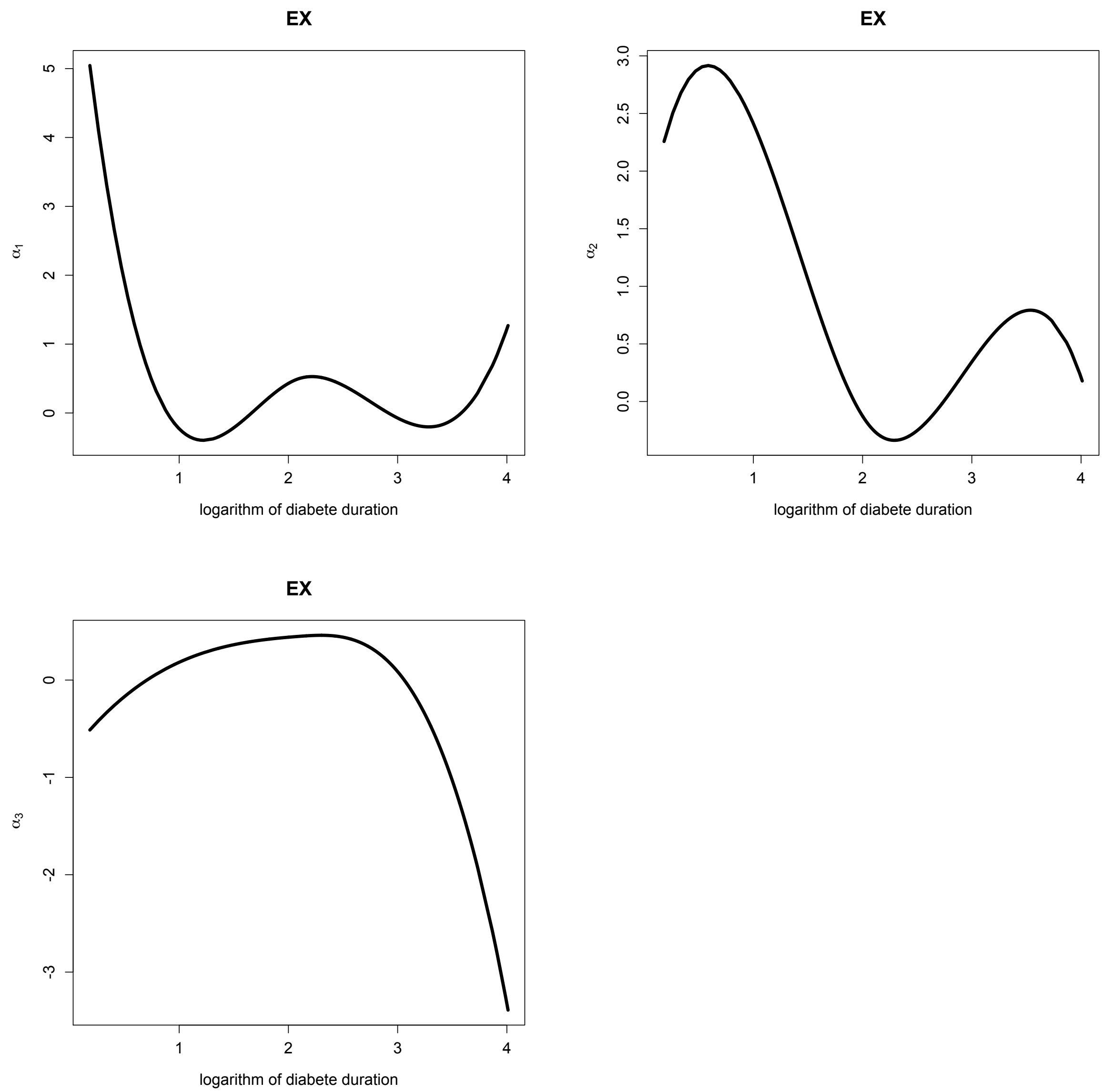
Figure 4.2: Plots of $\widehat{g}(\cdot)$ against the variables *body mass index* and *age at diabetes diagnosis*, respectively, under the IND and EX working correlation structures using the Wisconsin epidemiologic study.



We finally present the graphs of the estimated varying coefficient functions $\widehat{\alpha}_l(\cdot)$

$(l = 1, 2, 3)$ against *logarithm of diabetes duration* under the EX structure. Since the plots under the IND structure are similar to those under the EX structure, we omit them. The varying coefficient functions in Figure 4.3 exhibits strong nonlinear patterns. Specifically, $\widehat{\alpha}_1(\cdot)$ and $\widehat{\alpha}_2(\cdot)$ indicate that coefficients are largest when the diabetes duration is shortest, while $\widehat{\alpha}_3(\cdot)$ has the largest coefficient around the middle values of diabetes duration. Consequently, the associated coefficients for the variables *glycosylated hemoglobin level*, *systolic blood pressure*, and *diastolic blood pressure* are not constant across different durations.

Figure 4.3: Plots of $\widehat{\alpha}_l(\cdot)$, $l = 1, 2, 3$, against the *logarithm of diabetes duration* under the EX working correlation structure using the Wisconsin epidemiologic study.



4.6 Discussion

In this chapter, we introduce a generalized semiparametric model emerging from generalized partially linear single-index models and varying coefficient models with repeated measurements. For model estimation, we propose the profile QIF estimator for the regression parameters and the QIF spline estimators for the index function and varying coefficient functions. For model selections, penalized and group penalized estimation procedures are employed, respectively, for parametric and nonparametric functions. In addition, asymptotic consistency is studied for the resulting estimators, and asymptotic normality is further established for the parametric estimators for conducting statistical inference such as Wald test. Moreover, we demonstrate the oracle properties of the penalized estimators. Monte Carlo studies indicate that the proposed estimators perform well.

In practice, there are a few possible approaches to fit the data with model (1). Based on our limited experience, we propose the following procedures. First, place continuous variables into the single-index component and put discrete variables into either the varying coefficient component or the linear component. Second, for continuous variables, plot the estimated mean of the response variable (or the estimated single-index function) against each of them. If the plots of those variables do not depict the nonlinear pattern, one can put them into either the varying coefficient component or the linear component. Third, choose the varying coefficient index, which exhibits possible interaction effects with those variables assigned in the varying coefficient component.

To extend applications of the proposed generalized semiparametric model, we identify five future research topics. The first is to generalize the penalized quadratic inference function so that one is able to estimate and select the mean components and correlation components simultaneously. The second is to make inferences by testing the parametric and nonparametric components. The third is to adapt the approach of Stute (2005) and then develop a test for assessing the appropriateness of model (1). The fourth is to allow the nonparametric component to be a non-smooth function. Finally, we propose applying the proposed model to the areas of quantile regression and survival analysis. We believe that these efforts would broaden the usefulness of the proposed model.

Bibliography

[1] Barnhart, H. X. and Williamson, J. M. (1998). Goodness-of-fit tests for GEE modeling with binary responses. *Biometrics,* 54, 720–729.

[2] Boente, G., He, X., and Zhou, J. (2006). Robust estimates in generalized partially linear models. *The Annals of Statistics*, 34, 2856–2878.

[3] Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes*, Springer-Verlag, New York.

[4] Brillinger, D. R. (1983). A generalized linear model with "Gaussian" regression variables. In *A Festschrift for Erich L. Lehmann in Honor of His Sixty-Fifth Birthday* (Bickel, P. J., Doksum, K. A., and Hodges, J. L. eds), 97–114. Wadsworth, California.

[5] Cai, Z., Fan, J., and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association,* 95, 888–902.

[6] Cai, Z., Fan, J., and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association,* 95, 941–956.

[7] Carroll, R. J., Fan, J, Gijbels, I., and Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association,* 92, 477–489.

[8] Chen, R. and Tsay, R. S. (1993). Functional-coefficient autoregressive models. *Journal of the American Statistical Association,* 88, 289–308.

[9] Cleveland, W. S., Grosse, E., and Shyu, W. M. (1991). Local regression models. In *Statistical Models in S* (Chambers, J. M. and Hastie, T. J. eds), 309–376. Wadsworth & Brooks, California.

[10] Cui, X., Härdle, W. K. and Zhu, L. (2011). The EFM approach for single-index models. *The Annals of Statistics*, 39, 1658–1688.

[11] Davis, C. S. (2002). *Statistical Methods for the Analysis of Repeated Measurements,* Springer, New York.

[12] de Boor, C. (2001). *A Practical Guide to Splines, Vol. 27 of Applied Mathematical Sciences*, revised edn, Springer-Verlag, New York.

[13] DeVore, R. A. and Lorentz, G. G. (1993). *Constructive Approximation,* Springer-Berlin.

[14] Diggle, P. J., Heagerty, P., Liang, K. Y., and Zeger, S. (2002). *Analysis of Longitudinal Data, 2nd Ed.*, Oxford University Press, Oxford.

[15] Emrich, L. J. and Piedmonte, M. R. (1991). A method for generating high-dimensional multivariate binary variates. *The American Statistician*, 45, 302–304.

[16] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association,* 96, 1348–1360.

[17] Fan, J. and Li, R. (2004). New estimation and model selection procedures for semiparametric modelling in longitudinal data analysis. *Journal of the American Statistical Association,* 99, 710–723.

[18] Fan, J. and Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli,* 11, 1031–1057.

[19] Fan, J. and Zhang, W. (2008). Statistical methods with varying coefficient models. *Statistics and Its Interface,* 1, 179–195.

[20] Fan, J., Lin, H., and Zhou, Y. (2006) Local partial-likelihood estimation for lifetime data. *The Annals of Statistics*, 34, 290–325.

[21] Goldstein, M. and Smith, A. F. M. (1974). Ridge-type estimators for regression analysis. *Journal of the Royal Statistical Society, Series B,* 36, 284–291.

[22] Härdle, W., Liang, H., and Gao, J. (2000). *Partially Linear Models,* Physica-Verlag, Heidelberg.

[23] Hastie, T. J. and Tibshirani, R. J. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society, Series B* 55, 757–796.

[24] He, X., Fung, W. K., and Zhu, Z. (2005). Robust estimation in generalized partial linear models for clustered data. *Journal of the American Statistical Association*, 100, 1176–1184.

[25] Horowitz, J. L. (1998). *Semiparametric Methods in Econometrics*, Springer-Verlag, New York.

[26] Huang, J. Z., Wu, C. O., and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika,* 89, 111–128.

[27] Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single index models. *Journal of Econometrics*,58, 71–120.

[28] Klein, R., Klein, B. E. K., Moss, S. E., Davis, M. D., and DeMets, D. L. (1984). The Wisconsin Epidemiologic Study of Diabetic Retinopathy: II. Prevalence and risk of diabetic retinopathy when age at diagnosis is less than 30 years. *Archives of Opthalmology,* 102, 520–526.

[29] Li, R. and Liang, H. (2008). Variable selection in semiparametric regression modeling. *The Annals of Statistics*, 36, 261–286.

[30] Liang, H., Liu, X., Li, R., and Tsai, C. L. (2010). Estimation and testing for partially linear single-index models. *The Annals of Statistics*, 38, 3811-3836.

[31] Lian, H., Liang, H, and Wang, L. (2013). Generalized additive partial linear models for clustered data with diverging number of covariates using GEE. *Statistica Sinica,* 24, 173-196.

[32] Liang, K .Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika,* 73, 13–22.

[33] Lin, D. Y. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data (with discussion). *Journal of the American Statistical Association,* 96, 103–126.

[34] Lin, X. and Carroll, R. J. (2006). Semiparametric estimation in general repeated measure problems. *Journal of the Royal Statistical Society, Series B,* 68, 68–88.

[35] Ma, S. (2012). Two-step spline estimating equations for generalized additive partially linear models with large cluster sizes. *The Annals of Statistics*, 40, 2943–2972.

[36] Ma, S. and Song, P. (2014). Varying index coefficient models. *Journal of the American Statistical Association*, forthcoming.

[37] Ma, Y. and Zhu, L. (2013). Doubly robust and efficient estimators for heteroscedastic partially linear single-index models allowing high dimensional covariates. *Journal of the Royal Statistical Society, Series B,* 75, 305–322.

[38] McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models. 2nd Ed.,* Chapman and Hall, New York.

[39] Qu, A. and Li, R. (2006). Quadratic Inference Functions for Varying Coefficient Models with Longitudinal Data. *Biometrics,* 62, 379–391.

[40] Qu, A., Lindsay, B. G., and Li, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika,* 87, 823–836.

[41] Speckman, P. (1998). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, Series B,* 50, 413–436.

[42] Stute, W. and Zhu, L.-X. (2009). Nonparametric checks for single-index models. *The Annals of Statistics*, 33, 1048–1083.

[43] Tibshirani, R. J. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B* 58, 267–288.

[44] Wang, H. and Xia, Y. (2009). Shrinkage estimation of the varying coefficient model. *Journal of the American Statistical Association,* 104, 747–757.

[45] Wang, H., Li, R., and Tsai, C. L. (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika,* 94, 553–568.

[46] Wang, L., Chen, G. and Li, H. (2007). Group SCAD regression analysis for microarray time course gene expression data. *Journal of Bioinformatics,* 23, 1486–1494.

[47] Wang, L., Li, H., and Huang, J. (2008). Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. *Journal of the American Statistical Association*, 103, 1556–1569.

[48] Wang, N. (2003). Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika,* 90, 43–52.

[49] Xia, Y. C. and Härdle, W. (2006). Semi-parametric estimation of partially single-index models. *Journal of Multivariate Analysis* 97, 1162–1184.

[50] Xie, H. and Huang, J. (2009). SCAD-penalized regression in high-dimensional partially linear models. *The Annals of Statistics*, 37, 673–696.

[51] Xu, P. and Zhu, L. (2012). Estimation for a marginal generalized single-index longitudinal model. *Journal of Multivariate Analysis*, 105, 285-299.

[52] Xue, L., Qu, A. and Zhou, J. (2010). Consistent model selection for marginal generalized additive model for correlated data. *Journal of the American Statistical Association,* 105, 1518–1530.

[53] Yan, Y. and Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association,* 97, 1042–1054.

[54] Zeng, P., He, T. and Zhu, Y. (2012). A lasso-type approach for estimation and variable selection in single index models. *Journal of Computational and Graphical Statistics,* 21, 92–109.

[55] Zhang, W., Lee, S. Y., and Song, X. (2002). Local polynomial fitting in semivarying coefficient models. *Journal of Multivariate Analysis* 82, 166–188.

[56] Zhang, Y., Li, R. and Tsai, C. L. (2010). Regularization parameter selection via generalized information criterion. *Journal of the American Statistical Association,* 105, 312–323.

[57] Zhou, J. and Qu, A. (2012). In Semiparametric and nonparametric regression analysis of longitudinal data (with discussion). *Journal of the American Statistical Association,* 96, 103–126.

[58] Zhou, S, Shen, X. and Wolfe, D. A. (1998). Local asymptotics for regression splines and confidence regions. *Annals of Statistics,* 26, 1760–1782.

Chapter 5

Semiparametric Mixed Model Analysis for Nonlinear Gene-environment Interactions in Genome-wide Association Studies

5.1 Introduction

As we saw in Chapter 3, linear mixed models have demonstrated effectiveness in accounting for relatedness among samples and in controlling for population stratification. Consequently, there is an increasing interest in using linear mixed model to investigate associations between genes and diseases in genome-wide association studies (GWAS).

Kang (2008) proposed the efficient mixed model association (EMMA) methodology to account for pairwise relatedness between individuals. They applied the properties of singular value decomposition and estimated the variance components by maximizing the restricted likelihood function. Their strategy greatly improves the efficiency of mixed model method. However, when thousands of individuals are involved in GWAS, EMMA becomes computationally intractable due to the heavy computational burden in the estimation of variance parameters.

To make GWAS using linear mixed models possible for large data sets, two sophisticated approximate approaches have been suggested. Zhang (2010) used population parameters previously determined (P3D), which performed each test by simply using the pre-estimated variance components from the null model without estimating them repeatedly. Kang (2010) used a slightly different strategy; to test individual SNPs, they kept the heritability estimated from the null model fixed. They avoided estimating variance components repeatedly in this way. Their approach is implemented in the EMMA eXpedited (EMMAX) software. Both of these two approximations greatly reduce the computing time and maintain the statistical power at the same time.

On the other hand, environmental factors affect human health in important ways. It has been increasingly accepted that most human diseases are the result of a combination of genetic and environmental factors. Moreover, gene and environment (G $\times$ E) interactions play a pivotal role in the risk of developing human diseases, such as obesity (Hebebrand and Hinney (2009)), heart disease (Talmud (2007)), diabetes (Grarup and Andersen (2007)), and cancer (Song (2011)). Study of G $\times$ E interactions via statistical modeling is very important to improve the accuracy and precision when assessing genetic and environmental influences. Traditional analysis assumes linear G $\times$ E interactions, where the interaction effect is typically modeled as a product term. However, such modeling may not reflect the true nonlinear interactions between gene and environment, which could lead to a large estimation bias. To overcome this limitation, different non- and semi- parametric modeling methods have been proposed and developed. Such models include: profile likelihood-based semiparametric model in Chatterjee (2005) and Chen (2012); semiparametric model with Tukey's form of interaction in Maity (2009); a generalized likelihood ratio test for nonparametric effects in Wei (2011); and semiparametric Bayesian analysis in Lobach (2011) and Ahn (2013).

In this chapter, we aim to explore possible nonlinear G $\times$ E interactions to identify the genetic associations by considering hidden relatedness of the observations. From Zhou and Stephens (2012), a standard linear mixed model for GWAS can be written as

$$Y = \mathbf{T}\alpha + z\gamma + \xi + \epsilon, \tag{5.1}$$

where Y is the phenotype of interest, $\mathbf{T}$ are covariates, z is the SNP of interest, α is a vector of weights for the covariates, γ is the coefficient for the test SNP, ξ is a vector for unknown random polygenic effects, and ϵ is a vector of errors. $\xi\tilde{}\text{MVN}(0, \lambda\sigma^2\mathbf{K})$, $\epsilon\tilde{}\text{MVN}(0, \sigma^2\mathbf{I})$, where $\mathbf{K}$ is the kinship matrix calculated from either a set of genetic markers or pedigrees, $\mathbf{I}$ is the identity matrix, σ^2 is the variance of the residual errors, λ is the ratio between the two variance components, and MVN denotes multivariate normal distribution.

Considering the interplay of genetic and environmental factors, as mentioned in Saidou (2014), we may employ a parametric mixed model with both main and interaction effects given as

$$Y = \mathbf{T}\alpha + \mathbf{X}\beta + z\gamma + z\mathbf{X}\beta_1 + \xi + \epsilon, \tag{5.2}$$

where β are the effects of environmental factors, and β_1 are the effects for G $\times$ E interaction. By simple calculation, (5.2) can be written as

$$Y = \mathbf{T}\alpha + \mathbf{X}\beta + z(\gamma + \mathbf{X}\beta_1) + \xi + \epsilon.$$

Thus the contribution of the genetic factor z to the variation of Y is restricted to a linear function of $\mathbf{X}$. However, this restriction can be easily violated due to the underlying nonlinear pattern of the relationship between the response and explanatory variables. To allow for nonlinear G $\times$ E interactions, we can replace the coefficient of z with a smooth nonlinear function, which is of a linear combination of $\mathbf{X}$. Moreover, we can impose a nonlinear structure on the environmental term. Hence, we propose a partially linear single-index coefficient mixed model as

$$Y = \mathbf{T}\alpha + m_0(\mathbf{X}\beta) + m_1(\mathbf{X}\beta)z + \xi + \epsilon, \tag{5.3}$$

where ξ and ε are independent of $(z, \mathbf{X}, \mathbf{T})$. $m_0(.)$ and $m_1(.)$ are unknown smooth nonparametric functions with no specific functional form, so that model (5.3) can flexibly capture dynamic change patterns of the coefficient and intercept functions. We can assess whether a genetic marker is associated with the phenotype by testing $m_1(.) = 0$. Considering single-marker regression methods, we test each SNP and then apply a multiple testing correction procedure to select significant SNPs.

For estimation, we first approximate the nonparametric functions $m_0(.)$ and $m_1(.)$ by B-spline basis functions, then estimate parameters β and α by a profile maximum-likelihood method. To estimate the variance component λ and σ^2, a profile restricted maximum likelihood method is proposed. Moreover, we develop a score test for inferences on the coefficient function $m_1(.)$, which enables us to identify important genetic factors by testing $m_1(.) = 0$. The proposed testing procedure is easy to implement and fast to compute with the p-values or critical values obtained from the asymptotic distributions of the test statistics. In addition, it provides a useful inferential tool to identify genetic risk factors in my proposed model, which is more flexible than linear mixed models.

The proposed estimation procedure in model (5.3) is briefly described as follows. Assume λ is known, for given β and α, we approximate the intercept function $m_0(.)$ and coefficient function $m_1(.)$ by B-spline functions with coefficients obtained by maximizing a log-likelihood function. Thus the resulting estimators of $m_0(.)$ and $m_1(.)$ are functions of β and α. Estimation of the parameters β and α is achieved by replacing the true functions $m_0(.)$ and $m_1(.)$ with their spline estimators in the objective function. However, λ is unknown in reality; it can be estimated by maximizing the log-restricted likelihood function for given β and α. Finally, we propose a Newton-Raphson algorithm to estimate parameters β and α, nonparametric functions $m_l(.)$, λ and σ^2. Asymptotic normalities and consistency for estimators of both the parameters and nonparametric functions are established. When the data set consists of thousands of individuals, the estimation procedure becomes computationally intractable due to the heavy computational burden in the estimation of variance parameters. As described in Kang (2010), for most genetic association studies in humans, the effect of any given locus on the trait is very small, therefore, we only need to estimate the variance parameters once and globally apply them to each marker. Such strategy can markedly increase the speed of computations without decreasing the power of hypothesis testing.

The rest of this chapter is organized as follows. Section 2 introduces the profile log-likelihood estimation and presents the asymptotic properties of the proposed estimators. Section 3 develops score tests for nonparametric coefficient functions. Section 4 introduces the estimation of variance components through restricted likelihood. Section 5 proposes the computational algorithm. Section 6 evaluates the performance of the proposed estimation and inference procedures via simulation studies. Also we illustrate the proposed model and method through the analysis of the Framingham study to investigate gene associations with systolic blood pressure. All technical proofs are provided in Appendix B.

5.2 Estimation of Parameters and Nonparametric Functions

For the i^{th} subject, let Y_i be a quantitative trait, $\mathbf{Z}_i = (Z_{i1}, Z_{i2})^{\mathrm{T}}$ be a 2×1 vector, where $Z_{i1} = 1$ and Z_{i2} is the genetic factor, $\mathbf{X}_i = (X_{i1}, \ldots, X_{id_1})^{\mathrm{T}}$ be a $d_1 \times 1$ vector of environmental factors, $\mathbf{T}_i = (T_{i1}, ..., T_{id_2})^{\mathrm{T}}$ be a $d_2 \times 1$ vector of covariates, then we have the following proposed semiparametric model as

$$Y_i = \mathbf{T}_i^{\mathrm{T}} \alpha + \sum\nolimits_{\ell=1}^{2} m_\ell(\mathbf{X}_i^{\mathrm{T}} \beta) Z_{i\ell} + \xi_i + \varepsilon_i, \tag{5.4}$$

where $\alpha = (\alpha_1, ..., \alpha_{d_2})^{\mathrm{T}}$ are regression coefficients for $\mathbf{T}_i$, $\beta = (\beta_1, ..., \beta_{d_1})^{\mathrm{T}}$ are coefficient parameters for $\mathbf{X}_i$, and $m_l(.)$ denotes an unknown smooth nonparametric function. $\xi = (\xi_1, \ldots, \xi_n)^{\mathrm{T}}$ is an $n \times 1$ vector for the polygenic effects, $\xi \tilde{\ } \mathrm{MVN}_n \left(\mathbf{0}, \lambda\sigma^2 \mathbf{K}\right)$, in which $\mathbf{K}$ is an $n \times n$ marker-based kinship matrix. $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^{\mathrm{T}} \tilde{\ } \mathrm{MVN}_n \left(\mathbf{0}, \sigma^2 \mathbf{I}_n\right)$, in which $\mathbf{I}_n$ is the identity matrix. ξ_i and ε_i are independent of $(\mathbf{Z}_i\mathbf{X}_i, \mathbf{T}_i)$.

Denote $U_i(\beta) = \mathbf{X}_i^{\mathrm{T}}\beta$, we assume that $U_i(\beta)$ is distributed on a compact interval $[a, b]$. Without loss of generality, one can let $[a, b] = [0, 1]$. We estimate the unknown functions $m_\ell(\cdot)$ by B-spline functions. To this end, let $N = N_n$ be the number of interior knots. Divide $[0, 1]$ into $(N + 1)$ subintervals: $I_j = [\xi_j, \xi_{j+1})$ for $j = 0, \ldots, N - 1$ and $I_N = [\xi_N, 1]$, where $(\xi_j)_{j=1}^N$ is a sequence of interior knots that is given as

$$\xi_{-(q-1)} = \cdots = 0 = \xi_0 < \xi_1 < \cdots < \xi_N < 1 = \xi_{N+1} = \cdots = \xi_{N+q}.$$

For $0 \leq j \leq N_n$, let $h_j = \xi_{j+1} - \xi_j$ be the distance between neighboring knots and let $h = \max_{0\leq j\leq N_n} h_j$. Following Zhou (1998), to study the asymptotic properties for the spline estimator of $m_\ell(\cdot)$, we assume that $\max_{0\leq j\leq N_n-1} |h_{j+1} - h_j| = o\left(N^{-1}\right)$ and $h/\min_{0\leq j\leq N_n} h_j \leq M$, where $M > 0$ is a predetermined constant. Such an assumption assures that $M^{-1} < N_n h < M$, which is necessary for numerical implementation. Furthermore, define the q-th order normalized B-spline basis as $B(u) = \{B_j(u) : 1 \leq j \leq J_n\}^{\mathrm{T}}$(see de Boor (2001)), where $J_n = N + q$. Let $\mathcal{H}_n = \mathcal{H}_n^{(q-2)}$ be the space spanned by $B(u)$. Thus, there exists $\gamma_\ell = (\gamma_{1\ell}, \ldots, \gamma_{J_n\ell})^{\mathrm{T}}$ such that $m_\ell^0(u) = \sum_{j=1}^{J_n} B_j(u)\gamma_{j\ell} = B(u)^{\mathrm{T}}\gamma_\ell \in \mathcal{H}_n$, and under suitable smoothness assumptions, $m_\ell(u)$ can be well approximated by $m_\ell^0(u)$. Correspondingly, $\sum_{\ell=1}^2 m_\ell(\mathbf{X}_i^{\mathrm{T}}\beta)Z_{i\ell}$ can be approximated by

$$\sum\nolimits_{\ell=1}^2 m_\ell^0(\mathbf{X}_i^{\mathrm{T}}\beta)Z_{i\ell} = \sum\nolimits_{\ell=1}^2 B\left(\mathbf{X}_i^{\mathrm{T}}\beta\right)^{\mathrm{T}}\gamma_\ell Z_{i\ell} = \mathbf{Q}_i(\beta)^{\mathrm{T}}\gamma,$$

where $\gamma = \left(\gamma_1^{\mathrm{T}}, \gamma_2^{\mathrm{T}}\right)^{\mathrm{T}}$and $\mathbf{Q}_i(\beta) = \left[\left\{B\left(\mathbf{X}_i^{\mathrm{T}}\beta\right)^{\mathrm{T}} Z_{i1}, B\left(\mathbf{X}_i^{\mathrm{T}}\beta\right)^{\mathrm{T}} Z_{i2}\right\}^{\mathrm{T}}\right]_{2J_n\times 1}$.

5.2.1 Profile estimation of parameters β and α and nonparametric functions $m_\ell(\cdot)$.

Let $\mathbf{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$, $\mathbf{Q}(\beta) = \left[\{\mathbf{Q}_1(\beta), \ldots, \mathbf{Q}_n(\beta)\}^{\mathrm{T}}\right]_{n\times 2J_n}$ and

$$\mathbf{T} = \left[(\mathbf{T}_1, \ldots, \mathbf{T}_n)^{\mathrm{T}}\right]_{n\times d_2}.$$

For given β and α, the estimates of γ, λ and σ^2 can be obtained by maximizing the following log-likelihood function:

$$\begin{aligned} L_n\left(\gamma, \beta, \alpha, \lambda, \sigma^2\right) &= -\frac{n}{2}\log\left(\sigma^2\right) - \frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{H}| \\ &\quad -\frac{1}{2}\sigma^{-2}\left(\mathbf{Y} - \mathbf{T}\alpha - \mathbf{Q}(\beta)\gamma\right)^{\mathrm{T}}\mathbf{H}^{-1}\left(\mathbf{Y} - \mathbf{T}\alpha - \mathbf{Q}(\beta)\gamma\right), \end{aligned} \tag{5.5}$$

where $\mathbf{H} = \lambda\mathbf{K} + \mathbf{I}_n$. By assuming λ is known, the log-likelihood (5.5) is maximized at

$$\widetilde{\gamma}(\beta, \alpha) = \left\{\mathbf{Q}(\beta)^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{Q}(\beta)\right\}^{-1}\left\{\mathbf{Q}(\beta)^{\mathrm{T}}\mathbf{H}^{-1}(\mathbf{Y} - \mathbf{T}\alpha)\right\}.$$

To proceed estimation of β and α, to ensure identifiability we exclude the first component β_1 of β by letting $\beta_1 = \sqrt{1 - \|\beta_{-1}\|_2^2}$, where $\beta_{-1} = (\beta_2, \ldots, \beta_{d_1})^{\mathrm{T}}$, and reformulate the parameter space of β as follows:

$$\left\{\left(\sqrt{1 - \|\beta_{-1}\|_2^2}, \beta_2, \ldots, \beta_{d_1}\right)^{\mathrm{T}} : \|\beta_{-1}\|_2^2 < 1\right\}.$$

Replacing γ with $\widetilde{\gamma}(\beta, \alpha)$ in the objective function (5.5), β_{-1} and α are obtained by maximizing

$$\begin{aligned} L_n^*\left(\beta, \alpha, \sigma^2\right) &= -\frac{n}{2}\log\left(\sigma^2\right) - \frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{H}| \\ &\quad -\frac{1}{2}\sigma^{-2}\left(\mathbf{Y} - \mathbf{T}\alpha - \mathbf{Q}(\beta)\widetilde{\gamma}(\beta, \alpha)\right)^{\mathrm{T}}\mathbf{H}^{-1}\left(\mathbf{Y} - \mathbf{T}\alpha - \mathbf{Q}(\beta)\widetilde{\gamma}(\beta, \alpha)\right). \end{aligned} \tag{5.6}$$

The estimators of $\theta_{-1} = \left(\beta_{-1}^{\mathrm{T}}, \alpha^{\mathrm{T}}\right)^{\mathrm{T}}$ are given as $\widehat{\theta}_{-1} = \left(\widehat{\beta}_{-1}^{\mathrm{T}}, \widehat{\alpha}^{\mathrm{T}}\right)^{\mathrm{T}} = \arg\max_{(\beta_{-1},\alpha)} \left\{L_n^*\left(\beta, \alpha, \sigma^2\right)\right\}$. Thus $\widehat{\beta}_1 = \sqrt{1 - \left\|\widehat{\beta}_{-1}\right\|_2^2}$, $\widehat{\beta} = \left(\widehat{\beta}_1, \widehat{\beta}_{-1}^{\mathrm{T}}\right)^{\mathrm{T}}$. After we obtain the estimator $\widehat{\beta}$ and $\widehat{\alpha}$, the final estimators of the spline coefficients are given as

$$\widehat{\gamma} = \left\{\widehat{\gamma}_1^{\mathrm{T}}, \widehat{\gamma}_2^{\mathrm{T}}\right\}^{\mathrm{T}} = \widetilde{\gamma}(\widehat{\beta}, \widehat{\alpha}) = \arg\max_{\gamma} \left\{L_n\left(\gamma, \widehat{\beta}, \widehat{\alpha}, \lambda, \sigma^2\right)\right\},$$

and the ℓ^{th} nonparametric function $m_\ell(u)$ is estimated by

$$\widehat{m}_\ell(u) = \widetilde{m}_\ell\left(u, \widehat{\beta}, \widehat{\alpha}\right) = B(u)^{\mathrm{T}}\widehat{\gamma}_\ell = \sum\nolimits_{j=1}^{J_n} B_j(u)\,\widehat{\gamma}_{j\ell}.$$

In practice, λ is unknown. We propose a profile REML estimation procedure to estimate λ and σ^2. For these profile estimators, asymptotic consistency and normality are established.

5.2.2 Inference for the profile estimation

Consider the more general model as

$$Y_i = \mathbf{T}_i^{\mathrm{T}}\alpha + \sum\nolimits_{\ell=1}^{p} m_\ell(\mathbf{X}_i^{\mathrm{T}}\beta) Z_{i\ell} + \xi_i + \varepsilon_i, \tag{5.7}$$

where $Z_{i1} = 1$, $(Z_{i2}, ..., Z_{ip})$ are genetic factors for subject i, then our proposed model is a special case with $p = 2$. Let $\mathbf{J}(\beta) = \partial\beta/\partial\beta_{-1}^{\mathrm{T}}$ be the Jacobian matrix of size $d_1 \times (d_1 - 1)$, which is

$$\mathbf{J}(\beta) = \begin{pmatrix} -\beta_{-1}^{\mathrm{T}}/\sqrt{1 - \|\beta_{-1}\|_2^2} \\ \mathbf{I}_{d_1-1} \end{pmatrix}_{d_1\times(d_1-1)}.$$

Assume that the true parameters of β and α are β^0 and α^0 respectively, and the true parameter of θ_{-1} is θ_{-1}^0. Let $\mathbf{J} = \mathbf{J}\left(\beta^0\right)$. For any scalar or vector ξ_i, define

$$\mathrm{Proj}_{\mathcal{M}_n}\left(\xi_i^{\mathrm{T}}\right) = \mathbf{Q}_i\left(\beta^0\right)^{\mathrm{T}}\left\{\mathbf{Q}\left(\beta^0\right)^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{Q}\left(\beta^0\right)\right\}^{-1}\mathbf{Q}\left(\beta^0\right)^{\mathrm{T}}\mathbf{H}^{-1}\xi,$$

where $\xi = \left(\xi_1^{\mathrm{T}}, \ldots, \xi_n^{\mathrm{T}}\right)^{\mathrm{T}}$. It can be proved that

$$\begin{aligned}
& \mathbf{Q}_i\left(\beta^0\right)^{\mathrm{T}}\left\{\partial\left(\widetilde{\gamma}(\beta^0, \alpha^0) - \gamma\right)/\partial\theta_{-1}^{\mathrm{T}}\right\} \\
= \; & -\mathbf{Q}_i\left(\beta^0\right)^{\mathrm{T}}\left\{\mathbf{Q}\left(\beta^0\right)^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{Q}\left(\beta^0\right)\right\}^{-1}\mathbf{Q}\left(\beta^0\right)^{\mathrm{T}}\mathbf{H}^{-1}\left(\sum\nolimits_{\ell=1}^{p}\widetilde{m}_{\ell}'(\mathbf{X}_i^{\mathrm{T}}\beta^0)Z_{i\ell}\mathbf{X}_i^{\mathrm{T}}\mathbf{J}, \mathbf{T}_i^{\mathrm{T}}\right)_{i=1}^{n} \\
& + O_p\left(n^{-1/2}J_n^{1/2} + J_n^{-r+1}\right) \\
= \; & -\mathrm{Proj}_{\mathcal{M}_n}\left(\sum\nolimits_{\ell=1}^{p} m_{\ell}'(\mathbf{X}_i^{\mathrm{T}}\beta^0)Z_{i\ell}\mathbf{X}_i^{\mathrm{T}}\mathbf{J}, \mathbf{T}_i^{\mathrm{T}}\right) + O_p\left(n^{-1/2}J_n^{1/2} + J_n^{-r+1}\right),
\end{aligned}$$

$$\begin{aligned}
& \partial\left\{\mathbf{T}_i^{\mathrm{T}}\alpha^0 + \mathbf{Q}_i\left(\beta^0\right)^{\mathrm{T}}\widetilde{\gamma}\left(\beta^0, \alpha^0\right)\right\}/\partial\theta_{-1}^{\mathrm{T}} \\
= \; & \left(\sum\nolimits_{\ell=1}^{p} m_{\ell}'(\mathbf{X}_i^{\mathrm{T}}\beta^0)Z_{i\ell}\mathbf{X}_i^{\mathrm{T}}\mathbf{J}, \mathbf{T}_i^{\mathrm{T}}\right) - \mathrm{Proj}_{\mathcal{M}_n}\left(\sum\nolimits_{\ell=1}^{p} m_{\ell}'(\mathbf{X}_i^{\mathrm{T}}\beta^0)Z_{i\ell}\mathbf{X}_i^{\mathrm{T}}\mathbf{J}, \mathbf{T}_i^{\mathrm{T}}\right) \\
& + O_p\left(n^{-1/2}J_n^{1/2} + J_n^{-r+1}\right) \\
= \; & \left(\sum\nolimits_{\ell=1}^{p} m_{\ell}'(\mathbf{X}_i^{\mathrm{T}}\beta^0)Z_{i\ell}\widehat{\mathbf{X}}_i^{\mathrm{T}}\mathbf{J}, \widehat{\mathbf{T}}_i^{\mathrm{T}}\right) + O_p\left(n^{-1/2}J_n^{1/2} + J_n^{-r+1}\right),
\end{aligned}$$

where for any scalar or vector ξ_i, define $\widehat{\xi}_i^{\mathrm{T}} = \xi_i^{\mathrm{T}} - \mathrm{Proj}_{\mathcal{M}_n}\left(\xi_i^{\mathrm{T}}\right)$. Let

$$\mathbf{\Phi}\left(\beta\right) = \left[\mathrm{diag}\left\{\sum\nolimits_{\ell=1}^{p} m_{\ell}'(\mathbf{X}_1^{\mathrm{T}}\beta)Z_{1\ell}, \ldots, \sum\nolimits_{\ell=1}^{p} m_{\ell}'(\mathbf{X}_n^{\mathrm{T}}\beta)Z_{n\ell}\right\}\right]_{n\times n}.$$

Thus

$$\partial\left\{\mathbf{T}\alpha^{\mathbf{0}} + \mathbf{Q}\left(\beta^0\right)\widetilde{\gamma}\left(\beta^0, \alpha^0\right)\right\}/\partial\theta_{-1}^{\mathrm{T}} = \left(\mathbf{\Phi}\left(\beta^0\right)\widehat{\mathbf{X}}\mathbf{J}, \widehat{\mathbf{T}}\right) + O_p\left(n^{-1/2}J_n^{1/2} + J_n^{-r+1}\right),$$

where $\widehat{\mathbf{X}} = \left(\widehat{\mathbf{X}}_1, \ldots, \widehat{\mathbf{X}}_n\right)^{\mathrm{T}}$ and $\widehat{\mathbf{T}} = \left(\widehat{\mathbf{T}}_1, \ldots, \widehat{\mathbf{T}}_n\right)^{\mathrm{T}}$. Let $e_i = \xi_i + \varepsilon_i$ and $\mathbf{e} = (e_1, \ldots, e_n)^{\mathrm{T}}$. Now define the space $\mathcal{M}$ as a collection of functions with finite L_2 norm on $[0,1]\times R^p$ by

$$\mathcal{M} = \left\{\psi\left(u, \mathbf{x}\right) = \sum\nolimits_{\ell=1}^{p} g_{\ell}\left(u\right)x_{\ell}, Eg_{\ell}\left(U\right)^2 < \infty\right\},$$

where $\mathbf{x} = (x_1, \ldots, x_p)^{\mathrm{T}}$. For any vector $\xi = (\xi_1, \ldots, \xi_n)^{\mathrm{T}}$, let

$$\mathrm{Proj}_{\mathcal{M}}\left(\xi_i\right) = \Psi^*\left(U_i\left(\beta^0\right), \mathbf{X}_i\right) = \sum\nolimits_{\ell=1}^{p} g_{\ell}^*\left(U_i\left(\beta^0\right)\right)X_{i\ell} \in \mathcal{M}$$

be defined as the minimizer of

$$E\left\{\left(\xi-\Psi\left(\beta^0\right)\right)^{\mathrm{T}}\mathbf{H}^{-1}\left(\xi-\Psi\left(\beta^0\right)\right)\right\},$$

where $\Psi\left(\beta^0\right)=\left\{\Psi^*\left(U_1\left(\beta^0\right),\mathbf{X}_1\right),\ldots,\Psi^*\left(U_n\left(\beta^0\right),\mathbf{X}_n\right)\right\}^{\mathrm{T}}$, and let $\widetilde{\xi}_i=\xi_i-\mathrm{Proj}_{\mathcal{M}}\left(\xi_i\right)$. Denote $\widetilde{\mathbf{X}}_i=\left(\widetilde{\mathbf{X}}_{i1},\ldots,\widetilde{\mathbf{X}}_{ip}\right)^{\mathrm{T}}$, $\widetilde{\mathbf{X}}=\left(\widetilde{\mathbf{X}}_1,\ldots,\widetilde{\mathbf{X}}_n\right)^{\mathrm{T}}$, $\widetilde{\mathbf{T}}_i=\left(\widetilde{\mathbf{T}}_{i1},\ldots,\widetilde{\mathbf{T}}_{id_2}\right)^{\mathrm{T}}$ and $\widetilde{\mathbf{T}}=\left(\widetilde{\mathbf{T}}_1,\ldots,\widetilde{\mathbf{T}}_n\right)^{\mathrm{T}}$. Let $\mathbf{m}=(\sum_{l=1}^{p}m_l(\mathbf{X}_1^{\mathrm{T}}\beta^0)Z_{1l},\ldots\sum_{l=1}^{p}m_l(\mathbf{X}_n^{\mathrm{T}}\beta^0)Z_{nl})^{\mathrm{T}}$, it can be proved that

$$\partial L_n^*\left(\beta^0,\alpha^0,\sigma^2\right)/\partial\theta_{-1}^{\mathrm{T}}=\sigma^{-2}\left(\mathbf{Y}-\mathbf{T}\alpha^0-\mathbf{m}\right)^{\mathrm{T}}\mathbf{H}^{-1}\left(\mathbf{\Phi}\left(\beta^0\right)\widetilde{\mathbf{X}}\mathbf{J},\widetilde{\mathbf{T}}\right)+o_p\left(n^{1/2}\right),$$
$$\partial L_n^*\left(\beta^0,\alpha^0,\sigma^2\right)/\partial\theta_{-1}\partial\theta_{-1}^{\mathrm{T}}=\sigma^{-2}\left(\mathbf{\Phi}\left(\beta^0\right)\widetilde{\mathbf{X}}\mathbf{J},\widetilde{\mathbf{T}}\right)^{\mathrm{T}}\mathbf{H}^{-1}\left(\mathbf{\Phi}\left(\beta^0\right)\widetilde{\mathbf{X}}\mathbf{J},\widetilde{\mathbf{T}}\right)+o_p\left(n\right).$$

Let r with $r>3/2$ be the smoothness order of the coefficient functions $m_\ell\left(u\right)$ as given in Condition (C2) in Appendix B. Let $m_\ell'\left(u\right)$ be the first order derivative of $m_\ell\left(u\right)$, we have the following theorems, such that asymptotic consistency and normality are established for these profile parameter and nonparametric function estimators. Details of the theorems are given in Appendix B.

Theorem 5 *Under Conditions (C1)-(C6) in Appendix B,* $\max\left\{n^{1/(2r+2)},n^{1/(4r-2)}\right\}<<J_n<<n^{1/4}$*, we have (i) (consistency)* $\left\|\widehat{\theta}_{-1}-\theta_{-1}^0\right\|_2=O_p\left(n^{-1/2}\right)$*; and (ii) (asymptotic normality)*

$$\begin{aligned}\widehat{\theta}_{-1}-\theta_{-1}^0 &= \left[\left(\mathbf{\Phi}\left(\beta^0\right)\widetilde{\mathbf{X}}\mathbf{J},\widetilde{\mathbf{T}}\right)^T\mathbf{H}^{-1}\left(\mathbf{\Phi}\left(\beta^0\right)\widetilde{\mathbf{X}}\mathbf{J},\widetilde{\mathbf{T}}\right)\right]^{-1}\times\\ &\quad\left[\mathbf{e}^T\mathbf{H}^{-1}\left(\mathbf{\Phi}\left(\beta^0\right)\widetilde{\mathbf{X}}\mathbf{J},\widetilde{\mathbf{T}}\right)\right]+o_p\left(n^{-1/2}\right).\end{aligned}$$

Moreover $\Sigma_n^{-1/2}\left(\widehat{\theta}_{-1}-\theta_{-1}^0\right)\xrightarrow{d}\mathcal{N}\left(0,\mathbf{I}_{((d_1-1)+d_2)\times((d_1-1)+d_2)}\right)$*, as* $n\to\infty$*, where*

$$\Sigma_n=\sigma^2\left[\left(\mathbf{\Phi}\left(\beta^0\right)\widetilde{\mathbf{X}}\mathbf{J},\widetilde{\mathbf{T}}\right)^T\mathbf{H}^{-1}\left(\mathbf{\Phi}\left(\beta^0\right)\widetilde{\mathbf{X}}\mathbf{J},\widetilde{\mathbf{T}}\right)\right]^{-1}. \tag{5.8}$$

Both consistency and asymptotic normality of $\widehat{\beta}$ follow directly from Theorem 5 with an application of the multivariate delta-method. Thus we obtain

$$\left(\mathbf{J}\Sigma_{n,11}\mathbf{J}^{\mathrm{T}}\right)^{-1/2}\left(\widehat{\beta}-\beta^{0}\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbf{I}_{d_1 \times d_1}\right), n \to \infty,$$

where $\Sigma_{n,11}$ is the submatrix of Σ_n formed by the first $d_1 - 1$ rows and first $d_1 - 1$ columns of Σ_n. Define

$$\mathbf{B}(u) = \begin{bmatrix} B_1(u) & \cdots & B_{J_n}(u) & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & B_1(u) & \cdots & B_{J_n}(u) \end{bmatrix}_{p \times pJ_n}.$$

Theorem 6 *Under Conditions (C1)-(C6) in Appendix B,* $\max\left\{n^{1/(2r+2)}, n^{1/(4r-2)}\right\} << J_n << n^{1/4}$, *we have (i) consistency* $|\widehat{m}_\ell(u) - m_\ell(u)| = O_p\left(\sqrt{J_n/n} + J_n^{-r}\right)$ *uniformly in* $u \in [0,1]$*; and (ii) as* $n \to \infty$, *(asymptotic normality) for* $1 \le \ell \le p$,

$$\sigma_{\ell n}^{-1}(u)\left[\widehat{m}_\ell(u) - E\left\{\widehat{m}_\ell(u)|\mathbf{Z},\mathbf{X},\mathbf{T}\right\}\right] \to \mathcal{N}(0,1),$$

where

$$\sigma_{\ell n}^{2}(u) = e_\ell^T \mathbf{B}(u)\left\{\mathbf{Q}(\beta)^T \mathbf{H}^{-1}\mathbf{Q}(\beta)\right\}^{-1}\mathbf{B}(u)\, e_\ell, \tag{5.9}$$

and e_ℓ *is the* p*-dimensional vector with "1" as its* ℓ^{th} *element and "0" as other elements.*

Remark. Under the order assumption for the number of B-spline basis functions J_n in Theorems 5 and 6, the optimal order for the number of interior knots such that $N = J_n - q \asymp n^{1/(2r+1)}$ can be achieved by the proposed profile estimation procedure. Then the uniform convergence rate for $|\widehat{m}_\ell(u) - m_\ell(u)|$ is $O_p\left(J_n^{-r/(2r+1)}\right)$.

5.3 Hypothesis tests

Applying the profile estimation method described in Section 5.2, we propose the Rao-score-type hypothesis test to check the significance of genetic factor Z_2 by setting up the null and alternative hypotheses as $H_0 : m_2(.) = 0$ versus $H_1 : m_2(.) \neq 0$. Since each nonparametric function $m_2(u) \approx B(u)^{\mathrm{T}} \gamma_2$, the null and alternative hypotheses can be written as $H_0 : \gamma_2 = \mathbf{0}_{J_n}$ versus $H_1 : \gamma_2 \neq \mathbf{0}_{J_n}$. Let $\widehat{\gamma}^{\mathrm{N}} = \left\{ \left(\widehat{\gamma}_1^{\mathrm{N}}\right)^{\mathrm{T}}, \left(\widehat{\gamma}_2^{\mathrm{N}}\right)^{\mathrm{T}} \right\}^{\mathrm{T}}$ be the maximizer of $L_n\left(\gamma, \widehat{\beta}, \widehat{\alpha}, \lambda, \sigma^2\right)$ given in (5.5) under H_0, thus $\widehat{\gamma}_2^{\mathrm{N}} = \mathbf{0}$. Let

$$\begin{aligned}
\mathbf{Q}_{i,(1)}(\beta) &= \left[B\left(\mathbf{X}_i^{\mathrm{T}}\beta\right)\right]_{J_n \times 1}, \\
\mathbf{Q}_{i,(2)}(\beta) &= \left[B\left(\mathbf{X}_i^{\mathrm{T}}\beta\right) Z_{i2}\right]_{J_n \times 1}, \\
\mathbf{Q}_{(1)}(\beta) &= \left[\left\{\mathbf{Q}_{1,(1)}(\beta), \ldots, \mathbf{Q}_{n,(1)}(\beta)\right\}^{\mathrm{T}}\right]_{n \times J_n}, \\
\mathbf{Q}_{(2)}(\beta) &= \left[\left\{\mathbf{Q}_{1,(2)}(\beta), \ldots, \mathbf{Q}_{n,(2)}(\beta)\right\}^{\mathrm{T}}\right]_{n \times J_n}.
\end{aligned}$$

Define the score function as

$$\begin{aligned}
s_{2n}\left(\widehat{\gamma}^{\mathrm{N}}, \widehat{\beta}, \widehat{\alpha}\right) &= \partial L_n(\widehat{\gamma}^{\mathrm{N}}, \widehat{\beta}, \widehat{\alpha}, \lambda, \sigma^2)\sigma^2 / \partial \gamma_2 \\
&= \mathbf{Q}_{(2)}\left(\widehat{\beta}\right)^{\mathrm{T}} \mathbf{H}^{-1}\left(\mathbf{Y} - \mathbf{T}\widehat{\alpha} - \mathbf{Q}_{(1)}\left(\widehat{\beta}\right)\widehat{\gamma}_1^{\mathrm{N}}\right),
\end{aligned}$$

$$\mathbf{\Omega}_n = \mathbf{Q}\left(\beta^0\right)^{\mathrm{T}} \mathbf{H}^{-1} \mathbf{Q}\left(\beta^0\right) = \begin{pmatrix} \mathbf{\Omega}_{n,11} & \mathbf{\Omega}_{n,12} \\ \mathbf{\Omega}_{n,21} & \mathbf{\Omega}_{n,22} \end{pmatrix},$$

where $\mathbf{\Omega}_{n,kk'} = \mathbf{Q}_{(k)}\left(\beta^0\right)^{\mathrm{T}} \mathbf{H}^{-1} \mathbf{Q}_{(k')}\left(\beta^0\right)$, for $k, k' = 1, 2$.

Define the test statistic

$$\mathcal{T}_n = \sigma^{-2} s_{2n}\left(\widehat{\gamma}^{\mathrm{N}}, \widehat{\beta}, \widehat{\alpha}\right)^{\mathrm{T}} \mathbf{\Omega}_n^{22} s_{2n}\left(\widehat{\gamma}^{\mathrm{N}}, \widehat{\beta}, \widehat{\alpha}\right), \tag{5.10}$$

where $\mathbf{\Omega}_n^{22} = \left(\mathbf{\Omega}_{n,22} - \mathbf{\Omega}_{n,21}\mathbf{\Omega}_{n,11}^{-1}\mathbf{\Omega}_{n,12}\right)^{-1}$.

Theorem 7 *Under Conditions (C1)-(C6) in Appendix B, and* $n^{1/2r} << J_n << n^{1/4}$*, we have under* H_0*, as* $n \to \infty$*,*

$$\{2J_n\}^{-1/2} \{\mathcal{T}_n - J_n\} \to N\,(0{,}1)\,.$$

We can also extend this theorem to the more general case where several genetic factors are involved together as in model (5.7) in section 5.2.2. Details of the theorem is given in Appendix B.

Remark. Note that $\mathcal{T}_n$ defined in (5.10) contains population parameters, therefore, we cannot use it directly as a test statistic. To carry out the Rao-score-type hypothesis test, we instead use $\widehat{\mathcal{T}}_n$ as the test statistic. $\widehat{\mathcal{T}}_n$ has the same form as $\mathcal{T}_n$ with $\mathbf{\Omega}_{n,kk'}$ replaced by its consistent estimate $\mathbf{Q}_{(k)}\left(\widehat{\beta}\right)^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{Q}_{(k')}\left(\widehat{\beta}\right)$ and σ^2 replaced by consistent estimate $\widehat{\sigma}^2$. For implementation, the critical value of the test statistic is calculated from the chi-square distribution with $2J_n$ degrees of freedom.

5.4 Estimation of Variance Components

In Sections 5.2 and 5.3, we assume that λ is known when we estimate and make inference of the mean parameters. In this section, We propose a profile restricted maximum likelihood (REML) method to estimate λ. Denote $\widetilde{\mathbf{Y}} = \mathbf{Y} - \mathbf{T}\widehat{\alpha}$, $\mathbf{Q} = \mathbf{Q}\left(\widehat{\beta}\right)$ and $\mathbf{P_Q} = \mathbf{H}^{-1} - \mathbf{H}^{-1}\mathbf{Q}\left(\mathbf{Q}^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{Q}\right)^{-1}\mathbf{Q}^{\mathrm{T}}\mathbf{H}^{-1}$. For given $\widehat{\beta}$ and $\widehat{\alpha}$, the log-restricted likelihood function for estimating $\left(\lambda, \sigma^2\right)$ is given as

$$
\begin{aligned}
& L_r\left(\lambda, \sigma^2\right) \\
= & -\frac{1}{2} \log |\mathbf{H}|-\frac{1}{2} \log \left|\mathbf{Q}^{\mathrm{T}} \mathbf{H}^{-1} \mathbf{Q}\right|-\frac{1}{2 \sigma^2}\left(\widetilde{\mathbf{Y}}-\mathbf{Q} \widehat{\gamma}\right)^{\mathrm{T}} \mathbf{H}^{-1}\left(\widetilde{\mathbf{Y}}-\mathbf{Q} \widehat{\gamma}\right)-\frac{n-2 J_n}{2} \log \sigma^2 \\
= & -\frac{1}{2} \log |\mathbf{H}|-\frac{1}{2} \log \left|\mathbf{Q}^{\mathrm{T}} \mathbf{H}^{-1} \mathbf{Q}\right|-\frac{1}{2 \sigma^2} \widetilde{\mathbf{Y}}^{\mathrm{T}} \mathbf{P}_{\mathbf{Q}} \widetilde{\mathbf{Y}}-\frac{n-2 J_n}{2} \log \sigma^2 . \qquad (5.11)
\end{aligned}
$$

Thus for given λ, by maximizing (5.11) we obtain the estimator for σ^2 as

$$
\widetilde{\sigma}^2(\lambda)=\left(n-2 J_n\right)^{-1}\left(\widetilde{\mathbf{Y}}^{\mathrm{T}} \mathbf{P}_{\mathbf{Q}} \widetilde{\mathbf{Y}}\right) .
$$

Replacing σ^2 by $\widetilde{\sigma}^2(\lambda)$ in (5.11), we have

$$
L_r^*(\lambda)=-\frac{1}{2} \log |\mathbf{H}|-\frac{1}{2} \log \left|\mathbf{Q}^{\mathrm{T}} \mathbf{H}^{-1} \mathbf{Q}\right|-\frac{n-2 J_n}{2} \log \left(\widetilde{\mathbf{Y}}^{\mathrm{T}} \mathbf{P}_{\mathbf{Q}} \widetilde{\mathbf{Y}}\right)+\text{constant}. \qquad (5.12)
$$

Then the estimate $\widehat{\lambda}$ of λ is obtained by maximizing the objective function (5.12). Thus σ^2 is estimated by

$$
\widehat{\sigma}^2=\widetilde{\sigma}^2(\lambda)=\left(n-2 J_n\right)^{-1}\left(\widetilde{\mathbf{Y}}^{\mathrm{T}} \widehat{\mathbf{P}}_{\mathbf{Q}} \widetilde{\mathbf{Y}}\right), \qquad (5.13)
$$

where $\widehat{\mathbf{P}}_{\mathbf{Q}}$ is defined in the same way as $\mathbf{P}_{\mathbf{Q}}$ with λ replaced by $\widehat{\lambda}$. The objective function (5.12) requires determinant and inverse of $\mathbf{H}$ which is an $n \times n$ matrix, and thus the computation can be demanding when the sample size is large. We use eigenvalue decomposition to handle the kinship matrix $\mathbf{K}$, so that it can be decomposed as $\mathbf{K}=\boldsymbol{\Phi} \boldsymbol{\Delta} \boldsymbol{\Phi}^{\mathrm{T}}$, where $\boldsymbol{\Delta}=\operatorname{diag}(\delta_1, \ldots, \delta_n)$ is a diagonal matrix with eigenvalues and $\boldsymbol{\Phi}$ is an $n \times n$ matrix consisting of eigenvectors. It can be derived that $\log |\mathbf{H}|=\sum_{i=1}^n \log \left(\delta_i \lambda+1\right)$ and $\mathbf{H}^{-1}=\boldsymbol{\Phi}(\boldsymbol{\Delta} \lambda+\mathbf{I})^{-1} \boldsymbol{\Phi}^{\mathrm{T}}$. Let $\mathbf{q}_{2 J_n \times n}=\left(\mathbf{q}_1, \ldots, \mathbf{q}_n\right)=\mathbf{Q}^{\mathrm{T}} \boldsymbol{\Phi}$. Thus, we have

$$
\mathbf{Q}^{\mathrm{T}} \mathbf{H}^{-1} \mathbf{Q}=\mathbf{Q}^{\mathrm{T}} \boldsymbol{\Phi}(\boldsymbol{\Delta} \lambda+\mathbf{I})^{-1} \boldsymbol{\Phi}^{\mathrm{T}} \mathbf{Q}=\mathbf{a}(\boldsymbol{\Delta} \lambda+\mathbf{I})^{-1} \mathbf{a}^{\mathrm{T}}=\sum_{i=1}^n\left(\delta_i \lambda+1\right)^{-1} \mathbf{q}_i \mathbf{q}_i^{\mathrm{T}},
$$

which is a $2 J_n \times 2 J_n$ matrix.

5.5 Computational Algorithm

We propose the Newton-Raphson algorithm to estimate parameters β and α, nonparametric functions $m_\ell(\cdot)$, λ and σ^2. Let $\widehat{\lambda}^{(k)}$ and $\widehat{\sigma}^{2,(k)}$ be the k-th iterative estimate of λ and σ^2. The $(k+1)$-th estimates $\widehat{\beta}_{-1}^{(k+1)}$ and $\widehat{\alpha}^{(k+1)}$ of β_{-1} and α are obtained by maximizing the objective function (5.6) with σ^2 and $\mathbf{H}$ replaced by $\widehat{\sigma}^{2,(k)}$ and $\widehat{\mathbf{H}}^{(k)} = \widehat{\lambda}^{(k)}\mathbf{K} + \mathbf{I}_n$, and $\widehat{\beta}_1^{(k+1)} = \sqrt{1 - \left\|\widehat{\beta}_{-1}^{(k+1)}\right\|_2^2}$. Then $\widehat{\beta}^{(k+1)} = \left(\widehat{\beta}_1^{(k+1)}, \widehat{\beta}_{-1}^{(k+1)\mathrm{T}}\right)^{\mathrm{T}}$. The $(k+1)$-th estimate $\widehat{\gamma}^{(k+1)}$ of γ is given as

$$\widehat{\gamma}^{(k+1)} = \left\{\mathbf{Q}\left(\widehat{\beta}^{(k+1)}\right)^{\mathrm{T}}\left(\widehat{\mathbf{H}}^{(k)}\right)^{-1}\mathbf{Q}\left(\widehat{\beta}^{(k+1)}\right)\right\}^{-1}\left\{\mathbf{Q}\left(\widehat{\beta}^{(k+1)}\right)^{\mathrm{T}}\left(\widehat{\mathbf{H}}^{(k)}\right)^{-1}\left(\mathbf{Y} - \mathbf{T}\widehat{\alpha}^{(k+1)}\right)\right\}.$$

The $(k+1)$-th estimate $\widehat{\lambda}^{(k+1)}$ of λ is obtained by maximizing the objective function (5.12) and $\widehat{\sigma}^{2,(k+1)}$ is from (5.13) with β and α replaced by $\widehat{\beta}^{(k+1)}$ and $\widehat{\alpha}^{(k+1)}$. Let $\widehat{\theta}^{(k)} = \left(\widehat{\beta}^{(k)\mathrm{T}}, \widehat{\alpha}^{(k)\mathrm{T}}\right)^{\mathrm{T}}$. The iteration is stopped at the $(k+1)^{\text{th}}$ step if $\left\|\widehat{\theta}^{(k+1)} - \widehat{\theta}^{(k)}\right\| < \delta$ and $\left|\widehat{\lambda}^{(k+1)} - \widehat{\lambda}^{(k)}\right| < \delta$ for some small threshold value δ. The initial estimates of β and α are obtained by letting $\mathbf{H} = \mathbf{I}_n$.

The nonparametric functions $m_\ell(\cdot)$ are approximated by cubic spline ($q = 4$) with the number of interior knots N selected by minimizing the BIC criterion on the range $1 \le N \le \left[n^{1/2r}\right]$ given as

$$\mathrm{BIC}(N) = -2L_n\left(\widehat{\gamma}, \widehat{\beta}, \widehat{\alpha}, \widehat{\lambda}, \widehat{\sigma}^2\right) + p(N+q)(\log n).$$

Then one selects the optimal number of interior knots $\widehat{N} = \mathrm{argmin}_N \mathrm{BIC}(N)$. Commonly, the nonparametric functions are assumed to have the second order smoothness such that $r = 2$.

5.6 Numerical Examples

5.6.1 Simulation studies

Example 1. In this example, we wish to show the normality properties for the proposed estimators. We generate responses Y_i from the following model given as

$$Y_i = m_0(\mathbf{X}_i^{\mathrm{T}}\beta^0) + \sum\nolimits_{\ell=1}^{3} m_\ell(\mathbf{X}_i^{\mathrm{T}}\beta^0)Z_{i\ell} + \mathbf{T}_i^{\mathrm{T}}\alpha^0 + \xi_i + \varepsilon_i, \tag{5.14}$$

where $i = 1, \ldots, 500$, $\mathbf{X}_i = (X_{ik}, 1 \le k \le 3)^{\mathrm{T}}$ are simulated environmental effects, which are generated from independent uniform distributions on $[0, 1]$, $\mathbf{T}_i = (T_{is}, 1 \le s \le 3)^{\mathrm{T}}$ are covariates generated from the multivariate normal distribution with mean 0, marginal variance 1, and an AR-1 correlation matrix with autocorrelation coefficient 0.5, and $Z_{i\ell}$ $(1 \le \ell \le 3)$ are simulated genetic factors, which have three possible genotype categories represented by AA, Aa and aa and coded as 2, 1, and 0 with frequency $\left\{P_A^2, 2P_A(1-P_A), (1-P_A)^2\right\}$, with $P_A = 0.5$ which is the allele frequency for allele A. We use the empirical centered value of $Z_{i\ell}$ to generate Y_i. The error terms $\xi = (\xi_1, \ldots, \xi_n)^{\mathrm{T}} \tilde{} \mathrm{MVN}_n\left(\mathbf{0}, \lambda\sigma^2\mathbf{K}\right)$ and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^{\mathrm{T}} \tilde{} \mathrm{MVN}_n\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$, where $\sigma^2 = 0.5$, $\lambda = 1$ and $\mathbf{K}$ is an $n \times n$ AR-1 correlation matrix with autocorrelation coefficient 0.5. We let $\beta^0 = (3, 2, 1)/\sqrt{14}$, $\alpha^0 = (2, 1, -1)$, $m_0(u) = 1 + 5\sin(\pi u)$, and $m_\ell(u) = c \times \sin(2\pi u)$ for $1 \le \ell \le 3$, where c ranges from 0 to 0.5 with increment 0.1. We generate 200 simulated samples.

We compare the performance of the proposed estimators for β^0, α^0 and $m_\ell(\cdot), 0 \le \ell \le 3$, by estimating the correlation matrix $\mathbf{H}$ with the method given in Section 5.4 and by assuming $\mathbf{H} = \mathbf{I}_n$ so that correlations between subjects are ignored. We obtain very small sample bias for the estimates obtained by the estimated $\mathbf{H}$ and by letting $\mathbf{H} = \mathbf{I}_n$. For example, for $c = 0.5$, we obtain sample bias as 0.0051, -0.0047 and 0.0021 for estimates of β by the estimated $\mathbf{H}$ and 0.0089, -0.0058 and -0.0020 by letting $\mathbf{H} = \mathbf{I}_n$. It confirms that the resulting estimators are unbiased estimators even if $\mathbf{H}$ is misspecified. Tables 5.1 and 5.2 report the empirical standard errors (ESE) and the mean squared errors (MSE) $\times 10^{-2}$ defined as the sample mean of $\left\|\widehat{\beta} - \beta^0\right\|_2^2$ and $\left\|\widehat{\alpha} - \alpha^0\right\|_2^2$ among 200 replications by the estimated $\mathbf{H}$ and by assuming $\mathbf{H} = \mathbf{I}_n$ for $c = (0.5, 0.4, 0.3, 0.2, 0.1)$. We observe that the ESE and MSE values of the estimates obtained by the estimated $\mathbf{H}$ are smaller than the corresponding values of the estimates obtained by $\mathbf{H} = \mathbf{I}_n$ for all cases. This result demonstrates that the former estimators are more efficient and accurate than the latter ones. By ignoring possible correlations, the estimators may lose some efficiency. To evaluate the performance of the estimated nonparametric functions $m_\ell(\cdot)$ for $0 \leq l \leq 3$, we define the mean integrated squared error (MISE) as the average of

$$\mathrm{ISE}\left(\widehat{m}_\ell\right) = n^{-1} \sum\nolimits_{i=1}^{n} \left\{\widehat{m}_\ell\left(\mathbf{X}_i^{\mathrm{T}}\widehat{\beta}\right) - m_\ell(\mathbf{X}_i^{\mathrm{T}}\beta^0)\right\}^2,$$

for $0 \leq l \leq 3$ among the 200 replications. Table 5.3 reports the MISEs for the spline estimates $\widehat{m}_\ell(\cdot)$ for $0 \leq l \leq 3$ and $c = (0.5, 0.4, 0.3, 0.2, 0.1)$ by using the estimated $\mathbf{H}$ and by assuming $\mathbf{H} = \mathbf{I}_n$. It shows that the MISE values for the estimates obtained by the estimated $\mathbf{H}$ are smaller than those values for the estimates obtained by $\mathbf{H} = \mathbf{I}_n$. To evaluate the asymptotic normality results for the parameter estimators as established in Theorem 5, Table 5.4 and Table 5.5 report the average asymptotic standard errors (ASE) calculated based on (5.8) and empirical coverage probabilities of the 95% confidence intervals for the estimates of β^0 and α^0 by using the proposed method with the estimated $\mathbf{H}$. The empirical coverage probabilities are close to the nomimal confidence level 95% for most of the cases.

Comparing the ASEs with the ESEs reported in Table 5.1 and Table 5.2, we observe that values are very close for the corresponding cases, suggesting that the asymptotic covariance matrix is correctly derived. Moreover, Table 5.6 shows the average estimated values and the empirical standard errors for the variance components λ and σ^2. The average estimated values are close to the true values for all cases, which indicates that the proposed estimation procedure for variance components given in Section 5.4 performs well.

Table 5.1: The empirical standard errors (ESE) and the mean squared errors (MSE) $\times 10^{-2}$ for the estimates of β^0 among 200 replications by the estimated $\mathbf{H}$ and by assuming $\mathbf{H} = \mathbf{I}_n$ for $c = (0.5, 0.4, 0.3, 0.2, 0.1)$.

	Estimation by using the estimated $\mathbf{H}$				Estimation by assuming $\mathbf{H} = \mathbf{I}_n$			
	ESE			MSE	ESE			MSE
c	β_1^0	β_2^0	β_3^0		β_1^0	β_2^0	β_3^0	
0.5	0.0066	0.0086	0.0105	0.0226	0.0080	0.0107	0.0128	0.0342
0.4	0.0070	0.0101	0.0123	0.0302	0.0085	0.0118	0.0141	0.0409
0.3	0.0066	0.0096	0.0111	0.0260	0.0076	0.0110	0.0128	0.0344
0.2	0.0065	0.0097	0.0109	0.0255	0.0075	0.0113	0.0126	0.0342
0.1	0.0076	0.0103	0.0109	0.0282	0.0088	0.0117	0.0129	0.0382

Table 5.2: The empirical standard errors (ESE) and the mean squared errors (MSE) $\times 10^{-2}$ for the estimates of α^0 among 200 replications by the estimated $\mathbf{H}$ and by assuming $\mathbf{H} = \mathbf{I}_n$ for $c = (0.5, 0.4, 0.3, 0.2, 0.1)$.

	Estimation by using the estimated $\mathbf{H}$				Estimation by assuming $\mathbf{H} = \mathbf{I}_n$			
	ESE			MSE	ESE			MSE
c	α_1^0	α_2^0	α_3^0		α_1^0	α_2^0	α_3^0	
0.5	0.0409	0.0539	0.0444	0.6569	0.0504	0.0603	0.0510	0.8858
0.4	0.0458	0.0536	0.0442	0.6956	0.0541	0.0590	0.0536	0.9361
0.3	0.0452	0.0519	0.0490	0.7135	0.0517	0.0602	0.0561	0.9416
0.2	0.0450	0.0467	0.0453	0.6262	0.0530	0.0549	0.0504	0.8377
0.1	0.0441	0.0453	0.0455	0.6038	0.0512	0.0492	0.0540	0.7933

Table 5.3: The MISEs for the spline estimates $\widehat{m}_\ell(\cdot)$ of $m_\ell(\cdot)$ for $0 \le l \le 3$ and $c = (0.5, 0.4, 0.3, 0.2, 0.1)$ by using the estimated $\mathbf{H}$ and by assuming $\mathbf{H} = \mathbf{I}_n$.

	MISE by using the estimated $\mathbf{H}$				Estimation by assuming $\mathbf{H} = \mathbf{I}_n$			
c	$m_0(\cdot)$	$m_1(\cdot)$	$m_2(\cdot)$	$m_3(\cdot)$	$m_0(\cdot)$	$m_1(\cdot)$	$m_2(\cdot)$	$m_3(\cdot)$
0.5	0.0210	0.0181	0.0201	0.0211	0.0245	0.0245	0.0261	0.0281
0.4	0.0230	0.0208	0.0191	0.0192	0.0269	0.0282	0.0257	0.0255
0.3	0.0227	0.0193	0.0176	0.0178	0.0264	0.0274	0.0244	0.0242
0.2	0.0216	0.0174	0.0185	0.0169	0.0252	0.0243	0.0250	0.0238
0.1	0.0194	0.0181	0.0167	0.0174	0.0237	0.0247	0.0229	0.0225

Table 5.4: The average asymptotic standard errors (ASE) calculated from (5.8) and empirical coverage probabilities (CP) of 95% confidence intervals based on 200 replications for the estimates of $\beta^0 = \left(\beta_1^0, \beta_2^0, \beta_3^0\right)^{\mathrm{T}}$.

	β_1^0		β_2^0		β_3^0	
c	ASE	CP	ASE	CP	ASE	CP
0.5	0.0066	0.950	0.0094	0.965	0.0110	0.960
0.4	0.0067	0.930	0.0095	0.935	0.0110	0.940
0.3	0.0067	0.945	0.0095	0.940	0.0111	0.945
0.2	0.0067	0.945	0.0095	0.950	0.0111	0.945
0.1	0.0068	0.915	0.0096	0.915	0.0112	0.955

Table 5.5: The average asymptotic standard errors (ASE) calculated from (5.8) and empirical coverage probabilities (CP) of 95% confidence intervals based on 200 replications for the estimates of $\alpha^0 = \left(\alpha_1^0, \alpha_2^0, \alpha_3^0\right)^{\mathrm{T}}$.

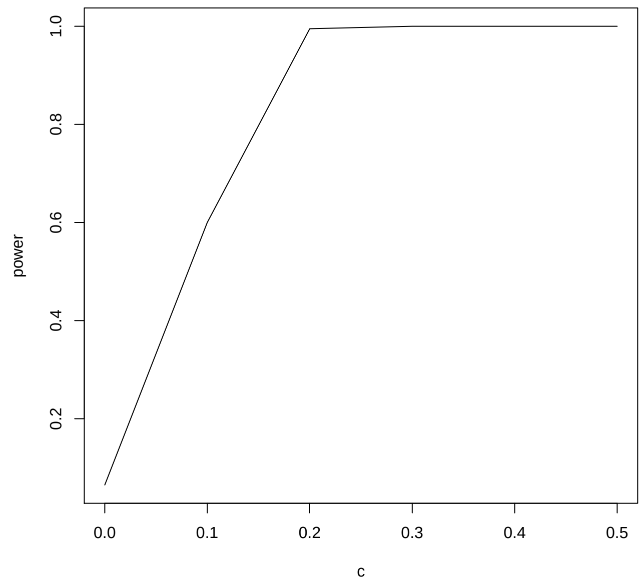
	α_1^0		α_2^0		α_3^0	
c	ASE	CP	ASE	CP	ASE	CP
0.5	0.0448	0.955	0.0500	0.930	0.0448	0.950
0.4	0.0445	0.935	0.0500	0.945	0.0447	0.940
0.3	0.0446	0.960	0.0500	0.940	0.0446	0.920
0.2	0.0444	0.950	0.0499	0.965	0.0446	0.945
0.1	0.0445	0.950	0.0499	0.965	0.0448	0.925

Table 5.6: The sample mean (Mean) and sample standard error (SE) of the estimates $\widehat{\lambda}$ and $\widehat{\sigma}^2$ based on 200 replications for $c = (0.5, 0.4, 0.3, 0.2, 0.1)$.

c		0.5	0.4	0.3	0.2	0.1	0
λ	Mean	1.0621	1.0012	1.0878	1.0791	1.0503	1.0545
	SE	0.2853	0.2904	0.3036	0.2909	0.2775	0.2791
σ^2	Mean	0.4940	0.4955	0.4887	0.4868	0.4899	0.4940
	SE	0.0519	0.0536	0.0535	0.0517	0.0519	0.0569

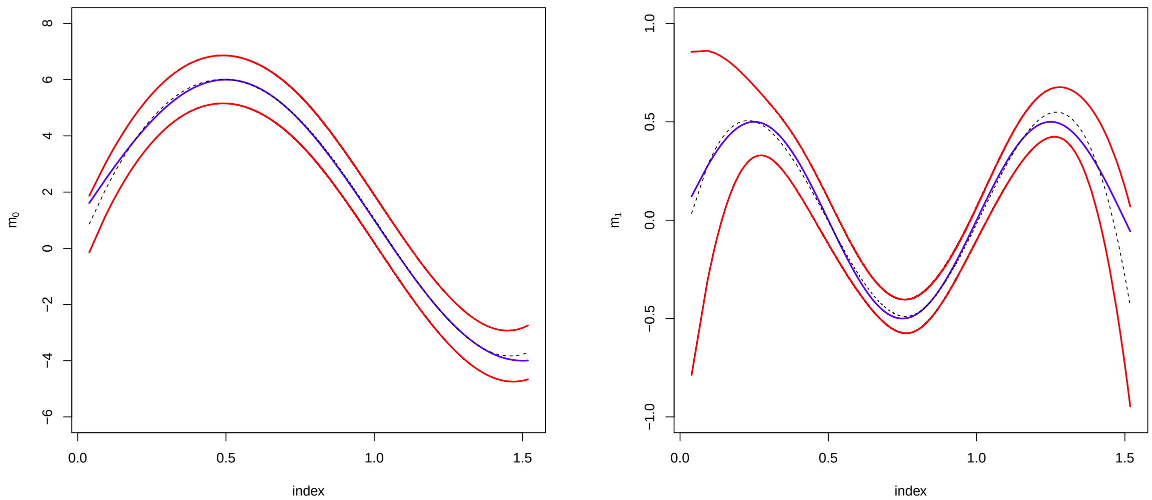
Next we perform the score test as described in Section 5.3 for testing $H_0 : m_\ell(\cdot) = 0$, for $\ell = 1, 2, 3$ versus $H_1 : m_\ell(\cdot) \neq 0$ for some $\ell \in (1, 2, 3)$ for $c = (0.5, 0.4, 0.3, 0.2, 0.1, 0)$. Figure 5.1 shows the plot of the power function against the c value at type I error 0.05. We observe that when $c = 0$, the empirical power size is 0.065 which is close to the type I error. Moreover, the power increases rapidly to 1 as the c value increases, which demostrates that the proposed score test is a powerful test. For visualization of the actual function estimates, Figure 5.2 shows the estimated curves $\widehat{m}_\ell(\cdot)$ (solid lines) against index, $\ell = 0, 1$, the true functions $m_\ell(\cdot)$ (dashed lines), and the upper and lower 95% pointwise confidence intervals (upper and lower solid lines) for $c = 0.5$, $n = 500$. We can observe that the proposed

Figure 5.1: Plot of the power function against the c value.



estimators perform well.

Figure 5.2: Plots of the estimated curves $\widehat{m}_\ell(\cdot)$ (solid lines) against index, $\ell = 0, 1$, the true functions $m_\ell(\cdot)$ (dashed lines), and the upper and lower 95% pointwise confidence intervals (upper and lower solid lines) for $c = 0.5$, $n = 500$.



Example 2. In this example, we wish to compare the score test performance in the proposed model with a parametric linear mixed model by assuming that the main effect of $\mathbf{X}$ as well as the interaction effects with $\mathbf{Z}$ are linear, and a semiparametric mixed model by assuming the main effect $m_0(.)$ of $\mathbf{X}$ is nonlinear but the interaction effects with $\mathbf{Z}$ are linear.

Suppose the proposed semiparametric mixed model with nonlinear interaction effects (SPNLMM) is given as

$$Y_i = m_0(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}^0) + m_1(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}^0)Z_i + \mathbf{T}_i^{\mathrm{T}}\boldsymbol{\alpha}^0 + \xi_i + \varepsilon_i, \tag{5.15}$$

the parametric linear mixed model (LMM) is given as

$$Y_i = a_0 + \sum_{k=1}^{3} b_k X_{ik} + (d + \sum_{k=1}^{3} d_k X_{ik})Z_i + \mathbf{T}_i^{\mathrm{T}}\boldsymbol{\alpha}^0 + \xi_i + \varepsilon_i, \tag{5.16}$$

and the semiparametric mixed model with linear interaction effects (SPLMM) is given as

$$Y_i = m_0(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}^0) + (d + \sum_{k=1}^{3} d_k X_{ik})Z_i + \mathbf{T}_i^{\mathrm{T}}\boldsymbol{\alpha}^0 + \xi_i + \varepsilon_i. \tag{5.17}$$

First, we generate responses Y_i from model (5.15) where $i = 1, ..., n$, $n = 2191$, $\mathbf{X}_i = (X_{ik}, 1 \le k \le 3)^{\mathrm{T}}$ are simulated environmental effects, which are generated from independent uniform distributions on [0,1], $\mathbf{T}_i = (T_{is}, 1 \le s \le 3)^{\mathrm{T}}$ are covariates generated from the multivariate normal distribution with mean 0, marginal variance 1, and an AR-1 correlation matrix with autocorrelation coefficient 0.5. Z_i are simulated genetic factors, which have three possible genotype categories represented by AA, Aa and aa , and coded as 1, 0 and -1 with allele frequency 0.5. We use the empirical centered value of Z_i to generate Y_i. The error terms $\boldsymbol{\xi} = (\xi_1, ..., \xi_n)^{\mathrm{T}} \sim \mathrm{MVN}_n(\mathbf{0}, \lambda\sigma^2\mathbf{K})$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_n)^{\mathrm{T}} \sim \mathrm{MVN}_n(\mathbf{0}, \sigma^2\mathbf{I})$, where $\sigma^2 = 0.5$, $\lambda = 1$ and $\mathbf{K}$ is a 2191×2191 correlation matrix calculated from the real SNPs in 22 chromosomes. We let $\boldsymbol{\beta}^0 = (3, 2, 1)/\sqrt{14}$, $\boldsymbol{\alpha}^0 = (2, 1, -1)$, $m_0(u) = 1 + 5\sin(\pi u)$ and $m_1(u) = c \times \sin(2\pi u)$, where c ranges from 0 to 0.5 with increment 0.1. 200 simulated samples are generated.

Then we perform the score test as described in Section 5.3 for testing $H_0 : m_1(\cdot) = 0$ versus $H_1 : m_1(\cdot) \neq 0$ for $c = (0, 0.1, 0.2, 0.3, 0.4, 0.5)$ in model (5.15). Meanwhile, we perform the score test for testing $H_0 : d, d_k = 0$ for $k = 1, 2, 3$ versus $H_1 : d \neq 0$, or $d_k \neq 0$ for some $k = 1, 2, 3$ in model (5.16) and model (5.17). Note that model (5.16) by assuming a linear main effect is misspecified, but the semiparametric model (5.17) is correctly specified under the null hypothesis H_0. Score tests are performed by assuming model (5.16) and model (5.17) as the alternative models, respectively. The critical value for both tests is calculated from the chi-square distribution with 4 degrees of freedom at significance level

0.05.

Table 5.7 reports the powers of the score tests for these three models SPNLMM, LMM and SPLMM for $c = (0, 0.1, 0.2, 0.3, 0.4, 0.5)$ with 200 simulated samples. Clearly we observe that for SPNLMM model, the power size at $c = 0$ (H_0 is true) is close to the nominal significance level 0.05, which confirms the asymptotic null distribution of the test statistic. The power increases to 1 rapidly as c increases. The results illustrate that the proposed score test is a powerful test. For the LMM model, the power is much larger than 0.05 when $c = 0$, since the model is misspecified under H_0. For the SPLMM model, the power is around 0.05 when $c = 0$, since the model is correctly specified under H_0, so that the score test works well under H_0. For both of these two misspecified models under H_1, the power increases very slowly as the value of c increases. This result indicates that when the actual nonlinear interaction effect is misspecified, the score test will become less powerful.

We also evaluate the performance of the proposed model when the true underlying interactions are linear. First, we generate Y_i from model (5.16) with $\boldsymbol{\alpha}^0 = (2, 1, -1)$, $a_0 = 0$, $(b_1, b_2, b_3) = (3, 2, 1)/\sqrt{14}$, $d = c \times 0.02$, $(d_1, d_2, d_3) = c \times (0.1, 0.2, 0.3)$, and c changes from 0 to 0.5 with increment 0.1, $i = 1, ..., 2191$. Table 5.8 reports the powers of the score tests for models SPNLMM and LMM, where we can see that as c increases, their powers increase to 1. Moreover, the powers for SPNLMM model are close to those for LMM model. Then, we generate Y_i from model (5.17) with $\boldsymbol{\alpha}^0 = (2, 1, -1)$, $m_0(u) = 1 + 5\sin(\pi u)$, $d = c \times 0.02$, $(d_1, d_2, d_3) = c \times (0.1, 0.2, 0.3)$ and c changes from 0 to 0.5 with increment 0.1, $i = 1, ..., 2191$. Table 5.9 reports the powers of the score tests for SPNLMM and SPLMM. From table 5.9, we can see that their powers increase to 1 as c increases, and the powers for SPNLMM model are very close to those for SPLMM model. The results indicate that even if the actual interactions between environment and gene follow a linear structure, the score test for our proposed model remains powerful.

Table 5.7: Powers of the score tests for the three models SPNLMM, LMM and SPLMM when the true model is SPNLMM based on 200 replications for different c.

c	0	0.1	0.2	0.3	0.4	0.5
SPNLMM	0.065	0.600	0.995	1.000	1.000	1.000
LMM	0.090	0.100	0.135	0.215	0.325	0.330
SPLMM	0.060	0.080	0.160	0.405	0.620	0.760

Table 5.8: Powers of the score tests for SPNLMM and LMM when the true model is LMM based on 200 replications for $c = (0, 0.1, 0.2, 0.3, 0.4, 0.5)$.

c	0	0.1	0.2	0.3	0.4	0.5
SPNLMM	0.030	0.135	0.425	0.850	0.980	1.000
LMM	0.065	0.150	0.440	0.885	0.990	1.000

Table 5.9: Powers of the score tests for SPNLMM and SPLMM when the true model is SPLMM based on 200 replications for $c = (0, 0.1, 0.2, 0.3, 0.4, 0.5)$.

c	0	0.1	0.2	0.3	0.4	0.5
SPNLMM	0.060	0.155	0.480	0.825	0.990	1.000
SPLMM	0.030	0.110	0.475	0.885	1.000	1.000

Example 3. Besides power, to appropriately evaluate the efficiency of the score test, false discovery rates (FDR) should be considered as well. In this simulation study, we wish to compare both powers and FDR of the score tests among SPNLMM, LMM and SPLMM models.

First, we generate responses Y_i from the following model given as

$$Y_i = m_0(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}^0) + \sum_{\ell=1}^{5} m_\ell(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}^0)Z_{i\ell} + \mathbf{T}_i^{\mathrm{T}}\boldsymbol{\alpha}^0 + \xi_i + \varepsilon_i, \tag{5.18}$$

where $i = 1, ..., 2191$, $\mathbf{X}_i = (X_{ik}, 1 \leq k \leq 3)^{\mathrm{T}}$ are simulated environmental effects, which are generated from independent uniform distributions on [0,1], $\mathbf{T}_i = (T_{is}, 1 \leq s \leq 3)^{\mathrm{T}}$ are covariates generated from the multivariate normal distribution with mean 0, marginal variance 1, and an AR-1 correlation matrix with autocorrelation coefficient 0.5, and $Z_{i\ell}(1 \leq \ell \leq 5)$ are five real SNPs coded as 1, 0, -1 chosen from chromosome 4, denoted by SNP_1, SNP_2, SNP_3, SNP_4 and SNP_5. We use the empirical centered value of Z_i to generate Y_i. The error terms $\boldsymbol{\xi} = (\xi_1, ..., \xi_n)^{\mathrm{T}} \sim \mathrm{MVN}_n(\mathbf{0}, \lambda\sigma^2\mathbf{K})$ and $\boldsymbol{\epsilon} = (\epsilon_1, ..., \epsilon_n)^{\mathrm{T}} \sim \mathrm{MVN}_n(\mathbf{0}, \sigma^2\mathbf{I})$, where $\sigma^2 = 0.5$, $\lambda = 1$ and $\mathbf{K}$ is a 2191×2191 correlation matrix calculated from the real SNPs in 22 chromosomes. We let $\boldsymbol{\beta}^0 = (3, 2, 1)/\sqrt{14}$, $\boldsymbol{\alpha}^0 = (2, 1, -1)$, $m_0(u) = 1+5\sin(\pi u)$, $m_1(u) = 0.1\sin(2\pi u)$, $m_2(u) = 0.2\sin(2\pi u)$, $m_3(u) = 0.3\sin(2\pi u)$, $m_4(u) = 0.4\sin(2\pi u)$, and $m_5(u) = 0.5\sin(2\pi u)$. 200 simulated samples are generated.

We test the significance of SNP_1, SNP_2, SNP_3, SNP_4, SNP_5 and g other SNPs chosen from chromosome 4 respectively. After obtaining the p-value (p_{value}) for each SNP, we apply the multiple testing correction procedure for GWAS given in Cheverud (2001) and Nyholt (2004). Therefore, H_0 is rejected when $p_{\mathrm{value}} < \alpha = \alpha_0/M_{\mathrm{eff}}$, where α_0 is the overall type I error for the study and we let it be 0.05, and M_{eff} is the Cheverud-Nyholt estimate of the effective number of tests calculated by $M_{\mathrm{eff}} = 1 + M^{-1}\sum_{j=1}^{M}\sum_{k=1}^{M}(1 - r_{jk}^2)$, where M is the total number of SNPs in the study (i.e., $M = 5 + g$), and r_{jk} are the correlation coefficients of these M SNPs. Let $g = 2000, 5000, 10000$ for comparison. The detection powers for SNP_1, SNP_2, SNP_3, SNP_4 and SNP_5 are the proportions of simulation runs with p_{value} less than α among 200 replications respectively. FDR are calculated using the times of falsely discovered SNPs divided by the total times of discovered SNPs in 200 replications. Table 5.10 reports the powers and FDR of the score tests for SPNLMM, LMM and SPLMM models under model (5.18). The result indicates that score test for SPNLMM has much higher power and much lower FDR than LMM and SPLMM when the underlying G $\times$ E interactions are nonlinear, which confirms the efficiency of our proposed score test. Moreover, the power of the score test increases as proportion of variance explained by the SNP increases.

Table 5.10: Powers and false discovery rates (FDR) of the score tests for three models SPNLMM, LMM and SPLMM estimated based on 200 replications when the true underlying model is semiparametric mixed model with nonlinear G $\times$ E interactions for $g = 2000, 5000, 10000$. Columns of SNP_i shows the power of each SNP, TP shows true positives and FP shows false positives respectively.

g	SNP_1	SNP_2	SNP_3	SNP_4	SNP_5	FDR	TP	FP
				SPNLMM				
2000	0.035	0.505	0.995	1.000	1.000	0.032	3.535	0.115
5000	0.035	0.500	1.000	1.000	1.000	0.030	3.535	0.110
10000	0.030	0.500	1.000	1.000	1.000	0.029	3.530	0.105
				LMM				
2000	0.000	0.000	0.000	0.000	0.065	0.594	0.065	0.095
5000	0.000	0.000	0.000	0.000	0.035	0.682	0.035	0.075
10000	0.000	0.000	0.000	0.000	0.040	0.636	0.040	0.070
				SPLMM				
2000	0.000	0.000	0.005	0.005	0.050	0.613	0.060	0.095
5000	0.000	0.000	0.000	0.010	0.045	0.656	0.055	0.105
10000	0.000	0.000	0.005	0.005	0.040	0.655	0.050	0.095

We also evaluate the powers and FDR for SPNLMM model when the underlying interactions are linear. First, we generate 200 simulated samples from model given as

$$Y_i = a_0 + \sum_{k=1}^{3} b_k X_{ik} + \sum_{\ell=1}^{5}(d_\ell + \sum_{k=1}^{3} c_k X_{ik}) Z_{i\ell} + \mathbf{T}_i^{\mathrm{T}} \boldsymbol{\alpha}^0 + \xi_i + \varepsilon_i \qquad (5.19)$$

with $a_0 = 0$, $(b_1, b_2, b_3) = (3, 2, 1)/\sqrt{14}$, $(c_1, c_2, c_3) = (0.1, 0.2, 0.3)$, $\boldsymbol{\alpha}^0 = (2, 1, -1)$, and $(d_1, d_2, d_3, d_4, d_5) = (0.02, 0.05, 0.08, 0.11, 0.14)$, $i = 1, ..., 2191$. Table 5.11 reports the powers and FDR of the score tests for models SPNLMM and LMM under model (5.19). Then, generate 200 simulated samples from model given as

$$Y_i = m_0(\mathbf{X}_i^{\mathrm{T}} \boldsymbol{\beta}^0) + \sum_{\ell=1}^{5} (d_\ell + \sum_{k=1}^{3} c_k X_{ik}) Z_{i\ell} + \mathbf{T}_i^{\mathrm{T}} \boldsymbol{\alpha}^0 + \xi_i + \varepsilon_i \tag{5.20}$$

with $\boldsymbol{\beta}^0 = (3, 2, 1)/\sqrt{14}$, $\boldsymbol{\alpha}^0 = (2, 1, -1)$, $m_0(u) = 1 + 5\sin(\pi u)$, $(c_1, c_2, c_3) = (0.1, 0.2, 0.3)$ and $(d_1, d_2, d_3, d_4, d_5) = (0.02, 0.05, 0.08, 0.11, 0.14)$, $i = 1, ..., 2191$. Table 5.12 reports the powers and FDR of the score tests for models SPNLMM and SPLMM under model (5.20). The results indicate that when the true G $\times$ E interaction structure is linear, powers in SPNLMM are slightly lower than those in SPLMM and LMM, FDR in SPNLMM is a little bit higher than the FDR in SPLMM and LMM, which is reasonable. Since the differences are not very large, our proposed score test is still efficient. Also, the results suggest that it is better to explore the pattern of G $\times$ E interactions before we conduct hypothesis testing for gene associations. If nonlinear G $\times$ E interactions exist, our proposed model is the most powerful and efficient approach to test significant genes. If G $\times$ E interactions are approximately linear, it is better to apply LMM or SPLMM.

Table 5.11: Powers and false discovery rates (FDR) of the score tests for two models SPNLMM and LMM estimated based on 200 replications when the true underlying model is linear mixed model for $g = 2000, 5000, 10000$. Columns of SNP_i shows the power of each SNP, TP shows true positives and FP shows false positives respectively.

g	SNP_1	SNP_2	SNP_3	SNP_4	SNP_5	FDR	TP	FP
SPNLMM								
2000	0.030	0.305	0.715	0.935	0.985	0.043	2.970	0.135
5000	0.035	0.310	0.710	0.925	0.980	0.042	2.960	0.130
10000	0.015	0.285	0.675	0.910	0.945	0.036	2.830	0.105
LMM								
2000	0.040	0.420	0.755	0.945	0.995	0.026	3.155	0.085
5000	0.040	0.385	0.740	0.935	0.990	0.024	3.090	0.075
10000	0.035	0.370	0.690	0.915	0.975	0.019	2.985	0.070

Table 5.12: Powers and false discovery rates (FDR) of the score tests for two models SPNLMM and SPLMM estimated based on 200 replications when the true underlying model is semiparametric mixed model with linear G × E interactions for $g = 2000, 5000, 10000$. Columns of SNP_i shows the power of each SNP, TP shows true positives and FP shows false positives respectively.

g	SNP_1	SNP_2	SNP_3	SNP_4	SNP_5	FDR	TP	FP
SPNLMM								
2000	0.060	0.415	0.815	0.975	0.995	0.024	3.260	0.080
5000	0.070	0.410	0.765	0.970	0.990	0.021	3.205	0.070
10000	0.055	0.380	0.715	0.945	0.990	0.018	3.085	0.055
SPLMM								
2000	0.090	0.420	0.845	0.985	0.995	0.021	3.335	0.070
5000	0.085	0.400	0.835	0.980	0.995	0.018	3.295	0.055
10000	0.050	0.385	0.785	0.955	0.990	0.014	3.165	0.040

5.6.2 Empirical example

Research on causes of hypertension has brought tremendous attention due to its risk of serious health problems, including heart attack and stroke. It is known that hypertension is not only related to genes but also to some environmental factors such as sleeping hours (Gottlieb (2006)) and physical activity (Kokkinos (2009)). Therefore, people may wonder how genetic and environmental factors together influence people's blood pressure. Now we illustrate our method using data from the Framingham Heart Study (Dawber (1951)) to investigate significant genetic factors on systolic blood pressure (SBP) in the context of possible nonlinear G $\times$ E interactions. Let X_1 = *sleeping hours per day*, X_2 = *hours of light activity per day*, and X_3 = *hours of moderate activity per day* be the environmental factors. SBP is used as the response variable and SNPs located in chromosome 1 are considered as the genetic factors. The three possible allele combinations are coded as $Z = (1, 0, -1)$. After eliminating SNPs departure from Hardy-Weinberg equilibrim, there are 31042 SNPs remaining in our study. In addition to genotypes, sleeping hour and activity hours, we use other 3 covariates as the linear part in the semiparametric mixed model, which are T_1 = *whether work (1=YES, 0=NO)*, T_2 = *health condition (1=EXCELLENT, 2=GOOD, 3=FAIR, 4=POOR)*, T_3 = *whether live with others (1=YES, 0=NO)*. After deleting missing data, 2191 subjects remain in our study. Then a semiparametric mixed model is fitted as

$$Y_i = m_0(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}) + m_1(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta})Z_i + \mathbf{T}_i^{\mathrm{T}}\boldsymbol{\alpha} + \xi_i + \epsilon_i, i = 1, ..., 2191, \tag{5.21}$$

where $\mathbf{X}_i = (X_{i1}, ..., X_{i3})^{\mathrm{T}}$, $\mathbf{T}_i = (T_{i1}, ..., T_{i3})^{\mathrm{T}}$ and Z_i are the observed values of the environmental factors, covariates and genetic factor, respectively, and Y_i is the standardized value of SBP for the i-th subject. When we estimate the parameters and nonparametric functions in model (5.21), we center and standardize all predictors.

The number of individuals included in this GWAS is huge, and from the results we get for a few SNPs, the estimated values of λ are very close to each other. Therefore, to dramatically reduce the computation time, we apply the strategy in Kang (2010). We estimate λ once with the method given in section 5.4 and then fix the λ for all the SNPs in chromosome 1 such that we do not need to estimate the variance component repeatedly. The same model is fitted for each SNP and after the estimation procedure, we apply the proposed score test statistic in section 5.3 to test genetic significance by setting up the null hypothesis $H_0 : m_1(.) = 0$ in model (5.21) to obtain the p_{value} for each SNP. Then multiple testing correction procedure for GWAS given in Cheverud (2001) and Nyholt (2004) is applied such that H_0 is rejected when $p_{\text{value}} < \alpha = \alpha_0/M_{\text{eff}}$, where α_0 is the overall type I error for the study and we let it be 0.05, and M_{eff} is the Cheverud-Nyholt estimate of the effective number of tests calculated by $M_{\text{eff}} = 1 + M^{-1}\sum_{j=1}^{M}\sum_{k=1}^{M}(1 - r_{jk}^2)$, where $M = 31042$, which is the number of SNPs in the study, and r_{jk} are the correlation coefficients of SNPs. Then we obtain $\alpha = 3\times10^{-6}$. As a result, the only SNP ss66457217 with $p_{\text{value}} = 2.8\times10^{-6}$ is selected by the procedure. The corresponding gene of this SNP is IGSF3. In addition, another SNP ss66343404 with $p_{\text{value}} = 6.6\times10^{-6}$ can also be considered important to SBP, the corresponding gene is LPPR5. It has been shown that both IGSF3 and LPPR5 have positive associations with SBP in medical research.

We then use these two SNPs (Z_1=ss66457217, Z_2=ss66343404) together to fit the model given as

$$Y_i = m_0(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}) + \sum_{\ell=1}^{2} m_\ell(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta})Z_{i\ell} + \mathbf{T}_i^{\mathrm{T}}\boldsymbol{\alpha} + \xi_i + \epsilon_i, i = 1, ..., 2191 \tag{5.22}$$

Table 5.13 presents the estimated values (EST) for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ by the profile log-likelihood method in section 5.2, the associated standard errors (SE) and the p-values for testing whether each parameter is zero or not by using the asymptotic normality distribution. We observe that all of the index parameters (β_1, β_2, β_3) are significantly different from zero with p-values much smaller than 0.05. This result indicates that the SBP is highly related to sleeping hours and physical activity hours, which confirms the finding in the literature (Gottlieb (2006) and Kokkinos (2009)). The small p-values for the parameters α_1 and α_2 indicate that work status and health condition are important factors to SBP, p-value for parameter α_3 indicates that whether living with others is slightly significant to SBP.

Table 5.13: The estimated values (EST) for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ in model (5.22), the associated standard errors (SE) and the p-values for testing whether the parameters are zero or not in the real data example.

	β_1	β_2	β_3
EST	0.3718	0.6875	0.6238
SE	0.0615	0.0296	0.0333
p-value	< 0.001	< 0.001	< 0.001
	α_1	α_2	α_3
EST	-0.1220	0.1214	-0.0362
SE	0.0212	0.0210	0.0210
p-value	< 0.001	< 0.001	0.084

To illustrate the change pattern of the estimated mean curve of SBP with index, we plot the estimated curve $\hat{m}_0(.)$ against the estimated index and sleeping hours. From Figure 5.3, we can see a nonlinear change pattern between $\hat{m}_0(.)$ and index, and sleeping hours as well. In the beginning, the value of $\hat{m}_0(.)$ increases with sleeping hours, then becomes smooth after about 8 hours. The patterns of $\hat{m}_0(.)$ against physical activity hours per day are similar. Next, we plot the estimated coefficient functions $\hat{m}_1(.)$ and $\hat{m}_2(.)$ in Figure 5.4 to illustrate the effect of a genetic factor interacting with the environmental factors. We also observe nonlinear pattern of the estimated functions $\hat{m}_1(.)$ and $\hat{m}_2(.)$.

Figure 5.3: Plots of the estimated function $\hat{m}_0(.)$ against the estimated index and sleeping hours per day.

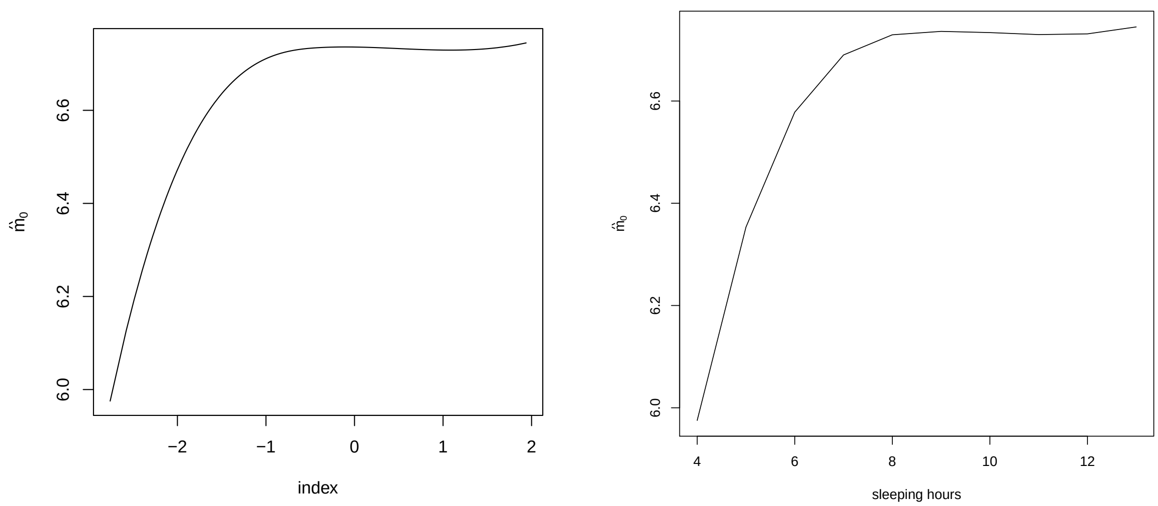
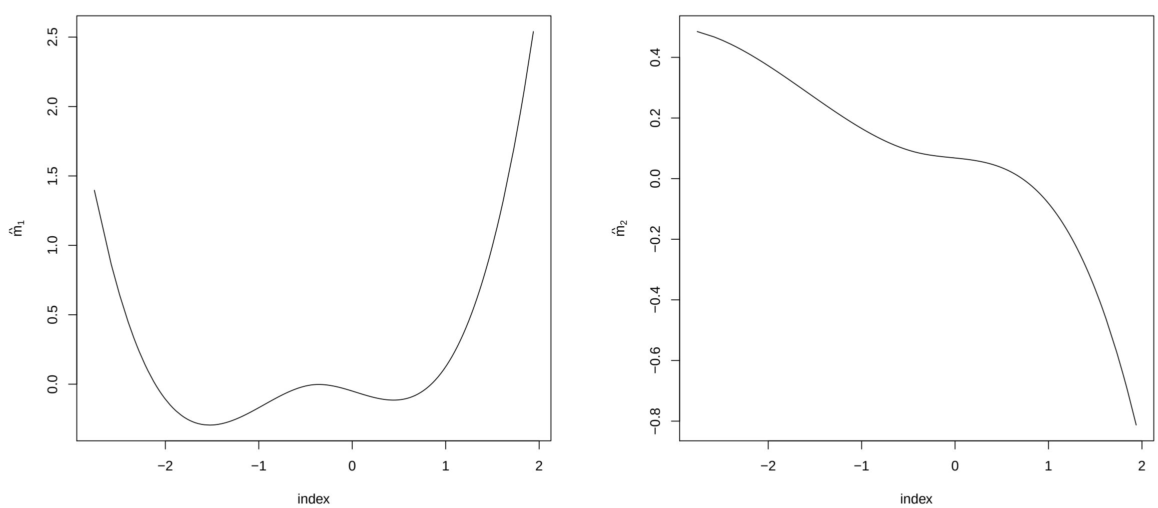


Figure 5.4: Plots of the estimated functions $\hat{m}_1(.)$ and $\hat{m}_2(.)$ against the estimated index.



Bibliography

[1] Ahn, J., Mukherjee, B., Gruber, S. B. and Ghosh, M. (2013). Bayesian semiparametric analysis for two-phase studies of gene-environment interaction. *Annals of Applied Statistics,* 7, 543-569.

[2] Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes.* Springer-Verlag, New York.

[3] Bray, G. A. and Popkin, B. M. (1998). Dietary fat intake does affect obesity! *The American Journal of Clinical Nutrition,* 68, 1157-1173.

[4] Cai, Z., Fan, J. and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association*, 95, 888-902.

[5] Carroll, R. J., Fan, J., Gijbels, I., and Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, 92, 477-489.

[6] Caspi, A. and Moffitt, T. E. (2006). Gene-environment interactions in psychiatry: joining forces with neuroscience. *Nature Reviews Neuroscience*, 7, 583-90.

[7] Chatterjee, N. and Carroll, R. (2005). Semiparametric maximum likelihood estimation exploiting gene-environment independence in case-control studies. *Biometrika,* 92, 399-418.

[8] Chen, J., Lin, D. and Hochner, H. (2012). Semiparametric maximum likelihood methods for analyzing genetic and environmental effects with case-control motherchild pair data. *Biometrics*, 3, 869-877.

[9] Cheverud, J. (2001). A simple correction for multiple comparisons in interval mapping genome scans. *Heredity*,87,52-28.

[10] Cui, X., Härdle, W. and Zhu, L. (2011). The EFM approach for single-index models. *The Annals of Statistics*, 39, 1658-1688.

[11] Dawber, T. R., Meadors, G. F. and Moore, F. E. (1951). Epidemiological approaches to heart disease: the framingham study. *Am. J. Public Health,* 41, 279-286

[12] de Boor, C. (2001). *A Practical Guide to Splines, Vol. 27 of Applied Mathematical Sciences*, revised edn, Springer-Verlag, New York.

[13] Demko, S. (1986). Spectral bounds for $\left|a^{-1}\right|_{\infty}$. *Journal of Approximation Theory*, 48, 207-212.

[14] DeVore, R. A. and Lorentz, G. G. (1993). *Constructive Approximation*. Berlin, Springer-Verlag.

[15] Fan, J. and Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli,* 11, 1031-1057.

[16] Grarup, N. and Andersen, G. (2007). Gene-environment interactions in the pathogenesis of type 2 diabetes and metabolism. *Current Opinion in Clinical Nutrition & Metabolic Care*, 10, 420-6.

[17] Gottlieb, D.J., Redline S., Nieto, F.J., Baldwin, C.M., Newman, A.B., Resnick, H.E. and Punjabi, N.M. (2006). Association of usual sleep duration with hypertension: the Sleep Heart Health Study. *sleep*, 29(8):1009-14

[18] Härdle, W., Liang, H. and Gao, J. (2000). *Partially Linear Models*. Physica Verlag, Heidelberg.

[19] Hebebrand, J. and Hinney, A. (2009). Environmental and genetic risk factors in obesity. *Child & Adolescent Psychiatric Clinics of North America,* 18, 83-94.

[20] Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S., Freimer, N. B., Sabatti, C. and Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics,* 42, 348-354.

[21] Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J. and Eskin, E. (2008). Efficient control of population structure in model organism association mapping. *Genetics,* 178, 1709-1723.

[22] Kokkinos P.F., Giannelou A., Manolis A. and Pittaras A. (2009). Physical activity in the prevention and management of high blood pressure. *Hellenic J Cardiol*, 50: 52-59.

[23] Liang, H., Liu, X., Li, R. and Tsai, C. L. (2010). Estimation and testing for partially linear single index models. *The Annals of Statistics*, 38, 3811-3836.

[24] Lobach, I., Mallick, B. and Carroll, R. (2011). Semiparametric Bayesian analysis of gene-environment interactions with error in measurement of environmental covariates and missing genetic data. *Statistics and its Interface*, 4, 305-316.

[25] Ma, S., Yang, L., Romero, R. and Cui, Y. (2011). Varying-coefficient models for gene-environment interaction: a non-linear look. *Bioinformatics*, 27, 2119-2126.

[26] Ma, S. and Xu, S. (2015). Semiparametric nonlinear regression for detecting gene and environment interactions. *Journal of Statistical Planning and Inference,* 156, 31-47

[27] Maity, A., Carroll, R., Mammen, E. and Chatterjee, N. (2009). Testing in semiparametric models with interaction, with applications to gene-environment interactions. *Journal of the Royal Statistical Society Series B,* 71, 75-96.

[28] McCullagh, P. (1983). Quasi-likelihood functions. *Annals of Statistics*, 11, 59-67.

[29] Nyholt, D. (2004). A simple correction for multiple testing for SNPs in linkage disequilibrium with each other. *The American Journal of Human Genetics*, 74, 756-769.

[30] Saidou, A., Thuillet, A., Couderc, M., Mariac, C. and Vigouroux, Y. (2014). Association studies including genotype by environment interactions: prospects and limits. *BMC Genetics*, 15:3.

[31] Schumaker, L. L. (1981). *Spline Functions*. Wiley, New York.

[32] Song, M., Lee, K. M. and Kang, D. (2011). Breast cancer prevention based on gene-environment interaction. *Molecular Carcinogenesis*, 50, 280-90.

[33] Talmud, P. J. (2007). Gene-environment interaction and its impact on coronary heart disease risk. *Nutrition, Metabolism & Cardiovascular Diseases,*17, 148-52.

[34] Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society, Series B,* 58, 267-288.

[35] Tsuang, M. T., Bar, J. L., Stone, W. S. and Faraone, S. V. (2004). Gene-environment interactions in mental disorders. *World Psychiatry*, 3, 73-83.

[36] Walker, S., Poston, C. and Foreyt, J. P. (1999). Obesity is an environmental issue. *Atherosclerosis*. 146, 201-209.

[37] Wei, J., Carroll, R. and Maity, A. (2011). Testing for constant nonparametric effects in general semiparametric regression models with interactions. *Statistics & Probability Letters*, 81, 717-723.

[38] Xia, Y. and Li, W. K. (1999). On single-index coefficient regression models. *Journal of the American Statistical Association*, 94, 1275-1285.

[39] Xia, Y. and Hardle, W. (2006). Semi-parametric estimation of partially single-index models. *J. Multivariate Analysis,* 97, 1162-1184.

[40] Zhang, Z., Ersoz, E., Lai, C., Todhunter, R. J., Tiwari, H. K., Gore, M. A., Bradbury, P. J., Yu, J., Arnett, D. K., Ordovas, J. M. and Buckler, E. S. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics*, 42, 355-360.

[41] Zhou, S., Shen, X. and Wolfe, D. A. (1998). Local asymptotics for regression splines and confidence regions. *The Annals of Statistics*, 26, 1760-1782.

[42] Zhou, X. and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, 44, 821-824.

Appendix A

Proof of Theorems in Chapter 4

We begin this appendix by introducing necessary notations used in the following proofs of theorems. For any positive numbers a_n and b_n, let $a_n \asymp b_n$ denote that $\lim_{n\to\infty} a_n/b_n = c$, where c is a positive constant, and let $a_n \sim b_n$ denote that $\lim_{n\to\infty} a_n/b_n = 1$. In addition, let $C^{(r)}(S) = \left\{\phi \left| \phi^{(r)} \in C(S)\right.\right\}$ be the set of the r-th order smooth functions ϕ on the support $\mathcal{S}$. For any vector $\zeta = (\zeta_1, \ldots, \zeta_s)^{\mathrm{T}} \in R^s$, denote $\|\zeta\|_\infty = \max(|\zeta_1| + \cdots + |\zeta_s|)$, and for any symmetric matrix $\mathbf{A}$, denote its L_r norm as $\|\mathbf{A}\|_r = \max_{\zeta\in R^s,\zeta\neq 0} \|\mathbf{A}\zeta\|_r \|\zeta\|_r^{-1}$. Moreover, for any matrix $\mathbf{A} = (A_{ij})_{i=1,j=1}^{s,t}$, denote $\|\mathbf{A}\|_\infty = \max_{1\leq i\leq s} \sum_{j=1}^{t} |A_{ij}|$ and $\mathbf{A}^{\otimes 2} = \mathbf{A}\mathbf{A}^{\mathrm{T}}$. To develop the theoretical results of the proposed estimators, we next present the following technical conditions.

A.1 Regularity Conditions

(C1) The density function $f_{\boldsymbol{\beta}^{\mathrm{T}} X_{ij}}\left(\boldsymbol{\beta}^{\mathrm{T}} x_{ij}\right)$ of random variable $\boldsymbol{\beta}^{\mathrm{T}} X_{ij}$ is bounded away from 0 on the support of $\boldsymbol{\beta}^{\mathrm{T}} X_{ij}$ for $\boldsymbol{\beta}$ in a neighborhood of $\boldsymbol{\beta}_0$, and the density function

$f_{T_{ij}}(t)$ of random variable T_{ij} is bounded away from 0 on the support of T_{ij}.

(C2) The true functions $g(u)$ and $\alpha_l(t)$ satisfy $g(u) \in C^{(r)}(S_U)$ and $\alpha_l(t) \in C^{(r)}(S_T)$ for $l = 1, \cdots, d_1$ and given integer $r > 3/2$, where $\mathcal{S}_U$ and $\mathcal{S}_T$ are the compact support sets of $U_{ij}(\boldsymbol{\beta}^0)$ and T_{ij}, respectively. In addition, the order of spline functions satisfies $q \geq r$.

(C3) The eigenvalues of $\mathbf{M}_k$, $1 \leq k \leq \kappa$ are bounded away from 0 and infinity. Let $\Gamma = (\Gamma_{k,k'})_{k,k'=1}^{\kappa} = (\Gamma_{j,j',k,k'})_{j,j'=1,k,k'=1}^{m,\kappa}$. For any $1 \leq j \leq m$, and any given vector $a = (a_k)_{k=1}^{\kappa} \in R^{\kappa}$, there exist constants $0 < c_\Gamma < C_\Gamma < \infty$, such that $c_\Gamma \sum_{k=1}^{\kappa} a_k^2 \leq \sum_{k,k'=1}^{\kappa} a_k a_{k'} \Gamma_{j,j,k,k'} \leq C_\Gamma \sum_{k=1}^{\kappa} a_k^2$.

(C4) The eigenvalues of $E\left((1, Z_{ij}^{(1)\mathrm{T}})^{\mathrm{T}}(1, Z_{ij}^{(1)\mathrm{T}}) \,|U_{ij}(\boldsymbol{\beta}^0) = u, T_{ij} = t\right)$ are uniformly bounded away from 0 and ∞ for all $u \in S_U$ and $t \in S_T$, where $Z_{ij}^{(1)} = (Z_{ij,1}, \ldots, Z_{ij,d_1})^{\mathrm{T}}$.

(C5) The eigenvalues of $\dot{\psi}_n(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0)$ and $\Psi_n(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0)$ are bounded away from 0 and infinity.

Conditions (C1) and (C2), which are given in [58], are typical assumptions in the nonparametric smoothing literature. Conditions (C3)-(C5) are needed for the convergence rates of the parametric and nonparametric estimators as well as the existence of asymptotic variances of the parametric estimators. It is worth noting that Condition (C1) ensures that the density functions are bounded away from 0 in their supports. In practice, we do not know the true support, and we may use the minimum and maximum as the bounded values of the support. In addition, the parameter estimators and their asymptotic properties may not be valid in the case that Conditions (C2)-(C5) are not satisfied.

A.2 Proofs of Theorems 1 and 2

Before proving both theorems, we demonstrate the three lemmas given below.

Lemma 8 *Under Conditions (C1) and (C4), for any* $\mathbf{a} \in R^{J_n}$*, there exist constants* $0 < c_1 < C_1 < \infty$ *such that for* $\forall \boldsymbol{\beta} \in \Theta$ *and for sufficiently large* n*,*

$$c_1 N^{-1} \|\mathbf{a}\|^2 \leq \mathbf{a}^T E\left\{\mathbf{Q}_i(\boldsymbol{\beta})^T \ \mathbf{Q}_i(\boldsymbol{\beta})\right\}\mathbf{a} \leq C_1 N^{-1}\|\mathbf{a}\|^2, \tag{A.1}$$

and

$$\begin{aligned}
&\max_{1\leq J,J'\leq N+q}\left|n^{-1}\sum\nolimits_{i=1}^{n} B_{1,J}(U_i(\boldsymbol{\beta}))^T B_{1,J'}(U_i(\boldsymbol{\beta})) - E\left\{B_{1,J}(U_i(\boldsymbol{\beta}))^T B_{1,J'}(U_i(\boldsymbol{\beta}))\right\}\right| \\
= \ & O_{a.s.}\left\{\sqrt{(\log n)/(nN)}\right\},
\end{aligned}$$

$$\begin{aligned}
&\max_{1\leq J,J'\leq N+q}\left|n^{-1}\sum\nolimits_{i=1}^{n} B_{2,J}(T_i)^T B_{2,J'}(T_i) - E\left\{B_{2,J}(T_i)^T B_{2,J'}(T_i)\right\}\right| \\
= \ & O_{a.s.}\left\{\sqrt{(\log n)/(nN)}\right\},
\end{aligned} \tag{A.2}$$

where

$$\begin{aligned}
B_J(U_i(\boldsymbol{\beta})) &= \left[\left\{B_{1,J}(U_{i1}(\boldsymbol{\beta})), \ldots, B_{1,J}(U_{im}(\boldsymbol{\beta}))\right\}^T\right]_{m\times 1}, \\
B_{2,J}(T_i) &= \left[\left\{B_{2,J}(T_{i1}), \ldots, B_{2,J}(T_{im})\right\}^T\right]_{m\times 1}.
\end{aligned}$$

Proof. By Theorem 5.4.2 of deVore and Lorentz (1993) and Condition (C1), we have that, for sufficiently large n and for any $\alpha = (\alpha_1, \ldots \alpha_{J_n})^{\mathrm{T}} \in R^{J_n}$, there exist constants $0 < c_1^* < C_1^* < \infty$ and $0 < c_2^* < C_2^* < \infty$ such that

$$c_1^* N^{-1}\sum\nolimits_{J=1}^{J_n}\alpha_J^2 \leq E\left\{\sum\nolimits_{J=1}^{J_n}\alpha_J B_{1,J}(U_{ij}(\boldsymbol{\beta}^0))\right\}^2 \leq C_1^* N^{-1}\sum\nolimits_{J=1}^{J_n}\alpha_J^2 \ \text{ and} \tag{A.3}$$

$$c_2^* N^{-1}\sum\nolimits_{J=1}^{J_n}\alpha_J^2 \leq E\left\{\sum\nolimits_{J=1}^{J_n}\alpha_J B_{2,J}(T_{ij})\right\}^2 \leq C_2^* N^{-1}\sum\nolimits_{J=1}^{J_n}\alpha_J^2. \tag{A.4}$$

In addition, Condition (C4) implies that, for any $\vartheta = (\vartheta_0, \vartheta_1, \ldots \vartheta_{d_1})^{\mathrm{T}} \in R^{d_1+1}$, there exist constants $0 < c_3^* < C_3^* < \infty$ such that

$$c_3^* \sum_{l=0}^{d_1} \vartheta_l^2 \leq E\{(\vartheta_0 + \sum_{l=1}^{d_1} \vartheta_l Z_{ij,l})^2 | U_{ij}(\boldsymbol{\beta}^0), T_{ij}\} \leq C_3^* \sum_{l=0}^{d_1} \vartheta_l^2. \quad \text{(A.5)}$$

Let $\mathbf{a} = (a_{J,l} : 1 \leq J \leq J_n, 0 \leq l \leq d_1)$. After algebraic simplification, we have

$$\begin{aligned}
& \mathbf{a}^{\mathrm{T}} E\left\{\mathbf{Q}_i\left(\boldsymbol{\beta}^0\right)^{\mathrm{T}} \mathbf{Q}_i\left(\boldsymbol{\beta}^0\right)\right\} \mathbf{a} \\
= & \sum_{j=1}^{m} \mathbf{a}^{\mathrm{T}} E\left\{Q_{ij}\left(\boldsymbol{\beta}^0\right) Q_{ij}\left(\boldsymbol{\beta}^0\right)^{\mathrm{T}}\right\} \mathbf{a} \\
= & \sum_{j=1}^{m} E[\{\sum_{J=1}^{J_n} a_{J,0} B_{1,J}(U_{ij}(\boldsymbol{\beta}^0))\} + \sum_{l=1}^{d_1}\{\sum_{J=1}^{J_n} a_{J,l} B_{2,J}\left(T_{ij}\right)\} Z_{ij,l}]^2 \\
\leq & \sum_{j=1}^{m} C_3^* [E\{\sum_{J=1}^{J_n} a_{J,0} B_{1,J}(U_{ij}(\boldsymbol{\beta}^0))\}^2 + \sum_{l=1}^{d_1} E\{\sum_{J=1}^{J_n} a_{J,l} B_{2,J}\left(T_{ij}\right)\}^2] \\
\leq & \sum_{j=1}^{m} C_3^* \{C_1^* N^{-1} \sum_{J=1}^{J_n} \alpha_J^2 + C_1^* N^{-1} \sum_{l=1}^{d_1} \sum_{J=1}^{J_n} a_{J,l}^2\},
\end{aligned}$$

where the first inequality of the above equation follows from (A.5) and the second inequality follows from (A.3) and (A.4). By letting $C_1 = m C_3^* C_1^*$, we then obtain that

$$\mathbf{a}^{\mathrm{T}} E\left\{\mathbf{Q}_i\left(\boldsymbol{\beta}^0\right)^{\mathrm{T}} \mathbf{Q}_i\left(\boldsymbol{\beta}^0\right)\right\} \mathbf{a} \leq C_1 N^{-1} \|\mathbf{a}\|^2.$$

Applying a similar approach, we can show that

$$\mathbf{a}^{\mathrm{T}} E\left\{\mathbf{Q}_i\left(\boldsymbol{\beta}^0\right)^{\mathrm{T}} \mathbf{Q}_i\left(\boldsymbol{\beta}^0\right)\right\} \mathbf{a} \geq c_1 N^{-1} \|\mathbf{a}\|^2.$$

This completes the proof of (A.1), and the result of (A.2) can be obtained by Bernstein's inequality from [2]. ■

Lemma 9 *Under Conditions (C1)-(C4), we have*

(1) $\left|\widetilde{g}\left(u, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right) - g\left(u\right)\right| = O_p\left(\sqrt{N/n} + N^{-r}\right)$ *and* $\left|\widetilde{g}'\left(u, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right) - g'\left(u\right)\right| = O_p\left(\sqrt{N^3/n} + N^{-r+1}\right)$ *uniformly in* $u \in S_U$*;*

(2) $\left|\widetilde{\alpha}_l\left(t,\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)-\alpha_l\left(t\right)\right| = O_p\left(\sqrt{N/n}+N^{-r}\right)$ *and* $\left|\widetilde{\alpha}_l'\left(t,\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)-\alpha_l'\left(t\right)\right| = O_p\left(\sqrt{N^3/n}+N^{-r+1}\right)$ *uniformly in* $t\in S_T$*, for* $1\leq l\leq d_1$*.*

Proof. For the sake of simplicity and with a slight abuse of notations, we denote $\widetilde{\boldsymbol{\gamma}}=\left(\widetilde{\boldsymbol{\gamma}}_0^{\mathrm{T}},\ldots,\ \widetilde{\boldsymbol{\gamma}}_{d_1}^{\mathrm{T}}\right)^{\mathrm{T}}=\widetilde{\boldsymbol{\gamma}}\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)$, $\widetilde{\phi}_n^0=\left\{\left(\widetilde{\phi}_{n,1}^0\right)^{\mathrm{T}},\ldots,\left(\widetilde{\phi}_{n,\kappa}^0\right)^{\mathrm{T}}\right\}^{\mathrm{T}}=\widetilde{\phi}_n\left(\boldsymbol{\gamma}^0,\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)$, and $\widetilde{C}_n^0=\widetilde{C}_n\left(\boldsymbol{\gamma}^0,\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)$. According to the result on page 149 of [12], for g and α_l satisfying Condition (C2), there exists $\boldsymbol{\gamma}_l^0\in R^{J_n}$ such that

$$\sup_{u\in S_U}\left|g\left(u\right)-g^0\left(u\right)\right|=O\left(N^{-r}\right)\quad\text{and}\quad\sup_{t\in S_T}\left|\alpha_l\left(t\right)-\alpha_l^0\left(t\right)\right|=O\left(N^{-r}\right),\tag{A.6}$$

where $g^0\left(u\right)=B_1\left(u\right)^{\mathrm{T}}\boldsymbol{\gamma}_0^0$ and $\alpha_l^0\left(t\right)=B_2\left(t\right)^{\mathrm{T}}\boldsymbol{\gamma}_l^0$. Let $\boldsymbol{\gamma}^0=\left(\boldsymbol{\gamma}_0^{0\mathrm{T}},\ldots,\boldsymbol{\gamma}_{d_1}^{0\mathrm{T}}\right)^{\mathrm{T}}$, and we then show that $\left\|\widetilde{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^0\right\|_\infty=o_{\text{a.s.}}\left(1\right)$. By the same arguments as given in [39], we know that the global minimum for $\widetilde{Q}_n\left(\boldsymbol{\gamma},\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)$ given in (4.7) exists. As a result, we only need to demonstrate that the minimizer $\widetilde{\boldsymbol{\gamma}}$ remains inside of $\mathcal{S}_{\boldsymbol{\gamma}^0}$, where $\mathcal{S}_{\boldsymbol{\gamma}^0}$ is any neighborhood of $\boldsymbol{\gamma}^0$.

Let

$$\widetilde{\varrho}_n\left(\boldsymbol{\gamma},\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)=\left\|n^{-1}\left\{E\widetilde{C}_n\left(\boldsymbol{\gamma},\boldsymbol{\beta}^0\ ,\boldsymbol{\alpha}^0\right)\right\}^{-1/2}\left\{E\widetilde{\phi}_n\left(\boldsymbol{\gamma},\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)\right\}\right\|;$$

it is noteworthy that $\widetilde{\varrho}_n\left(\boldsymbol{\gamma},\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)$ is a continuous function in $\boldsymbol{\gamma}$. By (A.1) and (A.6), we have

$$\left\|EC_n\left(\boldsymbol{\gamma}^0,\boldsymbol{\beta}^0\ ,\boldsymbol{\alpha}^0\right)^{1/2}\right\|_2\asymp\kappa\left\|\left[E\left\{\mathbf{Q}_i\left(\boldsymbol{\beta}^0\right)^{\mathrm{T}}\mathbf{Q}_i\left(\boldsymbol{\beta}^0\right)\right\}\right]^{1/2}\right\|_2\asymp N^{-1/2}$$

and $\left\|E\phi_n\left(\boldsymbol{\gamma},\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)\right\|_2=O\left(N^{-r-1/2}\right)$. Therefore,

$$\varrho_n\left(\boldsymbol{\gamma},\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)\leq\left\|EC_n\left(\boldsymbol{\gamma}^0,\boldsymbol{\beta}^0\ ,\boldsymbol{\alpha}^0\right)^{1/2}\right\|_2^{-1}\left\|E\phi_n\left(\boldsymbol{\gamma},\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)\right\|_2=O\left(N^{1/2}N^{-r-1/2}\right)=o\left(1\right).$$

Assume that $\widetilde{\boldsymbol{\gamma}} \in \mathcal{S}^C_{\boldsymbol{\gamma}^0}$, where $\mathcal{S}^C_{\boldsymbol{\gamma}^0}$ is complement of $\mathcal{S}_{\boldsymbol{\gamma}^0}$. Then, there exists a constant $0 < C < \infty$, such that

$$\widetilde{\varrho}_n\left(\widetilde{\boldsymbol{\gamma}}, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right) = \left\| n^{-1}\left\{E\widetilde{C}_n\left(\widetilde{\boldsymbol{\gamma}}, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)\right\}^{-1/2}\left\{E\widetilde{\phi}_n\left(\widetilde{\boldsymbol{\gamma}}, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)\right\}\right\| > C. \tag{A.7}$$

Since $\widetilde{\boldsymbol{\gamma}}$ is the minimizer of $\widetilde{Q}_n\left(\boldsymbol{\gamma}, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)$, we have that

$$\left\| n^{-1}\widetilde{C}_n\left(\widetilde{\boldsymbol{\gamma}}, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)^{-1/2}\widetilde{\phi}_n\left(\widetilde{\boldsymbol{\gamma}}, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)\right\| \leq \left\| n^{-1}\widetilde{C}_n\left(\boldsymbol{\gamma}^0, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)^{-1/2}\widetilde{\phi}_n\left(\boldsymbol{\gamma}^0, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)\right\|.$$

By the strong law of large numbers, we further obtain

$$\begin{aligned} & \left\| n^{-1}\widetilde{C}_n\left(\boldsymbol{\gamma}^0, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)^{-1/2}\widetilde{\phi}_n\left(\boldsymbol{\gamma}^0, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)\right\| \\ \to & \left\| n^{-1}\left(E\widetilde{C}^0_n\right)^{-1/2}\left(E\widetilde{\phi}^0_n\right)\right\| = o\left(1\right), \end{aligned}$$

almost surely. Thus, $\left\| n^{-1}\widetilde{C}_n\left(\widetilde{\boldsymbol{\gamma}}, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)^{-1/2}\widetilde{\phi}_n\left(\widetilde{\boldsymbol{\gamma}}, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)\right\| = o\left(1\right)$. Recall that

$$\widetilde{\phi}_n\left(\boldsymbol{\gamma}, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right) = n^{-1}\sum\nolimits_{i=1}^{n}\widetilde{\phi}_{in}\left(\boldsymbol{\gamma}, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)$$

as given in (4.6). It is also worth noting that $\widetilde{\phi}_{in}\left(\boldsymbol{\gamma}, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)$ is a continuous function of $\boldsymbol{\gamma}$, and, for all $\boldsymbol{\gamma} \in \mathcal{S}^C_{\boldsymbol{\gamma}^0}$, there exists $0 < C^* < \infty$ such that

$$\begin{aligned} & ||\widetilde{\phi}_n\left(\boldsymbol{\gamma}, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)|| \\ \leq\ & n^{-1}\{\sum\nolimits_{i=1}^{n}\widetilde{\phi}_{in}\left(\boldsymbol{\gamma}, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)^{\mathrm{T}}\widetilde{\phi}_{in}\left(\boldsymbol{\gamma}, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)\}^{1/2} \\ \leq\ & C^* n^{-1}(n\kappa J_n(1+d_1))^{1/2}. \end{aligned}$$

Then, by the uniform law of large numbers, we have

$$\sup_{\boldsymbol{\gamma}\in\mathcal{S}^C_{\boldsymbol{\gamma}^0}} ||\widetilde{\phi}_n\left(\boldsymbol{\gamma}, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right) - E\widetilde{\phi}_n\left(\boldsymbol{\gamma}, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)|| = o_{\text{a.s.}}\left(1\right).$$

This, together with the continuous mapping theorem, leads to

$$\left| n^{-1} \widetilde{C}_n \left(\widetilde{\boldsymbol{\gamma}} \ , \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0 \right)^{-1/2} \widetilde{\phi}_n \left(\widetilde{\boldsymbol{\gamma}}, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0 \right) - \widetilde{\varrho}_n \left(\widetilde{\boldsymbol{\gamma}}, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0 \right) \right| = o_{\text{a.s.}} (1) \, ,$$

which contradicts with (B.14). Consequently, $\widetilde{\boldsymbol{\gamma}}$ remains inside of $\mathcal{S}_{\boldsymbol{\gamma}^0}$.

Using the above result and the Taylor expansion, we have

$$\widetilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0 = - \left\{ \partial^2 \widetilde{Q}_n \left(\boldsymbol{\gamma}^0, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0 \right) / \partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^{\mathrm{T}} \right\}^{-1} \left\{ \partial \widetilde{Q}_n \left(\boldsymbol{\gamma}^0 \ , \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0 \right) / \partial \boldsymbol{\gamma} \right\} \left(1 + o_p (1) \right) .$$

Let

$$\Omega_n = \begin{pmatrix} \Omega_{n,1} \\ \vdots \\ \Omega_{n,\kappa} \end{pmatrix} = n^{-1} \sum\nolimits_{i=1}^{n} \begin{bmatrix} \mathbf{Q}_i \left(\boldsymbol{\beta}^0 \right)^{\mathrm{T}} \Delta_i \Lambda_1 \Delta_i \mathbf{Q}_i \left(\boldsymbol{\beta}^0 \right) \\ \vdots \\ \mathbf{Q}_i \left(\boldsymbol{\beta}^0 \right)^{\mathrm{T}} \Delta_i \Lambda_\kappa \Delta_i \mathbf{Q}_i \left(\boldsymbol{\beta}^0 \right) \end{bmatrix}_{\kappa J_n (1+d_1) \times J_n (1+d_1)} \quad \text{and}$$

$$\Xi_n = n^{-2} \sum\nolimits_{i=1}^{n} \begin{bmatrix} \mathbf{Q}_i \left(\boldsymbol{\beta}^0 \right)^{\mathrm{T}} \Delta_i \Gamma_{1,1} \Delta_i \mathbf{Q}_i \left(\boldsymbol{\beta}^0 \right) & \cdots & \mathbf{Q}_i \left(\boldsymbol{\beta}^0 \right)^{\mathrm{T}} \Delta_i \Gamma_{1,\kappa} \Delta_i \mathbf{Q}_i \left(\boldsymbol{\beta}^0 \right) \\ \vdots & \ddots & \vdots \\ \mathbf{Q}_i \left(\boldsymbol{\beta}^0 \right)^{\mathrm{T}} \Delta_i \Gamma_{\kappa,1} \Delta_i \mathbf{Q}_i \left(\boldsymbol{\beta}^0 \right) & \cdots & \mathbf{Q}_i \left(\boldsymbol{\beta}^0 \right)^{\mathrm{T}} \Delta_i \Gamma_{\kappa,\kappa} \Delta_i \mathbf{Q}_i \left(\boldsymbol{\beta}^0 \right) \end{bmatrix}$$

with dimension $\kappa J_n (1 + d_1) \times \kappa J_n (1 + d_1)$. By (A.6) and the weak law of large numbers, we have $C_n \left(\boldsymbol{\gamma}^0, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0 \right) = \Xi_n \left(1 + o_p (1) \right)$. Thus,

$$\begin{aligned} \partial Q_n \left(\boldsymbol{\gamma}^0, \boldsymbol{\beta}^0, \alpha^0 \right) / \partial \boldsymbol{\gamma} =& 2 \left\{ \partial \phi_n \left(\boldsymbol{\gamma}^0, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0 \right)^{\mathrm{T}} / \partial \boldsymbol{\gamma} \right\} \Xi_n^{-1} \phi_n \left(\boldsymbol{\gamma}^0, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0 \right) \left(1 + o_p (1) \right) \\ =& -2 \Omega_n^{\mathrm{T}} \Xi_n^{-1} \phi_n^0 \left(1 + o_p (1) \right) \quad \text{and} \end{aligned}$$

$$\partial^2 Q_n \left(\boldsymbol{\gamma}^0, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0 \right) / \partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^{\mathrm{T}} = 2 \Omega_n^{\mathrm{T}} \Xi_n^{-1} \Omega_n \left(1 + o_p (1) \right) . \tag{A.8}$$

As a result,

$$\widetilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0 = \left(n^{-1} \Omega_n^{\mathrm{T}} \Xi_n^{-1} \Omega_n \right)^{-1} \left(n^{-1} \Omega_n^{\mathrm{T}} \Xi_n^{-1} \phi_n^0 \right) \left(1 + o_p (1) \right) . \tag{A.9}$$

By (A.1), (A.2) and Condition (C3), it can be shown that, with probability approaching 1, $\|n\Xi_n\|_2 \asymp N^{-1}$ and $\sup_{1\le k\le \kappa} \|\Omega_{n,k}\|_2 \asymp N^{-1}$, and thus $\left\|n^{-1}\Xi_n^{-1}\right\|_2 \asymp N$. Moreover, by (A.1),

$$\begin{aligned}\left\|E\left(\Omega_n\right)^{\mathrm{T}} E\left(\Omega_n\right)\right\|_2 &= \left\|\sum_{k=1}^{\kappa} E\left\{\mathbf{Q}_i\left(\boldsymbol{\beta}^0\right)^{\mathrm{T}} \Delta_i \Lambda_k \Delta_i \mathbf{Q}_i\left(\boldsymbol{\beta}^0\right)\right\}^{\otimes 2}\right\|_2 \\ &\asymp \kappa\left\|E\left\{\mathbf{Q}_i\left(\boldsymbol{\beta}^0\right)^{\mathrm{T}} \mathbf{Q}_i\left(\boldsymbol{\beta}^0\right)\right\}\right\|_2^2 \asymp N^{-2}.\end{aligned}$$

This, together with (A.2), implies that, with probability approaching 1, $\left\|\Omega_n^{\mathrm{T}}\Omega_n\right\|_2 \asymp N^{-2}$. Accordingly, with probability approaching 1,

$$\left\|n^{-1}\Omega_n^{\mathrm{T}}\Xi_n^{-1}\Omega_n\right\|_2 \asymp N\left\|\Omega_n^{\mathrm{T}}\Omega_n\right\|_2 \asymp N^{-1} \text{ and } \left\|\left(n^{-1}\Omega_n^{\mathrm{T}}\Xi_n^{-1}\Omega_n\right)^{-1}\right\|_2 \asymp N. \tag{A.10}$$

Next, let $\mu_i = (\mu_{i1}, \ldots, \mu_{im})^{\mathrm{T}}$. By (4.6), $\phi_{n,k}^0$ can be decomposed into $\phi_{n,k,e}^0 + \phi_{n,k,\mu}^0$, where

$$\begin{aligned}\phi_{n,k,e}^0 &= n^{-1}\sum_{i=1}^{n} \mathbf{Q}_i\left(\boldsymbol{\beta}^0\right)^{\mathrm{T}} \widetilde{\Delta}_i\left(\boldsymbol{\gamma}^0, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)\Lambda_k\left(Y_i - \mu_i\right), \\ \phi_{n,k,\mu}^0 &= n^{-1}\sum_{i=1}^{n} \mathbf{Q}_i\left(\boldsymbol{\beta}^0\right)^{\mathrm{T}} \widetilde{\Delta}_i\left(\boldsymbol{\gamma}^0, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)\Lambda_k\left\{\mu_i - \widetilde{\mu}_i\left(\boldsymbol{\gamma}^0, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)\right\}.\end{aligned}$$

Denote $\phi_{n,e}^0 = \left\{\left(\phi_{n,1,e}^0\right)^{\mathrm{T}}, \ldots, \left(\phi_{n,\kappa,e}^0\right)^{\mathrm{T}}\right\}^{\mathrm{T}}$ and $\phi_{n,\mu}^0 = \left\{\left(\phi_{n,1,\mu}^0\right)^{\mathrm{T}}, \ldots, \left(\phi_{n,\kappa,\mu}^0\right)^{\mathrm{T}}\right\}^{\mathrm{T}}$. Accordingly, $\widetilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0 = \left(\widetilde{\boldsymbol{\gamma}}_e + \widetilde{\boldsymbol{\gamma}}_\mu\right)(1 + o_p(1))$, where $\widetilde{\boldsymbol{\gamma}}_e = \left(n^{-1}\Omega_n^{\mathrm{T}}\Xi_n^{-1}\Omega_n\right)^{-1}\left(n^{-1}\Omega_n^{\mathrm{T}}\Xi_n^{-1}\phi_{n,e}^0\right)$ and

$\widetilde{\boldsymbol{\gamma}}_\mu = \left(n^{-1}\Omega_n^{\mathrm{T}}\Xi_n^{-1}\Omega_n\right)^{-1}\left(n^{-1}\Omega_n^{\mathrm{T}}\Xi_n^{-1}\phi_{n,\mu}^0\right)$. Let $\mathbb{C} = \left(\mathbb{C}_{ij}^{\mathrm{T}}, 1 \le j \le m, 1 \le i \le n\right)^{\mathrm{T}}$. Then, for any vector $\mathbf{a} \in R^{J_n(1+d_1)}$ with $\|\mathbf{a}\| = 1$, $E\left(\mathbf{a}^{\mathrm{T}}\widetilde{\boldsymbol{\gamma}}_e\right) = 0$, and (A.10) leads to

$$E\left\{\left(\mathbf{a}^{\mathrm{T}}\widetilde{\boldsymbol{\gamma}}_e\right)^2 |\mathbb{C}\right\} \asymp \mathbf{a}^{\mathrm{T}}\left(\Omega_n^{\mathrm{T}}\Xi_n^{-1}\Omega_n\right)^{-1}\mathbf{a} \asymp Nn^{-1}.$$

Thus, by the weak law of large numbers, $\left|\mathbf{a}^{\mathrm{T}}\widetilde{\boldsymbol{\gamma}}_e\right| = O_p\left(N^{1/2}n^{-1/2}\right)$. Furthermore, with probability approaching 1, there exists a constant $0 < C < \infty$, such that

$$\begin{aligned}\left|\mathbf{a}^{\mathrm{T}}\widetilde{\boldsymbol{\gamma}}_\mu\right| &\le C\left\|\left(n^{-1}\Omega_n^{\mathrm{T}}\Xi_n^{-1}\Omega_n\right)^{-1}\right\|_2 \left\|n^{-1}\Xi_n^{-1}\right\|_2 \sup_{1\le k\le\kappa}\left\|\Omega_{n,k}\right\|_2 \left\|E\left\{Q_{ij}\left(\boldsymbol{\beta}^0\right)\right\}\right\| O\left(N^{-r}\right)\\ &= O\left(N^{-r}\right).\end{aligned}$$

The above results imply $\left|\mathbf{a}^{\mathrm{T}}\left(\widetilde{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^0\right)\right| = O_p\left(N^{1/2}n^{-1/2}+N^{-r}\right)$. This, in conjunction with (A.6), ensures that $\left|\widetilde{g}\left(u,\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)-g\left(u\right)\right| = O_p\left(N^{1/2}n^{-1/2}+N^{-r}\right)$ uniformly for every $u \in S_U$ and $\left|\widetilde{\alpha}_l\left(t,\boldsymbol{\beta}^0\ ,\alpha^0\right)-\alpha_l\left(t\right)\right| = O_p\left(N^{1/2}n^{-1/2}+N^{-r}\right)$ uniformly for every $t \in S_T$.

To show the second part of the lemma, we employ the results on page 116 of Davis C. S. (2002) and obtain that $\widetilde{g}'\left(u,\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right) = B_1^*\left(u\right)^{\mathrm{T}}\mathbf{D}_1\widetilde{\boldsymbol{\gamma}}_0\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)$ and $\widetilde{\alpha}_l'\left(t,\ \boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right) = B_2^*\left(t\right)^{\mathrm{T}}\mathbf{D}_1\widetilde{\boldsymbol{\gamma}}_l\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)$, where $B_1^*\left(u\right) = \left\{B_{1,J}^*\left(u\right): 1\le J\le N+q-1\right\}^{\mathrm{T}}$ is the $(q-1)$-th order B-spline basis, and

$\mathbf{D}_1 = \left[(q-1)\left\{\left(-\mathbf{D},\mathbf{0}_{(J_n-1)}\right)+\left(\ \mathbf{0}_{(J_n-1)},\mathbf{D}\right)\right\}\right]_{(J_n-1)\times J_n}$, $\mathbf{D}$ =diag$(d_1,\ldots,d_{N+q-1})$, $d_J = \left(\xi_{q-1+J}-\xi_J\right)^{-1}$ for $1\le J\le N+q-1$, and $\mathbf{0}_{(J_n-1)}$ is the $(N-1)$ dimensional vector with "0 " elements, and $B_2^*\left(t\right)$ is defined in the same way. It is easy to prove that $\left\|\mathbf{D}_1\right\|_\infty = O\left(N\right)$. Applying similar techniques to those used in the proofs for $\widetilde{g}\left(u,\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)$ and $\widetilde{\alpha}_l\left(t,\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)$, we have that $\left|\widetilde{g}'\left(u,\boldsymbol{\beta}^0\ ,\boldsymbol{\alpha}^0\right)-g'\left(u\right)\right| = O_p\left(\sqrt{N^3/n}+N^{-r+1}\right)$ uniformly in $u \in S_U$ and $\left|\widetilde{\alpha}_l'\left(t,\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)-\alpha_l'\left(t\right)\right| = O_p\left(\sqrt{N^3/n}+N^{-r+1}\right)$ uniformly in $t\in S_T$, for $1\le l\le d_1$, which completes the proof. ■

Lemma 10 *Under Conditions (C1)-(C4), we have that*

$$\frac{\partial\widehat{\eta}_{ij}\left(\boldsymbol{\beta}^0,\alpha^0\right)}{\partial\left(\boldsymbol{\beta}_{-1}^T,\ \alpha^T\right)^T} = \left\{\widetilde{g}'\left(X_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0,\boldsymbol{\beta}^0\ ,\alpha^0\right)\widehat{X}_{ij}^T\mathbf{J}^{\mathbf{0}},\widehat{Z}_{ij}^{(2)T}\right\}^T + O_p\left(N^{-r+1}+N^{3/2}n^{-1/2}\right).$$

Proof. By (4.8), we obtain

$$\frac{\partial \widehat{\eta}_{ij}\left(\boldsymbol{\beta}^{\mathbf{0}}, \boldsymbol{\alpha}^{\mathbf{0}}\right)}{\partial \left(\boldsymbol{\beta}_{-1}^{\mathrm{T}}, \boldsymbol{\alpha}^{\mathrm{T}}\right)^{\mathrm{T}}} = \left[\begin{array}{c} \widetilde{g}'\left(X_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{0}\ , \boldsymbol{\beta}^{0}, \boldsymbol{\alpha}^{0}\right)\widehat{X}_{ij}^{\mathrm{T}}\mathbf{J}^{\mathbf{0}} + \left\{Q_{ij}\left(\boldsymbol{\beta}\right)^{\mathrm{T}}\left(\partial \widetilde{\boldsymbol{\gamma}}\left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)/\partial\ \boldsymbol{\beta}_{-1}^{\mathrm{T}}\right)\right\}^{\mathrm{T}} \\ Z_{ij}^{(2)} + \left\{Q_{ij}\left(\boldsymbol{\beta}\right)^{\mathrm{T}}\left(\partial \widetilde{\boldsymbol{\gamma}}\left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)/\partial \boldsymbol{\alpha}^{\mathrm{T}}\right)\right\}^{\mathrm{T}} \end{array}\right].$$

From (A.9), it can be shown that

$$\begin{aligned}
& Q_{ij}\left(\boldsymbol{\beta}^{0}\right)^{\mathrm{T}}\left(\partial \widetilde{\boldsymbol{\gamma}}\left(\boldsymbol{\beta}^{0}, \boldsymbol{\alpha}^{0}\right)/\partial \boldsymbol{\beta}_{-1}^{\mathrm{T}}\right) \\
= \ & -Q_{ij}\left(\boldsymbol{\beta}^{0}\right)^{\mathrm{T}}\left(n^{-1}\Omega_n^{\mathrm{T}}\Xi_n^{-1}\Omega_n\right)^{-1}\left(n^{-1}\Omega_n^{\mathrm{T}}\Xi_n^{-1}\right)\times \\
& n^{-1}\sum\nolimits_{i=1}^{n}\mathbf{Q}_i\left(\boldsymbol{\beta}^{0}\right)^{\mathrm{T}}\Delta_i\Lambda_k\Delta_i\left\{g^{0\prime}\left(U_{ij}\left(\boldsymbol{\beta}^{0}\right)\right)X_{ij}, 1\leq j\leq m\right\}^{\mathrm{T}}\mathbf{J}^{\mathbf{0}} + \\
& O_p\left(N^{1/2}n^{-1/2} + N^{-r}\right) \\
= \ & -\left\{Q_{ij}\left(\boldsymbol{\beta}^{0}\right)^{\mathrm{T}}\widehat{\boldsymbol{\eta}}_s, 1\leq s\leq p\right\}\mathbf{J}^{\mathbf{0}} + O_p\left(N^{1/2}n^{-1/2} + N^{-r}\right),
\end{aligned}$$

where $g^{0\prime}\left(U_{ij}\left(\boldsymbol{\beta}^{0}\right)\right) = B_1'\left(U_{ij}\left(\boldsymbol{\beta}^{0}\right)\right)^{\mathrm{T}}\boldsymbol{\gamma}_0^{0}$ and

$$\begin{aligned}
\widehat{\boldsymbol{\zeta}}_s \ = \ & \left(n^{-1}\Omega_n^{\mathrm{T}}\Xi_n^{-1}\Omega_n\right)^{-1}\left(n^{-1}\Omega_n^{\mathrm{T}}\Xi_n^{-1}\right)\times \\
& n^{-1}\sum\nolimits_{i=1}^{n}\mathbf{Q}_i\left(\boldsymbol{\beta}^{0}\right)^{\mathrm{T}}\Delta_i\Lambda_k\Delta_i\left\{g^{0\prime}\left(U_{ij}\left(\boldsymbol{\beta}^{0}\right)\right)X_{ij,s}, 1\leq j\leq m\right\}^{\mathrm{T}}.
\end{aligned}$$

Furthermore, by Lemma 11, we have that

$$Q_{ij}\left(\boldsymbol{\beta}^{0}\right)^{\mathrm{T}}\widehat{\boldsymbol{\zeta}}_s = \widetilde{g}'\left(X_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{0}, \boldsymbol{\beta}^{0}, \boldsymbol{\alpha}^{0}\right)Q_{ij}\left(\boldsymbol{\beta}^{0}\right)^{\mathrm{T}}\widehat{\vartheta}_s + O_p\left(N^{-r+1} + N^{3/2}n^{-1/2}\right),$$

where

$$\widehat{\vartheta}_s = \left(n^{-1}\Omega_n^{\mathrm{T}}\Xi_n^{-1}\Omega_n\right)^{-1}\left(n^{-1}\Omega_n^{\mathrm{T}}\Xi_n^{-1}\right)n^{-1}\sum\nolimits_{i=1}^{n}\mathbf{Q}_i\left(\ \boldsymbol{\beta}^{0}\right)^{\mathrm{T}}\Delta_i\Lambda_k\Delta_i X_{i\cdot,s}, \tag{A.11}$$

and $X_{i\cdot,s} = \left\{X_{ij,s}, 1\leq j\leq m\right\}^{\mathrm{T}}$. Thus,

$$\begin{aligned}
& Q_{ij}\left(\boldsymbol{\beta}^{0}\right)^{\mathrm{T}}\left(\partial \widetilde{\boldsymbol{\gamma}}\left(\boldsymbol{\beta}^{0}, \boldsymbol{\alpha}^{0}\right)/\partial \boldsymbol{\beta}_{-1}^{\mathrm{T}}\right) \\
= \ & -\widetilde{g}'\left(X_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{0}, \boldsymbol{\beta}^{0}, \boldsymbol{\alpha}^{0}\right)\left\{Q_{ij}\left(\boldsymbol{\beta}^{0}\right)^{\mathrm{T}}\widehat{\vartheta}_s : 1\leq s\leq p\right\}\mathbf{J}^{\mathbf{0}} + O_p\left(N^{-r+1} + N^{3/2}n^{-1/2}\right).
\end{aligned}$$

Analogously, we can demonstrate that

$$\begin{aligned} & Q_{ij}\left(\boldsymbol{\beta}^{0}\right)^{\mathrm{T}}\left(\partial\widetilde{\boldsymbol{\gamma}}\left(\boldsymbol{\beta}^{0},\boldsymbol{\alpha}^{0}\right)/\partial\boldsymbol{\alpha}^{\mathrm{T}}\right) \\ = & -\left\{Q_{ij}\left(\boldsymbol{\beta}^{0}\right)^{\mathrm{T}}\widehat{\boldsymbol{\eta}}_{l}:d_{1}+1\leq l\leq d\right\}+O_{p}\left(N^{-r+1}+N^{3/2}n^{-1/2}\right). \end{aligned}$$

Accordingly,

$$\frac{\partial\widehat{\eta}_{ij}\left(\boldsymbol{\beta}^{0},\alpha^{0}\right)}{\partial\left(\boldsymbol{\beta}_{-1}^{\mathrm{T}},\ \alpha^{\mathrm{T}}\right)^{\mathrm{T}}}=\left\{\widetilde{g}'\left(X_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{0},\boldsymbol{\beta}^{0}\ ,\alpha^{0}\right)\widehat{X}_{ij}^{\mathrm{T}}\mathbf{J}^{\mathbf{0}},\widehat{Z}_{ij}^{(2)\mathrm{T}}\right\}^{\mathrm{T}}+O_{p}\left(N^{-r+1}+N^{3/2}n^{-1/2}\right),$$

which completes the proof. ■

Proof of Theorem 5. Let $\widehat{\boldsymbol{\theta}}=\left(\widehat{\boldsymbol{\beta}}_{-1}^{\mathrm{T}},\widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\right)^{\mathrm{T}}$ and $\theta^{0}=\left(\boldsymbol{\beta}_{-1}^{0\mathrm{T}},\boldsymbol{\alpha}^{0\mathrm{T}}\right)^{\mathrm{T}}$. Let $\mathcal{S}(\theta^{0})$ be any open set that include θ^{0}. We use the same technique given in the proofs of Lemma A.2 to show that $\widehat{\boldsymbol{\theta}}$ remains inside of $\mathcal{S}(\theta^{0})$, so that $||\widehat{\boldsymbol{\theta}}-\theta^{0}||=o_{\mathrm{a.s.}}(1)$. In the following, we demonstrate the asymptotic normality of $\widehat{\boldsymbol{\theta}}$. By the Taylor expansion, we have

$$\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^{0}=-\left\{\partial^{2}Q_{n}^{*}\left(\boldsymbol{\beta}^{0},\boldsymbol{\alpha}^{0}\right)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}\right\}^{-1}\left\{\partial Q_{n}^{*}\left(\boldsymbol{\beta}^{0},\alpha^{0}\right)/\partial\boldsymbol{\theta}\right\}\left\{1+o_{p}\left(1\right)\right\}.$$

Define $\dot{\psi}_{n}^{*}\left(\boldsymbol{\beta},\boldsymbol{\alpha}\right)=\left\{\dot{\psi}_{n,1}^{*}\left(\boldsymbol{\beta},\boldsymbol{\alpha}\right),\ldots,\dot{\psi}_{n,\kappa}^{*}\left(\boldsymbol{\beta},\boldsymbol{\alpha}\right)\right\}^{\mathrm{T}}$, where

$$\dot{\psi}_{n,k}^{*}\left(\boldsymbol{\beta},\boldsymbol{\alpha}\right)=n^{-1}\sum\nolimits_{i=1}^{n}\widehat{\mathbf{D}}_{i}^{\mathrm{T}}\left(\boldsymbol{\beta},\boldsymbol{\alpha}\right)\Delta_{i}\left(\boldsymbol{\beta},\boldsymbol{\alpha}\right)\Lambda_{k}\Delta_{i}\left(\boldsymbol{\beta},\boldsymbol{\alpha}\right)\widehat{\mathbf{D}}_{i}\left(\boldsymbol{\beta},\boldsymbol{\alpha}\right).$$

By the definition of $Q_{n}^{*}\left(\boldsymbol{\beta}^{0},\alpha^{0}\right)$ given in (4.9), it can be shown that

$$\begin{aligned} \partial Q_{n}^{*}\left(\boldsymbol{\beta}^{0},\boldsymbol{\alpha}^{0}\right)/\partial\boldsymbol{\theta} &= -2\dot{\psi}_{n}^{*}\left(\boldsymbol{\beta}^{0},\boldsymbol{\alpha}^{0}\right)^{\mathrm{T}}\Psi_{n}^{*}\left(\boldsymbol{\beta}^{0},\boldsymbol{\alpha}^{0}\right)^{-1}\psi_{n}^{*}\left(\boldsymbol{\beta}^{0},\boldsymbol{\alpha}^{0}\right)+O_{p}\left(n^{-1}\right)\ \text{and} \\ \partial^{2}Q_{n}^{*}\left(\boldsymbol{\beta}^{0},\boldsymbol{\alpha}^{0}\right)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}} &= 2\dot{\psi}_{n}^{*}\left(\boldsymbol{\beta}^{0},\boldsymbol{\alpha}^{0}\right)^{\mathrm{T}}\Psi_{n}^{*}\left(\boldsymbol{\beta}^{0},\boldsymbol{\alpha}^{0}\right)^{-1}\dot{\psi}_{n}^{*}\left(\boldsymbol{\beta}^{0},\boldsymbol{\alpha}^{0}\right)+o_{p}\left(1\right). \end{aligned}$$

By Lemmas 11 and 10, we have that

$$\dot{\psi}_{n}^{*}\left(\boldsymbol{\beta}^{0},\boldsymbol{\alpha}^{0}\right) = \dot{\psi}_{n}\left(\boldsymbol{\beta}^{0},\boldsymbol{\alpha}^{0}\right)+O_{p}\left(N^{3/2}n^{-1/2}+N^{-r+1}\right)\ \text{and} \tag{A.12}$$

$$n\Psi_{n}^{*}\left(\boldsymbol{\beta}^{0},\boldsymbol{\alpha}^{0}\right) = \Psi_{n}\left(\boldsymbol{\beta}^{0},\boldsymbol{\alpha}^{0}\right)+O_{p}\left(N^{3/2}n^{-1/2}+N^{-r+1}\right). \tag{A.13}$$

The above results imply that

$$\partial^2 Q_n^*\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}{=}2n\ \dot{\psi}_n\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)^{\mathrm{T}}\Psi_n\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)^{-1}\dot{\psi}_n\left(\boldsymbol{\beta}^0\ ,\boldsymbol{\alpha}^0\right)+o_p\left(n\right). \tag{A.14}$$

We next define $\psi_n\left(\boldsymbol{\beta},\boldsymbol{\alpha}\right)=\left\{\psi_{n,1}\left(\boldsymbol{\beta},\boldsymbol{\alpha}\right)^{\mathrm{T}},\ldots,\psi_{n,\kappa}\left(\boldsymbol{\beta},\boldsymbol{\alpha}\right)^{\mathrm{T}}\right\}^{\mathrm{T}}$, where

$$\psi_{n,k}\left(\boldsymbol{\beta},\boldsymbol{\alpha}\right)=n^{-1}\sum\nolimits_{i=1}^{n}\mathbf{D}_i^{\mathrm{T}}\left(\boldsymbol{\beta},\boldsymbol{\alpha}\right)\Delta_i\left(\boldsymbol{\beta},\boldsymbol{\alpha}\right)\Lambda_k\Delta_i\left(\boldsymbol{\beta},\boldsymbol{\alpha}\right)\left(Y_i-\mu_i\right).$$

Then, for $N^4n^{-1}=o\left(1\right)$, $N^{-4r+2}n=o\left(1\right)$ with $r>3/2$, and $1\leq k\leq\kappa$, we employ Lemma 11 and obtain

$$\begin{aligned}
& \psi_{n,k}^*\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)-\psi_{n,k}\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)\\
= & \ n^{-1}\sum\nolimits_{i=1}^{n}\left\{\widehat{\mathbf{D}}_i^{\mathrm{T}}\left(\boldsymbol{\beta},\boldsymbol{\alpha}\right)-\mathbf{D}_i^{\mathrm{T}}\left(\boldsymbol{\beta},\boldsymbol{\alpha}\right)\right\}\Delta_i\left(\boldsymbol{\beta},\boldsymbol{\alpha}\right)\Lambda_k\left(Y_i-\widehat{\mu}_i\left(\ \boldsymbol{\beta},\boldsymbol{\alpha}\right)\right)\\
& +n^{-1}\sum\nolimits_{i=1}^{n}\mathbf{D}_i^{\mathrm{T}}\left(\ \boldsymbol{\beta},\boldsymbol{\alpha}\right)\Delta_i\left(\boldsymbol{\beta},\boldsymbol{\alpha}\right)\Lambda_k\left(\mu_i\left(\boldsymbol{\beta},\boldsymbol{\alpha}\right)-\widehat{\mu}_i\left(\boldsymbol{\beta},\boldsymbol{\alpha}\right)\right)\\
= & \ O_p\left(N^{3/2}n^{-1/2}+N^{-r+1}\right)O_p\left(n^{-1/2}+N^{1/2}n^{-1/2}+N^{-r}\right)\\
& +O_p\left(n^{-1/2}\right)O_p\left(N^{1/2}n^{-1/2}+N^{-r}\right)\\
= & \ o_p\left(n^{-1/2}\right).
\end{aligned} \tag{A.15}$$

By (A.12), (A.13) and (A.15),

$$\partial Q_n^*\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)/\partial\theta=-2n\dot{\psi}_n\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)^{\mathrm{T}}\Psi_n\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)^{-1}\psi_n\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)+o_p\left(n^{1/2}\right). \tag{A.16}$$

This, together with (A.14), leads to

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0 \ = & \ \left\{\widetilde{\dot{\psi}}_n\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)^{\mathrm{T}}\widetilde{\Psi}_n\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)^{-1}\widetilde{\dot{\psi}}_n\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)\right\}^{-1}\times\\
& \left\{\widetilde{\dot{\psi}}_n\left(\boldsymbol{\beta}^0\ ,\alpha^0\right)^{\mathrm{T}}\widetilde{\Psi}_n\left(\boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)^{-1}\widetilde{\psi}_n\left(\ \boldsymbol{\beta}^0,\boldsymbol{\alpha}^0\right)\right\}+o_p\left(n^{-1/2}\right).
\end{aligned}$$

By the Lindeberg-Feller Central Limit Theorem and Condition (C5), we then obtain the asymptotic normality of $\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0$ presented in Theorem 5. ∎

Proof of Theorem 2. Applying Lemma 11 and the fact that

$$\left\| \left(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}, \widehat{\boldsymbol{\alpha}}^{\mathrm{T}} \right)^{\mathrm{T}} - \left(\boldsymbol{\beta}^{0\mathrm{T}}, \boldsymbol{\alpha}^{0\mathrm{T}} \right)^{\mathrm{T}} \right\| = O_p\left(n^{-1/2}\right),$$

we are able to prove this theorem straightforwardly. ■

A.3 Proof of Theorem 3

We consider the three steps given below to show the oracle properties of the PQIF estimators.

Step I: Find the convergence rate of $\left\{ \left(\widehat{\boldsymbol{\beta}}_{-1}^{\mathrm{PQIF}} \right)^{\mathrm{T}}, \left(\widehat{\boldsymbol{\alpha}}^{\mathrm{PQIF}} \right)^{\mathrm{T}} \right\}^{\mathrm{T}}$. Let $\widetilde{\boldsymbol{\beta}}_{-1} = \boldsymbol{\beta}_{-1}^0 + n^{-1/2}\mathbf{v}_{-1} = \left(\widetilde{\beta}_2, \ldots, \widetilde{\beta}_p\right)^{\mathrm{T}}$, $\widetilde{\beta}_1 = \sqrt{1 - \left\|\widetilde{\boldsymbol{\beta}}_{-1}\right\|^2}$, $\widetilde{\boldsymbol{\beta}} = \left(\widetilde{\beta}_1, \widetilde{\boldsymbol{\beta}}_{-1}^{\mathrm{T}}\right)^{\mathrm{T}}$, and $\widetilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha}^0 + n^{-1/2}\mathbf{w} = (\widetilde{\alpha}_{d_1+1}, \ldots, \widetilde{\alpha}_d)^{\mathrm{T}}$, where $\mathbf{v} = \left(v_1, \mathbf{v}_{-1}^{\mathrm{T}}\right)^{\mathrm{T}} = (v_1, \ldots, v_p)^{\mathrm{T}}$, $\mathbf{w} = (w_{d_1+1}, \ldots, w_d)^{\mathrm{T}}$, and $\|\mathbf{v}\|_2 = \|\mathbf{w}\|_2 = C$ for some positive constant C. Denote $\widetilde{\boldsymbol{\theta}} = \left(\widetilde{\boldsymbol{\beta}}_{-1}^{\mathrm{T}}, \widetilde{\boldsymbol{\alpha}}^{\mathrm{T}}\right)^{\mathrm{T}}$, $\theta = \left(\boldsymbol{\beta}_{-1}^{\mathrm{T}}, \boldsymbol{\alpha}^{\mathrm{T}}\right)^{\mathrm{T}}$, $\dot{Q}_n^*(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \partial Q_n^*(\boldsymbol{\beta}, \boldsymbol{\alpha})/\partial \boldsymbol{\theta}$ and $\ddot{Q}_n^*(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \partial^2 Q_n^*(\boldsymbol{\beta}, \boldsymbol{\alpha})/\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^{\mathrm{T}}$. Then,

$$Q_n^*\left(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\alpha}}\right) - Q_n^*\left(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right) = \left(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\right)^{\mathrm{T}} \dot{Q}_n^*\left(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right) + \frac{1}{2}\left(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\right)^{\mathrm{T}} \ddot{Q}_n^*\left(\boldsymbol{\beta}^*, \boldsymbol{\alpha}^*\right)\left(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\right), \tag{A.17}$$

for some $\left(\boldsymbol{\beta}^{*\mathrm{T}}, \boldsymbol{\alpha}^{*\mathrm{T}}\right)^{\mathrm{T}}$ that lies between $\left(\boldsymbol{\beta}^{0\mathrm{T}}, \boldsymbol{\alpha}^{0\mathrm{T}}\right)^{\mathrm{T}}$ and $\left(\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}, \widetilde{\boldsymbol{\alpha}}^{\mathrm{T}}\right)^{\mathrm{T}}$. By (A.14) and (A.16), we have, with probability approaching 1,

$$\begin{aligned}
&\left(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\right)^{\mathrm{T}} \ddot{Q}_n^*\left(\boldsymbol{\beta}^*, \boldsymbol{\alpha}^*\right)\left(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\right) \\
=\;& \left(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\right)^{\mathrm{T}} \ddot{Q}_n^*\left(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)\left(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\right) + O\left(C^3 n^{-1/2}\right) \\
\asymp\;& 2n\left(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\right)^{\mathrm{T}} \dot{\psi}_n\left(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)^{\mathrm{T}} \Psi_n\left(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)^{-1} \dot{\psi}_n\left(\boldsymbol{\beta}^0, \alpha^0\right)\left(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\right) + o\left(C^2\right) + O\left(C^3 n^{-1/2}\right) \\
\asymp\;& C^2 + O\left(C^3 n^{-1/2}\right)
\end{aligned}$$

and $\left(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}^0\right)^{\mathrm{T}} \dot{Q}_n^*\left(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)=O_p(C)$.

Next, let

$$
\begin{aligned}
a_n &= \max_{2 \leq s \leq p}\left\{\left|p_{\lambda_{n1}}^{\prime}\left(\left|\beta_s^0\right|\right)\right|, \beta_s^0 \neq 0\right\} \\
b_n &= \max_{2 \leq s \leq p}\left\{\left|p_{\lambda_{n1}}^{\prime\prime}\left(\left|\beta_s^0\right|\right)\right|, \beta_s^0 \neq 0\right\} \\
c_n &= \max_{d_1+1 \leq l \leq d}\left\{\left|p_{\lambda_{n2}}^{\prime}\left(\left|\alpha_l^0\right|\right)\right|, \alpha_l^0 \neq 0\right\} \\
d_n &= \max_{d_1+1 \leq l \leq d}\left\{\left|p_{\lambda_{n2}}^{\prime\prime}\left(\left|\alpha_l^0\right|\right)\right|, \alpha_l^0 \neq 0\right\}.
\end{aligned}
$$

Under the assumptions that $\lambda_{n1} \rightarrow 0$ and $\lambda_{n2} \rightarrow 0$, we have that $a_n=0$ and $c_n=0$. From the Taylor expansion and the Cauchy-Schwarz inequality, as $n \rightarrow \infty$, we further have that

$$
\begin{aligned}
&-\left\{n \sum_{s=2}^p p_{\lambda_{n1}}\left(\left|\widetilde{\beta}_s\right|\right)-n \sum_{s=2}^p p_{\lambda_{n1}}\left(\left|\beta_s^0\right|\right)\right\}-\left\{n \sum_{l=d_1+1}^d p_{\lambda_{n2}}\left(\left|\widetilde{\alpha}_l\right|\right)-n \sum_{l=d_1+1}^d p_{\lambda_{n2}}\left(\left|\alpha_l^0\right|\right)\right\} \\
&\quad \leq -n \sum_{s=2}^{p_1}\left\{p_{\lambda_{n1}}\left(\left|\widetilde{\beta}_s\right|\right)-p_{\lambda_{n1}}\left(\left|\beta_s^0\right|\right)\right\}-n \sum_{l=d_1+1}^{d_1+d_{20}}\left\{p_{\lambda_{n2}}\left(\left|\widetilde{\alpha}_l\right|\right)-p_{\lambda_{n2}}\left(\left|\alpha_l^0\right|\right)\right\} \\
&\quad \leq n\left(n^{-1/2} \sqrt{p_1} a_n\left\|\mathbf{v}_{-1}\right\|_2+n^{-1} b_n\left\|\mathbf{v}_{-1}\right\|_2^2+n^{-1/2} \sqrt{d_{20}} c_n\|\mathbf{w}\|_2+n^{-1} d_n\|\mathbf{w}\|_2^2\right) \\
&\quad \leq C^2\left(b_n+d_n\right). \qquad \text{(A.18)}
\end{aligned}
$$

When $b_n \rightarrow 0$, $d_n \rightarrow 0$, and C is sufficiently large, the second term on the right-hand side of (A.17) dominates its first term and (A.18). Accordingly, for any give $\nu>0$, there exists a large constant $\tilde{C}$ such that,

$$
P\left\{\inf_{V_{12}} \mathcal{L}_n^*\left(\boldsymbol{\beta}^0+n^{-1/2} \mathbf{v}, \alpha^0+n^{-1/2} \mathbf{w}\right)>\mathcal{L}_n^*\left(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)\right\} \geq 1-\nu,
$$

as $n \rightarrow \infty$, where $V_{12}=\left\{\left(\mathbf{v}^{\mathrm{T}}, \mathbf{w}^{\mathrm{T}}\right)^{\mathrm{T}}:\|\mathbf{v}\|=\tilde{C} \text{ and } \|\mathbf{w}\|=\tilde{C}\right\}$. Consequently, the rate of convergence of $\left\{\left(\widehat{\boldsymbol{\beta}}_{-1}^{\mathrm{PQIF}}\right)^{\mathrm{T}},\left(\widehat{\boldsymbol{\alpha}}^{\mathrm{PQIF}}\right)^{\mathrm{T}}\right\}^{\mathrm{T}}$ is $O_p\left(n^{-1/2}\right)$.

Step II: Demonstrate the sparsity of $\left\{\left(\widehat{\boldsymbol{\beta}}_{-1}^{\mathrm{PQIF}}\right)^{\mathrm{T}}, \left(\widehat{\boldsymbol{\alpha}}^{\mathrm{PQIF}}\right)^{\mathrm{T}}\right\}^{\mathrm{T}}$. Assume that $\boldsymbol{\beta}_{(1)} = \left\{\beta_1, \left(\boldsymbol{\beta}_{(1),-1}\right)^{\mathrm{T}}\right\}^{\mathrm{T}}$ and $\boldsymbol{\alpha}_{(1)}$ satisfy $\left\|\boldsymbol{\beta}_{(1)} - \boldsymbol{\beta}_{(1)}^0\right\| = O_p\left(n^{-1/2}\right)$ and $\left\| \boldsymbol{\alpha}_{(1)} - \boldsymbol{\alpha}_{(1)}^0\right\| = O_p\left(n^{-1/2}\right)$, respectively. We then show, with probability tending to 1, that

$$\mathcal{L}_n^*\left\{\begin{pmatrix}\boldsymbol{\beta}_{(1)}\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\boldsymbol{\alpha}_{(1)}\\ \mathbf{0}\end{pmatrix}\right\} = \min_{\mathcal{C}} \mathcal{L}_n^*\left\{\begin{pmatrix}\boldsymbol{\beta}_{(1)}\\ \boldsymbol{\beta}_{(2)}\end{pmatrix}, \begin{pmatrix}\boldsymbol{\alpha}_{(1)}\\ \boldsymbol{\alpha}_{(2)}\end{pmatrix}\right\}, \tag{A.19}$$

as $n \to \infty$, where $\mathcal{C} = \left\{(\boldsymbol{\beta}_{(2)}^{\mathrm{T}},\ \boldsymbol{\alpha}_{(2)}^{\mathrm{T}})^{\mathrm{T}} : \left\| \boldsymbol{\beta}_{(2)}\right\| \leq C^* n^{-1/2} \text{ and } \left\|\boldsymbol{\alpha}_{(2)}\right\| \leq C^* n^{-1/2}\right\}$ and C^* is a positive constant.

When $\beta_s \neq 0$, one has $\partial \mathcal{L}_n^*\left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)/\partial \beta_s = \partial Q_n^*\left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)/\partial\beta_s + np'_{\lambda_{n1}}\left(|\beta_s|\right)\operatorname{sgn}\left(\beta_s\right)$. By (A.16), it can be shown that $\partial Q_n^*\left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)/\partial\beta_s = O_p\left(n^{1/2}\right)$. Thus,

$$\partial \mathcal{L}_n^*\left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)/\partial \beta_s = n\lambda_{n1}\left\{\lambda_{n1}^{-1} n^{-1/2} + \lambda_{n1}^{-1} p'_{\lambda_{n1}}\left(|\beta_s|\right)\operatorname{sgn}\left(\beta_s\right)\right\}.$$

Using the fact that $\liminf_{n\to\infty} \liminf_{\beta_s \to 0^+} \lambda_{n1}^{-1} p'_{\lambda_{n1}}\left(|\beta_s|\right) > 0$ and $n^{-1/2}\lambda_{n1}^{-1} \to 0$, we further obtain $\partial \mathcal{L}_n^*\left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)/\partial\beta_s > 0$ for $\beta_s > 0$ and $\partial \mathcal{L}_n^*\left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)/\partial\beta_s < 0$ for $\beta_s < 0$. Analogously, we can demonstrate that $\partial \mathcal{L}_n^*\left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)/\partial\alpha_l > 0$ for $\alpha_l > 0$ and $\partial \mathcal{L}_n^*\left(\boldsymbol{\beta},\ \boldsymbol{\alpha}\right)/\partial\alpha_l < 0$ for $\alpha_l < 0$. Consequently, the minimum of $\mathcal{L}_n^*\left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)$ is attained at $\boldsymbol{\beta}_{(2)} = 0$ and $\boldsymbol{\alpha}_{(2)} = 0$, which proves (A.19). This, together with the result of Step I, implies that, with probability tending to 1, $\widehat{\boldsymbol{\beta}}_{(2)}^{\mathrm{PQIF}} = 0$ and $\widehat{\boldsymbol{\alpha}}_{(2)}^{\mathrm{PQIF}} = 0$, as $n \to \infty$. This completes the proof of part (i) in Theorem 3.

Step III: Demonstrate the asymptotic normality of $\widehat{\boldsymbol{\beta}}_{(1),-1}^{\mathrm{PQIF}}$ and $\widehat{\boldsymbol{\alpha}}_{(1)}^{\mathrm{PQIF}}$. Define

$$\begin{aligned}
R_{\lambda_{n1}} &= \left\{p'_{\lambda_{n1}}\left(\left|\beta_2^0\right|\right)\operatorname{sgn}\left(\beta_2^0\right), \ldots, p'_{\lambda_{n1}}\left(\left|\beta_{p_1}^0\right|\right)\operatorname{sgn}\left(\beta_{p_1}^0\right)\right\}^{\mathrm{T}},\\
\Sigma_{\lambda_{n1}} &= \operatorname{diag}\left\{p''_{\lambda_{n1}}\left(\left|\beta_2^0\right|\right), \ldots, p''_{\lambda_{n1}}\left(\left|\beta_{p_1}^0\right|\right)\right\},\\
R_{\lambda_{n2}} &= \left\{p'_{\lambda_{n2}}\left(\left|\alpha_{d_1+1}^0\right|\right)\operatorname{sgn}\left(\alpha_{d_1+1}^0\right), \ldots, p'_{\lambda_{n2}}\left(\left|\alpha_{d_1+d_{20}}^0\right|\right)\operatorname{sgn}\left(\alpha_{d_1+d_{20}}^0\right)\right\}^{\mathrm{T}}, \text{ and}
\end{aligned}$$

$$\Sigma_{\lambda_{n2}} = \operatorname{diag}\left\{p''_{\lambda_{n2}}\left(\left|\alpha_{d_1+1}^0\right|\right), \ldots, p''_{\lambda_{n2}}\left(\left|\alpha_{d_1+d_{20}}^0\right|\right)\right\}. \tag{A.20}$$

By (A.19), with probability tending to 1, $\widehat{\boldsymbol{\beta}}_{(1),-1}^{\mathrm{PQIF}}$ and $\widehat{\boldsymbol{\alpha}}_{(1)}^{\mathrm{PQIF}}$ are obtained by minimizing

$$\mathcal{L}_n^*\left(\boldsymbol{\beta}_{(1)}, \boldsymbol{\alpha}_{(1)}\right) = \frac{1}{2}Q_n^*\left(\boldsymbol{\beta}_{(1)}, \boldsymbol{\alpha}_{(1)}\right) + n\sum\nolimits_{s=2}^{p_1} p_{\lambda_{n1}}\left(|\beta_s|\right) + n\sum\nolimits_{l=d_1+1}^{d_1+d_{20}} p_{\lambda_{n2}}\left(|\alpha_l|\right),$$

where $Q_n^*\left(\boldsymbol{\beta}_{(1)}, \alpha_{(1)}\right)$ is defined similar to $Q_n^*\left(\boldsymbol{\beta}, \boldsymbol{\alpha}\right)$ by using their nonzero components. We then have

$$\begin{aligned}
\mathbf{0} &= \begin{pmatrix} \partial\mathcal{L}_n^*\left(\widehat{\boldsymbol{\beta}}_{(1)}^{\mathrm{PQIF}}, \widehat{\boldsymbol{\alpha}}_{(1)}^{\mathrm{PQIF}}\right)/\partial\boldsymbol{\beta}_{(1),-1} \\ \partial\mathcal{L}_n^*\left(\widehat{\boldsymbol{\beta}}_{(1)}^{\mathrm{PQIF}}, \widehat{\boldsymbol{\alpha}}_{(1)}^{\mathrm{PQIF}}\right)/\partial\boldsymbol{\alpha}_{(1)} \end{pmatrix} \\
&= \frac{1}{2}\dot{Q}_n^*\left(\widehat{\boldsymbol{\beta}}_{(1)}^{\mathrm{PQIF}}, \widehat{\boldsymbol{\alpha}}_{(1)}^{\mathrm{PQIF}}\right) + n\begin{pmatrix} R_{\lambda_{n1}} \\ R_{\lambda_{n2}} \end{pmatrix} + n\Sigma_\lambda \begin{pmatrix} \widehat{\boldsymbol{\beta}}_{(1),-1}^{\mathrm{PQIF}} - \boldsymbol{\beta}_{(1),-1}^0 \\ \widehat{\boldsymbol{\alpha}}_{(1)}^{\mathrm{PQIF}} - \boldsymbol{\alpha}_{(1)}^0 \end{pmatrix} + O_p(1),
\end{aligned}$$

where $\Sigma_\lambda = \begin{pmatrix} \Sigma_{\lambda_{n1}} & \mathbf{0} \\ \mathbf{0}^{\mathrm{T}} & \Sigma_{\lambda_{n2}} \end{pmatrix}$. Subsequently, applying similar techniques to those used the proof of Theorem 5, we obtain that

$$\sqrt{n}\left(\Sigma_{n1}^{(1)}\right)^{-1/2}\left(\Sigma_{n1}^{(1)} + \Sigma_\lambda\right)\left\{\begin{pmatrix} \widehat{\boldsymbol{\beta}}_{(1),-1}^{\mathrm{PQIF}} - \boldsymbol{\beta}_{(1),-1}^0 \\ \widehat{\boldsymbol{\alpha}}_{(1)}^{\mathrm{PQIF}} - \boldsymbol{\alpha}_{(1)}^0 \end{pmatrix} + \left(\Sigma_{n1}^{(1)} + \Sigma_\lambda\right)^{-1}\begin{pmatrix} R_{\lambda_{n1}} \\ R_{\lambda_{n2}} \end{pmatrix}\right\} \to N\left(\mathbf{0}, \mathbf{I}_{(p_1+d_{20})}\right).$$

Finally, under the assumptions that $\lambda_{n1} \to 0$ and $\lambda_{n2} \to 0$, and the fact that $\sqrt{n}\Sigma_\lambda = \sqrt{n}R_{\lambda_{n1}} = \sqrt{n}R_{\lambda_{n2}} = \mathbf{0}$, we complete the proof of part (ii) in Theorem 3.

A.4 Proof of Theorem 4

Assume that the true parameters $\left(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)$ in model (4.1) are known. Then, the resulting penalized estimator of $\boldsymbol{\gamma}$, $\widetilde{\boldsymbol{\gamma}}^{\mathrm{PQIF}} = \left\{\left(\widetilde{\gamma}_l^{\mathrm{PQIF}}\right)^{\mathrm{T}}, 0 \le l \le d_1 + 1\right\}^{\mathrm{T}}$, is obtained by minimizing the following penalized-QIF:

$$\mathcal{L}_n(\boldsymbol{\gamma}) = \frac{1}{2} Q_n\left(\boldsymbol{\gamma}, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right) + n \sum\nolimits_{l=1}^{d_1} p_{\lambda_{n3}}\left(\| \boldsymbol{\gamma}_l \|\right).$$

Define $\widetilde{g}^{\mathrm{PQIF}}(u) = B_1(u)^{\mathrm{T}} \widetilde{\boldsymbol{\gamma}}_0^{\mathrm{PQIF}}$ and $\widetilde{\alpha}_l^{\mathrm{PQIF}}(t) = B_2(t)^{\mathrm{T}} \widetilde{\boldsymbol{\gamma}}_l^{\mathrm{PQIF}}$. In the following, we will show the convergence rate for $\widetilde{g}^{\mathrm{PQIF}}(\cdot)$ and $\widetilde{\alpha}_l^{\mathrm{PQIF}}(\cdot)$ as well as demonstrate the sparsity of $\widetilde{\boldsymbol{\gamma}}^{\mathrm{PQIF}}$.

Let $\widetilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma}^0 + \tilde{\varrho}_n \mathbf{v} = \left\{ (\widetilde{\boldsymbol{\gamma}}_l)^{\mathrm{T}}, 0 \le l \le d_1 + 1 \right\}^{\mathrm{T}}$, where $\mathbf{v} = \left\{ (\mathbf{v}_l)^{\mathrm{T}}, 0 \le l \le d_1 + 1 \right\}^{\mathrm{T}}$, $\mathbf{v}_l = (v_{1,l}, \ldots, v_{N+q,l})^{\mathrm{T}}$, and $\|\mathbf{v}\| = C$ for some positive constant C. In addition, let $\dot{Q}_n\left(\boldsymbol{\gamma}, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right) = \partial Q_n^*(\boldsymbol{\beta}, \boldsymbol{\alpha}) / \partial \boldsymbol{\gamma}$ and $\ddot{Q}_n\left(\boldsymbol{\gamma}, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right) = \partial^2 Q_n^*(\boldsymbol{\beta}, \boldsymbol{\alpha}) / \partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^{\mathrm{T}}$. Then, we obtain

$$\begin{aligned} & Q_n\left(\widetilde{\boldsymbol{\gamma}}, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right) - Q_n\left(\boldsymbol{\gamma}^0, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right) \\ = & \left(\widetilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0\right)^{\mathrm{T}} \dot{Q}_n\left(\boldsymbol{\gamma}^0, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right) + \frac{1}{2}\left(\widetilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0\right)^{\mathrm{T}} \ddot{Q}_n\left(\boldsymbol{\gamma}^*, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right)\left(\widetilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0\right), \end{aligned} \quad \text{(A.21)}$$

where $\boldsymbol{\gamma}^*$ lies between $\widetilde{\boldsymbol{\gamma}}$ and $\boldsymbol{\gamma}^0$. By (A.8) and (A.10), with probability approaching 1,

$$\begin{aligned} \left(\widetilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0\right)^{\mathrm{T}} \ddot{Q}_n\left(\boldsymbol{\gamma}^*, \boldsymbol{\beta}^0, \alpha^0\right)\left(\widetilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0\right) & \asymp 2\left(\widetilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0\right)^{\mathrm{T}} \Omega_n^{\mathrm{T}} \Xi_n^{-1} \Omega_n \left(\widetilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0\right) + O\left(nC^3 \tilde{\varrho}_n^3\right) \\ & \asymp C^2 \tilde{\varrho}_n^2 \left(nN^{-1}\right) + nC^3 \tilde{\varrho}_n^3. \end{aligned}$$

Furthermore, by the weak law of large numbers and (A.6), there exist constants $0 < C_1 < \infty$ and $0 < C_2 < \infty$ such that

$$\begin{aligned} \left(\widetilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0\right)^{\mathrm{T}} \dot{Q}_n\left(\boldsymbol{\gamma}^0, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right) & \le C \tilde{\varrho}_n \left\| \dot{Q}_n\left(\boldsymbol{\gamma}^0, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0\right) \right\| \\ & \le 2CC_1 n \tilde{\varrho}_n \left\| \phi_n^0 \right\| \le 2CC_1C_2 n \tilde{\varrho}_n \left(n^{-1/2} + N^{-r-1/2} \right). \end{aligned}$$

Accordingly, by the assumption $N \asymp n^{1/(2r+1)}$ and taking $\tilde{\varrho}_n = \sqrt{N} n^{-r/(2r+1)}$, we obtain, with probability approaching 1, that

$$\left(\widetilde{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^{0}\right)^{\mathrm{T}} \ddot{Q}_{n}\left(\boldsymbol{\gamma}^{*}, \boldsymbol{\beta}^{0},\ \boldsymbol{\alpha}^{0}\right)\left(\widetilde{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^{0}\right) \asymp C^{2} N$$

and $\left(\widetilde{\boldsymbol{\gamma}}-\ \boldsymbol{\gamma}^{0}\right)^{\mathrm{T}} \dot{Q}_{n}\left(\boldsymbol{\gamma}^{0}\ , \boldsymbol{\beta}^{0}, \boldsymbol{\alpha}^{0}\right)=O\left(CN\right)$.

Next, let

$$\begin{aligned}
\tilde{a}_{n} &= \max_{1 \leq l \leq d_{1}}\left\{\left|p_{\lambda_{n3}}^{\prime}\left(\left\|\boldsymbol{\gamma}_{l}^{0}\right\|\right)\right|, \boldsymbol{\gamma}_{l}^{0} \neq \mathbf{0}\right\} \\
\tilde{b}_{n} &= \max_{1 \leq l \leq d_{1}}\left\{\left|p_{\lambda_{n3}}^{\prime\prime}\left(\left\|\boldsymbol{\gamma}_{l}^{0}\right\|\right)\right|,\ \boldsymbol{\gamma}_{l}^{0} \neq \mathbf{0}\right\}.
\end{aligned}$$

Under the assumptions that $\lambda_{n3} \rightarrow 0$, we have that $\tilde{a}_{n}=0$. By the Taylor expansion and the Cauchy-Schwarz inequality, as $n \rightarrow \infty$, we further have that

$$\begin{aligned}
& -\left\{n \sum_{l=1}^{d_{1}} p_{\lambda_{n3}}\left(\|\widetilde{\gamma}_{l}\|_{\mathbf{W}n}\right)-n \sum_{l=1}^{d_{1}} p_{\lambda_{n3}}\left(\|\ \boldsymbol{\gamma}_{l}^{0}\|_{\mathbf{W}n}\right)\right\} \\
\leq &\ -n \sum_{l=1}^{d_{10}}\left\{p_{\lambda_{n3}}\left(\|\widetilde{\gamma}_{l}\|_{\mathbf{W}n}\right)-p_{\lambda_{n3}}\left(\left\|\boldsymbol{\gamma}_{l}^{0}\right\|_{\mathbf{W}n}\right)\right\} \\
\leq &\ n \tilde{b}_{n} \sum_{l=1}^{d_{10}}\left\|\ \widetilde{\gamma}_{l}-\boldsymbol{\gamma}_{l}^{0}\right\|_{\mathbf{W}n}^{2} \asymp n \tilde{b}_{n} C^{2} \tilde{\varrho}_{n}^{2} N^{-1}=C^{2} \tilde{b}_{n} N. \qquad \text{(A.22)}
\end{aligned}$$

When $\tilde{b}_{n} \rightarrow 0$ and C is sufficiently large, the second term on the right-hand side of (A.21) dominates its first term and (A.22). Thus, for any given $\nu>0$, there exists a large constant C such that,

$$P\left\{\inf_{V} \mathcal{L}_{n}\left(\boldsymbol{\gamma}^{0}+\tilde{\varrho}_{n} \mathbf{v}\right)>\mathcal{L}_{n}\left(\boldsymbol{\gamma}^{0}\right)\right\} \geq 1-\nu,$$

as $n \rightarrow \infty$, where $V=\{\mathbf{v}:\|\mathbf{v}\|=C\}$. As a result, $\left\|\widetilde{\boldsymbol{\gamma}}^{\mathrm{PQIF}}-\boldsymbol{\gamma}^{0}\right\|=O_{p}\left(\tilde{\varrho}_{n}\right)$, which leads to

$$\left\|\widetilde{g}^{\mathrm{PQIF}}\left(\cdot\right)-g\left(\cdot\right)\right\| \asymp N^{-1/2}\left\|\widetilde{\boldsymbol{\gamma}}_{0}^{\mathrm{PQIF}}-\boldsymbol{\gamma}_{0}^{0}\right\|=O_{p}\left(N^{-r/(2r+1)}\right)$$

and $\left\|\widetilde{\alpha}_{l}^{\mathrm{PQIF}}\left(\cdot\right)-\alpha_{l}\left(\cdot\right)\right\| \asymp N^{-1/2}\left\|\widetilde{\boldsymbol{\gamma}}_{l}^{\mathrm{PQIF}}-\boldsymbol{\gamma}_{l}^{0}\right\|=O_{p}\left(N^{-r/(2r+1)}\right)$.

Finally, let $\boldsymbol{\gamma} = \left\{ \left(\boldsymbol{\gamma}_{(1)}^{\mathrm{T}}\right)_{(d_{10}+1)\times 1}, \left(\boldsymbol{\gamma}_{(2)}^{\mathrm{T}}\right)_{(d_1-d_{10})\times 1} \right\}^{\mathrm{T}}$. We then show that, with probability tending to 1,

$$\mathcal{L}_n \left\{ \left(\boldsymbol{\gamma}_{(1)}^{\mathrm{T}}, \boldsymbol{\gamma}_{(2)}^{\mathrm{T}} \right)^{\mathrm{T}} \right\} = \min_{\mathcal{C}} \mathcal{L}_n \left\{ \left(\boldsymbol{\gamma}_{(1)}^{\mathrm{T}}, \mathbf{0}^{\mathrm{T}} \right)^{\mathrm{T}} \right\},$$

as $n \to \infty$, where $\mathcal{C} = \left\{ \left\| \ \boldsymbol{\gamma}_{(2)} \right\| \le C^* \varrho_n \right\}$ and C^* is a positive constant. When $\|\boldsymbol{\gamma}_l\| \neq 0$, there exists a constant $0 < c < \infty$ such that, with probability approaching 1,

$$\begin{aligned}
\partial \mathcal{L}_n (\boldsymbol{\gamma}) / \partial \ \boldsymbol{\gamma}_l &= \partial Q_n \left(\boldsymbol{\gamma}, \boldsymbol{\beta}^0, \ \boldsymbol{\alpha}^0 \right) / \partial \boldsymbol{\gamma}_l + n p'_{\lambda_{n3}} \left(\|\boldsymbol{\gamma}_l\|_{\mathbf{W}_n} \right) \|\boldsymbol{\gamma}_l\|_{\mathbf{W}_n}^{-1} \mathbf{W}_n \boldsymbol{\gamma}_l \\
&\asymp \partial Q_n \left(\boldsymbol{\gamma}, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0 \right) / \partial \boldsymbol{\gamma}_l + c N^{-1} n p'_{\lambda_{n3}} \left(\|\boldsymbol{\gamma}_l\|_{\mathbf{W}_n} \right) \boldsymbol{\gamma}_l \\
&= \partial Q_n \left(\boldsymbol{\gamma}, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0 \right) / \partial \boldsymbol{\gamma}_l + c n^{2r/(2r+1)} p'_{\lambda_{n3}} \left(\|\boldsymbol{\gamma}_l\|_{\mathbf{W}_n} \right) \boldsymbol{\gamma}_l .
\end{aligned}$$

By (A.8), it can be shown that $\partial Q_n \left(\ \boldsymbol{\gamma}, \boldsymbol{\beta}^0, \boldsymbol{\alpha}^0 \right) / \partial \boldsymbol{\gamma}_l = O_p \left(n^{r/(2r+1)} \right)$. As a result,

$$\partial \mathcal{L}_n (\boldsymbol{\gamma}) / \partial \boldsymbol{\gamma}_l = n^{2r/(2r+1)} \lambda_{n3} \left\{ \lambda_{n3}^{-1} n^{-r/(2r+1)} + \lambda_{n3}^{-1} p'_{\lambda_{n3}} \left(\|\boldsymbol{\gamma}_l\|_{\mathbf{W}_n} \right) \boldsymbol{\gamma}_l \right\}.$$

Using the fact that $\lambda_{n3}^{-1} n^{-r/(2r+1)} \to 0$ and $\liminf_{n\to\infty} \liminf_{\|\boldsymbol{\gamma}_l\|_{\mathbf{W}_n} \to 0^+} \lambda_{n3}^{-1} p'_{\lambda_{n3}} \left(\|\boldsymbol{\gamma}_l\|_{\mathbf{W}_n} \right) > 0$, we further obtain $\partial \mathcal{L}_n (\boldsymbol{\gamma}) / \partial \gamma_{J,l} > 0$ for $\gamma_{J,l} > 0$ and $\partial \mathcal{L}_n (\boldsymbol{\gamma}) / \partial \gamma_{J,l} < 0$ for $\gamma_{J,l} < 0$. Consequently, the minimum of $\partial \mathcal{L}_n (\boldsymbol{\gamma})$ is attained at $\boldsymbol{\gamma}_l = \mathbf{0}$ for $(d_{10}+1) \le l \le d_1$. This implies, with probability tending to 1, $\widetilde{\boldsymbol{\gamma}}_l^{\mathrm{PQIF}} = \mathbf{0}$ for $(d_{10}+1) \le l \le d_1$. Subsequently, using the fact that $\left\| \left\{ \left(\widehat{\boldsymbol{\beta}}^{\mathrm{PQIF}} \right)^{\mathrm{T}}, \left(\widehat{\boldsymbol{\alpha}}^{\mathrm{PQIF}} \right)^{\mathrm{T}} \right\}^{\mathrm{T}} - \left(\boldsymbol{\beta}^{0\mathrm{T}}, \boldsymbol{\alpha}^{0\mathrm{T}} \right)^{\mathrm{T}} \right\| = O_p \left(n^{-1/2} \right)$ and those assumptions given in Theorem 4, the above results of convergence rate and sparsity can be applied to the penalized estimators $\widehat{\boldsymbol{\gamma}}^{\mathrm{PQIF}}$, $\widehat{g}^{\mathrm{PQIF}}(\cdot)$, and $\widehat{\alpha}_l^{\mathrm{PQIF}}(\cdot)$. This completes the proof.

Appendix B

Proof of Theorems in Chapter 5

We begin this appendix by presenting some notation that will be used in the proofs of theorems. For any positive numbers a_n and b_n, let $a_n \asymp b_n$ denote that $\lim_{n\to\infty} a_n/b_n = c$, where c is a positive constant, and let $a_n \sim b_n$ denote that $\lim_{n\to\infty} a_n/b_n = 1$. Denote the space of the r-th order smooth functions ϕ as $C^{(r)}\left([0,1]\right) = \left\{\phi \left| \phi^{(r)} \in [0,1] \right.\right\}$. For any vector $\zeta = (\zeta_1, \ldots, \zeta_s)^{\mathrm{T}} \in R^s$, denote $\|\zeta\|_\infty = \max\left(|\zeta_1| + \cdots + |\zeta_s|\right)$. For any symmetric matrix $\mathbf{A}$, denote its L_r norm as $\|\mathbf{A}\|_r = \max_{\zeta\in R^s, \zeta\neq 0} \|\mathbf{A}\zeta\|_r \|\zeta\|_r^{-1}$. For any matrix $\mathbf{A} = (A_{ij})_{i=1,j=1}^{s,t}$, denote $\|\mathbf{A}\|_\infty = \max_{1\leq i\leq s} \sum_{j=1}^{t} |A_{ij}|$ and $\mathbf{A}^{\otimes 2} = \mathbf{A}\mathbf{A}^{\mathrm{T}}$. To develop the theoretical results of the proposed estimators, we next present the following technical conditions.

B.1 Regularity Conditions

(C1) The density function $f_{\beta^{\mathrm{T}} X_{ij}}\left(\beta^{\mathrm{T}} x_{ij}\right)$ of random variable $\beta^{\mathrm{T}} X_{ij}$ is bounded away from 0 on S_β for β in a neighborhood of β_0 and satisfies the Lipschitz condition of order 1 on S_β, where $S_\beta = \left\{\beta^{\mathrm{T}} x_{ij};\, x_{ij} \in S\right\}$ and S is a compact support set of X_{ij}. Without loss of generality, we assume $S_\beta = [0,1]$. Similarly, the density function $f_{T_{ij}}(t)$ of random variable T_{ij} is bounded away from 0 on $[0,1]$ and satisfies the Lipschitz condition of order 1 on $[0,1]$.

(C2) The coefficient functions $m_\ell(u)$ satisfy $m_\ell(u) \in C^{(r)}([0,1])$ for given integer $r > 3/2$. Spline order satisfies $q \geq r$.

(C3) The eigenvalues of $E\left(Z_{ij}Z_{ij}^{\mathrm{T}} | T_{ij} = t\right)$ are uniformly bounded away from 0 and ∞ for all $t \in [0,1]$.

(C4) The eigenvalues of $\widetilde{\dot{\psi}}_n\left(\beta^0, \alpha^0\right)$ and $\widetilde{\Psi}_n\left(\beta^0, \alpha^0\right)$ are bounded away from 0 and infinity.

(C5) The eigenvalues of $\mathbf{H}$ are bounded.

(C6) $h_\ell(u) \in C^{(1)}([0,1])$.

B.2 Proofs of Theorems 5 and 6

By maximizing the log-likelihood function, we have

$$\widetilde{\gamma}(\beta^0, \alpha^0) = \left\{\mathbf{Q}(\beta^0)^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{Q}(\beta^0)\right\}^{-1}\left\{\mathbf{Q}(\beta^0)^{\mathrm{T}}\mathbf{H}^{-1}(\mathbf{Y} - \mathbf{T}\alpha^0)\right\}.$$

$\widetilde{\gamma}(\beta^0, \alpha^0)$ can be decomposed into $\widetilde{\gamma} = \widetilde{\gamma}_m + \widetilde{\gamma}_e$, where

$$\begin{aligned}\widetilde{\gamma}_m(\beta^0, \alpha^0) &= \left\{\mathbf{Q}(\beta^0)^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{Q}(\beta^0)\right\}^{-1}\mathbf{Q}(\beta^0)^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{m} \\ \widetilde{\gamma}_e(\beta^0, \alpha^0) &= \left\{\mathbf{Q}(\beta^0)^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{Q}(\beta^0)\right\}^{-1}\mathbf{Q}(\beta^0)^{\mathrm{T}}\mathbf{H}^{-1}(\mathbf{Y} - \mathbf{T}\alpha^0 - \mathbf{m}) \qquad \text{(B.1)}\end{aligned}$$

Define $\hat{\mathbf{\Omega}}_n(\beta) = n^{-1}\mathbf{Q}(\beta)^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{Q}(\beta)$. Under condition (C5), it can be proved by Theorem 5.4.2 of [14] and Berstein's inequality in [2] that for large enough n and with probability tending to 1, there are constants $0 < c_R \leq C_R < \infty$ such that

$$c_R J_n \leq \left\| \hat{\mathbf{\Omega}}_n (\beta)^{-1} \right\|_2 \leq C_R J_n, \tag{B.2}$$

and by the above result and [13], we have $\left\| \hat{\mathbf{\Omega}}_n (\beta)^{-1} \right\|_\infty = O_p (J_n)$ for $\forall \beta \in \Theta$. The first lemma below presents the convergence rates of $\widetilde{m}_\ell \left(u, \beta^0, \alpha^0\right)$ and $\widetilde{m}'_\ell \left(u, \beta^0, \alpha^0\right)$ to $m_\ell (u)$ and $m'_\ell (u)$ given the true parameters β^0 and α^0. The results will be used in the proof of Theorem 5.

Lemma 11 *Under Conditions (C1)-(C5), and $J_n \to \infty$ and $J_n^3 n^{-1} = o(1)$, we have*

(1) $\left|\widetilde{m}_\ell \left(u, \beta^0, \alpha^0\right) - m_\ell (u)\right| = O_p \left(\sqrt{J_n/n} + J_n^{-r}\right)$ *and* $\left|\widetilde{m}'_\ell \left(u, \beta^0, \alpha^0\right) - m'_\ell (u)\right| = O_p \left(\sqrt{J_n^3/n} + J_n^{-r+1}\right)$ *uniformly in* $u \in [0,1]$, *and*

(2) for $1 \leq \ell \leq p$, $\sigma_{\ell n}^{-1} (u) \left[\widetilde{m}_\ell \left(u, \beta^0, \alpha^0\right) - E\left\{m_\ell (u) | \mathbb{Z}, \mathbb{X}, \mathbb{T}\right\}\right] \to N(0,1)$, *where* $\sigma_{\ell n}^2 (u)$ *is defined in (5.9).*

Proof. Let

$$\widetilde{\gamma}_e(\beta^0, \alpha^0) = \left\{\widetilde{\gamma}_{1,e}(\beta^0, \alpha^0)^{\mathrm{T}}, ..., \widetilde{\gamma}_{p,e}(\beta^0, \alpha^0)^{\mathrm{T}}\right\}^{\mathrm{T}},$$

$$\widetilde{\gamma}_m(\beta^0, \alpha^0) = \left\{\widetilde{\gamma}_{1,m}(\beta^0, \alpha^0)^{\mathrm{T}}, ..., \widetilde{\gamma}_{p,m}(\beta^0, \alpha^0)^{\mathrm{T}}\right\}^{\mathrm{T}},$$

Thus

$$\widetilde{m}_l(u, \beta^0, \alpha^0) = \widetilde{m}_{l,e}(u, \beta^0, \alpha^0) + \widetilde{m}_{l,m}(u, \beta^0, \alpha^0) \tag{B.3}$$

where

$$\widetilde{m}_{l,e}(u, \beta^0, \alpha^0) = B(u)^{\mathrm{T}} \widetilde{\gamma}_{l,e}(\beta^0, \alpha^0) = e_\ell^{\mathrm{T}} \mathbf{B}(u) \widetilde{\gamma}_e(\beta^0, \alpha^0),$$

$$\widetilde{m}_{l,m}(u, \beta^0, \alpha^0) = B(u)^{\mathrm{T}} \widetilde{\gamma}_{l,m}(\beta^0, \alpha^0) = e_\ell^{\mathrm{T}} \mathbf{B}(u) \widetilde{\gamma}_m(\beta^0, \alpha^0) \tag{B.4}$$

According to the result on page 149 of [12], for m_ℓ satisfying Condition (C2), there is a function $m_\ell^0(u) = B(u)^{\mathrm{T}} \gamma_\ell^0 \in \mathcal{H}_n$ such that

$$\sup_{u\in[0,1]} \left| m_\ell^0(u) - m_\ell(u) \right| = O\left(J_n^{-r}\right). \tag{B.5}$$

By Berstein's inequality in Bosq(1998), it can be proved that $\left\| n^{-1}\mathbf{Q}(\beta^0)^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{1}_n \right\|_\infty = O_p(J_n^{-1})$. Thus by, for every $u \in [0,1]$,

$$\begin{aligned}
&\left|\widetilde{m}_{\ell,m}\left(u,\beta^0,\alpha^0\right) - m_\ell^0\left(u,\beta^0,\alpha^0\right)\right| \\
= \;& \left| n^{-1} e_\ell^{\mathrm{T}} \mathbf{B}(u)\, \hat{\boldsymbol{\Omega}}_n\left(\beta^0\right)^{-1} \mathbf{Q}(\beta^0)^{\mathrm{T}}\mathbf{H}^{-1}\left\{\mathbf{m} - \mathbf{Q}(\beta^0)\gamma^0\right\} \right| \\
\leq \;& \left| \sum_{j=1}^{J_n} B_s(u) \right| \left\| \hat{\boldsymbol{\Omega}}_n\left(\beta^0\right)^{-1} \right\|_\infty \left\| n^{-1}\mathbf{Q}(\beta^0)^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{1}_n \right\|_\infty O(J_n^{-r}) \\
= \;& O_p(J_n) O_p(J_n^{-1}) O(J_n^{-r}) = O_p(J_n^{-r})
\end{aligned} \tag{B.6}$$

Moreover, for every $u \in [0,1]$, by, with probability approaching 1,

$$\begin{aligned}
& E\left\{\widetilde{m}_{l,e}(u,\beta^0,\alpha^0) | \mathbf{X},\mathbf{T},\mathbf{Z}\right\}^2 \\
= \;& n^{-2} e_\ell^{\mathrm{T}} \mathbf{B}(u) \hat{\boldsymbol{\Omega}}_n\left(\beta^0\right)^{-1} \mathbf{Q}(\beta^0)^{\mathrm{T}}\mathbf{H}^{-1} E(\mathbf{e}\mathbf{e}^{\mathrm{T}}|\mathbf{X},\mathbf{T},\mathbf{Z}) \mathbf{H}^{-1}\mathbf{Q}(\beta^0) \hat{\boldsymbol{\Omega}}_n\left(\beta^0\right)^{-1} \mathbf{B}(u)^{\mathrm{T}} e_\ell \\
= \;& n^{-1}\sigma^2 e_\ell^{\mathrm{T}} \mathbf{B}(u) \hat{\boldsymbol{\Omega}}_n\left(\beta^0\right)^{-1} \mathbf{B}(u)^{\mathrm{T}} e_\ell \\
\leq \;& n^{-1}\sigma^2 \left\| \mathbf{B}(u)^{\mathrm{T}} e_\ell \right\|_2^2 \left\| \hat{\boldsymbol{\Omega}}_n\left(\beta^0\right)^{-1} \right\|_\infty = O_p(J_n/n)
\end{aligned} \tag{B.7}$$

Thus by the weak law of large numbers, we have for every $u \in [0,1]$, $\widetilde{m}_{l,e}(u,\beta^0,\alpha^0) = O_p(J_n^{-1/2} n^{-1/2})$. Therefore, we have for every $1 \leq \ell \leq p$, $\left|\widetilde{m}_\ell\left(u,\beta^0,\alpha^0\right) - m_\ell^0\left(u,\beta^0,\alpha^0\right)\right| = O_p\left(\sqrt{J_n/n} + J_n^{-r}\right)$ uniformly in $u \in [0,1]$. According to de Boor (2001, page 116),

$$\widetilde{m}_\ell'\left(u,\beta^0,\alpha^0\right) = B_{q-1}(u)^{\mathrm{T}} \mathbf{D}_1 \widetilde{\gamma}_\ell\left(\beta^0,\alpha^0\right),$$

where $B_{q-1}(u) = \{B_{j,q-1}(u) : 2 \leq j \leq J_n\}^{\mathrm{T}}$ is the $(q-1)$-th order normalized B-spline basis, and $\mathbf{D}_1 = \begin{pmatrix} \frac{-1}{\xi_1-\xi_{2-q}} & \frac{1}{\xi_1-\xi_{2-q}} & 0 & \cdots & 0 \\ 0 & \frac{-1}{\xi_2-\xi_{3-q}} & \frac{1}{\xi_2-\xi_{3-q}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{-1}{\xi_{N+q-1}-\xi_N} & \frac{1}{\xi_{N+q-1}-\xi_N} \end{pmatrix}_{(J_n-1)\times J_n}$. Similarly, $\widetilde{m}'_\ell(u,\beta^0,\alpha^0)$ can be written as $\widetilde{m}'_{\ell,m}(u,\beta^0,\alpha^0)+\widetilde{m}'_{\ell,e}(u,\beta^0,\alpha^0)$, where $\widetilde{m}'_{\ell,m}(u,\beta^0,\alpha^0) = B_{q-1}(u)^{\mathrm{T}}\mathbf{D}_1\widetilde{\gamma}_{\ell,m}(\beta^0,\alpha^0)$ and $\widetilde{m}'_{\ell,e}(u,\beta^0,\alpha^0) = B_{q-1}(u)^{\mathrm{T}}\mathbf{D}_1\widetilde{\gamma}_{\ell,e}(\beta^0,\alpha^0)$. It is easy to prove that $\|\mathbf{D}_1\|_\infty = O(J_n)$. Following similar reasoning as the proof for $\widetilde{m}_\ell(u,\beta^0,\alpha^0)$, one can prove that

$$\widetilde{m}'_\ell(u,\beta^0,\alpha^0) - m'_\ell(u) = O_p\left(J_n^{3/2}n^{-1/2} + J_n^{-r+1}\right),$$

uniformly for every $u \in [0,1]$. ■

Lemma 12 *Under Conditions (C1)-(C6), for* $r > 3/2$, $n^{-1}J_n^4 = o(1)$, $nJ_n^{-2r-2} = o(1)$ *and* $nJ_n^{-4r+2} = o(1)$, *we have*

$$\partial L_n^*(\beta^0,\alpha^0,\sigma^2)/\partial\theta_{-1}^T \;=\; \sigma^{-2}\left(\mathbf{Y}-\mathbf{T}\alpha^0-\mathbf{m}\right)^T\mathbf{H}^{-1}\left(\mathbf{\Phi}(\beta^0)\widetilde{\mathbf{X}}\mathbf{J},\widetilde{\mathbf{T}}\right) + o_p\left(n^{1/2}\right).$$

Proof. By Taylor expansion,

$$\mathbf{Q}_i(\beta^0)^{\mathrm{T}}\left\{\widetilde{\gamma}(\beta^0,\alpha^0)-\gamma^0\right\} = \mathbf{Q}_i(\beta^0)^{\mathrm{T}}\left\{\hat{\gamma}_e(\beta^0,\alpha^0)+\hat{\gamma}_m(\beta^0,\alpha^0)\right\} + O_p\left(\sqrt{J_n/n}+J_n^{-r}\right), \tag{B.8}$$

where

$$\hat{\gamma}_m(\beta^0,\alpha^0) = \left\{\mathbf{Q}(\beta^0)^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{Q}(\beta^0)\right\}^{-1}\mathbf{Q}(\beta^0)^{\mathrm{T}}\mathbf{H}^{-1}(\mathbf{m}-\mathbf{Q}(\beta^0)\gamma^0)$$

$$\hat{\gamma}_e(\beta^0,\alpha^0) = \left\{\mathbf{Q}(\beta^0)^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{Q}(\beta^0)\right\}^{-1}\mathbf{Q}(\beta^0)^{\mathrm{T}}\mathbf{H}^{-1}(\mathbf{Y}-\mathbf{T}\alpha^0-\mathbf{m}) \tag{B.9}$$

By the weak law of large numbers, it can be proved that

$$\begin{aligned}
& \mathbf{Q}_i\left(\beta^0\right)^{\mathrm{T}}\left\{\partial\hat{\gamma}_e(\beta^0,\alpha^0)/\partial\theta_{-1}^{\mathrm{T}}\right\} \\
= & -\mathbf{Q}_i\left(\beta^0\right)^{\mathrm{T}}\left\{\mathbf{Q}(\beta^0)^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{Q}(\beta^0)\right\}^{-1}\mathbf{Q}(\beta^0)^{\mathrm{T}}\mathbf{H}^{-1}\left(\sum\nolimits_{\ell=1}^{p} m_\ell'(\mathbf{X}_i^{\mathrm{T}}\beta^0)Z_{i\ell}\mathbf{X}_i^{\mathrm{T}}\mathbf{J},\mathbf{T}_i^{\mathrm{T}}\right)_{i=1}^{n} \\
& +O_p\left(n^{-1/2}J_n^{1/2}\right) \\
= & -\mathrm{Proj}_{\mathcal{M}_n}\left(\sum\nolimits_{\ell=1}^{p} m_\ell'(\mathbf{X}_i^{\mathrm{T}}\beta^0)Z_{i\ell}\mathbf{X}_i^{\mathrm{T}}\mathbf{J},\mathbf{T}_i^{\mathrm{T}}\right)+O_p\left(n^{-1/2}J_n^{1/2}\right).
\end{aligned}$$

Moreover,

$$\begin{aligned}
& \left\|\mathbf{Q}_i\left(\beta^0\right)^{\mathrm{T}}\left\{\partial\hat{\gamma}_m(\beta^0,\alpha^0)/\partial\theta_{-1}^{\mathrm{T}}\right\}\right\|_\infty \\
\leq & \sum\nolimits_{j=1,\ell=1}^{J_n,p}\left|Q_{j\ell,i}\left(\beta^0\right)\right|\left\|\hat{\mathbf{\Omega}}_n\left(\beta^0\right)^{-1}\right\|_\infty\left\|n^{-1}\mathbf{Q}(\beta^0)^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{1}_n\right\|_\infty O\left(J_n^{-r+1}\right) \\
= & O_p\left(J_n\right)O_p\left(J_n^{-1}\right)O\left(J_n^{-r+1}\right)=O_p\left(J_n^{-r+1}\right).
\end{aligned} \tag{B.10}$$

Therefore, we have

$$\begin{aligned}
& \mathbf{Q}_i\left(\beta^0\right)^{\mathrm{T}}\left\{\partial\left(\widetilde{\gamma}(\beta^0,\alpha^0)-\gamma\right)/\partial\theta_{-1}^{\mathrm{T}}\right\} \\
= & -\mathrm{Proj}_{\mathcal{M}_n}\left(\sum\nolimits_{\ell=1}^{p} m_\ell'(\mathbf{X}_i^{\mathrm{T}}\beta^0)Z_{i\ell}\mathbf{X}_i^{\mathrm{T}}\mathbf{J},\mathbf{T}_i^{\mathrm{T}}\right)+O_p\left(n^{-1/2}J_n^{1/2}+J_n^{-r+1}\right),
\end{aligned}$$

By B-spline properties and Condition (C6), it can be proved that

$$\begin{aligned}
& \left|\mathrm{Proj}_{\mathcal{M}_n}\left(\sum\nolimits_{\ell=1}^{p} m_\ell'(\mathbf{X}_i^{\mathrm{T}}\beta^0)Z_{i\ell}\mathbf{X}_i^{\mathrm{T}}\mathbf{J},\mathbf{T}_i^{\mathrm{T}}\right)-\mathrm{Proj}_{\mathcal{M}}\left(\sum\nolimits_{\ell=1}^{p} m_\ell'(\mathbf{X}_i^{\mathrm{T}}\beta^0)Z_{i\ell}\mathbf{X}_i^{\mathrm{T}}\mathbf{J},\mathbf{T}_i^{\mathrm{T}}\right)\right| \\
= & O_p\left(n^{-1/2}J_n^{1/2}+J_n^{-1}\right).
\end{aligned}$$

Then we have

$$
\begin{aligned}
& \partial\left\{\mathbf{T}_i^{\mathrm{T}}\alpha^0+\mathbf{Q}_i\left(\beta^0\right)^{\mathrm{T}}\widetilde{\gamma}\left(\beta^0,\alpha^0\right)\right\}/\partial\theta_{-1}^{\mathrm{T}} \\
= & \left(\sum\nolimits_{\ell=1}^{p} m_\ell'(\mathbf{X}_i^{\mathrm{T}}\beta^0)Z_{i\ell}\mathbf{X}_i^{\mathrm{T}}\mathbf{J},\mathbf{T}_i^{\mathrm{T}}\right)-\mathrm{Proj}_{\mathcal{M}}\left(\sum\nolimits_{\ell=1}^{p} m_\ell'(\mathbf{X}_i^{\mathrm{T}}\beta^0)Z_{i\ell}\mathbf{X}_i^{\mathrm{T}}\mathbf{J},\mathbf{T}_i^{\mathrm{T}}\right) \\
& +O_p\left(n^{-1/2}J_n^{1/2}+J_n^{-r+1}+J_n^{-1}\right) \\
= & \left(\sum\nolimits_{\ell=1}^{p} m_\ell'(\mathbf{X}_i^{\mathrm{T}}\beta^0)Z_{i\ell}\widetilde{\mathbf{X}}_i^{\mathrm{T}}\mathbf{J},\widetilde{\mathbf{T}}_i^{\mathrm{T}}\right)+O_p\left(n^{-1/2}J_n^{1/2}+J_n^{-r+1}+J_n^{-1}\right) \\
= & \left(\mathbf{\Phi}(\beta^0)\widetilde{\mathbf{X}}\mathbf{J},\widetilde{\mathbf{T}}\right)+O_p\left(n^{-1/2}J_n^{1/2}+J_n^{-r+1}+J_n^{-1}\right). \qquad \text{(B.11)}
\end{aligned}
$$

By (B.8), (B.10) and (B.11) and Lemma 11, for $r > 3/2$, $n^{-1}J_n^4 = o(1)$, $nJ_n^{-2r-2} = o(1)$ and $nJ_n^{-4r+2} = o(1)$,

$$
\begin{aligned}
& \partial L_n^*(\beta^0,\alpha^0,\sigma^2)/\partial\theta_{-1} \\
= & \ \sigma^{-2}\left[(\mathbf{Y}-\mathbf{T}\alpha^0-\mathbf{m})^{\mathrm{T}}+O_p\left(n^{-1/2}J_n^{1/2}+J_n^{-r}\right)\right] \\
& \mathbf{H}^{-1}\times\left[\left(\mathbf{\Phi}\left(\beta^0\right)\widetilde{\mathbf{X}}\mathbf{J},\widetilde{\mathbf{T}}\right)+O_p\left(n^{-1/2}J_n^{1/2}+J_n^{-1}+J_n^{-r+1}\right)\right] \\
= & \ \sigma^{-2}(\mathbf{Y}-\mathbf{T}\alpha^0-\mathbf{m})^{\mathrm{T}}\mathbf{H}^{-1}\left(\mathbf{\Phi}\left(\beta^0\right)\widetilde{\mathbf{X}}\mathbf{J},\widetilde{\mathbf{T}}\right) \\
& +O_p\left(n^{-1/2}J_n^{1/2}+J_n^{-1}+J_n^{-r+1}\right)\left\{O_p\left(n^{1/2}\right)+O_p\left(n^{1/2}J_n^{1/2}+nJ_n^{-r}\right)\right\} \\
= & \ \sigma^{-2}(\mathbf{Y}-\mathbf{T}\alpha^0-\mathbf{m})^{\mathrm{T}}\mathbf{H}^{-1}\left(\mathbf{\Phi}\left(\beta^0\right)\widetilde{\mathbf{X}}\mathbf{J},\widetilde{\mathbf{T}}\right)+o_p\left(n^{1/2}\right).
\end{aligned}
$$

■

Proof of Theorem 5. The consistency of the parametric estimator $\widehat{\theta}_{-1}$ can be proved following similar arguments as in Lemma 11, and thus omitted. By Lemma 12, it is straightforward to prove that

$$
\partial L_n^*\left(\beta^0,\alpha^0,\sigma^2\right)/\partial\theta_{-1}\partial\theta_{-1}^{\mathrm{T}}=\sigma^{-2}\left(\mathbf{\Phi}\left(\beta^0\right)\widetilde{\mathbf{X}}\mathbf{J},\widetilde{\mathbf{T}}\right)^{\mathrm{T}}\mathbf{H}^{-1}\left(\mathbf{\Phi}\left(\beta^0\right)\widetilde{\mathbf{X}}\mathbf{J},\widetilde{\mathbf{T}}\right)+o_p\left(n\right).
$$

Thus by Taylor expansion, Lemma 12 and the above result, we have

$$\begin{aligned}
\widehat{\theta}_{-1}-\theta_{-1}^{0} &= \left\{\partial L_{n}^{*}(\beta^{0},\alpha^{0},\sigma^{2})/\partial\theta_{-1}\partial\theta_{-1}^{\mathrm{T}}\right\}^{-1}\left\{\partial L_{n}^{*}(\beta^{0},\alpha^{0},\sigma^{2})/\partial\theta_{-1}\right\}+o_{p}\left(n^{-1/2}\right)\\
&= \left[\left(\mathbf{\Phi}\left(\beta^{0}\right)\widetilde{\mathbf{X}}\mathbf{J},\widetilde{\mathbf{T}}\right)^{\mathrm{T}}\mathbf{H}^{-1}\left(\mathbf{\Phi}\left(\beta^{0}\right)\widetilde{\mathbf{X}}\mathbf{J},\widetilde{\mathbf{T}}\right)\right]^{-1}\times\\
&\quad \left[\mathbf{e}^{\mathrm{T}}\mathbf{H}^{-1}\left(\mathbf{\Phi}\left(\beta^{0}\right)\widetilde{\mathbf{X}}\mathbf{J},\widetilde{\mathbf{T}}\right)\right]+o_{p}\left(n^{-1/2}\right).
\end{aligned}$$

Theorem 5 can be proved by Lindeberg-Feller Central Limit Theorem. ■

Proof of Theorem 6. The results in Theorem 6 follow from $\left\|\widehat{\theta}-\theta^{0}\right\|_{2}=O_{p}\left(n^{-1/2}\right)$ and Lemma 11. ■

B.3 Proof of Theorem 7

Consider the general case, to test whether genetic factors $Z_{\ell}, 2\leq\ell\leq p$, are important to the phenotype. We set up the null and alternative hypotheses as $H_{0}: m_{\ell}(\cdot)=0$, for $\ell=2,\ldots,p$ versus $H_{1}: m_{\ell}(\cdot)\neq 0$ for some $\ell\in(2,\ldots,p)$. Since each nonparametric function $m_{\ell}(u)\approx B\left(u\right)^{\mathrm{T}}\gamma_{\ell}$, the null and alternative hypotheses can be written as $H_{0}:\gamma_{\ell}=\mathbf{0}_{J_{n}}$, for $\ell=2,\ldots,p$ versus $H_{1}:\gamma_{\ell}\neq\mathbf{0}$ for some $\ell\in(2,\ldots,p)$. Let $\widehat{\gamma}^{\mathrm{N}}=\left\{\left(\widehat{\gamma}_{1}^{\mathrm{N}}\right)^{\mathrm{T}},\ldots,\left(\widehat{\gamma}_{p}^{\mathrm{N}}\right)^{\mathrm{T}}\right\}^{\mathrm{T}}$ be the maximizer of $L_{n}\left(\gamma,\widehat{\beta},\widehat{\alpha},\lambda,\sigma^{2}\right)$ given in (5.5) under H_{0}. Thus $\widehat{\gamma}_{\ell}^{\mathrm{N}}=\mathbf{0}$, for $\ell=2,\ldots,p$. Let $\widehat{\gamma}_{(2)}^{\mathrm{N}}=\left\{\left(\widehat{\gamma}_{\ell}^{\mathrm{N}}\right)^{\mathrm{T}}:2\leq\ell\leq p\right\}^{\mathrm{T}}$ and $\gamma_{(2)}=\left(\gamma_{\ell}^{\mathrm{T}}:2\leq\ell\leq p\right)^{\mathrm{T}}$. Let

$$\begin{aligned}
\mathbf{Q}_{i,(1)}\left(\beta\right) &= \left[B\left(\mathbf{X}_{i}^{\mathrm{T}}\beta\right)Z_{i1}\right]_{J_{n}\times 1}\\
\mathbf{Q}_{i,(2)}\left(\beta\right) &= \left[\left\{B\left(\mathbf{X}_{i}^{\mathrm{T}}\beta\right)^{\mathrm{T}}Z_{i2},\ldots,B\left(\mathbf{X}_{i}^{\mathrm{T}}\beta\right)^{\mathrm{T}}Z_{ip}\right\}^{\mathrm{T}}\right]_{J_{n}(p-1)\times 1}\\
\mathbf{Q}_{(1)}\left(\beta\right) &= \left[\left\{\mathbf{Q}_{1,(1)}\left(\beta\right),\ldots,\mathbf{Q}_{n,(1)}\left(\beta\right)\right\}^{\mathrm{T}}\right]_{n\times J_{n}}\\
\mathbf{Q}_{(2)}\left(\beta\right) &= \left[\left\{\mathbf{Q}_{1,(2)}\left(\beta\right),\ldots,\mathbf{Q}_{n,(2)}\left(\beta\right)\right\}^{\mathrm{T}}\right]_{n\times J_{n}(p-1)}.
\end{aligned}$$

Define the score function as

$$\begin{aligned} s_{2n}\left(\widehat{\gamma}^{\mathrm{N}},\widehat{\beta},\widehat{\alpha}\right) &= \partial L_n(\widehat{\gamma}^{\mathrm{N}},\widehat{\beta},\widehat{\alpha},\lambda,\sigma^2)\sigma^2/\partial\gamma_{(2)} \\ &= \mathbf{Q}_{(2)}\left(\widehat{\beta}\right)^{\mathrm{T}}\mathbf{H}^{-1}\left(\mathbf{Y}-\mathbf{T}\widehat{\alpha}-\mathbf{Q}_{(1)}\left(\widehat{\beta}\right)\widehat{\gamma}_1^{\mathrm{N}}\right). \end{aligned}$$

Define

$$\boldsymbol{\Omega}_n = \mathbf{Q}\left(\beta^0\right)^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{Q}\left(\beta^0\right) = \begin{pmatrix} \boldsymbol{\Omega}_{n,11} & \boldsymbol{\Omega}_{n,12} \\ \boldsymbol{\Omega}_{n,21} & \boldsymbol{\Omega}_{n,22} \end{pmatrix},$$

where $\boldsymbol{\Omega}_{n,kk'} = \mathbf{Q}_{(k)}\left(\beta^0\right)^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{Q}_{(k')}\left(\beta^0\right)$, for $k,k'=1,2$. Define the test statistic

$$\mathcal{T}_n = \sigma^{-2}s_{2n}\left(\widehat{\gamma}^{\mathrm{N}},\widehat{\beta},\widehat{\alpha}\right)^{\mathrm{T}}\boldsymbol{\Omega}_n^{22}s_{2n}\left(\widehat{\gamma}^{\mathrm{N}},\widehat{\beta},\widehat{\alpha}\right), \tag{B.12}$$

where $\boldsymbol{\Omega}_n^{22} = \left(\boldsymbol{\Omega}_{n,22}-\boldsymbol{\Omega}_{n,21}\boldsymbol{\Omega}_{n,11}^{-1}\boldsymbol{\Omega}_{n,12}\right)^{-1}$.

Then we have under Conditions (C1)-(C6) in the Appendix, and $n^{1/2r} << J_n << n^{1/4}$, we have under H_0, as $n\to\infty$,

$$\{2J_n(p-1)\}^{-1/2}\{\mathcal{T}_n - J_n(p-1)\} \to N(0,1).$$

Proof of Theorem 7. Consider the general case. Let $\widetilde{\gamma}^{\mathrm{N}} = \left\{\left(\widetilde{\gamma}_1^{\mathrm{N}}\right)^{\mathrm{T}},\ldots,\left(\widetilde{\gamma}_p^{\mathrm{N}}\right)^{\mathrm{T}}\right\}^{\mathrm{T}}$ be the maximizer of $L_n(\gamma,\beta^0,\alpha^0)$ under H_0. Thus $\widetilde{\gamma}_\ell^N = \mathbf{0}$ for $\ell = 2,...,p$. Let $\widetilde{\gamma}_{(2)}^{\mathrm{N}} = \left\{\left(\widetilde{\gamma}_\ell^{\mathrm{N}}\right)^{\mathrm{T}}: 2\le\ell\le p\right\}^{\mathrm{T}}$. We will show the asymptotic results for $\widetilde{\gamma}^{\mathrm{N}}$ and $s_{2n}\left(\widetilde{\gamma}^{\mathrm{N}},\beta^0,\alpha^0\right)$. The same asymptotic results for $\widehat{\gamma}^{\mathrm{N}}$ and $s_{2n}\left(\widehat{\gamma}^{\mathrm{N}},\widehat{\beta},\widehat{\alpha}\right)$ can be obtained by the fact that $\left\|\widehat{\theta}-\theta^0\right\|_2 = O_p\left(n^{-1/2}\right)$. Following the same reasoning as the proof for Theorem 6, it can be proved that

$$\widetilde{\gamma}_1^{\mathrm{N}} - \gamma_1^0 = \boldsymbol{\Omega}_{n,11}^{-1}\mathbf{Q}_{(1)}(\beta^0)^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{e} + O_p\left(J_n^{-r}\right), \tag{B.13}$$

and

$$\left|B(u)^{\mathrm{T}}\left(\widetilde{\gamma}_\ell^{\mathrm{N}}-\gamma_\ell^0\right)\right| = O_p\left(\sqrt{J_n/n}+J_n^{-r}\right), \tag{B.14}$$

uniformly in $u \in [0,1]$. Let $\eta_i^0(\gamma) = \sum_{\ell=1}^p B\left(\mathbf{X}_i^{\mathrm{T}}\beta^0\right)^{\mathrm{T}} \gamma_\ell Z_{i\ell}$, $\boldsymbol{\eta}^0(\gamma) = \left(\eta_1^0(\gamma), ..., \eta_n^0(\gamma)\right)^{\mathrm{T}}$. Then

$$
\begin{aligned}
s_{2n}\left(\widetilde{\gamma}^{\mathrm{N}}, \beta^0, \alpha^0\right) &= \partial L_n(\widetilde{\gamma}^{\mathrm{N}}, \beta^0, \alpha^0)\sigma^2/\partial\gamma_{(2)} \\
&= \mathbf{Q}_{(2)}(\beta^0)^{\mathrm{T}}\mathbf{H}^{-1}\left(\mathbf{Y} - \mathbf{T}\alpha^0 - \mathbf{Q}_{(1)}(\beta^0)\widetilde{\gamma}_1^{\mathrm{N}} - \mathbf{Q}_{(2)}(\beta^0)\widetilde{\gamma}_{(2)}^{\mathrm{N}}\right) \\
&= \mathbf{Q}_{(2)}(\beta^0)^{\mathrm{T}}\mathbf{H}^{-1}\left(\mathbf{Y} - \mathbf{T}\alpha^0 - \boldsymbol{\eta}^0(\widetilde{\gamma}^{\mathrm{N}})\right).
\end{aligned}
$$

Let $s_{2n}\left(\widetilde{\gamma}^{\mathrm{N}}, \beta^0, \alpha^0\right) = s_{2n,1} + s_{2n,2} + s_{2n,3}$, where

$$
\begin{aligned}
s_{2n,1} &= \mathbf{Q}_{(2)}(\beta^0)^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{e}, \\
s_{2n,2} &= \mathbf{Q}_{(2)}(\beta^0)^{\mathrm{T}}\mathbf{H}^{-1}(\mathbf{m} - \boldsymbol{\eta}^0(\gamma^0)), \\
s_{2n,3} &= \mathbf{Q}_{(2)}(\beta^0)^{\mathrm{T}}\mathbf{H}^{-1}(\boldsymbol{\eta}^0(\gamma^0) - \boldsymbol{\eta}^0(\widetilde{\gamma}^{\mathrm{N}})).
\end{aligned}
$$

By Berstein's inequality in [12], it can be proved that

$$
\|s_{2n,2}\|_\infty \asymp O\left(J_n^{-r}\right)\left\|\mathbf{Q}_{(2)}\left(\beta^0\right)^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{1}_n\right\|_\infty = O_p\left(nJ_n^{-r-1}\right).
$$

By (B.13),

$$
\left\|s_{2n,3} + \boldsymbol{\Omega}_{n,21}\boldsymbol{\Omega}_{n,11}^{-1}\mathbf{Q}_{(1)}(\beta^0)^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{e}\right\|_\infty = O_p\left(nJ_n^{-r-1}\right).
$$

Therefore,

$$
\left\|s_{2n}\left(\widetilde{\gamma}^{\mathrm{N}}, \beta^0, \alpha^0\right) - \widetilde{s}_{2n}\left(\widetilde{\gamma}^{\mathrm{N}}, \beta^0, \alpha^0\right)\right\|_\infty = O_p\left(nJ_n^{-r-1}\right) + O_p(1), \tag{B.15}
$$

where

$$
\widetilde{s}_{2n}\left(\widetilde{\gamma}^{\mathrm{N}}, \beta^0, \alpha^0\right) = \left\{\mathbf{Q}_{(2)}\left(\beta^0\right)^{\mathrm{T}}\mathbf{H}^{-1} - \boldsymbol{\Omega}_{n,21}\boldsymbol{\Omega}_{n,11}^{-1}\mathbf{Q}_{(1)}\left(\beta^0\right)^{\mathrm{T}}\mathbf{H}^{-1}\right\}\mathbf{e}.
$$

Moreover, $E\left\{\widetilde{s}_{2n}\left(\widetilde{\gamma}^{\mathrm{N}},\beta^0,\alpha^0\right)\right\}=0$ and

$$\begin{aligned}
&\operatorname{var}\left\{\widetilde{s}_{2n}\left(\widetilde{\gamma}^{\mathrm{N}},\beta^0,\alpha^0\right)|\mathbb{Z},\mathbb{X},\mathbb{T}\right\}\\
=\;&\sigma^2\left\{\mathbf{Q}_{(2)}\left(\beta^0\right)^{\mathrm{T}}\mathbf{H}^{-1}-\mathbf{\Omega}_{n,21}\mathbf{\Omega}_{n,11}^{-1}\mathbf{Q}_{(1)}\left(\beta^0\right)^{\mathrm{T}}\mathbf{H}^{-1}\right\}\mathbf{H}\\
&\times\left\{\mathbf{Q}_{(2)}\left(\beta^0\right)^{\mathrm{T}}\mathbf{H}^{-1}-\mathbf{\Omega}_{n,21}\mathbf{\Omega}_{n,11}^{-1}\mathbf{Q}_{(1)}\left(\beta^0\right)^{\mathrm{T}}\mathbf{H}^{-1}\right\}^{\mathrm{T}}\\
=\;&\sigma^2\left\{\mathbf{Q}_{(2)}\left(\beta^0\right)^{\mathrm{T}}-\mathbf{\Omega}_{n,21}\mathbf{\Omega}_{n,11}^{-1}\mathbf{Q}_{(1)}\left(\beta^0\right)^{\mathrm{T}}\right\}\mathbf{H}^{-1}\left\{\mathbf{Q}_{(2)}\left(\beta^0\right)-\mathbf{Q}_{(1)}\left(\beta^0\right)\mathbf{\Omega}_{n,11}^{-1}\mathbf{\Omega}_{n,12}\right\}\\
=\;&\sigma^2(\mathbf{\Omega}_{n,22}-\mathbf{\Omega}_{n,21}\mathbf{\Omega}_{n,11}^{-1}\mathbf{\Omega}_{n,12})=\sigma^2(\mathbf{\Omega}_n^{22})^{-1}.
\end{aligned}$$

Thus, by Lindeberg-Feller Central Limit Theorem, for any $\mathbf{a}\in R^{J_n(p-1)}$ with $\|\mathbf{a}\|_2=1$,

$$\mathbf{a}^{\mathrm{T}}\sigma^{-1}(\mathbf{\Omega}_n^{22})^{1/2}\widetilde{s}_{2n}\left(\widetilde{\gamma}^{\mathrm{N}},\beta^0,\alpha^0\right)\xrightarrow{d}N\left(0{,}1\right).$$

Therefore, $\sigma^{-2}\widetilde{s}_{2n}\left(\widetilde{\gamma}^{\mathrm{N}},\beta^0,\alpha^0\right)^{\mathrm{T}}\mathbf{\Omega}_n^{22}\widetilde{s}_{2n}\left(\widetilde{\gamma}^{\mathrm{N}},\beta^0,\alpha^0\right)$ has the asymptotic distribution as $\chi^2_{J_n(p-1)}$, with $J_n(p-1)\to\infty$ as $n\to\infty$, and hence,

$$\left\{2J_n\left(p-1\right)\right\}^{-1/2}\left\{\sigma^{-2}\widetilde{s}_{2n}\left(\widetilde{\gamma}^{\mathrm{N}},\beta^0,\alpha^0\right)^{\mathrm{T}}\mathbf{\Omega}_n^{22}\widetilde{s}_{2n}\left(\widetilde{\gamma}^{\mathrm{N}},\beta^0,\alpha^0\right)-J_n\left(p-1\right)\right\}\to N\left(0{,}1\right).$$

By (B.15) and Berstein's inequality, we have

$$\begin{aligned}
&\Big|\sigma^{-2}s_{2n}\left(\widetilde{\gamma}^{\mathrm{N}},\beta^0,\alpha^0\right)^{\mathrm{T}}\mathbf{\Omega}_n^{22}s_{2n}\left(\widetilde{\gamma}^{\mathrm{N}},\beta^0,\alpha^0\right)-\\
&\sigma^{-2}\widetilde{s}_{2n}\left(\widetilde{\gamma}^{\mathrm{N}},\beta^0,\alpha^0\right)^{\mathrm{T}}\mathbf{\Omega}_n^{22}\widetilde{s}_{2n}\left(\widetilde{\gamma}^{\mathrm{N}},\beta^0,\alpha^0\right)\Big|=O_p\left(n^{1/2}J_n^{-r+1/2}\right).
\end{aligned}$$

Thus for $n^{1/2}J_n^{-r}=o\left(1\right)$, Theorem 7 follows from Slutsky theorem. ■