TOPICS IN PARTIALLY LINEAR SINGLE-INDEX MODELS FOR LONGITUDINAL DATA

A Dissertation

by

QUAN CAI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Suojin Wang
Committee Members,	Michael T. Longnecker
	Mohsen Pourahmadi
	Jianxin Zhou
Head of Department,	Valen Johnson

August 2017

Major Subject: Statistics

ABSTRACT

The partially linear single-index model is a semiparametric model proposed to the case when some predictors are linearly associated with the response variable, while some other predictors are nonlinearly associated with the response variable. It is widely used for its flexibility in statistical modeling. Furthermore, its generalized version is a generalization of some popular models such as the generalized linear model, the partially linear model and the single-index model. However, the proper estimation in partially linear single-index models for longitudinal data, where multiple measurements are observed for each subject, is still open to discussion. Our main purpose is to establish an unified estimation method for the longitudinal partially linear single-index model and its generalized version.

With this question in mind, we propose a new iterative three-stage estimation method in partially linear single-index models and generalized partially linear single-index models for longitudinal data. With the proposed method, the within-subject correlation is properly taken into consideration in the estimation of both the parameters and the nonparametric single-index function. The parameter estimators are shown to be asymptotically semiparametric efficient. The asymptotic variance of the single-index function estimator is shown to be generally less than that of existing estimators. Simulation studies are performed to demonstrate the finite sample performance. Three real data examples are also analyzed to illustrate the methodology.

DEDICATION

To my parents

ACKNOWLEDGMENTS

I would like to gratefully thank Dr. Suojin Wang for his patient guidance and consistent support during the past five years. Never taken for granted, I learned from Dr. Wang both in academics and in life. His intellect and experiences led me to the world of statistics and the world of scientific research. He helped me to be a better statistician, a better researcher and a better person. I also thank Dr. Michael T. Longnecker, Dr. Mohsen Pourahmadi and Dr. Jianxin Zhou for coursework and willingness to serve on my advisory committee. They provided me with valuable advice for my research studies and presentation skills.

I would like to thank my parents and my girlfriend. They provided me with consistent support all the time even they are on the other side of the world. They were the source of power and enlightened me when I ran into trouble. Finally I would like to thank all my friends for the support and help all the way.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professors Suojin Wang (advisor), Michael T. Longnecker and Mohsens Pourahmadi of the Department of Statistics and Professor Jianxin Zhou of the Department of Mathematics.

The algorithms and proofs in Chapter 3 and 4 were coauthored with Professor Suojin Wang.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a graduate assistantship from Department of Statistics, Texas A&M University.

NOMENCLATURE

AMSE	Averaged Mean Squared Error
GEE	Generalized Estimating Equation
GLM	Generalized Linear Model
GPLSIM	Generalized Partially Linear Single-Index Models
ICHS	Indonesian Children's Health Study
i.i.d.	Independent and Identically Distributed
MLE	Maximum Likelihood Estimation
NCHS	National Center for Health Statistics
PLSIM	Partially Linear Single-Index Models
SCAD	Smoothly Clipped Absolute Deviation
SE	Standard Error
SEE	Semiparametric Estimating Equation
SMGEE	Semiparametric Marginal Generalized Estimating Equation
SGEE	Semiparametric Generalized Estimating Equation
SWSE	Sandwich Standard Error
WI	Working Independence

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

1. INTRODUCTION

Longitudinal data differs from cross sectional data in that each subject is observed for multiple times and the characteristics of each subject are measured repeatedly over time. We usually assume that the observations are independent from different subjects, but observations from the same subject are potentially correlated. Longitudinal data analysis is popular in a variety of fields such as epidemiology, clinic trials and economics. There is extensive literature discussing problems and developing theory and methodology in longitudinal data analysis; see Laird and Ware (1982), Liang and Zeger (1986), Zeger et al. (1985), and Zeger and Liang (1986), etc. The main problems lie in the estimation of the mean function, the within-subject covariance function and the regression parameters for some specified models. There are generally three approaches to model the longitudinal data: marginal models, random effect models and transition (Markov) models; see Diggle et al. (2013). The mean and the covariance estimations are generally more well addressed for balanced longitudinal cases, but they are often more difficult to handle in unbalanced cases, where subjects have different number of observations over time.

Semiparametric models are popular in recent years since they enjoy the advantages of incorporating both the parametric and nonparametric components. They can be defined by $\{\mathcal{P}_{\boldsymbol{\theta},\varphi} : \boldsymbol{\theta} \in \boldsymbol{\Theta}, \varphi \in \mathcal{F}\}$, where $\boldsymbol{\theta}$ is a finite-dimensional vector in the finite dimensional vector space $\boldsymbol{\Theta}$, while φ is a function in an infinite-dimensional space $\mathcal{F}$. Semiparametric models have many specific forms; see e.g., Ichimura (1993), Chen and Shiau (1994) and Carroll et al. (1997). One of the applications of semiparametric models is the partially linear single-index model (PLSIM)

$$Y_i = \mathbf{X}_i\boldsymbol{\beta} + \phi(\mathbf{Z}_i\boldsymbol{\theta}) + \epsilon_i \tag{1.1}$$

for $i = 1, \ldots, n$. Here $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are parameters associated with covariates $\mathbf{X}$ and $\mathbf{Z}$ respectively. The response variable Y is continuous and $\phi(\cdot)$ is an unknown function. The error term ϵ_i has mean zero. This model can be appropriately applied when some covariates ($\mathbf{X}$) are linearly associated with the response while some other covariates ($\mathbf{Z}$) are nonlinearly associated with the response.

The generalized partially linear single-index model (GPLSIM) is the generalized version of (1.1). It is more flexible and more general in that we can model both categorical response and transformation-necessary response such as heavy-tailed variable with multiple covariates, especially when some covariates are parametrically correlated with the response and the others are nonparametrically correlated to the response. Suppose we have an univariate response variable Y and possibly multi-dimensional covariates $\mathbf{X}$ and $\mathbf{W}$, The GPLSIM has the form

$$\begin{aligned} &\mathrm{E}(Y|\mathbf{X} = \mathbf{x}, \mathbf{W} = \mathbf{w}) = \mu(\mathbf{x}, \mathbf{w}), \\ &g^{-1}\{\mu(\mathbf{x}, \mathbf{w})\} = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + \gamma(\mathbf{w}^{\mathrm{T}}\boldsymbol{\alpha}), \end{aligned} \tag{1.2}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are possibly multi-dimensional parameters associated with predictors $\mathbf{X}$ and $\mathbf{W}$ respectively, $\mathbf{x}$ and $\mathbf{w}$ are the realizations of $\mathbf{X}$ and $\mathbf{W}$ respectively and $\gamma(\cdot)$ is an unknown function, referred by the single-index function hereafter. Besides, $g^{-1}(\cdot)$ is assumed to be a known monotonic and differentiable link function. The goal is to estimate the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ and the single-index function $\gamma(\cdot)$ in (1.2).

The GPLSIM in (1.2) is the generalized model of several types of models. When the single-index function is the identity function, it becomes the generalized linear model (Nelder and Baker (2004)). When there are no covariates $\mathbf{X}$, it becomes the generalized single-index model (Ichimura (1993)). When $\mathbf{W}$ is one-dimensional, it becomes the generalized partially linear model (Chen and Shiau (1994)). When the link function $g^{-1}(\cdot)$

is the identity function and the response variable is continuous, it becomes the PLSIM in (1.1); see Carroll et al. (1997) and Chen et al. (2015). Therefore, efficient estimation of GPLSIM is of major interest in that it can unify the estimation of several important models above and has broad applications.

Carroll et al. (1997) proposed and discussed estimation, testing and theoretical results of PLSIM and GPLSIM for independent and identically distributed (i.i.d.) data. In the i.i.d. case, Liang et al. (2010) proposed the profile least-squares method to obtain the semiparametrically efficient parameter estimators. Besides, the smoothly clipped absolute deviation penalty (SCAD) approach is applied for variable selection. Chen and Parker (2014) specifically calculated semiparametric information bound for PLSIM with the method of Severini and Tripathi (2001). Hu et al. (2004) and Li et al. (2010) discussed the inferences with PLSIM for longitudinal data.

However, the proper estimation of the parameters and the single-index function in PLSIM and GPLSIM for longitudinal data continues to receive considerable attentions. The most frequently used methods treat each observation independently which indicates that the working independence is assumed in both the parameters estimation step and the single-index estimation step. It does not lead to the efficient estimation in longitudinal PLSIM and GPLSIM. Therefore, we investigate the research problems in the dissertation listed as follows:

1. How to efficiently estimate the parameters in PLSIM and GPLSIM for unbalanced longitudinal data;

2. How to estimate the single-index function by properly taking into consideration within-subject correlation in the model;

3. Whether the parameter estimators are semiparametrically efficient.

Our objective in the dissertation is to provide answers to the above questions by proposing a more efficient methodology than the current estimation methods both theoretically and numerically.

The following chapters are organized as follows. In Chapter 2, we review the recent development of longitudinal data analysis and semiparametric models. Particularly, we focus on some classic literature in recent years, especially in the estimation of semiparametric models for longitudinal data. In Chapter 3, we propose a new iterative method to efficiently estimate the parameters and the single-index function in PLSIM for longitudinal data. We also investigate its analytic properties and provide empirical studies and real data analysis in support of the methodology. In Chapter 4, we extend the result in Chapter 3 to inference with GPLSIM. Theoretical results, simulation studies and real data examples are performed as well. We conclude the dissertation in Chapter 5.

2. LITERATURE REVIEW

2.1 Introduction

In this chapter, we summarize and review the key concepts and relevant models in longitudinal data analysis, semiparametric models and semiparametric models for longitudinal data. In Section 2.2, we introduce the characteristics of longitudinal data and review some key literature in the field of estimating the mean function (Section 2.2.1) and the covariance matrices (Section 2.2.2) for longitudinal data. In Section 2.3, we go over the concepts of semiparametric models. We introduce the semiparametric efficiency theories first in Section 2.3.1. Then in Section 2.3.2, we particularly discuss partially linear single-index models (PLSIM) and generalized partially linear single-index models (GPLSIM). The applications of these two models for longitudinal data will be our main focus in Chapters 3 and 4 respectively. In Section 2.4, some related research in semiparametric models in longitudinal data, especially the longitudinal PLSIM and GPLSIM (Section 2.4.1) is examined. Finally we conclude the whole chapter in Section 2.5.

2.2 Longitudinal Data Analysis

Longitudinal data, sometimes also named panel data, is widely applicable in many fields, including economics, medical research, epidemiological studies and clinical trials. Different from cross sectional studies when we assume subjects are independently observed or measured for only once, in longitudinal data studies, subjects are observed or measured for multiple times. The key feature of longitudinal data is that observations within a subject are more likely to be associated than observations from different subjects. Therefore, measurements within a subject are correlated, whereas measurements from separate subject are usually treated as independent. Due to the difference in the data structure from independent data, we have particular approaches for analyzing longitudinal data. There are typically three types of approaches for analyzing regression-based models for longitudinal data. They are marginal models, transition models and random/mixed effects models; see Diggle et al. (2013). In this dissertation, we focus on the approach of marginal models to analyze longitudinal data. The fundamental assumption for marginal longitudinal models (Pepe and Anderson (1994)) is

$$\mathrm{E}(Y_{ij}|\mathbf{X}_i) = \mathrm{E}(Y_{ij}|\mathbf{X}_{ij}), \tag{2.1}$$

where Y_{ij} is the univariate response variable, $\mathbf{X}_{ij}$ is the possibly multivariate explanatory variables for the j^{th} measurement of subject i and $\mathbf{X}_i = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{im_i})$, $i = 1, \ldots, n$, $j = 1, \ldots, m_i$. We usually assume in longitudinal data that the number of subjects $n \to \infty$. The number of measurements m_i could be different. Dense longitudinal data is that $m_i \geq M_n$ for some sequence $M_n \to \infty$ as $n \to \infty$. On the other side, if $\max_i m_i \leq M$ for some positive constant M as $n \to \infty$, we denote the longitudinal data sparse. When $m_i = m$ across all subjects, the longitudinal data is balanced. However, in general case where m_i may not be all the same for all i, we call the longitudinal data unbalanced or irregular.

There are two main goals in longitudinal data analysis. The first one is to properly estimate the mean function. The mean function could be a (generalized) linear combination of explanatory variables in (generalized) linear models, or more generally speaking, parametric models. It could also be a nonparametric function in nonparametric regression models, partially linear models, etc. The second goal is to properly estimate the within-subject covariance matrices/function, especially for unbalanced longitudinal data. Furthermore, the accuracy for the covariance estimation determines the efficiency of the mean function estimator.

2.2.1 Mean Function Estimation for Longitudinal Data

The estimation of the mean function in longitudinal data can be classified into two groups:

- The mean function estimation in longitudinal parametric models;
- The mean function estimation in longitudinal nonparametric models.

For the study of longitudinal parametric models, it dates back to 1980s. Laird and Ware (1982) proposed two-stage random effects models for longitudinal data when the response variable is approximately normal. For generalized case when the response variable is non-normal, Liang and Zeger (1986) extended the estimation method in generalized linear models from cross sectional data to longitudinal data. They proposed the generalized estimating equations (GEE) method for the generalized linear model (GLM)

$$\begin{aligned} \mathrm{E}(Y_{ij}) &= \mu_{ij}, \\ g^{-1}(\mu_{ij}) &= \mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}, \end{aligned} \tag{2.2}$$

where μ_{ij} is the mean of the response variable Y_{ij} and $\mathbf{X}_{ij}$ are the explanatory variables of dimension p. Parameters $\boldsymbol{\beta}$ of dimension p are to be estimated and $g^{-1}(\cdot)$ is a known, monotonic link function. Here the marginal model assumption in (2.1) is assumed. We introduce the idea behind the generalized estimating equation as follows to estimate the parameters $\boldsymbol{\beta}$ in generalized linear model in (2.2). There are two key features for GLM:

- The expected value of the response $\mu_{ij} = \mathrm{E}(Y_{ij})$ is linked to a linear combination of the covariates $\mathbf{X}_{ij}$ and regression parameter $\boldsymbol{\beta}$ through a proper link function $g(\cdot)$:

$$\mathrm{E}(Y_{ij}) = g(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}).$$

- The variance of the response might be represented as a function of the mean of the response:

$$\text{var}(Y_{ij}) = \phi V\{\text{E}(Y_{ij})\} = \phi V\{g(\mathbf{X}_{ij}^{\text{T}}\boldsymbol{\beta})\}, \tag{2.3}$$

where V is a function of the mean response and ϕ is a constant. For simplicity, here we assume that the constant is the same for all i and j.

We can see that the GLM requires the specification of a model for the mean and covariance of the response variable. In many scenarios, it is not appropriate or there is not always sufficient information to specify a certain probability distribution, e.g., multivariate normal distribution. Therefore, the method of maximum likelihood estimation (MLE) is not applicable for estimation and testing of the parameters.

We generally do not know the probability distribution for the response variable in the GLM (2.2). However, we can start with two special cases in GLM to formulate the estimating equations. Then we can extend them to the general case, i.e., the GEE.

First we define

$$\begin{aligned}
\mathbf{Y}_i &= (Y_{i1}, \ldots, Y_{im_i})^{\text{T}}, \\
\boldsymbol{\mu}_i &= (\mu_{i1}, \ldots, \mu_{im_i})^{\text{T}}, \\
\mathbf{X}_i &= (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{im_i})^{\text{T}}, \\
\text{var}(\mathbf{Y}_i) &= \boldsymbol{\Sigma}_i
\end{aligned}$$

for $i = 1, \ldots, n$. Suppose $\mathbf{Y}_i$ follows a multivariate normal distribution. Then from the MLE theory, the estimator for $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}} = (\sum_{i=1}^{n} \mathbf{X}_i^{\text{T}} \widehat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{X}_i)^{-1} \sum_{i=1}^{n} \mathbf{X}_i^{\text{T}} \widehat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{Y}_i,$$

where $\widehat{\boldsymbol{\Sigma}}_i$ is the estimator of covariance matrix $\boldsymbol{\Sigma}_i$. We call it the working covariance matrix. The above formula can be written as the estimating equation

$$\sum_{i=1}^{n} \mathbf{X}_i^{\mathrm{T}} \widehat{\boldsymbol{\Sigma}}_i^{-1} (\mathbf{Y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}}) = \mathbf{0}. \tag{2.4}$$

In the case of generalized linear models for the i.i.d. data (i.e., with $m_i = 1$ in (2.2)), similarly we solve equations for $\boldsymbol{\beta}$ as $\widehat{\boldsymbol{\beta}}$ that satisfies the following

$$\sum_{i=1}^{n} \frac{g^{(1)}(\mathbf{X}_i \widehat{\boldsymbol{\beta}})}{\widehat{V}\{g(\mathbf{X}_i \widehat{\boldsymbol{\beta}})\}} \mathbf{X}_i^{\mathrm{T}} \left\{ Y_i - g(\mathbf{X}_i \widehat{\boldsymbol{\beta}}) \right\} = \mathbf{0}. \tag{2.5}$$

Here $g^{(1)}$ is the first order derivative of $g(\cdot)$ with respect to its argument and $\widehat{V}$ is the estimated variance function defined in (2.3). The method of iteratively reweighted least squares or Fisher scoring methods are usually used to solve (2.5).

There is a similar pattern when comparing estimating equations (2.4) and (2.5): The estimating equations are formulated by the linear functions of the difference between the observed response values and their estimated means with weights. The weights are associated with the inverse of the variance or the working covariances of the response.

From the above observations, a natural generalized approach for fitting longitudinal data is to solve for an estimating equation consisting of p equations for $\boldsymbol{\beta}$ which satisfies the following structure:

1. The estimating equations are a linear function of $\mathbf{Y}_i - \boldsymbol{\mu}_i$;

2. Similarly to (2.4) and (2.5), the weights in the proposed estimating equation are associated with the inverse of the variance or the working covariance matrices $\widehat{\boldsymbol{\Sigma}}_i$.

These results lead to the consideration of the following equations to be solved for $\boldsymbol{\beta}$ in longitudinal data:

$$\sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}_i^{\mathrm{T}}}{\partial \boldsymbol{\beta}} \widehat{\boldsymbol{\Sigma}}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) = \mathbf{0}. \tag{2.6}$$

The estimating equation (2.6) is the GEE.

Liang and Zeger (1986) showed that the GEE estimators of the parameters are asymptotically consistent as long as the mean function μ_{ij} is correctly specified, even if the within-subject working correlation matrix $\widehat{\mathbf{R}}_i$ is misspecified for $i = 1, \ldots, n$, where

$$\widehat{\boldsymbol{\Sigma}}_i = \mathbf{S}_i \widehat{\mathbf{R}}_i \mathbf{S}_i.$$

Here $\mathbf{S}_i$ is the diagonal matrix of the estimated standard deviation of the response variable for subject i. The working correlation matrix $\widehat{\mathbf{R}}_i$ could be specified to a particular form such as AR(1) or ARMA(1,1). Then the estimation of $\mathbf{R}_i$ is equivalent to the estimation of the association parameters $\boldsymbol{\rho}$ in $\mathbf{R}_i = \mathbf{R}_i(\boldsymbol{\rho})$. Prentice (1988) extended the GEE approach to estimate association parameters for binary data by specifying a second set of estimating equations. The method is useful for simultaneous inference about the mean and association parameters .

For the study of nonparametric models for longitudinal data, Lin and Carroll (2000) considered local polynomial kernel estimating equations for the nonparametric function estimation. Suppose that the data consist of n subjects with the i^{th} subject having m_i observations as before. Considering the longitudinal nonparametric function model. It has the form

$$\begin{aligned} \mathrm{E}(Y_{ij}|X_{ij}) &= \mu_{ij}, \\ g^{-1}(\mu_{ij}) &= \phi(X_{ij}), \end{aligned} \tag{2.7}$$

where $\phi(\cdot)$ is an unknown nonparametric smooth function to be estimated and X_{ij} is the univariate predictor variable. Other notation such as Y_{ij}, $g^{-1}(\cdot)$ and μ_{ij} are defined the same way as in (2.2). Similarly to the parametric longitudinal models, the marginal model assumption in (2.1) is also assumed.

For independent data, local polynomial smoothing method, especially the local linear smoothing, has been widely used in nonparametric regression. Lin and Carroll (2000) extended local polynomial kernel smoothing to model (2.7) for correlated data. When constructing estimating equation for $\phi(\cdot)$, first assuming that $\phi(\cdot)$ is a parametric k^{th} order polynomial function which satisfies

$$\phi(\cdot) = \mathbf{L}_k(\cdot)^{\mathrm{T}}\boldsymbol{\beta}$$

with

$$\mathbf{L}_k(a) = (1, a, \ldots, a^k)^{\mathrm{T}},$$
$$\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)^{\mathrm{T}}.$$

Define $\mathbf{L}_{ik} = \{\mathbf{L}_k(X_{i1}), \ldots, \mathbf{L}_k(X_{im_i})\}^{\mathrm{T}}$. The GEE of Liang and Zeger (1986) has the form

$$\sum_{i=1}^{n} \mathbf{L}_{ik}^{\mathrm{T}} \boldsymbol{\Delta}_i \widehat{\boldsymbol{\Sigma}}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) = 0, \tag{2.8}$$

where j^{th} component of $\boldsymbol{\mu}_i$ is

$$\mu_{ij} = \mu\{\mathbf{L}_k^{\mathrm{T}}(X_{ij})\boldsymbol{\beta}\}$$

and

$$\boldsymbol{\Delta}_i = \text{diag}\big[\mu^{(1)}\{\mathbf{L}_k^{\mathrm{T}}(X_{ij})\boldsymbol{\beta}\}\big].$$

Here $\mu^{(1)}(\cdot)$ is the first derivative of $\mu(\cdot)$.

When $\phi(\cdot)$ is a nonparametric function, the parametric GEE (2.8) can be applied to (2.7) with the kernel smoothing method. First we define the kernel function $K(\cdot)$ to be symmetric and continuous. It also satisfies

$$\int_{-\infty}^{+\infty} K(x)dx = 1, \qquad \int_{-\infty}^{+\infty} x^2 K(x)dx < \infty.$$

We also denote

$$K_h(x) = \frac{1}{h}K(\frac{x}{h}),$$

where h is a bandwidth. It controls the smoothness of the kernel estimators.

To approximate $\phi(\cdot)$ at any given u within the domain of X with the local polynomial smoothing approach, we have

$$\phi(X) = \{\mathbf{L}_k(X-u)\}^{\mathrm{T}}\boldsymbol{\beta}.$$

Define $\mathbf{L}_{ik}(u) = \{\mathbf{L}_k(X_{i1}-u), \ldots, \mathbf{L}_k(X_{im_i}-u)\}^{\mathrm{T}}$. To apply the kernel estimation of the longitudinal nonparametric function, we need to incorporate the kernel weight function $K_h(\cdot)$ in (2.8). Depending on where the kernel weight is placed, there are two kernel estimating equations for estimating $\phi(u)$:

$$\sum_{i=1}^{n} \mathbf{L}_{ik}(u)^{\mathrm{T}}\boldsymbol{\Delta}_i(u)\widehat{\boldsymbol{\Sigma}}_i^{-1}(u)\mathbf{K}_{ih}(u)\{\mathbf{Y}_i - \boldsymbol{\mu}_i(u)\} = 0, \tag{2.9}$$

and

$$\sum_{i=1}^{n} \mathbf{L}_{ik}(u)^{\mathrm{T}} \boldsymbol{\Delta}_i(u) \mathbf{K}_{ih}^{1/2}(u) \widehat{\boldsymbol{\Sigma}}_i^{-1}(u) \mathbf{K}_{ih}^{1/2}(u) \{\mathbf{Y}_i - \boldsymbol{\mu}_i(u)\} = 0, \tag{2.10}$$

where $\mathbf{K}_{ih}(u)$ is the diagonal kernel matrix for subject i with the j^{th} diagonal element being $\mathbf{K}_h(X_{ij} - u)$. Equation (2.10) weights the residuals $\{\mathbf{Y}_i - \boldsymbol{\mu}_i(u)\}$ symmetrically, while Equation (2.9) does not. As a result, they provide different estimators for $\phi(u)$. An application of the Fisher scoring algorithm to (2.9) shows that the estimator of $\boldsymbol{\beta}$ can be updated using iteratively re-weighted least squares. This type of estimating equations (2.9) and (2.10) is called the kernel GEE or the profile kernel GEE.

Unlike the parametric GEE estimator in (2.8), if $\phi(u)$ is a nonparametric function estimated by the kernel GEE in (2.9) or (2.10), the asymptotically most efficient estimators are obtained by ignoring the within-subject correlation entirely; that is, assuming working independence $\widehat{\mathbf{R}}_i = \mathbf{I}_i$ for all subjects, where $\mathbf{I}$ is the identity matrix. Correctly specifying the correlation matrices in fact results in a less efficient estimator of $\phi(u)$ asymptotically; see Lin and Carroll (2000).

Wang (2003) reviewed the profile kernel GEE methods. Using asymptotic theory, she explained why this type of estimating equations requires working independence to achieve more efficiency for estimating the nonparametric function $\phi(\cdot)$.

First denote the $(k,l)^{th}$ element of $\widehat{\boldsymbol{\Sigma}}_i^{-1}$ by v_i^{kl} and u is within the domain of X_{ij} for $i = 1, \ldots, n$, $j = 1, \ldots, m_i$. By examining on the estimation equation (2.10), if the density of X_{ij} is bounded away from 0 and $K(\cdot)$ has compact support, when $h \to 0$, the term $\mathbf{K}_{ih}$ has only one nonzero element asymptotically. As a result, the following two findings were obtained:

1. Asymptotically, each subject has only one observation to contribute for estimating $\phi(u)$ in (2.10);

2. From 1, suppose the only nonzero observation in the i^{th} subject is the j^{th} measurement. When we "sandwitch" multiply $(\mathbf{K}_{ih})^{1/2}$ to $\widehat{\boldsymbol{\Sigma}}_i^{-1}$, there is also only one nonzero term in $\mathbf{L}_{ik}(u)^{\mathrm{T}}\boldsymbol{\Delta}_i(u)\mathbf{K}_{ih}^{1/2}(u)\widehat{\boldsymbol{\Sigma}}_i^{-1}(u)\mathbf{K}_{ih}^{1/2}(u)$ in (2.10). The nonzero term is $K_h(X_{ij}-u)(v^i)^{jj}$.

From the two implications above, when $n \to \infty$, the linear term $\mathbf{Y}_{ij} - \boldsymbol{\mu}_{ij}(u)$ is weighted by $K_h(X_{ij}-u)(v^i)^{jj}$. As we assume that the observations from different subjects are uncorrelated, to obtain the function estimator with the smallest asymptotic variance by the kernel GEE, the weighted term satisfies

$$v_i^{jj} = \text{var}(Y_{ij}|X_{ij}).$$

This result indicates that the estimated covariance $\widehat{\boldsymbol{\Sigma}}_i$ has to be diagonal with its k^{th} diagonal element being σ_i^{kk}. Therefore, the most efficient nonparametric function estimator for kernel GEE type method is obtained by assuming working independence. The similar result can be derived from kernel GEE estimating equation (2.9) as well.

For the kernel GEE method in (2.9) and (2.10), the kernel weights are used for the purpose of reducing the bias for estimating the nonparametric function. However, it also has the potential risk of eliminating the contributions of all correlated measurements for each subject asymptotically. Therefore, the kernel GEE does not make use of all the information with the repeated measurements by assuming working independence. Therefore, the kernel GEE method is not optimal regarding to the asymptotic variance of the nonparametric function estimator. To control the asymptotic variance, Wang (2003) proposed the marginal kernel regression method. It is a two-step algorithm to control the bias and variance at a certain level simultaneously.

Suppose the j^{th} observation in subject i is within h distance of u, the kernel function incorporates and weights this observation in estimating $\phi(u)$. Then we use all observations of subject i for estimating $\phi(u)$. This behavior leads to estimation bias. To accommodate this, all except the j^{th} observations contribute to estimate $\phi(u)$ by multiplying the residuals calculated in step 1.

To be more specific, let $\mathbf{M}_{ij} = [\mathbf{1}_j, \mathbf{1}_j(u - X_{ij})/h]$ be the $m_i \times 2$ matrix, where $\mathbf{1}_j$ denotes the indicator vector with the j^{th} entry equal to 1, and 0 elsewhere. Also, let $\widetilde{\phi}(u)$ be the working independence estimator. Define $\{\widehat{\phi}(u), \widehat{\phi}^{(1)}(u)\} = (b_1, b_2)$, where (b_1, b_2) solves the kernel-weighted estimating equation

$$0 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(u - X_{ij})\mu^{(1)}\{b_1 + b_2(u - X_{ij})\}(\mathbf{M}_{ij})^{\mathrm{T}}\widehat{\mathbf{\Sigma}}_i^{-1} \left[Y_i - \mu_{*j}\{X^i, b_1, b_2, \widetilde{\phi}(X_i)\}\right], \quad (2.11)$$

where the k^{th} element of $\mu_{*j}\{X^i, b_1, b_2, \widetilde{\phi}(X_i)\}$ is

$$\mathrm{I}(k = j)\mu\{b_1 + b_2(u - X_{ik})/h\} + \mathrm{I}(k \neq j)\mu\{\widetilde{\phi}(X_{ik})\},$$

where $\mathrm{I}(\cdot)$ is the indicator function.

To see the difference between the marginal kernel GEE estimator displayed in (2.11) and the working independence estimator, consider the simplest case when the link function $g(\cdot)$ is the identity function. Then (2.7) has the form

$$Y_{ij} = \phi(X_{ij}) + \epsilon_{ij}.$$

In this case, the asymptotic form of the marginal kernel GEE estimator $\widehat{\phi}(u)$ is

$$\widehat{\phi}(u) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(X_{ij} - u)\left\{(v_i^{jj}Y_{ij} + \sum_{k\neq j}^{m_i} v_i^{jk}(Y_{ik} - \widetilde{\phi}_{ik})\right\}}{\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(X_{ij} - u)v_i^{jj}}.$$

In the meantime, the asymptotic form of the working independence estimator $\widetilde{\phi}(u)$ is

$$\widetilde{\phi}(u) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i} v_i^{jj} K_h(X_{ij}-u)Y_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{m_i} v_i^{jj} K_h(X_{ij}-u)}.$$

Comparing the two estimators above, the marginal kernel GEE estimator has one extra term – the kernel and covariance weighted residuals obtained from the working independence step. The marginal kernel GEE function estimator is proved to have smaller asymptotic variance than the working independence function estimator in general.

The idea of Wang (2003) is applied to the estimation of semiparametric models, e.g., partially linear models in Wang et al. (2005), the estimation of the covariance in Li (2011) and the proposed methods in Chapters 3 and 4.

2.2.2 Covariance Estimation for Longitudinal Data

Covariance estimation plays an important role in multivariate data analysis, longitudinal data analysis and spatial data analysis. The main concerns with estimating the covariance or the inverse of covariance (precision) matrices lie in high dimensionality and positive definiteness. When the dimension (p) is relatively large to the sample size (n), the sample covariance is not a good estimator of the true covariance matrix. When the need for the precision matrix is stronger than the covariance matrix, even if we have the estimated covariance matrix, it is still computationally expensive to do the inversion since it generally takes $O(p^3)$ time. Also the inversion of a covariance matrix sometimes distorts the structure of the precision matrix. The other main concern is that the estimated covariance or precision matrix should be positive definite. For a $p \times p$ covariance or precision matrix, it has as many as $p \times (p+1)/2$ constrained parameters to be estimated, which is usually not an easy problem.

In longitudinal data analysis, we have repeated measurements for each subject. It is common to assume that there is no correlation between subjects while correlation exists within each subject. The correlated observations are usually ordered by time. There is extensive literature discussing the estimation of the within-subject covariance. Meanwhile, the precision estimation is also important since in most classical longitudinal data analysis, the GEE (Zeger et al. (1988)) is applied to estimate the mean function where the precision matrix should be estimated. This finding leads to the idea of estimation of the precision matrix and covariance matrix separately. Pourahmadi (1999) proposed a method for covariance and precision matrix estimation based on the modified Cholesky decomposition. This method decomposes a positive definite matrix into a unique lower triangular matrix and a unique diagonal matrix. It provides an unconstrained and statistically interpretable re-parameterization of covariances or precision matrices. A series of papers based on the modified Cholesky decomposition were proposed to efficiently estimate the covariances or precision matrices; see Pourahmadi (1999), Pourahmadi (2000), Wu and Pourahmadi (2003). However, the method of Pourahmadi (1999) is not designed to handle the unbalanced longitudinal case. Huang et al. (2012) adopted the EM algorithm for missing data analysis to investigate this problem. However, this method has the limitation that it applies to only mildly unbalanced cases where most subjects have similar observations. There are also approaches to estimate working covariance with variance-correlation decomposition. Fan et al. (2007) and Fan and Wu (2008) proposed semiparametric estimation of within-subject covariance for longitudinal data. There are also nonparametric approaches to estimate the working covariance; see Li (2011).

We review the semiparametric estimation of longitudinal within-subject covariance by Fan et al. (2007) since it is used in our development. Fan et al. (2007) estimated the covariance function in longitudinal data by variance-correlation decomposition. To be specific, the marginal variances are estimated by a kernel smoothing method. A specific correlation model is assumed for the correlation matrix such as the AR(1) model. Then the association parameters attached to the correlation model are estimated by minimizing the empirical asymptotic variance of the parameter estimators in the mean function.

The varying-coefficient partially linear model is considered in Fan et al. (2007), it has the form

$$Y_i(t_{ij}) = \mathbf{X}_i^{\mathrm{T}}(t_{ij})\boldsymbol{\beta}_i(t_{ij}) + \mathbf{Z}_i^{\mathrm{T}}(t_{ij})\boldsymbol{\theta} + \epsilon_i(t_{ij}), \tag{2.12}$$

where $i = 1, \ldots, n$. In model (2.12), $\boldsymbol{\beta}_i(t_{ij})$ is a vector of smooth functions and $\boldsymbol{\theta}$ are parameters, both to be estimated. Covariates $\mathbf{X}$ and $\mathbf{Z}$ vary with time t. The error term satisfies

$$\mathrm{E}\{\epsilon_i(t_{ij})|\mathbf{X}_i(t_{ij}), \mathbf{Z}_i(t_{ij})\} = 0.$$

To estimate the working covariance matrices, Fan et al. (2007) assume that there exists a q-dimensional association parameters $\boldsymbol{\phi}$. The association parameters are associated with the working correlation matrix, such that $\mathbf{R}_i = \mathbf{R}_i(\boldsymbol{\phi})$. For example, for $t \neq s$, assuming the AR(1) correlation structure, we have

$$\mathrm{cor}\{\epsilon_i(t_{ij}), \epsilon_i(t_{ik})\} = \rho^{|j-k|},$$

where $\boldsymbol{\phi} = \rho$. If we assume the ARMA(1,1) correlation structure,

$$\mathrm{cor}\{\epsilon_i(t_{ij}), \epsilon_i(t_{ik})\} = \begin{cases} \gamma\rho^{|j-k|}, & j \neq k, \\ 1, & j = k, \end{cases}$$

where $\boldsymbol{\phi} = (\gamma, \rho)$. The variance-correlation decomposition is applied: $\boldsymbol{\Sigma}_i = \mathbf{S}_i^{1/2}\mathbf{R}_i(\phi)\mathbf{S}_i^{1/2}$, where $\boldsymbol{\Sigma}_i$ is the covariance matrix of residuals. Here $\mathbf{S}_i$ is a diagonal matrix which satisfies $\mathbf{S}_i = \mathrm{diag}\{\sigma^2(t_{i1}), \ldots, \sigma^2(t_{im_i})\}$, where $\sigma^2(t_{ij}) = \mathrm{E}\{\epsilon_i^2(t_{ij})|\mathbf{X}_i(t_{ij}), \mathbf{Z}_i(t_{ij})\}$.

To estimate the variance component, Fan et al. (2007) proposed the Nadaraya-Watson estimator. Specifically, the kernel-smoothing-based estimator

$$\widehat{\sigma}^2(t) = \frac{\frac{1}{n}\sum_1^n \sum_{j=1}^{m_i} \widehat{\epsilon}_i^2(t_{ij}) K_h(t - t_{ij})}{\frac{1}{n}\sum_1^n \sum_{j=1}^{m_i} K_h(t - t_{ij})}$$

is applied to estimate $\sigma^2(t)$.

In order to estimate $\boldsymbol{\phi}$, the minimum generalized variance method by Fan et al. (2007) is applied, which is chosen to minimize the asymptotic covariance of the parameter estimators $\widehat{\boldsymbol{\theta}}$. Denote the generalized variance of $\widehat{\boldsymbol{\theta}}$ by $\text{cov}(\widehat{\boldsymbol{\theta}})$. To obtain $\widehat{\boldsymbol{\phi}}$, the generalized minimum variance method is proposed to minimize the determinant of $\text{cov}(\widehat{\boldsymbol{\theta}})$. Therefore, we can choose the working covariance matrix as $\widehat{\boldsymbol{\Sigma}}_i(\widehat{\boldsymbol{\phi}})$.

Chen et al. (2015) modified the Nadaraya-Watson estimation of the variance function by Fan et al. (2007). They proposed a nonparametric approach to estimate the variance function $\sigma^2(t)$ by taking log transformation. This method can deal with heavy tail errors.

Let $\epsilon^2(t_{ij})$ be the squared residuals of the model (2.12). Since it is a random variable, we can find another random variable $\delta(t_{ij})$ so that

$$\epsilon^2(t_{ij}) = \sigma^2(t_{ij})\delta^2(t_{ij}) \quad \text{and} \quad \mathrm{E}\{\delta^2(t_{ij})\} = 1.$$

By taking log on both sides of the above equation, we have

$$\log\{\epsilon^2(t_{ij})\} = \log\{\psi\sigma^2(t_{ij})\} + \log\{\psi^{-1}\delta^2(t_{ij})\} = \sigma_*^2(t_{ij}) + \delta_*(t_{ij}),$$

where ψ is a positive constant to make $\mathrm{E}\{\delta_*(t_{ij})\} = 0$. Hence, $\delta_*(t_{ij})$ can be treated as an residual term with expectation zero. Using local linear approximation, $\widehat{\sigma}_*^2(t)$ can be nonparametrically estimated. On the other hand, we have the equation

$$\frac{\delta^2(t_{ij})}{\psi} = r^2(t_{ij}) \exp\{-\sigma_*^2(t_{ij})\}$$

with $\mathrm{E}\{\delta^2(t_{ij})\} = 1$. Hence, we can estimate ψ by

$$\widehat{\psi} = \Big[\frac{1}{\sum_{i=1}^n m_i} \sum_{i=1}^n \sum_{j=1}^{m_i} \widehat{r}^2(t_{ij}) \exp\big\{ - \widehat{\sigma}_*^2(t_{ij})\big\}\Big]^{-1}.$$

Therefore, we have the estimate of $\sigma^2(t)$ by

$$\widehat{\sigma}^2(t) = \frac{\exp \widehat{\sigma}_*^2(t)}{\widehat{\psi}}.$$

It is shown that the semiparametric estimators for the association parameters are consistent when the covariance structure is correctly specified.

The semiparametric regression estimator has the possible weakness of loss of efficiency when the within-subject correlation structure is misspecified. Therefore, there is literature discussing nonparametric estimation of the covariance matrices; see Wu and Pourahmadi (2003) and Li (2011). Li et al. (2010) proposed a nonparametric covariance function estimation method with the ideas in spatial statistics. This approach is compared with the semiparametric covariance function estimation approach of Fan et al. (2007) and Chen et al. (2015) in PLSIM in Chapter 3.

2.3 Semiparametric Models

Semiparametric models are flexible in statistical modeling with their advantages of incorporating both the parametric and nonparametric components. It is popular in economics, biomedical science and many other research fields.

The estimation of semiparametric models date back to Severini and Staniswalis (1994) when the quasi-likelihood method was proposed. A series of quasi-likelihood-based methods for different semiparametric models were then studies; see Chen and Jin (2006), Fan and Li (2004) and He et al. (2005). The general approach for parameter estimation in these models is likelihood based. The smoothing back-fitting algorithm method (Buja et al. (1989)) is often applied for estimating the nonparametric functions.

To study semiparametric models, first we define the model $\mathcal{P}$ as a set of probability distributions for the observed data. A statistical model P_θ is a collection of probability measures defined by $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$. It specifies the distribution of a random variable X. For parametric models, Θ is a parameter in the finite dimensional Euclidean space. While for nonparametric models, Θ is in an infinite dimensional Euclidean space. Semiparametric models lie in between parametric and nonparametric models which can be denoted by

$$\mathcal{P} = \{P_{\boldsymbol{\theta},\varphi} : \boldsymbol{\theta} \in \boldsymbol{\Theta}, \varphi \in \mathcal{F}\}, \tag{2.13}$$

where $\boldsymbol{\theta}$ is in a finite dimensional space $\boldsymbol{\Theta}$ and is usually of primary interest. Meanwhile, φ is in an infinite dimensional space $\mathcal{F}$ and is usually of secondary interest. There may be other components in $\mathcal{P}$ which are not of our interest. Hence they are not parameterized in (2.13). It is possible that $\boldsymbol{\theta}$ and φ may have multiple subcomponents as well. The goals of semiparametric inferences are primarily but not limited to:

- Select an appropriate model of (2.13) with given data;
- Derive the semiparametric score function and information bound for $\boldsymbol{\theta}$;
- Estimate one or more subcomponents of $\boldsymbol{\theta}$ or φ;
- Make inference (e.g., confidence intervals) for the parameters of interest;

- Obtain semiparametric efficient estimators for the parameters $\boldsymbol{\theta}$.

Some of the common semiparametric models are partially linear models (Chen and Shiau (1994)), single-index models (Hardle et al. (1993)), Cox proportional hazard models (Lin and Wei (1989)) and GPLSIM (Carroll et al. (1997)).

In Chapters 3 and 4, we focus on PLSIM and GPLSIM for longitudinal data. We derive the score function and information bound for estimating parameters. Both components $\boldsymbol{\theta}$ and φ in the finite and infinite spaces are estimated. By deriving their asymptotic distributions, we can make inferences for both components. The proposed parameter estimators are proved to be semiparametrically efficient.

2.3.1 Semiparametric Efficiency

To define the semiparametric efficiency for parameters in a semiparametric model, we start with an asymptotic theorem with a parametric model. Suppose there is an estimator T_n for θ based on the data $X_1, \ldots, X_n$. The estimator T_n is called regular when the limiting distribution of $\sqrt{n}(T_n - \theta)$ does not depend on θ. Suppose that $\mathrm{E}\{\sqrt{n}(T_n - \theta)\} \to 0$ in probability. Then to efficiently estimate θ, the limiting variance of $\sqrt{n}(T_n - \theta)$ has to be the smallest among all regular estimators for θ. The inverse of the limiting variance is the information bound for T_n.

For a semiparametric model $\mathcal{P}$ in (2.13), if the information for the regular estimator T_n for θ equals the minimum of the information bounds for all parametric submodels of $\mathcal{P}$, then T_n is semiparametric efficient for estimating θ. Besides, the associated information for T_n is call the "efficient information" (also called the efficient information for $\mathcal{P}$). Sometimes it is important for us to consider a collection of many one-dimensional submodels surrounding a representative $\mathcal{P}$. Each submodel is represented by a score function $\mathcal{S}$. The score function is defined in terms of the Fréchet derivatives; see Van der Vaart (2000). The collection of score functions $\dot{\mathcal{P}}_P$ is called a tangent set. When the tangent set is closed under linear combinations, it is called a tangent space. The tangent space for model $\mathcal{P}$ is the closed linear span of all score functions of regular parametric submodels of $\mathcal{P}$. The general definitions of score function and tangent space can be used to describe the projection method below. More details can be found in Bickel et al. (1993).

Bickel et al. (1993) proposed the projection method to find the semiparametric efficient score and the semiparametric information bound. Consider a regular semiparametric model $\mathcal{P} = \mathcal{P}_{\boldsymbol{\theta},\varphi}$, where $\boldsymbol{\theta}$ is the finite dimensional parameters and φ is the nonparametric part with infinite dimension. Let $\mathcal{S}_{\boldsymbol{\theta}}$ be the score function with respect to $\boldsymbol{\theta}$ in submodel $\mathcal{P}_{\boldsymbol{\theta}}$ which is $\mathcal{M}$ with the true function $\boldsymbol{\phi}_0$ given. Besides, let $\dot{\mathcal{P}}_\varphi$ be the tangent space for submodel $\mathcal{P}_\varphi$ which is model $\mathcal{P}$ evaluated at the true parameters values $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Consider $\mathcal{S}_{\boldsymbol{\theta}}$ as an element in the Hilbert space and $\dot{\mathcal{P}}_\varphi$ as a subset of the same Hilbert space with inner product $\mathrm{E}(\eta_1^{\mathrm{T}}\eta_2)$, where η_1 and η_2 are two elements in $\dot{\mathcal{P}}_\varphi$. Then the residual from the projection of $\mathcal{S}_{\boldsymbol{\theta}}$ on $\dot{\mathcal{P}}_\varphi$ exists and there is a unique vector $\mathcal{S}_{\mathrm{e}}$ satisfying

$$\mathcal{S}_{\boldsymbol{\theta}} - \mathcal{S}_{\mathrm{e}} \in \dot{\mathcal{P}}_\varphi \quad \text{and} \quad \mathrm{E}(\mathcal{S}_{\mathrm{e}}^{\mathrm{T}}\mathbf{w}) = 0 \quad \text{for all} \quad \mathbf{w} \in \dot{\mathcal{P}}_\varphi. \tag{2.14}$$

If the likelihood function is regular with score function $\mathcal{S}_{\boldsymbol{\theta}}$ and $\mathrm{E}(\mathcal{S}_{\mathrm{e}}\mathcal{S}_{\mathrm{e}}^{\mathrm{T}})$ is nonsingular, then the semiparametric information bound is $\Omega = \{\mathrm{E}(\mathcal{S}_{\mathrm{e}}\mathcal{S}_{\mathrm{e}}^{\mathrm{T}})\}^{-1}$ and the semiparametric efficient score is $\mathcal{S}_{\mathrm{e}}$. The project method is useful for obtaining the semiparametric efficiency of parameters for PLSIM and GPLSIM with longitudinal data in Chapters 3 and 4.

2.3.2 Partially Linear Single-Index Models

Semiparametric models such as partially linear models are flexible in statistical modeling. However, they still have limitations. For example, the partially linear model with one multivariate nonparametric function suffers from the "curse of dimensionality" when the number of covariates in the nonlinear component is large. On the other side, a semiparametric model with multiple univariate nonparametric functions such as partially linear nonparametric additive models does not take into account the interaction effects. Carroll et al. (1997) first proposed GPLSIM. It has the form:

$$\begin{aligned} &\mathrm{E}(Y|\mathbf{X}_i, \mathbf{Z}_i) = \mu(\mathbf{X}_i, \mathbf{Z}_i), \\ &g^{-1}\{\mu(\mathbf{X}_i, \mathbf{Z}_i)\} = \mathbf{X}_i\boldsymbol{\beta} + \gamma(\mathbf{Z}_i\boldsymbol{\theta}), \end{aligned} \tag{2.15}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are possibly multi-dimensional parameters associated with predictors $\mathbf{X}$ and $\mathbf{Z}$ respectively, $\gamma(\cdot)$ is an unknown single-index function and $g^{-1}(\cdot)$ is a known monotonic and differentiable link function. Model (2.15) is a generalization of several models such as generalized linear models, partially linear models and single-index models. It has the advantage of modeling when some covariates, $\mathbf{X}$, are parametrically associated with the response Y and some other covariates, $\mathbf{Z}$, are nonparametrically associated with the response variable.

With such a complicated model in (2.15), the first problem is to make constraints to make it identifiable. Lin and Kulasekera (2007) discussed this problem in detail. The first requirement is $||\boldsymbol{\theta}|| = 1$, where $||\boldsymbol{\theta}||$ indicates the Euclidean norm of $\boldsymbol{\theta}$. Besides, the first element of $\boldsymbol{\theta}$ should be positive. Depending on the estimation methods, other assumptions are required to make the parameters and the single-index function estimable. For example, the "delete-one-component" method is used in Yu and Ruppert (2002) and Zhu and Xue (2006) where partial derivatives are used as a component of the estimators.

For the estimation of PLSIM, the back-fitting algorithm for the quasi-likelihood method proposed by Carroll et al. (1997) is not stable and computationally expensive. Furthermore, undersmoothing of the single-index function is required to reduce the bias of the parameter estimators. Yu and Ruppert (2002) proposed the penalized spline method and Xia and Härdle (2006) proposed the minimum average variance method to accommodate this problem. However, the parameter estimators and the single-index function estimator of these methods are still not efficient. Liang et al. (2010) approached the model with a iterative estimation idea. They applied the profile least-squares procedure to obtain the efficient estimators with i.i.d. data. Li et al. (2010) proposed bias-corrected block empirical likelihood inference for GPLSIM and provided confidence regions for the parameters. Methods based on the least squares and kernel smoothing or spline smoothing are further proposed and semiparametric efficient estimators are obtained.

2.4 Semiparametric Models for Longitudinal Data

It is more difficult to efficiently estimate in semiparametric models with correlated data. The main difficulty is the proper incorporation of the within-subject correlation in both the parametric component estimation step and the nonparametric component estimation step.

For this problem, there are typically two types of approaches to consider. The first type is the combination of the kernel-based method and the GEE. Wang et al. (2005) studied the efficient estimation of generalized partially linear models for longitudinal data. They extended the results of Wang (2003) for the longitudinal nonparametric function estimation. The second type is the spline-based method. Yu and Ruppert (2002) proposed the penalized spline method for PLSIM on i.i.d. data. Huang et al. (2007) further investigated the spline-based additive models for partially linear models. They also studied the efficient estimation of generalized partially linear models for longitudinal data. Cheng et al. (2014) extended the results of Huang et al. (2007) to generalized partially linear additive models. They also showed that the estimated parameters are semiparametric efficient. In the following section, we particularly reviewed the literature in the estimation of PLSIM for longitudinal data.

2.4.1 Partially Linear Single-Index Models for Longitudinal Data

In PLSIM for longitudinal data, the following marginal model is considered:

$$Y_{ij} = \mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \gamma(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}) + \epsilon_{ij} \tag{2.16}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m_i$. For the i^{th} subject, the continuous response variable Y_{ij} and covariates $\{\mathbf{X}_{ij}, \mathbf{Z}_{ij}\}$ are observed for the j^{th} time. The parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ and the single-index function $\gamma(\cdot)$ are to be estimated.

A list of recent works on this topic are summarized here. Ping et al. (2010) used the generalized penalized spline least squares method and assumed working correlation matrices to estimate the parameters and the single-index function. Lai et al. (2013) and Ma et al. (2014) proposed the bias-corrected quadratic inference function method (Qu et al. (2000)) to estimate the parameters in model (2.16) by accounting for the within-subject correlation. Chen et al. (2015) proposed a unified semiparametric GEE analysis of (2.16) for both sparse and dense unbalanced longitudinal data. We particularly review their work and relate it to our proposed methods in Chapters 3 and 4. Chen et al. (2015) showed that the convergence rate and the limiting variance/covariance are different for the sparse and dense longitudinal cases with the proposed method; see Kim and Zhao (2012). However, their parameter estimators are generally not semiparametrically efficient. For the estimation of the single-index function, they applied local linear approximation adjusted by the number of measurements for each subject. To be specific, given $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, they estimated the single-index function $\gamma(\cdot)$ and its first derivative $\gamma'(\cdot)$ at point u by applying local linear approximation, which is to minimize the following loss function

$$\sum_{i=1}^{n}\left[\omega_i \sum_{j=1}^{m_i}\left\{y_{ij} - \mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} - a - b(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)\right\}^2 K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)\right] \tag{2.17}$$

where $\omega_i = 1/(nm_i)$ since the number of observations may not be relatively similar across the subjects; see Li and Hsing (2010).

We pay a special attention to the choices of weight ω_i when estimating the single-index function since the weight in (2.17) is not the most efficient one for every type of longitudinal data. In other words, the simply weighted local linear approximation approach in (2.17) does not fully take into consideration the within-subject covariance. The two examples of the weights

1. $\omega_i = 1/\sum_{i=1}^{n} m_i$

2. $\omega_i = 1/(nm_i)$

correspond to the independence within-subject and perfect correlation within-subject cases respectively. While in real scenarios, there is some correlation in within subject measurements. Therefore, a more proper weighting approach is desired in (2.17).

One idea to efficient estimate the parameters in PLSIM and GPLSIM is to make use of the GEE in parameter estimation and marginal kernel method in the single-index function estimation. Inspired by the novel marginal kernel method of Wang (2003) in the nonparametric function estimation, we properly take into consideration the within subject correlation in sparse longitudinal data setting. We will particularly introduce our proposed approach in Chapters 3 and 4.

For the estimation of GPLSIM for longitudinal data, Grace et al. (2009a) and Grace et al. (2009b) proposed marginal estimation of semiparamtric methods for correlated binary response data. The mean response parameters as well as the association parameters in odds ratio (Lipsitz et al. (1991)) are efficiently estimated and the theoretical results of the estimators are established. Chowdhury and Sinha (2015) considered the same problem with the second-order GEE method where similar results are obtained. For correlated count data, Wang et al. (2015) proposed the profile MLE for GPLSIM. The zero-inflated

Poisson approach is applied as the base model. In our model development, we would like to propose a general estimation method for longitudinal GPLSIM. The response variable can be Bernoulli, Poisson, Gamma or other distributions in the exponential family. The detailed descriptions on this topic is in Chapter 4.

2.5 Summary

From the literature reviews above, we observe that the efficient estimation of PLSIM and GPLSIM for longitudinal data is still a challenging work. It consists of several sub-problems:

1. Efficient estimation of the parameters;
2. Semiparametric efficiency of the parameter estimators;
3. Optimal asymptotic properties of the single-index function estimator;
4. Covariance (precision) matrix estimation;
5. Comparison of the proposed method with the existing methods.

In the rest of the dissertation, we focus on the analysis of longitudinal PLSIM and GPLSIM on the sub-problems listed above. By properly taken into consideration the within-subject correlation, the proposed methods effectively estimate the parameters and the single-index function. The semiparametric efficiency theories discussed in this chapter are applied. We will also compare the proposed method with the existing methods reviewed above.

3. PARTIALLY LINEAR SINGLE-INDEX MODELS FOR LONGITUDINAL DATA

In this chapter, we discuss the estimation of partially linear single-index models for longitudinal data which may be unbalanced. In particular, a new iterative five-step method is established to estimate the parametric components with the GEE and the single-index function by the marginal kernel method. The resulting estimators properly account for the within-subject correlation. The parameter estimators are shown to be asymptotically semiparametric efficient. The asymptotic variance of the single-index function estimator is shown to be minimized when the working error covariance matrices are correctly specified. The new estimators are more efficient than estimators in the existing literature. These asymptotic results are obtained without assuming normality. Simulation studies are performed to demonstrate the finite sample performance. To illustrate the proposed method, two real data examples are appropriately analyzed.

3.1 Introduction

Longitudinal data analysis is popular in a variety of fields such as biology, epidemiology and economics. There is an extensive literature discussing problems and developing theory and methodology in longitudinal data analysis; see Laird and Ware (1982), Liang and Zeger (1986), Zeger et al. (1985), and Zeger and Liang (1986), etc. The main problems lie in the estimation of the mean function, the within-subject covariance function and the regression parameters for some specified models. These problems are generally more well addressed for balanced longitudinal cases, but they are much more difficult to handle in unbalanced cases.

Semiparametric models are popular in recent years since they enjoy the advantages of incorporating both the parametric and nonparametric components. They have many applications in longitudinal data; see, e.g., Ichimura (1993), Chen and Shiau (1994) and Carroll et al. (1997). One of the applications of semiparametric models is PLSIM

$$Y_i = \mathbf{X}_i\boldsymbol{\beta} + \phi(\mathbf{Z}_i\boldsymbol{\theta}) + \epsilon_i$$

for $i = 1, \ldots, n$, where $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are parameters and $\phi(\cdot)$ is an unknown single-index function, all to be estimated. The residual term ϵ_i has mean zero. This model enjoys the advantages when some covariates are linearly related to the response while some other covariates are nonlinearly related to the response. Carroll et al. (1997) proposed and discussed estimation, testing and theoretical results of the above models for the i.i.d. case. For the i.i.d. data, Liang et al. (2010) proposed the profile least-squares method to obtain the semiparametrically efficient parameter estimators. Besides, the smoothly clipped absolute deviation penalty (SCAD) approach to is applied for variable selection. Chen and Parker (2014) specifically calculated semiparametric information bound for PLSIM with the method of Severini and Tripathi (2001). Hu et al. (2004) and Li et al. (2010) discussed how PLSIM are applied to longitudinal data.

However, how to properly estimate the parameters and single-index function in PLSIM in the longitudinal data setting continues to receive considerable attention. Recently, Chen et al. (2015) proposed a unified semiparametric generalized estimating equation (GEE) analysis in PLSIM for both sparse and dense unbalanced longitudinal data. Hereafter the method is referred as SGEE. They pointed out that the convergence rate and the limiting variance/covariance are different for the sparse and dense longitudinal cases with the proposed method. However, their parameter estimators are generally not semiparametric efficient. For the estimation of the single-index function, they applied local linear approximation adjusted by the number of measurements for each subject. This method does not fully take into consideration the within-subject covariance.

Considering within-subject correlation for nonparametric function estimation for sparse longitudinal data, Lin and Carroll (2000, 2001) proposed the nonparametric profile-kernel GEE for partially linear models. They showed that among all profile-kernel GEEs, the consistent and most efficient estimator can be obtained by completely ignoring the within-subject correlation and undersmoothing the nonparametric single-index function. Wang (2003) re-examined the profile-kernel GEE methods. She pointed out that when kernel weights are used to control the bias for single-index function estimation, they also eliminate the contributions of correlated measurements for each subject asymptotically. Therefore, the profile-kernel GEE does not use all the information provided by the repeated measurements for the single-index function estimation. As a result, the profile-kernel GEE method is not optimal regarding to the asymptotic variance of the single-index function estimator. To control the asymptotic variance, she proposed marginal kernel regression which is a two-step algorithm to control the bias and variance simultaneously.

In this chapter we focus on the estimation efficiency of the parameters as well as the single-index function in PLSIM with a general unbalanced sparse longitudinal data setting. Specifically, first we use the working independence (WI) kernel GEE for the single-index function estimation by fixing the parameters as known and use the WI least square estimation for the parameters by fixing the single-index function as known. After convergence in the iteration, we estimate the within-subject covariance semiparametrically with variance-correlation decomposition. In the refined iterated estimation step, we estimate the unknown single-index function by using the marginal kernel regression as in Wang (2003) and estimate the parameters with GEE. We show that the proposed refined estimators for the parameters are semiparametric efficient. Furthermore, we also show that the proposed single-index function estimator is more efficient than the single-index function estimator in SGEE. It is important to note that in our proposed methodology and its corresponding theory no distributional assumptions such as multivariate normality are needed for $\mathbf{X}_i$, $\mathbf{Z}_i$

or ϵ_i.

The rest of the chapter is organized as follows. In Section 3.2 we describe an estimation procedure to obtain new estimators for the parameters and the single-index function in longitudinal PLSIM. Section 3.3 provides some asymptotic results for the estimators where the asymptotic variance, asymptotic bias and convergence rates are presented. We show that the proposed single-index function estimator is more efficient than the working independence estimator. Moreover, both the parameter and single-index function estimators reach minimum asymptotic variances when the covariances are correctly specified in which case the parameter estimators are further shown to be semiparametric efficient. Regularity conditions and some technical arguments are also given after each theorem. Section 3.4 gives some finite sample simulations to compare the proposed method with some existing ones. In Section 3.5 we apply the proposed method in two real data examples. In Section 3.6 we conclude the chapter and discuss some possible extensions for future research.

3.2 Methodology

Rewrite the longitudinal PLSIM as follows:

$$Y_{ij} = \mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \phi(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}) + \epsilon_{ij} \tag{3.1}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m_i$. In the sparse longitudinal case m_i is bounded for all i while $n \to \infty$. Parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ of dimension p and q respectively are to be estimated. The univariate unknown single-index function $\phi(\cdot)$ is also to be estimated. The continuous response variable Y_{ij} and covariates $(\mathbf{X}_{ij}^{\mathrm{T}}, \mathbf{Z}_{ij}^{\mathrm{T}})$ are observed at time t_{ij} and ϵ_{ij} is a random error. For simplicity, in this chapter we assume that $\mathbf{Z}_{ij}$ are continuous random variables and $\mathbf{X}_{ij}$ can be either continuous or categorical. As in most longitudinal studies, we assume that the subjects are mutually independent, while there is a within-subject correlation for each subject. Let $\boldsymbol{\Sigma}_i = \text{cov}(\mathbf{E}_i)$ with $\mathbf{E}_i^{\mathrm{T}} = (\epsilon_{i1}, \ldots, \epsilon_{im_i})$.

The main task is to estimate the true parameters $\boldsymbol{\Theta}_0 = (\boldsymbol{\beta}_0^{\mathrm{T}}, \boldsymbol{\theta}_0^{\mathrm{T}})^{\mathrm{T}}$ and the true unknown nonparametric single-index function $\phi_0(\cdot)$ in Equation (3.1). To guarantee identifiability, we assume that the Frobenius norm of $\boldsymbol{\theta}_0$ is 1 with the first element of $\boldsymbol{\theta}_0$ being positive as in Liang et al. (2010). To simplify the notation, we denote $N = \sum_{i=1}^n m_i$, $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{im_i})^{\mathrm{T}}$, $\mathbf{X}_i = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{im_i})^{\mathrm{T}}$, $\mathbf{Z}_i = (\mathbf{Z}_{i1}, \ldots, \mathbf{Z}_{im_i})^{\mathrm{T}}$, $\phi(\mathbf{Z}_i\boldsymbol{\theta}) = (\phi(\mathbf{Z}_{i1}^{\mathrm{T}}\boldsymbol{\theta}), \ldots, \phi(\mathbf{Z}_{im_i}^{\mathrm{T}}\boldsymbol{\theta}))^{\mathrm{T}}$ and $\boldsymbol{\Theta} = (\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\theta}^{\mathrm{T}})^{\mathrm{T}}$. Besides, let $K(\cdot)$ be a symmetric kernel density function and $K_h(x) = h^{-1}K(x/h)$, where h is the bandwidth. The proposed estimation procedure has the following steps:

1. Let u be a point in $\mathcal{G}$, the domain of $\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}$ defined in Assumption 3.3.4. Given $\boldsymbol{\Theta}$, the one-step estimate of $\phi(\cdot)$ and $\phi^{(1)}(\cdot)$, denoted by WI kernel GEE, is obtained by solving the following estimating equation

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \mathbf{L}^i(u)^{\mathrm{T}} \mathbf{K}_h^i(u) \big\{ \mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{L}^i(u)\mathbf{a} \big\} = 0, \tag{3.2}$$

   where $\mathbf{L}^i(u)$ is an $m_i \times 2$ matrix with the $(j,k)^{th}$ element $(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)^{k-1}$ and $\mathbf{K}_h^i(u) = \text{diag}\{K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)\}$ with the $(l,l)^{th}$ entry being $K_h(\mathbf{Z}_{il}^{\mathrm{T}}\boldsymbol{\theta} - u)$. Here h is a proper bandwidth to be discussed later. The obtained estimates are $\widetilde{\mathbf{a}}(u, \boldsymbol{\Theta}) = \{\widetilde{\phi}(u, \boldsymbol{\Theta}), \widetilde{\phi}^{(1)}(u, \boldsymbol{\Theta})\}^{\mathrm{T}}$.

2. With the estimated single-index function, the one-step estimates of $\boldsymbol{\Theta}$ which ignore the within-subject correlation structure are obtained by minimizing

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \big\{ \mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \widetilde{\phi}(\mathbf{Z}_i\boldsymbol{\theta}, \boldsymbol{\Theta}) \big\}^{\mathrm{T}} \big\{ \mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \widetilde{\phi}(\mathbf{Z}_i\boldsymbol{\theta}, \boldsymbol{\Theta}) \big\}. \tag{3.3}$$

   So the initial estimates $\widetilde{\boldsymbol{\Theta}}$ and $\widetilde{\phi}(\cdot)$ are obtained by iterating between Steps 1 and 2 until convergence.

3. Let $\mathbf{V}_i$ be the estimated working covariance matrix for subject i, $i = 1, \ldots, n$. From the initial estimates, we get the residual term $\widetilde{e}_{ij} = Y_{ij} - \mathbf{X}_{ij}^{\mathrm{T}}\widetilde{\boldsymbol{\beta}} - \widetilde{\phi}(\mathbf{Z}_{ij}^{\mathrm{T}}\widetilde{\boldsymbol{\theta}})$ as an estimate of e_{ij}. A semiparametric variance correlation decomposition approach is applied to estimate the true covariance $\boldsymbol{\Sigma}_i$.

4. Re-estimate $\phi(\cdot)$. Given $\boldsymbol{\Theta}$, let $\mathrm{M}_{ij} = [\mathbf{1}_j, \mathbf{1}_j(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)]$ be the $m_i \times 2$ matrix, where $\mathbf{1}_j$ denotes the indicator vector with j^{th} entry equal to 1, and 0 elsewhere. Define $\{\widehat{\phi}(u, \boldsymbol{\Theta}), \widehat{\phi}^{(1)}(u, \boldsymbol{\Theta})\} = (b_1, b_2)$, where (b_1, b_2) solves the kernel-weighted estimating equation

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \sum_{j=1}^{m_i} K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)\mathrm{M}_{ij}^{\mathrm{T}}\mathbf{V}_i^{-1} \Big[\mathbf{Y}_i - \boldsymbol{\mu}^*\big\{u, \mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\Theta}, \widehat{\phi}_c(\cdot), b_1, b_2\big\}\Big] = 0, \qquad (3.4)$$

where $\boldsymbol{\mu}^*\{u, \mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\Theta}, \widehat{\phi}_c(\cdot), b_1, b_2\}$ is an $m_i \times 1$ vector with the k^{th} element being

$$\mathbf{X}_{ik}^{\mathrm{T}}\boldsymbol{\beta} + \mathrm{I}(k = j)\big\{b_1 + b_2(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)\big\} + \mathrm{I}(k \neq j)\widehat{\phi}_c(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}, \boldsymbol{\Theta})$$

and h is another proper bandwidth. Here $\widehat{\phi}_c(\cdot)$ is the current estimate of $\phi(\cdot)$ and $\mathrm{I}(\cdot)$ is the indicator function.

5. Re-estimate the parameters $\boldsymbol{\Theta}$ using the following GEE

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \frac{\partial\big\{\mathbf{X}_i\boldsymbol{\beta} + \widehat{\phi}(\mathbf{Z}_i\boldsymbol{\theta}, \boldsymbol{\Theta})\big\}^{\mathrm{T}}}{\partial\boldsymbol{\Theta}} \mathbf{V}_i^{-1}\big\{\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta} - \widehat{\phi}(\mathbf{Z}_i\boldsymbol{\theta}, \boldsymbol{\Theta})\big\} = 0. \qquad (3.5)$$

The solutions of (3.5) are updated parameter estimates. We then assign the updated parameter estimates values to Step 4 for iteration. The final estimates of the parameters $\widehat{\boldsymbol{\Theta}} = (\widehat{\boldsymbol{\beta}}^{\mathrm{T}}, \widehat{\boldsymbol{\theta}}^{\mathrm{T}})^{\mathrm{T}}$ and the single-index function estimate $\widehat{\phi}(u, \widehat{\boldsymbol{\Theta}}) = \widehat{\phi}(u)$ are obtained by iterating between Steps 4 and 5 until convergence.

The estimation procedure above can be generally separated into three stages. The first stage (Steps 1 and 2) is to apply the WI kernel GEE and the WI least square methodology to obtain the initial estimates of the parameters, the unknown single-index function and the residuals. The resulting estimates are $\sqrt{n}$–consistent (Lin and Carroll (2000)). The second stage (Step 3) is to obtain proper covariance estimates. The last stage (Steps 4 and 5) is to obtain the refined estimates. By plugging in the working covariance matrices estimated in the second stage, the within-subject correlation is taken into consideration in both the parameter and single-index function estimation steps. More efficiency is expected of the refined estimators. We will investigate the efficiency issue both theoretically and empirically in later sections.

Remark 3.2.1. The unknown single-index function $\phi_0(u)$ in PLSIM does not depend on parameters $\boldsymbol{\Theta}_0$. However, while in the estimation steps of the unknown single-index function, the estimated single-index function depends on the parameter estimates in the current step. When the iteration stops, we have the final estimate for the single-index function $\widehat{\phi}(u, \widehat{\boldsymbol{\Theta}}) = \widehat{\phi}(u)$. For clarity, we use $\widehat{\phi}(u, \boldsymbol{\Theta})$ to emphasize its dependence on $\boldsymbol{\Theta}$ in the estimation procedure and use $\widehat{\phi}(u)$ for simplicity when no confusion arises.

Remark 3.2.2. Given the residual estimates $\widetilde{e}_{ij}$ from the initial estimation step, there are several ways to obtain the working covariance matrix estimates $\mathbf{V}_i$ for longitudinal data, such as Wu and Pourahmadi (2003). One of the covariance modeling techniques is based on variance-correlation decomposition. Some recent works include Fan et al. (2007), Fan and Wu (2008) and Li (2011). In Step 3 above we choose to follow the method employed by Chen et al. (2015). The main idea is based on a variance-correlation decomposition.

First we estimate the variance term by taking a log transformation to accommodate possibly nonstationary error variance and use a local linear approximation to obtain the estimates. Then we assume a common specific correlation structure for all subjects such as compound symmetry or AR(d) with unknown parameters to be estimated. Finally we estimate the parameters in the correlation structure by minimizing the determinant of the asymptotic variance of the parameter estimators introduced by Fan et al. (2007). The details of the covariance estimation can be found in Section 4 of Chen et al. (2015).

3.3 Theoretical Properties

In our theoretical development, we assume that the number of subjects n goes to ∞ and the number of measurements m_i is bounded for $i = 1, \ldots, n$. Wang et al. (2005) proposed the semiparametric efficient estimators for the marginal generalized partially linear models when the multivariate normality is assumed. Their theoretical and empirical investigations were focused on the parameter estimation part. Inspired by Lin and Carroll (2001) and Wang et al. (2005), we start with the semiparametric efficient score function and the information bound for longitudinal PLSIM but without making any distributional assumptions. We then show that when the covariance matrices are correctly specified, the proposed parameter estimators achieve the semiparametric information bound. Moreover, the asymptotic results for the nonparametric single-index function are also presented, showing that the asymptotic variance of the estimator is minimized when the covariance matrices are correctly specified. These results are formally presented in Theorems 3.3.1–3.3.6 with some detailed proofs.

3.3.1 Assumptions

Before we provide all theoretical results, we need the following essential assumptions.

Assumption 3.3.1. *Kernel function $K(\cdot)$ is bounded and symmetric with a compact support. It also has continuous first-order derivative $K^{(1)}(\cdot)$.*

Assumption 3.3.2. *The residuals are unbiased and bounded for the second-order moment, i.e.,* $\mathrm{E}(e_{ij}) = 0$, $\mathrm{E}(e_{ij}^2) \leq M$ *for some* $M > 0$, $i = 1, \ldots, n$ *and* $j = 1, \ldots, m_i$.

Assumption 3.3.3. $\phi(\cdot)$ *has second continuous derivative.*

Assumption 3.3.4. *For* $i = 1, \ldots, n$, $j = 1, \ldots, m_i$, $\mathbf{Z}_{ij}$ *is bounded with a compact support. The density of* $\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}$*, denoted by* $f_{ij}(u)$*, is twice continuously differentiable and positive for all* $u \in \mathcal{G}$ *with* $\mathcal{G} = \{u = \mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} : \mathbf{Z}_{ij} \in \mathcal{Z}, \boldsymbol{\theta} \in \Xi\}$*. Here* Ξ *is the a compact parameter space for* $\boldsymbol{\theta}$ *and* $\mathcal{Z}$ *is a compact support for* $\mathbf{Z}_{ij}$*. In addition, the joint density of* $(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}, \mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta})$ *has first partial derivatives.*

Assumption 3.3.5. $\mathrm{E}(\mathrm{X}_{ij}|\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} = u)$ *and* $\mathrm{E}(\mathbf{Z}_{ij}|\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} = u)$ *are smooth functions of* u *with continuous derivatives up to the second order. In addition,* $\sup_{u\in\mathcal{G}} \mathrm{E}(||\mathbf{X}_{ij}||^2|\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} = u)$ *and* $\sup_{u\in\mathcal{G}} \mathrm{E}(||\mathbf{Z}_{ij}||^2|\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} = u)$ *are bounded for all* $i = 1, \ldots, n$, $j = 1, \ldots, m_i$.

Assumption 3.3.6. $h \to 0$, $nh^8 \to 0$ *and* $nh/log(1/h) \to \infty$.

Assumption 3.3.1 lists some regularity conditions for the kernel function. Assumption 3.3.2 is imposed for the consistency and asymptotic normality of our estimators. Assumption 3.3.3 is the smoothness restriction for the unknown single-index function. Assumption 3.3.4 ensures that the denominator of the kernel estimator for the single-index function in Steps 1 and 4 in Section 3.2 is meaningful and that some relevant asymptotic expansions are valid. Assumption 3.3.5 is a commonly used moments condition for predictors in PLSIM. Assumption 3.3.6 imposes some bandwidth conditions to allow the optimal bandwidth to be included. Existing works impose various bandwidth smoothness conditions. The general conditions given in Assumption 3.3.6 are sufficient for the properties obtained in this chapter.

3.3.2 Theory for Semiparametric Efficiency

First we define the L_2 norm of a square-integrable function $f(u)$ for all $u \in \mathcal{G}$ by $[\int_{\mathcal{G}}\{f(u)\}^2 du]^{1/2}$, where $\mathcal{G}$ is defined in Assumption 4 in Section 3.3.1. Besides, we denote $(\mathbf{X}_i, \underline{\mathbf{Z}}_i) = [\mathbf{X}_i, \{\phi_0^{(1)}(\mathbf{Z}_i\boldsymbol{\theta}_0) \otimes \mathbf{1}_q^{\mathrm{T}}\} \odot \mathbf{Z}_i]$, where $\otimes$ is the Kronecker product and $\odot$ is the component-wise product. Using the notation in Proof 3.3.2, we have

$$\widetilde{\mathbf{X}}_{i,\mathrm{e}} = \mathbf{X}_i - \psi_{n,\boldsymbol{\beta}}(\mathbf{Z}_i\boldsymbol{\theta}_0), \quad \widetilde{\mathbf{Z}}_{i,\mathrm{e}} = \underline{\mathbf{Z}}_i - \psi_{n,\boldsymbol{\theta}}(\mathbf{Z}_i\boldsymbol{\theta}_0), \tag{3.6}$$

where $\psi_{n,\boldsymbol{\beta}}(\mathbf{Z}_i\boldsymbol{\theta}_0) = \{\psi_{n,\beta_1}(\mathbf{Z}_i\boldsymbol{\theta}_0), \ldots, \psi_{n,\beta_p}(\mathbf{Z}_i\boldsymbol{\theta}_0)\}$ is an $m_i \times p$ matrix with $\psi_{n,\beta_l}(\cdot) \in L_2$, $l = 1, \ldots, p$. It satisfies

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\Big\{\widetilde{\mathbf{X}}_{i,\mathrm{e}}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\kappa_n(\mathbf{Z}_i\boldsymbol{\theta}_0)\Big\} = 0 \tag{3.7}$$

for all $\kappa_n(\cdot) \in L_2$. Here $\kappa_n^{\mathrm{T}}(\mathbf{Z}_i\boldsymbol{\theta}_0) = \{\kappa_n(\mathbf{Z}_{i1}^{\mathrm{T}}\boldsymbol{\theta}_0), \ldots, \kappa_n(\mathbf{Z}_{im_i}^{\mathrm{T}}\boldsymbol{\theta}_0)\}$. Besides, $\psi_{n,\boldsymbol{\theta}}(\mathbf{Z}_i\boldsymbol{\theta}_0) = \{\psi_{n,\theta_1}(\mathbf{Z}_i\boldsymbol{\theta}_0), \ldots, \psi_{n,\theta_q}(\mathbf{Z}_i\boldsymbol{\theta}_0)\}$ is an $m_i \times q$ matrix with $\psi_{n,\theta_l}(\cdot) \in L_2$, $l = 1, \ldots, q$. It also satisfies

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\Big\{\widetilde{\mathbf{Z}}_{i,\mathrm{e}}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\kappa_n(\mathbf{Z}_i\boldsymbol{\theta}_0)\Big\} = 0 \tag{3.8}$$

for all $\kappa_n(\cdot) \in L_2$. Similarly to the arguments in Lemma A4 in Huang et al. (2007), the semiparametric efficient score function of $\boldsymbol{\Theta}_0$ is (see Proof 3.3.2)

$$\mathbf{S}_{\mathrm{e}} = \sum_{i=1}^{n}(\widetilde{\mathbf{X}}_{i,\mathrm{e}}, \widetilde{\mathbf{Z}}_{i,\mathrm{e}})^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\Big\{\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}_0 - \phi_0(\mathbf{Z}_i\boldsymbol{\theta}_0)\Big\}. \tag{3.9}$$

We now present our first theorem below.

Theorem 3.3.1. *If Assumptions 3.3.4–3.3.5 hold, the semiparametric information bound* *of* $\mathbf{\Theta}_0$ *is*

$$\mathbf{U}_{\rm e} = \lim_{n\to\infty} \big\{\frac{1}{n}{\rm E}(\mathbf{S}_{\rm e}\mathbf{S}_{\rm e}^{\rm T})\big\}^{-1} = \lim_{n\to\infty} \left[\frac{1}{n}\sum_{i=1}^{n} {\rm E}\big\{(\widetilde{\mathbf{X}}_{i,{\rm e}}, \widetilde{\mathbf{Z}}_{i,{\rm e}})^{\rm T}\mathbf{\Sigma}_i^{-1}(\widetilde{\mathbf{X}}_{i,{\rm e}}, \widetilde{\mathbf{Z}}_{i,{\rm e}})\big\}\right]^{-1}. \quad (3.10)$$

From Kress et al. (1989), each element of $\psi_{n,\boldsymbol{\beta}}(\cdot)$ and $\psi_{n,\boldsymbol{\theta}}(\cdot)$ solves the Fredholm integral equation of the second kind, which is shown in Equations (3.26) and (3.28). We use these equations to show that the proposed parameter estimators reach the semiparametric information bound in the proof below.

Proof. We apply the projection method (see Bickel et al. (1993), Chapter 3) to find the semiparametric efficient score and the semiparametric information bound. Consider a regular semiparametric model $\mathcal{M} = \mathcal{M}_{\boldsymbol{\alpha},\boldsymbol{\phi}}$, where $\boldsymbol{\alpha}$ is the finite dimensional parameters and $\boldsymbol{\phi}$ is the nonparametric part with infinite dimension. Let $\mathcal{S}_{\boldsymbol{\alpha}}$ be the score function with respect to $\boldsymbol{\alpha}$ in submodel $\mathcal{M}_{\boldsymbol{\alpha}}$ which is $\mathcal{M}$ with the true function $\boldsymbol{\phi}_0$ given. Besides, let $\dot{\mathcal{M}}_{\boldsymbol{\phi}}$ be the tangent space for submodel $\mathcal{M}_{\boldsymbol{\phi}}$ which is model $\mathcal{M}$ evaluated at the true parameters values $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$. Consider $\mathcal{S}_{\boldsymbol{\alpha}}$ as an element in the Hilbert space and $\dot{\mathcal{M}}_{\boldsymbol{\phi}}$ as a subset of the same Hilbert space with inner product ${\rm E}(\eta_1^{\rm T}\eta_2)$, where η_1 and η_2 are two elements in $\dot{\mathcal{M}}_{\boldsymbol{\phi}}$. Then the residual from the projection of $\mathcal{S}_{\boldsymbol{\alpha}}$ on $\dot{\mathcal{M}}_{\boldsymbol{\phi}}$ exists and there is a unique vector $\mathcal{S}_{\rm e}$ satisfying

$$\mathcal{S}_{\boldsymbol{\alpha}} - \mathcal{S}_{\rm e} \in \dot{\mathcal{M}}_{\boldsymbol{\phi}} \quad \text{and} \quad {\rm E}(\mathcal{S}_{\rm e}^{\rm T}\mathbf{w}) = 0 \quad \text{for all} \quad \mathbf{w} \in \dot{\mathcal{M}}_{\boldsymbol{\phi}}. \quad (3.11)$$

If the likelihood function is regular with score function $\mathcal{S}_{\boldsymbol{\alpha}}$ and ${\rm E}(\mathcal{S}_{\rm e}\mathcal{S}_{\rm e}^{\rm T})$ is nonsingular, then the semiparametric information bound is $\mathbf{U} = \{{\rm E}(\mathcal{S}_{\rm e}\mathcal{S}_{\rm e}^{\rm T})\}^{-1}$ and the semiparametric efficient score is $\mathcal{S}_{\rm e}$.

Now we denote the longitudinal PLSIM in Equation (3.1) by M. This is a special case of the general framework above. Model M has three unknown parts: $\mathbf{\Theta}_0$, $\phi_0(\cdot)$ and the joint distribution of $(\mathbf{Y}_{ij}, \mathbf{X}_{ij}, \mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0)$ for $i = 1, \ldots, n$, $j = 1, \ldots, m_i$. To derive the efficient score for $\boldsymbol{\Theta}_0$, consider three submodels of model M:

M_1: Model M with only $\boldsymbol{\Theta}_0$ unknown;

M_2: Model M with only $\phi_0(\cdot)$ unknown;

M_3: Model M with both $\boldsymbol{\Theta}_0$ and $\phi_0(\cdot)$ known.

Let $\mathbf{S}_{\boldsymbol{\Theta}}$ be the score function in submodel M_1 and $\dot{M}_k$ be the tangent space for submodel M_k, $k = 2, 3$. By applying the projection method in (3.11), the semiparametric efficient score for model M is

$$\begin{aligned}\mathbf{S}_{\mathrm{e}} &= \mathbf{S}_{\boldsymbol{\Theta}} - \Pi(\mathbf{S}_{\boldsymbol{\Theta}}|\dot{M}_2 + \dot{M}_3) \\ &= \mathbf{S}_{\boldsymbol{\Theta}} - \Pi(\mathbf{S}_{\boldsymbol{\Theta}}|\dot{M}_3) - \Pi(\mathbf{S}_{\boldsymbol{\Theta}}|\Pi_{\dot{M}_3^{\perp}}\dot{M}_2),\end{aligned} \tag{3.12}$$

where $M^{\perp}$ is the perpendicular space of model M, $\Pi(\mathbf{S}_{\boldsymbol{\Theta}}|R)$ is the projection of score function $\mathbf{S}_{\boldsymbol{\Theta}}$ on space R and $\Pi_R T$ is the projection of space T on space R.

Let M_4 be the submodel of M with only $\phi_0(\cdot)$ known. Then M_1 and M_3 are two subspaces of M_4 corresponding to the finite dimensional part and infinite dimensional part, respectively. From (3.11) $\mathbf{S}_{\boldsymbol{\Theta}} - \Pi(\mathbf{S}_{\boldsymbol{\Theta}}|\dot{M}_3)$ is the efficient score for $\boldsymbol{\Theta}_0$ in model M_4. According to Lemma A4 in Huang et al. (2007),

$$\Pi(\mathbf{S}_{\boldsymbol{\Theta}}|\dot{M}_3^{\perp}) = \sum_{i=1}^{n}(\mathbf{X}_i, \underline{\mathbf{Z}}_i)^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\{\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}_0 - \phi_0(\mathbf{Z}_i\boldsymbol{\theta}_0)\}.$$

Similarly, by considering the parametric submodels of M_2, together with Lemma A4 in Huang et al. (2007), we have

$$\Pi_{\dot{M}_3^{\perp}}\dot{M}_2 = \Big\{\sum_{i=1}^{n}\kappa_n^{\mathrm{T}}(\mathbf{Z}_i\boldsymbol{\theta}_0)\boldsymbol{\Sigma}_i^{-1}\{\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}_0 - \phi_0(\mathbf{Z}_i\boldsymbol{\theta}_0)\},\ \kappa_n(\cdot) \in L_2\Big\}.$$

Projecting the score function $\mathbf{S}_{\boldsymbol{\Theta}}$ on space $\Pi_{\dot{M}_3^\perp}\dot{M}_2$ and putting the above two equations back to (3.12), we have

$$\mathbf{S}_{\rm e} = \sum_{i=1}^{n}(\widetilde{\mathbf{X}}_{i,\rm e}, \widetilde{\mathbf{Z}}_{i,\rm e})^{\rm T}\boldsymbol{\Sigma}_i^{-1}\{\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}_0 - \phi_0(\mathbf{Z}_i\boldsymbol{\theta}_0)\},$$

where $\widetilde{\mathbf{X}}_{i,\rm e} = \mathbf{X}_i - \psi_{n,\boldsymbol{\beta}}(\mathbf{Z}_i\boldsymbol{\theta}_0)$, $\widetilde{\mathbf{Z}}_{i,\rm e} = \underline{\mathbf{Z}}_i - \psi_{n,\boldsymbol{\theta}}(\mathbf{Z}_i\boldsymbol{\theta}_0)$ for some unique $\psi_{n,\boldsymbol{\beta}}(\cdot) \in L_2$ and $\psi_{n,\boldsymbol{\theta}}(\cdot) \in L_2$ by (3.11).

According to (3.11), $\mathbf{S}_{\rm e}$ is orthogonal to any member in $\dot{M}_2 + \dot{M}_3$. Besides, $\Pi_{\dot{M}_3^\perp}\dot{M}_2$ is a subset of $\dot{M}_2+\dot{M}_3$. Therefore, (3.11) implies Equations (3.7) and (3.8) for all $\kappa_n(\cdot) \in L_2$. The existence and uniqueness of the semiparametric score function $\mathbf{S}_{\rm e}$ for model M are guaranteed by (3.11). Therefore, we have the semiparametric efficient score function as shown in Equation (3.9). Finally, we have the semiparametric information bound $\mathbf{U}_{\rm e}$ given in Equation (3.10), completing the proof of Theorem 3.3.1. □

3.3.3 Theory for the Single-Index Function Estimator

To establish the asymptotic distribution theory for the single-index function estimator $\widehat{\phi}(u)$, we first define the following notation. Let $f_{ij}(\cdot)$ be the density of $\mathbf{Z}_{ij}^{\rm T}\boldsymbol{\theta}$ and $c_{i,j} = \int u^i K^j(u)du$. Of course, as a special case we may assume a common density function, i.e., $f_{ij}(\cdot) = f(\cdot)$. Furthermore, define

$$Q_1(u) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} v_i^{jj} f_{ij}(u), \quad Q_2(u) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} \xi_{i,jj} f_{ij}(u),$$
$$Q_3(u) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} \sigma_i^{jj} f_{ij}(u), \quad Q_4(u) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} \sigma_{i,jj}^{-1} f_{ij}(u),$$

where v_i^{jj}, $\xi_{i,jj}$, σ_i^{jj} and $\sigma_{i,jj}$ are the $(j,j)^{th}$ element of $\mathbf{V}_i^{-1}$, $\mathbf{V}_i^{-1}\boldsymbol{\Sigma}_i\mathbf{V}_i^{-1}$, $\boldsymbol{\Sigma}_i^{-1}$ and $\boldsymbol{\Sigma}_i$ respectively. These Q functions are simply the limits of some weighted averages. In the special case when $\mathbf{V}_i = \boldsymbol{\Sigma}_i$ for all i, $Q_1(u) = Q_2(u) = Q_3(u)$.

We now present the following theorem whose proof is given in Proof 3.3.3.

Theorem 3.3.2. *If Assumptions 3.3.1–3.3.6 hold, we have*

$$\sqrt{nh}\{\widehat{\phi}(u) - \phi_0(u) - c_{2,1}b(u)h^2\} \xrightarrow{D} N(0, \sigma_\xi^2(u)), \tag{3.13}$$

where $b(u)$ *satisfies*

$$b(u) + \int b(w)\eta(u,w)dw = \frac{1}{2}\phi_0^{(2)}(u)$$

with

$$\eta(u,w) = \frac{\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k\neq j} v_i^{jk} f_{ijk}(u,w)}{\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^{m_i} v_i^{jj} f_{ij}(w)}$$

and $\sigma_\xi^2(u) = c_{0,2}Q_2(u)/Q_1^2(u)$.

Proof. First denote $\mu_{ij} = \mathrm{E}(Y_{ij}) = \mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0 + \phi_0(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0)$. From Equation (3.2), we have

$$\frac{1}{n}\sum_{i=1}^n \sum_{j=1}^{m_i} K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)\Big[\begin{pmatrix} 1 \\ \mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u \end{pmatrix}\{Y_{ij} - \mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} - \widetilde{\phi}(u) - \widetilde{\phi}^{(1)}(u)(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)\}\Big] = 0.$$

It is equivalent to

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^n \sum_{j=1}^{m_i} K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)\Big\{\begin{pmatrix} 1 \\ \mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u \end{pmatrix}[Y_{ij} - \mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} - \phi_0(u) - \phi_0^{(1)}(u)(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u) \\ &- \{\widetilde{\phi}(u) - \phi_0(u)\} - \{\widetilde{\phi}^{(1)}(u) - \phi_0^{(1)}(u)\}(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)]\Big\} = 0. \end{aligned}$$

So we have the following exact form:

$$\widetilde{\phi}(u) - \phi_0(u) = \begin{bmatrix} 1 & 0 \end{bmatrix} Q^{-1}\boldsymbol{\delta} = \frac{q_{22}\delta_1 - q_{12}\delta_2}{q_{11}q_{22} - q_{12}^2},$$

where

$$Q = \begin{pmatrix} q_{11} & q_{12} \\ q_{12} & q_{22} \end{pmatrix} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u) \begin{pmatrix} 1 & \mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u \\ \mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u & (\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)^2 \end{pmatrix},$$

$$\begin{aligned}\boldsymbol{\delta} =& \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)\begin{pmatrix} 1 \\ \mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u \end{pmatrix}\Big\{Y_{ij} - \mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} - \phi_0(u) \\ & - \phi_0^{(1)}(u)(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)\Big\}.\end{aligned}$$

Since

$$\begin{aligned}\phi_0(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}) =& \phi_0(u) + \phi_0^{(1)}(u)(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u) + \frac{1}{2}\phi_0^{(2)}(u)(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)^2 \\ & + \frac{1}{6}\phi_0^{(3)}(u^*)(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)^3,\end{aligned}$$

where $u^* \in \{\min(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}, u), \max(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}, u)\}$,

$$\begin{aligned}\mathrm{E}\{K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)\} &= \frac{1}{h}\int K(\frac{x-u}{h})f_{ij}(x)dx \\ &= \int K(y)f_{ij}(hy+u)dy \\ &= \int K(y)\Big\{f_{ij}(u) + hyf_{ij}^{(1)}(u) + \frac{1}{2}h^2y^2f_{ij}^{(2)}(u) + o(h^2)\Big\}dy \\ &= f_{ij}(u) + \frac{1}{2}c_{2,1}h^2f_{ij}^{(2)}(u) + o(h^2),\end{aligned}$$

$$\mathrm{E}\{K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)\} = c_{2,1}h^2f_{ij}^{(1)}(u) + o(h^2),$$

$$\mathrm{E}\{K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)^2\} = c_{2,1}h^2f_{ij}(u) + o(h^2),$$

and

$$\mathrm{E}\{K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}-u)(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}-u)^3\}=o(h^2),$$

we have

$$\begin{aligned}
q_{11} &= c_1 + O_p\{h^2+(nh)^{-1/2}\},\\
q_{12} &= O_p\{h^2+(nh)^{-1/2}\},\\
q_{22} &= O_p\{h^2+(nh)^{-1/2}\},\\
\delta_1 &= O_p\{h^2+(nh)^{-1/2}\},\\
\delta_2 &= o_p(\delta_1) = o_p\{h^2+(nh)^{-1/2}\},
\end{aligned}$$

where $c_1 > 0$ is a constant. Therefore,

$$\begin{aligned}
\widetilde{\phi}(u)-\phi_0(u) &= \frac{q_{22}\delta_1 - q_{12}\delta_2}{q_{11}q_{22}-q_{12}^2} = q_{11}^{-1}\delta_1 + o_p\{h^2+(nh)^{-1/2}\}\\
&= \{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} f_{ij}(u)\}^{-1}\delta_1 + o_p\{h^2+(nh)^{-1/2}\}
\end{aligned}$$

since

$$q_{11}^{-1} = \{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} f_{ij}(u)\}^{-1}\Big[1+O_p\{h^2+(nh)^{-1/2}\}\Big].$$

By calculating the mean and variance of $\widetilde{\phi}(u)-\phi_0(u)$, when the parameters are evaluated at true values, we have

$$\widetilde{\phi}(u,\boldsymbol{\Theta}_0)-\phi_0(u) = \{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} f_{ij}(u)\}^{-1}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0-u)(Y_{ij}-\mu_{ij})$$

$$+\frac{1}{2}c_{2,1}b^0(u)h^2+o_p\{h^2+(nh)^{-1/2}\}. \tag{3.14}$$

Now we derive the asymptotic properties of $\widehat{\phi}(\cdot)$. To simplify the notation, we denote

$$\begin{aligned}
a_n(u)&=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}v_i^{jj}f_{ij}(u),\\
J_n(u_1,u_2)&=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k\neq j}v_i^{jk}\mathrm{E}\{a_n^{-1}(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0)\}f_{ijk}(u_1,u_2),\\
b_n^r(u)&=b^0(u)-a_n^{-1}(u)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}f_{ij}(u)\sum_{k\neq j}v_i^{jk}b_n^{r-1}(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0),
\end{aligned}$$

where $b^0(u)=b_n^0(u)=\phi_0^{(2)}(u)$ and f_{ijk} is the joint density of $(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta},\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta})$. For the first step updated estimating equation from Equation (3.4), we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}-u)\Big[\begin{pmatrix}1\\ \mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}-u\end{pmatrix}v_i^{jj}\{Y_{ij}-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}-\widehat{\phi}(u)-\widehat{\phi}^{(1)}(u)(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}-u)\}\\
&\qquad+\sum_{k\neq j}\begin{pmatrix}1\\ \mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}-u\end{pmatrix}v_i^{jk}\{Y_{ik}-\mathbf{X}_{ik}^{\mathrm{T}}\boldsymbol{\beta}-\widetilde{\phi}(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta})\}\Big]=0.
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\widehat{\phi}^1(u)-\phi_0(u)=\begin{bmatrix}1 & 0\end{bmatrix}R^{-1}\boldsymbol{\tau}&=\frac{r_{22}\tau_1-r_{12}\tau_2}{r_{11}r_{22}-r_{12}^2}=r_{11}^{-1}\tau_1+o_p\{h^2+(nh)^{-1/2}\}\\
&=a_n^{-1}(u)\tau_1+o_p\{h^2+(nh)^{-1/2}\},
\end{aligned}$$

where

$$R=\begin{pmatrix}r_{11} & r_{12}\\ r_{12} & r_{22}\end{pmatrix}=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}-u)v_i^{jj}\begin{pmatrix}1 & \mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}-u\\ \mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}-u & (\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}-u)^2\end{pmatrix},$$

and

$$\boldsymbol{\tau} = \begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)\Big[\begin{pmatrix} 1 \\ \mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u \end{pmatrix} v_i^{jj}\{Y_{ij} - \mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} - \phi_0(u) - \phi_0^{(1)}(u)(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)\} + \sum_{k\neq j}\begin{pmatrix} 1 \\ \mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u \end{pmatrix} v_i^{jk}\{Y_{jk} - \mathbf{X}_{ik}^{\mathrm{T}}\boldsymbol{\beta} - \widetilde{\phi}(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta})\}\Big].$$

Similarly to (3.14), the one-step update of $\widetilde{\phi}$, defined as $\widehat{\phi}^1$, has the following asymptotic expansion

$$\widehat{\phi}^1(u, \boldsymbol{\Theta}_0) - \phi_0(u) = a_n^{-1}(u)(B_{1n} + B_{2n} + B_{3n}) + o_p\{h^2 + (nh)^{-1/2}\}, \qquad (3.15)$$

where

$$B_{1n} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0 - u) v_i^{jj}\Big\{Y_{ij} - \mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0 - \phi_0(u) - \phi_0^{(1)}(u)(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0 - u)\Big\},$$
$$B_{2n} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0 - u)\Big\{\sum_{k\neq j} v_i^{jk}(Y_{ik} - \mu_{ik})\Big\},$$
$$B_{3n} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0 - u)\Big[\sum_{k\neq j} v_i^{jk}\{\widetilde{\phi}(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0) - \phi_0(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0)\}\Big].$$

By plugging (3.14) into B_{3n}, we have

$$\widehat{\phi}^1(u, \boldsymbol{\Theta}_0) - \phi_0(u) = D_{1n}(u) + D_{2n}(u) + \frac{1}{2}c_{2,1}b_n^1(u)h^2 + o_p\{h^2 + (nh)^{-1/2}\}, \quad (3.16)$$

where

$$D_{1n}(u) = a_n^{-1}(u)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0 - u)\{\sum_{k=1}^{m_i} v_i^{jk}(Y_{ik} - \mu_{ik})\},$$
$$D_{2n}(u) = -a_n^{-1}(u)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} v_i^{jj} J_n(u, \mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0)(Y_{ij} - \mu_{ij}).$$

For the second iteration step, we have an equation similar to (3.15) except that $\widehat{\phi}(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0)-\phi_0(u)$ in B_{3n} is replaced by $\widehat{\phi}^1(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0)-\phi_0(u)$, which can be obtained from (3.16). Thus we have

$$\widehat{\phi}^2(u,\boldsymbol{\Theta}_0)-\phi_0(u)=a_n^{-1}(u)(B_{1n}+B_{2n}+B_{4n})+o_p\{h^2+(nh)^{-1/2}\},$$

where

$$\begin{aligned}B_{4n}=&-\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0-u)\Big(\sum_{k\neq j}v_i^{jk}[D_{1n}(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0)+D_{2n}(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0)\\&+\frac{1}{2}c_{2,1}b_n^1(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0)h^2]\Big).\end{aligned}$$

The bias term is then

$$\frac{1}{2}c_{2,1}h^2\Big\{b^0(u)-a_n^{-1}(u)\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k\neq j}v_i^{jk}b_n^1(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0)f_{ij}(u)\Big\}=\frac{1}{2}c_{2,1}b_n^2(u).$$

Define

$$H_n(J_n;u_1,u_2)=-\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k\neq j}v_i^{jk}\mathrm{E}\{a_n^{-1}(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0)J_n(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0,u_2)\}f_{ij}(u_1).$$

Then we have

$$\begin{aligned}&a_n^{-1}(u)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0-u)\sum_{k\neq j}v_i^{jk}D_{1n}(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0)\\&=a_n^{-1}(u)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0-u)\{\sum_{k\neq j}v_i^{jk}a_n^{-1}(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0)\\&\quad\frac{1}{n}\sum_{i'=1}^{n}\sum_{j'=1}^{m_i}K_h(\mathbf{Z}_{i'j'}^{\mathrm{T}}\boldsymbol{\theta}_0-\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0)\sum_{k'=1}^{m_i}v_i^{j'k'}(Y_{i'k'}-\mu_{i'k'})\}\end{aligned}$$

$$= a_n^{-1}(u)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} J_n(u, \mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0)\sum_{k=1}^{m_i} v_i^{jk}(Y_{ik}-\mu_{ik}) + o_p\{(nh)^{-1/2}\},$$

$$\begin{aligned}
&a_n^{-1}(u)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0 - u)\sum_{k\neq j} v_i^{jk} D_{2n}(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0)\\
&\quad = -\, a_n^{-1}(u)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0 - u)\sum_{k\neq j} v_i^{jk}\{a_n^{-1}(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0)\\
&\qquad \frac{1}{n}\sum_{i'=1}^{n}\sum_{j'=1}^{m_i} v_{i'}^{j'j'} J_n(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0, \mathbf{Z}_{i'j'}^{\mathrm{T}}\boldsymbol{\theta}_0)(Y_{i'j'}-\mu_{i'j'})\}\\
&\quad = a_n^{-1}(u)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} v_i^{jj} H_n(J_n; u, \mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0)(Y_{ij}-\mu_{ij}) + o_p\{(nh)^{-1/2}\}.
\end{aligned}$$

Similarly to the second iteration step, for the r^{th} $(r \geq 2)$ iteration step, we have

$$\widehat{\phi}^r(u, \boldsymbol{\Theta}_0) - \phi_0(u) = D_{1n} + E_{1n}^r + E_{2n}^r + \frac{1}{2}c_{2,1}b_n^r(u)h^2 + o_p\{h^2 + (nh)^{-1/2}\}, \quad (3.17)$$

where

$$\begin{aligned}
E_{1n}^r &= a_n^{-1}(u)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} J_{n,1}^r(u, \mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0)\{\sum_{k=1}^{m_i} v_i^{jk}(Y_{ik}-\mu_{ik})\},\\
E_{2n}^r &= a_n^{-1}(u)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} v_i^{jj} J_{n,2}^r(u, \mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0)(Y_{ij}-\mu_{ij}).
\end{aligned}$$

Here $J_{n,1}^1(u_1,u_2) = 0$, $J_{n,1}^r(u_1,u_2) = -J_n(u_1,u_2) + H_n(J_{n,1}^{r-1}; u_1, u_2)$, $J_{n,2}^1(u_1,u_2) = -J_n(u_1,u_2)$, and $J_{n,2}^r(u_1,u_2) = H_n(J_{n,2}^{r-1}; u_1, u_2)$. At convergence, $\widehat{\phi}(u) - \phi_0(u)$ has a form similar to Equation (3.17) except that b_n^r, $J_{n,1}^r$, $J_{n,2}^r$ are respectively replaced by their limits $b_n(u)$, $J_{n,1}$ and $J_{n,2}$ which satisfy the following equations:

$$
\begin{aligned}
&b_n(u) = \phi_0^{(2)}(u) - a_n^{-1}(u)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k\neq j} v_i^{jk}\mathrm{E}\{b_n(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0)\}f_{ij}(u),\\
&J_{n,1}(u_1,u_2) = -J_n(u_1,u_2) + H_n(J_{n,1};u_1,u_2),\\
&J_{n,2}(u_1,u_2) = H_n(J_{n,2};u_1,u_2).
\end{aligned}
\tag{3.18}
$$

Since $\mathrm{E}(E_{1n}^r) = \mathrm{E}(E_{2n}^r) = 0$ and the variances of E_{1n}^r and E_{2n}^r are of order $O(n^{-1}) = o\{(nh)^{-1}\}$, we have $E_{1n}^r = E_{2n}^r = o_p\{(nh)^{-1/2}\}$. Thus, Equation (3.17) can be simplified as

$$
\widehat{\phi}^r(u,\boldsymbol{\Theta}_0) - \phi_0(u) = D_{1n} + \frac{1}{2}c_{2,1}b_n^r(u)h^2 + o_p\{h^2 + (nh)^{-1/2}\}.
$$

Therefore, we have

$$
\begin{aligned}
&\widehat{\phi}(u) - \phi_0(u)\\
&= a_n^{-1}(u)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0 - u)\Big\{\sum_{k=1}^{m_i} v_i^{jk}(Y_{ik} - \mu_{ik})\Big\}\\
&\quad + c_{2,1}\Big[\frac{h^2\phi_0^{(2)}(u)}{2} - h^2 a_n^{-1}(u)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k\neq j} v_i^{jk}\mathrm{E}\{b(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0)\}f_{ij}(u)\Big]\\
&\quad + o_p\{h^2 + (nh)^{-1/2}\},
\end{aligned}
$$

where $b(\cdot)$ is defined in Theorem 3.3.2. Some standard calculations show that the asymptotic bias for $\widehat{\phi}(u)$ is $c_{2,1}b(u)h^2$, where $b(u)$ satisfies

$$
b(u) + \int b(w)\eta(u,w)dw = \frac{1}{2}\phi_0^{(2)}(u)
$$

with $\eta(u,w)$ given in Theorem 3.3.2. The asymptotic variance of $\sqrt{nh}\{\widehat{\phi}(u) - \phi_0(u)\}$ is readily seen to be $\sigma_\xi^2(u) = c_{0,2}Q_2(u)/Q_1^2(u)$. In fact, the one-step update $\widehat{\phi}^1(u)$ also has the same asymptotic variance of $\sigma_\xi^2(u)$. That is, further iteration steps beyond the first update do not change the asymptotic variance.

It has been shown in Theorem 3.3.3 that the asymptotic variance of $\widehat{\phi}(u)$ is minimized when $\mathbf{V}_i = \mathbf{\Sigma}_i$. Therefore, when the covariance matrices are correctly specified, the asymptotic variance of $\widehat{\phi}(u)$ is $\sigma_\phi^2(u)$ as shown in Theorem 3.3.3. □

In the following we show that the asymptotic variance of $\widehat{\phi}(u)$ is minimized if and only if $\mathbf{V}_i = \mathbf{\Sigma}_i$, $i = 1, \ldots, n$. Define $\mathbf{F}_i^{1/2}(u)$ as a $m_i \times m_i$ diagonal matrix with the $(j,j)^{th}$ element being $f_{ij}^{1/2}(u)$ and $\mathbf{F}_i^{-1/2}(u)$ as the inverse matrix of $\mathbf{F}_i^{1/2}(u)$. Further define $\mathbf{G}_i^{-1}(u) = \mathbf{F}_i^{1/2}(u)\mathbf{V}_i^{-1}\mathbf{F}_i^{1/2}(u)$ and $\mathbf{H}_i(u) = \mathbf{F}_i^{-1/2}(u)\mathbf{\Sigma}_i\mathbf{F}_i^{-1/2}(u)$. Then the problem of minimizing $\sigma_\xi^2(u)$ is asymptotically equivalent to minimizing

$$\frac{\sum_{i=1}^n \text{tr}\{\mathbf{G}_i^{-1}(u)\mathbf{H}_i(u)\mathbf{G}_i^{-1}(u)\}}{[\sum_{i=1}^n \text{tr}\{\mathbf{G}_i^{-1}(u)\}]^2},$$

where tr($\mathbf{A}$) is the trace of a matrix $\mathbf{A}$. The extended Cauchy-Schwarz inequality (Magnus and Neudecker (1995), p. 227, Theorem 2) shows that

$$\{\text{tr}(\mathbf{A}^\text{T}\mathbf{B})\}^2 \le \{\text{tr}(\mathbf{A}^\text{T}\mathbf{A})\}\{\text{tr}(\mathbf{B}^\text{T}\mathbf{B})\}$$

for any square matrices $\mathbf{A}$ and $\mathbf{B}$ with equality if and only if $\mathbf{A} = c\mathbf{B}$ for some constant c. Now let $\mathbf{A}$ and $\mathbf{B}$ be two $N \times N$ block diagonal matrices

$$\mathbf{A} = \text{diag}(\mathbf{A}_1, \ldots, \mathbf{A}_n) \quad \text{and} \quad \mathbf{B} = \text{diag}(\mathbf{B}_1, \ldots, \mathbf{B}_n),$$

where $\mathbf{A}_i = \mathbf{H}_i^{-1/2}(u)$ and $\mathbf{B}_i = \mathbf{H}_i^{1/2}(u)\mathbf{G}_i^{-1}(u)$ for $i = 1, \ldots, n$. Then we have

$$[\sum_{i=1}^{n} \text{tr}\{\mathbf{G}_i^{-1}(u)\}]^2 \leq \sum_{i=1}^{n} \text{tr}\{\mathbf{H}_i^{-1}(u)\} \sum_{i=1}^{n} \text{tr}\{\mathbf{G}_i^{-1}(u)\mathbf{H}_i(u)\mathbf{G}_i^{-1}(u)\}.$$

It is equivalent to

$$\frac{\sum_{i=1}^{n} \text{tr}\{\mathbf{G}_i^{-1}(u)\mathbf{H}_i(u)\mathbf{G}_i^{-1}(u)\}}{[\sum_{i=1}^{n} \text{tr}\{\mathbf{G}_i^{-1}(u)\}]^2} \geq \frac{1}{\sum_{i=1}^{n} \text{tr}\{\mathbf{H}_i^{-1}(u)\}}.$$

Without loss of generality, let $c = 1$. Then the equality holds if and only if $\mathbf{A} = \mathbf{B}$. It leads to $\mathbf{H}_i(u) = \mathbf{G}_i(u)$, which is equivalent to $\mathbf{V}_i = \boldsymbol{\Sigma}_i$ for $i = 1, \ldots, n$. The result is formally presented in the following theorem.

Theorem 3.3.3. *If Assumptions 3.3.1–3.3.6 in Section 3.3.1 hold, the asymptotic variance of $\widehat{\phi}(u)$ is minimized if and only if $\mathbf{V}_i = \boldsymbol{\Sigma}_i$ i.e., when the working covariance matrices are correctly specified. In this case, the estimated single-index function $\widehat{\phi}(u)$ has the following asymptotic normality property*

$$\sqrt{nh}\{\widehat{\phi}(u) - \phi_0(u) - c_{2,1}b(u)h^2\} \xrightarrow{D} N(0, \sigma_\phi^2(u)), \tag{3.19}$$

where $\sigma_\phi^2(u) = c_{0,2}Q_3^{-1}(u)$.

Our method extends SGEE of Chen et al. (2015) by updating the estimation for the unknown single-index function. To see the advantages of the proposed single-index function estimator, we compare the asymptotic variances of the single-index function estimators. When the working covariance matrices are correctly specified in SGEE, the asymptotic variance of unknown single-index function estimator is (see Chen et al. (2015))

$$\text{var}(\widetilde{\phi}(u)) = \frac{1}{nh}c_{0,2}Q_4^{-1}(u). \tag{3.20}$$

Comparing (3.19) and (3.20), the SGEE estimator and the proposed estimator of the single-index function share the same order of pointwise convergence rate of $O(\sqrt{nh})$, but the asymptotic variances are different. To show that the proposed estimator is more efficient, it is sufficient to show $\sigma_i^{jj} \geq 1/\sigma_{i,jj}$ for all i, j. We consider $j = 1$ without loss of generality. Writing $\mathbf{\Sigma}_i$ in the block matrix form, we have

$$\mathbf{\Sigma}_i = \begin{pmatrix} \sigma_{i,11} & \Sigma_{i,12} \\ \Sigma_{i,12}^{\mathrm{T}} & \Sigma_{i,22} \end{pmatrix} \geq 0.$$

By the extended Cauchy-Schwartz inequality,

$$\sigma_i^{11} = (\sigma_{i,11} - \Sigma_{i,12}\Sigma_{i,22}^{-1}\Sigma_{i,12}^{\mathrm{T}})^{-1} \geq (\sigma_{i,11})^{-1}.$$

Therefore, when the working covariance matrices are correctly specified, the proposed single-index function estimator has the asymptotic variance less than or equal to the asymptotic variance of SGEE estimator. Therefore, it is in general more efficient. The result is formally presented in the following theorem.

Theorem 3.3.4. *If Assumptions 3.3.1–3.3.6 hold, when the working covariance matrices are correctly specified, the asymptotic variance of the proposed estimator $\widehat{\phi}(u)$ has a variance less than or equal to that of the SGEE estimator for the single-index function.*

Remark 3.3.1. While here we have used the same bandwidth for easy comparisons of different methods, in order to obtain the optimal numerical performance for each method, our limited empirical experience suggests that it seems slightly helpful to select a different bandwidth in each estimation step by cross validation.

3.3.4 Theory for the Parameter Estimators

Now we investigate the asymptotic distribution of the estimators $\widehat{\boldsymbol{\Theta}}$ and prove that they reach the semiparametric information bound. Define

$$\begin{aligned}\varphi_{n,\boldsymbol{\beta}}(u, \widehat{\boldsymbol{\Theta}}) &= -\frac{\partial\widehat{\phi}(u, \boldsymbol{\Theta})}{\partial\boldsymbol{\beta}^{\mathrm{T}}}|_{\boldsymbol{\Theta}=\widehat{\boldsymbol{\Theta}}} \to \varphi_{\boldsymbol{\beta}}(u),\\ \varphi_{n,\boldsymbol{\theta}}(u, \widehat{\boldsymbol{\Theta}}) &= -\frac{\partial\widehat{\phi}(u, \boldsymbol{\Theta})}{\partial\boldsymbol{\theta}^{\mathrm{T}}}|_{\boldsymbol{\Theta}=\widehat{\boldsymbol{\Theta}}} \to \varphi_{\boldsymbol{\theta}}(u)\end{aligned} \tag{3.21}$$

in probability as $n \to \infty$. Moreover, by Assumption 3.3.3 on sufficient smoothness of $\phi(\cdot)$, it is readily seen that the convergence is uniform over u in the compact domain $\mathcal{G}$. Further define

$$\begin{aligned}\boldsymbol{\Omega}_0 &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\Big\{(\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{Z}}_i)^{\mathrm{T}}\mathbf{V}_i^{-1}(\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{Z}}_i)\Big\},\\ \boldsymbol{\Omega}_1 &= \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\Big\{(\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{Z}}_i)^{\mathrm{T}}\mathbf{V}_i^{-1}\boldsymbol{\Sigma}_i\mathbf{V}_i^{-1}(\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{Z}}_i)\Big\},\end{aligned} \tag{3.22}$$

where $(\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{Z}}_i) = \{\mathbf{X}_i - \varphi_{\boldsymbol{\beta}}(\mathbf{Z}_i\boldsymbol{\theta}_0), \underline{\mathbf{Z}}_i - \varphi_{\boldsymbol{\theta}}(\mathbf{Z}_i\boldsymbol{\theta}_0)\}$. Assume that $\boldsymbol{\Omega}_0$ and $\boldsymbol{\Omega}_1$ are non-negative definite matrices. Then we have the following theorem a proof of which is given in Proof 3.3.4.

Theorem 3.3.5. *If Assumptions 3.3.1–3.3.6 hold, we have*

$$n^{1/2}(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Omega}_0^{-1}\boldsymbol{\Omega}_1\boldsymbol{\Omega}_0^{-1}) \tag{3.23}$$

where $\xrightarrow{D}$ *denotes convergence in distribution.*

Proof. To derive the asymptotic properties for the parameter estimators, the uniform consistency of the single-index function estimator is required. However, the only difference between the uniform consistency and pointwise consistency of the single-index function estimator is a different order rate. It changes from $o_p\{h^2+(nh)^{-1/2}\}$ to $o_p\{h^2+[\log(n)/nh]^{1/2}\}$. It does not affect the asymptotic results for the parameter estimators. Now denote $\boldsymbol{\Delta}_i\{\boldsymbol{\beta},\widehat{\phi}(\mathbf{Z}_i\boldsymbol{\theta},\boldsymbol{\Theta})\}^{\mathrm{T}}=\partial\{\mathbf{X}_i\boldsymbol{\beta}+\widehat{\phi}(\mathbf{Z}_i\boldsymbol{\theta},\boldsymbol{\Theta})\}^{\mathrm{T}}/\partial\boldsymbol{\Theta}$. From the GEE in Equation (3.5), the parameter estimates satisfy the equation

$$n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\Delta}_i\{\widehat{\boldsymbol{\beta}},\widehat{\phi}(\mathbf{Z}_i\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\Theta}})\}^{\mathrm{T}}\mathbf{V}_i^{-1}\Big\{\mathbf{Y}_i-\mathbf{X}_i\widehat{\boldsymbol{\beta}}-\widehat{\phi}(\mathbf{Z}_i\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\Theta}})\Big\}=0.$$

By expending the equation above at $\boldsymbol{\Theta}=\boldsymbol{\Theta}_0$ and after some further algebra, we have the following equation

$$\begin{aligned}n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\Delta}_i\{\boldsymbol{\beta}_0,\widehat{\phi}(\mathbf{Z}_i\boldsymbol{\theta}_0)\}^{\mathrm{T}}\mathbf{V}_i^{-1}\Big[&\mathbf{Y}_i-\mathbf{X}_i\boldsymbol{\beta}_0-\widehat{\phi}(\mathbf{Z}_i\boldsymbol{\theta}_0)\\ &-\boldsymbol{\Delta}_i\big\{\boldsymbol{\beta}_0,\widehat{\phi}(\mathbf{Z}_i\boldsymbol{\theta}_0)\big\}(\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}_0)\Big]+o_p(1)=0,\end{aligned}$$

where we recall that $\widehat{\phi}(\mathbf{Z}_i\boldsymbol{\theta}_0)=\widehat{\phi}(\mathbf{Z}_i\boldsymbol{\theta}_0,\widehat{\boldsymbol{\Theta}})$. Thus we have

$$\begin{aligned}&n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\Delta}_i\big\{\boldsymbol{\beta}_0,\widehat{\phi}(\mathbf{Z}_i\boldsymbol{\theta}_0)\big\}^{\mathrm{T}}\mathbf{V}_i^{-1}\boldsymbol{\Delta}_i\big\{\boldsymbol{\beta}_0,\widehat{\phi}(\mathbf{Z}_i\boldsymbol{\theta}_0)\big\}(\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}_0)\\ &\quad=n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\Delta}_i\{\boldsymbol{\beta}_0,\widehat{\phi}(\mathbf{Z}_i\boldsymbol{\theta}_0)\}^{\mathrm{T}}\mathbf{V}_i^{-1}\Big[\mathbf{Y}_i-\mathbf{X}_i\boldsymbol{\beta}_0-\phi_0(\mathbf{Z}_i\boldsymbol{\theta}_0)\\ &\qquad-\big\{\widehat{\phi}(\mathbf{Z}_i\boldsymbol{\theta}_0)-\phi_0(\mathbf{Z}_i\boldsymbol{\theta}_0)\big\}\Big]+o_p(1).\end{aligned}$$

Then it is readily seen that

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}(\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{Z}}_i)^{\mathrm{T}}\mathbf{V}_i^{-1}(\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{Z}}_i)\big\{n^{1/2}(\widehat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}_0)\big\}\\ &=n^{-1/2}\sum_{i=1}^{n}(\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{Z}}_i)^{\mathrm{T}}\mathbf{V}_i^{-1}[\mathbf{Y}_i-\mathbf{X}_i\boldsymbol{\beta}_0-\phi_0(\mathbf{Z}_i\boldsymbol{\theta}_0)\\ &\quad-\{\widehat{\phi}(\mathbf{Z}_i\boldsymbol{\theta}_0)-\phi_0(\mathbf{Z}_i\boldsymbol{\theta}_0)\}]+o_p(1).\end{aligned}\tag{3.24}$$

By Equations (3.17) and (3.18) in the proof of Theorem 3.3.2 and a second-order bias expansion, we have

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}(\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{Z}}_i)^{\mathrm{T}}\mathbf{V}_i^{-1}\{\widehat{\phi}(\mathbf{Z}_i\boldsymbol{\theta}_0)-\phi_0(\mathbf{Z}_i\boldsymbol{\theta}_0)\} \\
=& n^{-1/2}\frac{h^2}{2}\Big[\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}v_i^{jk}(\widetilde{\mathbf{X}}_{ij},\widetilde{\mathbf{Z}}_{ij})^{\mathrm{T}}\big\{b(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0)+hb_1(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0)+O_p(h^2)\big\}\Big] \\
&+\Bigg(n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}v_i^{jk}(\widetilde{\mathbf{X}}_{ij},\widetilde{\mathbf{Z}}_{ij})^{\mathrm{T}}\Big[a_n^{-1}(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0)\frac{1}{n}\sum_{i'=1}^{n}\sum_{j'=1}^{m}v_{i'}^{j'j'} \\
&\times\Big\{K_h(\mathbf{Z}_{i'j'}^{\mathrm{T}}\boldsymbol{\theta}_0-\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0)\sum_{l=1}^{m_i}v_{i'}^{j'l}(Y_{i'l}-\mu_{i'l}) \\
&+J_{n,2}(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0,\mathbf{Z}_{i'j'}^{\mathrm{T}}\boldsymbol{\theta}_0)(Y_{i'j'}-\mu_{i'j'}) \\
&+J_{n,1}(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0,\mathbf{Z}_{i'j'}^{\mathrm{T}}\boldsymbol{\theta}_0)\sum_{l=1}^{m_i}v_{i'}^{j'l}(Y_{i'l}-\mu_{i'l})\Big\}\Big]\Bigg)+o_p(1) \\
=& T_1+T_2+o_p(1),
\end{aligned}
$$

where b_1 is the next higher-order bias expansion of $\widehat{\phi}$. For T_2, rewrite it as $T_2 = T_{21} + T_{22} + T_{23}$. We then have

$$
\begin{aligned}
T_{21} =& n^{-\frac{1}{2}}\sum_{i'=1}^{n}\sum_{j'=1}^{m_i}v_{i'}^{j'j'}\Big\{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}K_h(\mathbf{Z}_{i'j'}^{\mathrm{T}}\boldsymbol{\theta}_0-\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0)(\widetilde{\mathbf{X}}_{ij},\widetilde{\mathbf{Z}}_{ij})^{\mathrm{T}}v_i^{jk} \\
& a_n^{-1}(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0)\Big\}\sum_{l=1}^{m_i}v_{i'}^{j'l}(Y_{i'l}-\mu_{i'l}).
\end{aligned}
$$

We now note that it is easy to see that when working covariances $\mathbf{V}_i$ are used in place of $\boldsymbol{\Sigma}_i$, Equations (3.7) and (3.8) are asymptotically equivalent to the following equation:

$$
\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\big\{(\widetilde{\mathbf{X}}_i,\widetilde{\mathbf{Z}}_i)^{\mathrm{T}}\mathbf{V}_i^{-1}g_n(\mathbf{D}_i)\odot\mathbf{f}_i(\mathbf{D}_i)|\mathbf{D}_i=\mathbf{d}_i\big\}=\mathbf{0} \tag{3.25}
$$

for any function $g_n(\cdot) \in L_2$, where $\mathbf{D}_i = \mathbf{Z}_i\boldsymbol{\theta}$, $\mathbf{d}_i = (d_{11}, \ldots, d_{im_i})^{\mathrm{T}}$ and $\mathbf{f}_i(\mathbf{d}_i) = (f_{i1}(d_{i1}), \ldots, f_{im_i}(d_{im_i}))^{\mathrm{T}}$ for $d_{ij} \in \mathcal{G}$, $j = 1, \ldots, m_i$.

Similarly to the derivations in Section A.4 of Wang et al. (2005), we can obtain that $T_1 = o_p(1)$ if $nh^8 \to 0$. Moreover, the sample average term inside the braces in T_{21} is asymptotically equal to

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}\mathrm{E}\big\{(\widetilde{\mathbf{X}}_{ij}, \widetilde{\mathbf{Z}}_{ij})^{\mathrm{T}} v_i^{jk} a_n^{-1}(D_{ik})|D_{ik} = d\big\} f_{ik}(d)|_{d=\mathbf{Z}_{i'j'}^{\mathrm{T}}\boldsymbol{\theta}_0}$$

with $D_{ik} = \mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0$, which is 0 in probability by (3.25). Therefore, we obtain that $T_{21} = o_p(1)$. Using similar arguments, we can also show that T_{22} and T_{23} are of order $o_p(1)$.

As a result, with Assumptions 3.3.1–3.3.6, $T_1 + T_2 = o_p(1)$. It follows from Equation (3.24) that

$$n^{\frac{1}{2}}(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0) = \boldsymbol{\Omega}_0^{-1} n^{-\frac{1}{2}}\sum_{i=1}^{n}(\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{Z}}_i)^{\mathrm{T}}\mathbf{V}_i^{-1}\big\{\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}_0 - \phi_0(\mathbf{Z}_i\boldsymbol{\theta}_0)\big\} + o_p(1).$$

This directly leads to Theorem 3.3.5. $\square$

Furthermore, in the following we show that when the covariance matrices are correctly specified, the asymptotic covariance of the parameter estimators is minimized. That is, for all estimated working covariance $\mathbf{V}_i$, $\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Omega}_1\boldsymbol{\Omega}_0^{-1} - \boldsymbol{\Omega}_1^{-1}$ is a semi-positive definite matrix.

First denote $\mathbf{A} \geq 0$ when $\mathbf{A}$ is a semi-positive definite matrix. By the extended Cauchy-Schwarz inequality, if

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \geq 0,$$

then $\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \geq 0$. Further denote $\mathbf{U}_i = (\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{Z}}_i)$. Then the problem of minimizing $\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Omega}_1\boldsymbol{\Omega}_0^{-1}$ is asymptotically equivalent to minimizing

$$(\sum_{i=1}^{n}\mathbf{U}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{U}_i)^{-1}(\sum_{i=1}^{n}\mathbf{U}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\boldsymbol{\Sigma}_i\mathbf{V}_i^{-1}\mathbf{U}_i)(\sum_{i=1}^{n}\mathbf{U}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{U}_i)^{-1}.$$

Since

$$\begin{bmatrix} \sum_{i=1}^{n}\mathbf{U}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\boldsymbol{\Sigma}_i\mathbf{V}_i^{-1}\mathbf{U}_i & \sum_{i=1}^{n}\mathbf{U}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{U}_i \\ \sum_{i=1}^{n}\mathbf{U}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{U}_i & \sum_{i=1}^{n}\mathbf{U}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{U}_i \end{bmatrix} = \sum_{i=1}^{n}\mathbf{U}_i^{\mathrm{T}}\begin{pmatrix}\mathbf{V}_i^{-1}\\ \boldsymbol{\Sigma}_i^{-1}\end{pmatrix}\boldsymbol{\Sigma}_i(\mathbf{V}_i^{-1},\boldsymbol{\Sigma}_i^{-1})\mathbf{U}_i \geq 0,$$

we have

$$\sum_{i=1}^{n}\mathbf{U}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\boldsymbol{\Sigma}_i\mathbf{V}_i^{-1}\mathbf{U}_i - (\sum_{i=1}^{n}\mathbf{U}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{U}_i)(\sum_{i=1}^{n}\mathbf{U}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{U}_i)^{-1}(\sum_{i=1}^{n}\mathbf{U}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{U}_i) \geq 0,$$

which leads to

$$(\sum_{i=1}^{n}\mathbf{U}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{U}_i)^{-1}(\sum_{i=1}^{n}\mathbf{U}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\boldsymbol{\Sigma}_i\mathbf{V}_i^{-1}\mathbf{U}_i)(\sum_{i=1}^{n}\mathbf{U}_i^{\mathrm{T}}\mathbf{V}_i^{-1}\mathbf{U}_i)^{-1} - (\sum_{i=1}^{n}\mathbf{U}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\mathbf{U}_i)^{-1} \geq 0.$$

The inequality becomes equality if and only if $\mathbf{V}_i = c\boldsymbol{\Sigma}_i$ for some $c > 0$. Without loss of generality, we set $c = 1$ here and in the proof of Theorem 3.3.3.

Moreover, the asymptotic covariance of the parameter estimators reaches the semiparametric information bound when $\mathbf{V}_i = \boldsymbol{\Sigma}_i$. We formally include this result in the following theorem.

Theorem 3.3.6. *Under Assumptions 3.3.1–3.3.6, if the covariance matrices are correctly specified, then we have*

$$\psi_{n,\boldsymbol{\beta}}(u) \to \varphi_{\boldsymbol{\beta}}(u) \quad \text{and} \quad \psi_{n,\boldsymbol{\theta}}(u) \to \varphi_{\boldsymbol{\theta}}(u)$$

in probability as $n \to \infty$ for every $u \in \mathcal{G}$, where $\varphi_{\boldsymbol{\beta}}(u)$ and $\varphi_{\boldsymbol{\theta}}(u)$ are defined in Equation (3.21). Furthermore, the proposed parameter estimators reach the semiparametric information bound and are thus semiparametric efficient.

A proof of Theorem 3.3.6 is given in Proof 3.3.4 below. It is worth noting that the centering part of the asymptotic variance of the parameter estimators from Chen et al. (2015) is the conditional mean of $\mathbf{X}$ and $\mathbf{Z}$ given the single-index part. Therefore, the asymptotic properties in Theorem 3.3.6 generally do not hold for their parameter estimators. This implies that their parameter estimators are generally not semiparametric efficient.

Proof. Denote the l^{th} element of $\mathbf{X}_{ij}$ and $\underline{\mathbf{Z}}_{ij}$ by X_{ijl} and $\underline{Z}_{ijl}$, respectively. Similarly to Wang et al. (2005), by (3.7) we can obtain that $\psi_{n,\beta_l}(\cdot)$ solves the Fredholm integral equation of the second kind:

$$\psi_{n,\beta_l}(u_1) = r_{n,\beta_l}(u_1) - \int W_n(u_1, u_2)\psi_{n,\beta_l}(u_2)du_2, \tag{3.26}$$

where

$$W_n(u_1, u_2) = \frac{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k\neq j}\sigma_i^{jk} f_{ijk}(u_2, u_1)}{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sigma_i^{jj} f_{ij}(u_1)} \tag{3.27}$$

and

$$r_{n,\beta_l}(u) = \frac{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}\sigma_i^{jk}\mathrm{E}(X_{ikl}) f_{ij}(u)}{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sigma_i^{jj} f_{ij}(u)}$$

for $l = 1, \ldots, p$. Similarly, from Equation (3.8) we can obtain that $\psi_{n,\theta_l}(\cdot)$ solves the Fredholm integral equation of the second kind:

$$\psi_{n,\theta_l}(u_1) = r_{n,\theta_l}(u_1) - \int W_n(u_1, u_2)\psi_{n,\theta_l}(u_2)du_2, \tag{3.28}$$

where

$$r_{n,\theta_l}(u) = \frac{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}\sigma_i^{jk}\mathrm{E}(\underline{Z}_{ikl})f_{ij}(u)}{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sigma_i^{jj}f_{ij}(u)}$$

for $l = 1, \ldots, q$. If we take the limit of Equations (3.26) and (3.28) as $n \to \infty$ we see that $\psi_{\boldsymbol{\beta}}(u) = \lim_{n\to\infty}\psi_{n,\boldsymbol{\beta}}(u)$ and $\psi_{\boldsymbol{\theta}}(u) = \lim_{n\to\infty}\psi_{n,\boldsymbol{\theta}}(u)$ satisfy the Fredholm integral equations of the second kind in the corresponding limiting form, respectively.

In order to show the semiparametric efficiency of the proposed estimators, it is sufficient to show that $\varphi_{\boldsymbol{\beta}}(u)$ and $\varphi_{\boldsymbol{\theta}}(u)$ satisfy the Fredholm integral equations of the second kind as same as $\psi_{\boldsymbol{\beta}}(u)$ and $\psi_{\boldsymbol{\theta}}(u)$, respectively. By Equation (3.4), when $\mathbf{V}_i = \boldsymbol{\Sigma}_i$ for $i = 1, \ldots, n$, we have

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sigma_i^{jj}K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)\big\{Y_{ij} - \mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} - \widehat{\phi}(u, \boldsymbol{\Theta}) - \widehat{\phi}^{(1)}(u, \boldsymbol{\Theta})(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)\big\}$$
$$+ \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k\neq j}\sigma_i^{jk}K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta} - u)\big\{Y_{ik} - \mathbf{X}_{ik}^{\mathrm{T}}\boldsymbol{\beta} - \widehat{\phi}(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}, \boldsymbol{\Theta})\big\} = 0. \quad (3.29)$$

Taking derivative with respect to $\boldsymbol{\theta}$ on both sides and evaluating at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, we get

$$L_1 + L_2 + L_3 + L_4 = 0,$$

where

$$L_1 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sigma_i^{jj}\frac{1}{h}K_h^{(1)}(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0 - u)\mathbf{Z}_{ij}$$
$$\big\{Y_{ij} - \mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0 - \widehat{\phi}(u, \boldsymbol{\Theta}_0) - \widehat{\phi}^{(1)}(u, \boldsymbol{\Theta}_0)(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0 - u)\big\},$$
$$L_2 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k\neq j}\sigma_i^{jj}\frac{1}{h}K_h^{(1)}(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0 - u)\mathbf{Z}_{ij}$$
$$\big\{Y_{ij} - \mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0 - \widehat{\phi}(u, \boldsymbol{\Theta}_0) - \widehat{\phi}^{(1)}(u, \boldsymbol{\Theta}_0)(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0 - u)\big\},$$

$$L_3 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sigma_i^{jj}K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0 - u)\big\{\varphi_{n,\boldsymbol{\theta}}(u,\boldsymbol{\Theta}_0) - \frac{\partial\widehat{\phi}^{(1)}(u,\boldsymbol{\Theta}_0)}{\partial\boldsymbol{\Theta}_0}(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0 - u)$$

$$- \widehat{\phi}^{(1)}(u,\boldsymbol{\Theta}_0)\mathbf{Z}_{ij}\big\} = L_{31} + L_{32} + L_{33},$$

$$L_4 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k\neq j}\sigma_i^{jk}K_h(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0 - u)\big\{\varphi_{n,\boldsymbol{\theta}}(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0,\boldsymbol{\Theta}_0) - \widehat{\phi}^{(1)}(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0,\boldsymbol{\Theta}_0)\mathbf{Z}_{ik}\big\}.$$

With Assumptions 3.3.1–3.3.6, when $n \to \infty$ and taking expectations, some standard calculations lead to $L_1 = o_p(1)$, $L_2 = o_p(1)$ and $L_{32} = o_p(1)$. Then we get

$$\varphi_{\boldsymbol{\theta}}(u)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sigma_i^{jj}f_{ij}(u) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}\sigma_i^{jk}\mathrm{E}\big\{\mathbf{Z}_{ik}\phi_0^{(1)}(\mathbf{Z}_{ij}^{\mathrm{T}}\boldsymbol{\theta}_0)\big\}f_{ij}(u)$$
$$-\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k\neq j}\int\sigma_i^{jk}f_{ijk}(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0,u)\varphi_{\boldsymbol{\theta}}(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0)d(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\theta}_0) + o_p(1)$$

as $n \to \infty$. By taking the limit of both (3.28) and the equation above, it is easily seen that these two equations have the same limiting equations and thus are asymptotically equivalent. Similarly, taking derivative with respect to $\boldsymbol{\beta}$ on both sides of Equation (3.29) and applying similar steps, we can obtain Equation (3.26) as $\psi_{n,\boldsymbol{\beta}}$, r_{n,β_l} and W_n are replaced by $\varphi_{\boldsymbol{\beta}}$, $r_{\beta_l} = \lim_{n\to\infty} r_{n,\beta_l}$ and $W = \lim_{n\to\infty} W_n$ respectively when $n \to \infty$. The uniqueness is guaranteed by the projection method in (3.11). Therefore, the parameter estimators reach the semiparametric information bound, completing the proof of Theorem 3.3.6. □

3.4 Simulation Studies

3.4.1 Simulation Setup

In our simulation studies, we considered PLSIM in (3.1). The true parameter settings are similar to Chen et al. (2015). Parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are of dimensions 2 and 3 with true values $\boldsymbol{\beta}_0 = (2,1)^{\mathrm{T}}$ and $\boldsymbol{\theta}_0 = (2/3, 1/3, 2/3)^{\mathrm{T}}$. The covariates $\mathbf{X}$ and $\mathbf{Z}$ are jointly generated from multivariate normal distribution with mean zero, standard deviation 1 and correlation 0.1. The true single-index function is $\phi_0(x) = \exp(x)/2$. The observation times t_{ij} are generated in the same way as in Fan et al. (2007): for each subject, there is a set of time points $\{1, 2, \ldots, T\}$ where each time point has a 0.2 probability of being missing. Then the simulated observation time is the sum of the non-missing time point and a random number from the uniform $[0, 1]$ distribution. Here T is set to be 12 to obtain an average number of observations $\bar{m}_i = 10$. The number of subjects was set to be $n = 30$, 50 and 100, respectively.

For residuals, three different scenarios were considered respectively:

1. For each i, the within-subject correlation structure is AR(1) with $\gamma = 0.75$ so that for $\epsilon_i(t_j) = \epsilon_{ij}$, cor$(\epsilon_i(t_1), \epsilon_i(t_2)) = \gamma^{|t_2 - t_1|}, t_1 \neq t_2$. Besides, the variance is set to be 1 for each observation time.

2. The within-subject correlation structure is AR(1) with $\gamma = 0.75$. For each i, the residual terms ϵ_{ij} are generated from a Gaussian process with mean 0, variance function var$\{\epsilon_i(t)\} = 0.25 \exp(t/12)$. Therefore, the residuals have nonstationary heteroskedastic variances.

3. The true within-subject correlation structure is ARMA(1,1) with $(\gamma, \rho) = (0.75, 0.6)$, where cor$(\epsilon_i(t_1), \epsilon_i(t_2)) = \rho\gamma^{|t_2 - t_1|}, t_1 \neq t_2$, but we model it as an AR(1) correlation structure. The residuals are generated with var$\{\epsilon_i(t)\} = 0.25 \exp(t/12)$.

3.4.2 Simulation Results

Under the above settings, we compared the proposed semiparametric marginal GEE method, denoted as SMGEE, with the commonly used WI method and the SGEE method. Li (2011) proposed nonparametric covariance estimation under the partially linear model setting, hereafter referred as GEE-NC. In order to measure the sensitivity of within-subject covariance estimation to the parameters and the single-index function estimation, we compared the proposed method under the current covariance estimation method, the nonparametric estimation method (GEE-NC) and when the true covariance is assumed (GEE-TC). Due to running time limitation, we limited this comparison to a particular case: $n = 30$, $\bar{m} = 10$ and the residual term follows the second case in Section 3.4.1.

The simulation results indicate that the squared bias term is negligible relative to the variance term. When presenting our numerical results we used the standard error (SE) and sandwich standard error (SWSE) as the comparison criteria, where SE is the Monte Carlo standard error and SWSE is the empirical average of the asymptotic standard error of the parameter estimates. The performance of single-index function estimation $\widehat{\phi}(\cdot)$ for $\phi_0(\cdot)$ is evaluated by averaged mean squared error (AMSE) evaluated at the observed data points:

$$\text{AMSE} = \sum_{b=1}^{B} \text{MSE}_b,$$

where

$$\text{MSE}_b = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left\{ \widehat{\phi}_b(\mathbf{Z}_{ij}^{\mathrm{T}} \widehat{\boldsymbol{\theta}}) - \phi_0(\mathbf{Z}_{ij}^{\mathrm{T}} \widehat{\boldsymbol{\theta}}) \right\}^2.$$

The Epanechnikov kernel was used and the bandwidths were selected with the leave-one-subject-out cross validation. Since there are three bandwidths involved in Steps 1, 3 and 4, the iterative and sequential bandwidth selections to choose the optimal bandwidths were compared in 20 replications. Here the iterative bandwidth selection is defined as to select all optimal bandwidths simultaneously and the sequential bandwidth selection is to choose one optimal bandwidth in each step. Simulation results from the case with $n = 30$, $\bar{m} = 10$ indicate that there are no significant differences between the iterative and sequential bandwidth selection methods: the iterative method selected bandwidths 0.61, 0.75 and 0.68, while the sequential method selected 0.62, 0.74 and 0.65, respectively. We also compared the empirical relative efficiency, which is the ratio of the numerical variances of the parameter estimates under the iterative and sequential bandwidth selections. The empirical relative efficiencies are (0.99, 0.98, 0.99, 1.02, 0.98) for $(\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\theta}^{\mathrm{T}})$. Therefore, in all the simulation studies we used the leave-one-subject-out cross validation to choose the optimal bandwidths sequentially due to the huge savings in computation time.

Remark 3.4.1. Chen et al. (2015) used an iterative method to choose the optimal bandwidths – the first two bandwidths appeared in our proposed method. In order to compare the results with that chapter, we also implemented the iterative method to choose the optimal values for the first two bandwidths and then used the chosen bandwidth in Step 1 as the same bandwidth in Step 4. The differences are negligible.

As in Chen et al. (2015) the simulations were repeated for 200 times. The results are given in Tables 3.1–3.4 with SE, SWSE and AMSE values being in percentage. In Table 3.1, we compare the estimators under the true covariances, nonparametrically estimated covariances and semiparametrically estimated covariances in the case of the sample size $n = 30$, $\bar{m} = 10$. Besides, the nonstationary residual variance (residual type 2) is assumed. We observe that the covariance estimation is not that sensitive to the estimation results: SE, SWSE and AMSE of the estimated parameters and single-index function are relatively close to those for the case when the true covariance is assumed. In Tables 3.2–3.4 we see that both SGEE and SMGEE outperform the WI estimators in all simulation settings. SMGEE is clearly more efficient than SGEE regarding the single-index function estimation and is also overall more efficient than SGEE in the parameter estimation in these settings.

3.5 Real Data Analysis

We now apply the proposed method to analyze two longitudinal datasets, one of which studies the relationship between smoking and CD4 percentage and the other studies bond

Table 3.1: Performance comparisons of the estimates of the parameters and single-index function under different covariance estimation methods while all other estimation steps are the same. The PLSIM with nonstationary residual variance and AR(1) correlation ($\gamma = 0.75$) is assumed. GEE-TC, GEE-NC and SMGEE are estimation methods with the true covariance, nonparametrically estimated covariance (Li (2011)) and semiparametrically estimated covariance (proposed) respectively. All the values are in percentage. SE stands for Monte Carlo standard error, SWSE stands for empirical asymptotic standard error and AMSE stands for averaged mean squared error.

Sample Size	Models	β_1		β_2		θ_1		θ_2		θ_3		$\phi(\cdot)$
		SE	SWSE	SE	SWSE	SE	SWSE	SE	SWSE	SE	SWSE	AMSE
$n = 30$ $\bar{m} = 10$	GEE-TC	2.68	2.72	2.65	2.66	1.75	1.76	2.02	2.05	1.86	1.85	1.14
	GEE-NC	2.80	2.84	2.73	2.76	1.77	1.80	2.08	2.08	1.95	1.93	1.40
	SMGEE	2.73	2.80	2.71	2.76	1.74	1.75	2.09	2.07	1.91	1.88	1.23

maturity firms from the period 1980 to 1989.

3.5.1 CD4 Data Analysis

In the CD4 dataset, there are a total of 283 homosexual males who were HIV positive from 1984 and 1991. They were scheduled to have the measurements but had different numbers of repeated measurements which ranges from 1 to 14 because of missing or rescheduled appointments. This dataset has also been analyzed by Qu and Li (2006). In our model, the response variable Y is the CD4 cell percentage over time. The covariates are patient's measuring time Z_1 calculated as the difference between the stopping time and starting time, patient's age Z_2, the CD4 cell percentage before infection Z_3, and smoking status X which is a binary variable.

First we perform some exploratory data analysis of the CD4 data described above. In Figure 3.1, we examined the relationship of CD4 percentage with each predictor variable. Since the predictors measuring time, patients' age and CD4 percentage before infection are continuous, we presented the scatterplot together with its local linear smoothing fit for the CD4 percentage with each predictor variable. There are nonlinear patterns for the three

Table 3.2: Performance comparisons of the estimates of the parameters and single-index function for estimation methods WI, SGEE and SMGEE. The data were generated using PLSIM with constant residual variance and AR(1) correlation ($\gamma = 0.75$). All the values are in percentage. SE stands for Monte Carlo standard error, SWSE stands for empirical asymptotic standard error and AMSE stands for averaged mean squared error.

Sample Size	Models	β_1		β_2		θ_1		θ_2		θ_3		$\phi(\cdot)$
		SE	SWSE	SE	SWSE	SE	SWSE	SE	SWSE	SE	SWSE	AMSE
$n = 30$ $\bar{m} = 10$	WI	3.82	3.89	4.12	4.14	2.72	2.75	3.12	3.15	2.66	2.70	1.77
	SGEE	2.77	2.79	2.69	2.71	1.73	1.72	2.14	2.13	1.87	1.89	1.73
	SMGEE	2.65	2.67	2.65	2.70	1.73	1.73	2.12	2.13	1.85	1.87	1.23
$n = 50$ $\bar{m} = 10$	WI	3.06	3.11	2.71	2.79	1.86	1.92	2.20	2.24	1.89	1.94	1.48
	SGEE	2.61	2.62	2.10	2.13	1.17	1.19	1.06	1.04	0.96	0.99	1.34
	SMGEE	2.54	2.56	2.10	2.14	1.15	1.16	1.03	1.03	0.96	0.98	0.90
$n = 100$ $\bar{m} = 10$	WI	2.19	2.21	2.29	2.30	1.61	1.65	1.74	1.76	1.48	1.50	1.07
	SGEE	1.58	1.58	1.60	1.50	0.87	0.90	1.17	1.18	1.13	1.14	0.99
	SMGEE	1.53	1.55	1.58	1.50	0.87	0.89	1.17	1.18	1.13	1.12	0.72

sub-figures, indicating that multiple linear regression may not be the most appropriate approach to analyze this data. For smoking status, since it is a binary variable, we provided with a side-by-side boxplot to compare the distribution of CD4 percentage in each smoking category. The difference of distribution is not clear for each smoking status, but we can see that there are some outliers of CD4 percentage for the smoking group.

After the exploratory data analysis, we attempt to analyze the CD4 data by multiple linear regression for longitudinal data with GEE approach proposed by Liang and Zeger (1986). From the diagnostic analysis, the relationship of the residual with each predictor variable appears to have similar patterns to Figure 3.1. As a consequence, we apply PLSIM in (3.1) to the data with $\mathbf{Z}_{ij} = (Z_{1ij}, Z_{2ij}, Z_{3ij})^{\mathrm{T}}$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^{\mathrm{T}}$.

Using the proposed method SMGEE and assuming the AR(1) working correlation structure, we compared the parameter estimates with the existing methods WI and SGEE. The results are listed in Table 3.5. Furthermore, the estimated single-index function $\widehat{\phi}(\cdot)$ evaluated at $\mathbf{Z}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\theta}}$ and its pointwise 95% bootstrap confidence band are shown in Figure 3.2

Table 3.3: Performance comparisons of the estimates of the parameters and single-index function for estimation methods WI, SGEE and SMGEE. The data were generated using PLSIM with nonstationary residual variance and AR(1) correlation ($\gamma = 0.75$). All the values are in percentage. SE stands for Monte Carlo standard error, SWSE stands for empirical asymptotic standard error and AMSE stands for averaged mean squared error.

Sample Size	Models	β_1		β_2		θ_1		θ_2		θ_3		$\phi(\cdot)$
		SE	SWSE	SE	SWSE	SE	SWSE	SE	SWSE	SE	SWSE	AMSE
$n = 30$ $\bar{m} = 10$	WI	3.98	4.02	4.17	4.20	2.84	2.86	3.19	3.19	2.74	2.77	1.81
	SGEE	2.87	2.95	2.73	2.81	1.78	1.82	2.18	2.20	1.91	1.96	1.75
	SMGEE	2.73	2.80	2.71	2.76	1.74	1.75	2.09	2.07	1.91	1.88	1.23
$n = 50$ $\bar{m} = 10$	WI	3.04	3.08	2.79	2.83	1.88	1.92	2.25	2.24	2.03	2.02	1.46
	SGEE	2.67	2.70	2.15	2.20	1.20	1.23	1.01	1.04	0.99	1.00	1.37
	SMGEE	2.61	2.56	2.11	2.13	1.19	1.20	0.98	1.01	0.97	0.99	0.92
$n = 100$ $\bar{m} = 10$	WI	2.21	2.23	2.33	2.34	1.68	1.69	1.83	1.83	1.52	1.54	1.03
	SGEE	1.60	1.66	1.50	1.52	0.91	0.92	1.22	1.24	1.15	1.15	0.98
	SMGEE	1.56	1.59	1.47	1.50	0.91	0.92	1.21	1.23	1.14	1.15	0.74

produced by WI, SGEE and SMGEE respectively. The parameter estimates in Table 3.5 are similar for the three methods. However, the standard errors are generally smaller with the proposed SMGEE method. This finding supports our theoretical results developed in Section 3.3. By comparing the estimated single-index function in Figure 3.2, the decreasing patterns are similar for all three methods.

From the linear component estimates in Table 3.5, we observe that the estimated coefficient for variable smoking is positive. However, the standard error compared with the coefficient estimate indicates that the smoking status is not a significant factor for CD4 cell percentage. This conclusion agrees with previous findings that smoking is not an inducing factor for HIV; see Uppal et al. (2003) and Qu and Li (2006). To study the relationship between CD4 cell percentage and measuring time, patient's age and CD4 cell percentage before infection, we can look at the parameter estimates together with the nonparametric single-index function estimate. Since the general trend for the single-index function is decreasing in Figure 3.2 by SMGEE, together with the sign and magnitude of the parameter

Table 3.4: Performance comparisons of the estimates of the parameters and single-index function for estimation methods WI, SGEE and SMGEE. The data were generated using PLSIM with nonstationary residual variance and ARMA(1,1) correlation ($\gamma = 0.75$, $\nu = 0.6$). The misspecified AR(1) was applied to model the correlation structure. All the values are in percentage. SE stands for Monte Carlo standard error, SWSE stands for empirical asymptotic standard error and AMSE stands for averaged mean squared error.

Sample Size	Models	β_1		β_2		θ_1		θ_2		θ_3		$\phi(\cdot)$
		SE	SWSE	SE	SWSE	SE	SWSE	SE	SWSE	SE	SWSE	AMSE
$n = 30$ $\bar{m} = 10$	WI	4.20	4.24	4.26	4.29	3.44	3.53	3.93	4.09	2.95	2.95	2.09
	SGEE	3.01	3.06	2.89	2.85	2.06	2.14	2.70	2.81	2.28	2.35	1.88
	SMGEE	2.97	3.05	2.84	2.75	2.02	1.99	2.69	2.77	2.31	2.36	1.35
$n = 50$ $\bar{m} = 10$	WI	3.15	3.13	2.95	3.04	2.75	2.79	2.83	2.90	2.36	2.46	1.54
	SGEE	2.83	2.85	2.20	2.28	1.62	1.70	1.40	1.48	2.04	2.10	1.39
	SMGEE	2.80	2.79	2.21	2.26	1.63	1.68	1.42	1.48	2.07	2.11	0.94
$n = 100$ $\bar{m} = 10$	WI	2.66	2.69	2.56	2.64	1.85	1.93	2.08	2.11	1.88	1.84	1.30
	SGEE	1.92	2.01	1.70	1.72	1.11	1.14	1.68	1.72	1.34	1.36	1.02
	SMGEE	1.94	1.98	1.68	1.70	1.07	1.12	1.70	1.72	1.33	1.35	0.76

estimates by SMGEE, the CD4 cell percentage is negatively related to the measuring time and patient's age, but is positively related to the CD4 cell percentage before infection. It means that as time goes by, for patients with older age and less CD4 cell percentage before infection, the HIV condition is generally worse. The CD4 cell percentage drops slowly at first, stays stable for a while afterwards and then drops sharply.

Table 3.5: Parameter estimates by WI, SGEE and SMGEE and their standard errors for the CD4 data.

Model	WI		SGEE		SMSEE	
Estimates	Estimate	SE	Estimate	SE	Estimate	SE
β	0.044	0.058	0.050	0.055	0.046	0.052
θ_1	0.880	0.036	0.842	0.033	0.827	0.033
θ_2	0.065	0.013	0.061	0.009	0.057	0.008
θ_3	−0.470	0.041	−0.536	0.040	−0.559	0.040

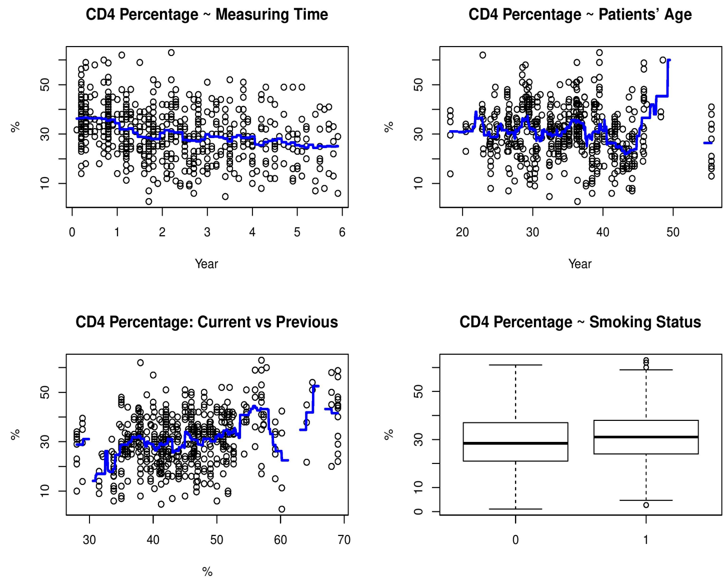

Figure 3.1: Scatterplots associated with the local linear approximation and a side-by-side boxplot for the CD4 Percentage versus each individual predictor variable. Top left: Scatterplot of the CD4 Percentage vs. Measuring Time; Top right: Scatterplot of the CD4 Percentage vs. Patients' Age; Bottom left: Scatterplot of the CD4 Percentage vs. the CD4 Percentage before Infection; Bottom right: Side-by-side boxplot of the CD4 Percentage vs. Smoking Status.

3.5.2 Debt Maturity Data Analysis

The debt maturity data previously analyzed by Ma et al. (2014) with the quadratic inference function method have 328 unregulated firms with indexes observed annually for 10 years from 1980 to 1989. The response variable is the log transformation of the debt maturity index of the firms since the index is highly right skewed. Similar to the exploratory data analysis of CD4 data, when judging from the scatter plots of the response versus each predictor, we select two binary variables X_1 and X_2 as the covariates in the linear component. Here X_1 = Low bond being 1 only if the firm has a rating of CCC or unrated. Similarly, X_2 = High bond being 1 only if the firm has a rating of AA or higher. There are four covariates in the nonlinear single-index component. They are $Z_1 =$

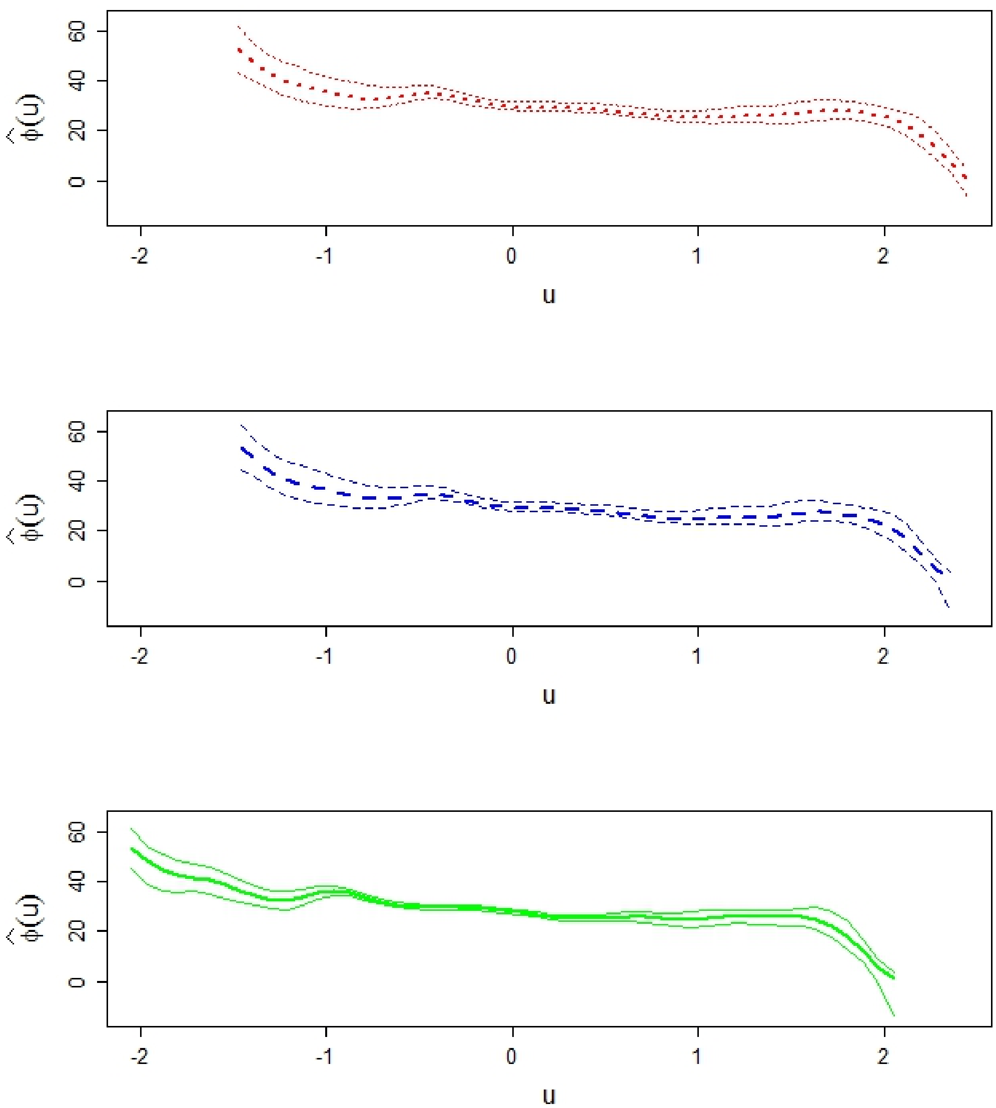

Figure 3.2: Estimated single-index function and its 95% bootstrap confidence band for the CD4 data by WI, SGEE and SMGEE. The red dotted curves are for WI. The blue dashed curves are for SGEE. The green solid curves are for SMGEE.

Leverage of a firm defined as the ratio of total debt to the market value, Z_2 = Asset Maturity which is the value-weighted average of the maturities of current assets and net property, plant, and equipment, Z_3 = MV/BV defined as the market value of the firm scaled by the assets value, and $Z_4 =$ VAR, the ratio of the standard deviation of the first difference in earnings before interest, depreciation, and taxes to the average of assets over the ten year period. Model (3.1) is fitted to the data with $i = 1, \ldots, 328$ and $j = 1, \ldots, 10$, $\boldsymbol{\beta} = (\beta_1, \beta_2)^{\mathrm{T}}$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)^{\mathrm{T}}$.

Assuming the AR(1) working correlation structure, we compared the parameter estimates by SMGEE with those by WI and SGEE. The results are listed in Table 3.6. Moreover, the estimated single-index function evaluated at $\mathbf{Z}_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\theta}}$ and its pointwise 95% bootstrap confidence band are shown in Figure 3.3. They are obtained with WI, SGEE and SMGEE respectively. Table 3.6 shows again that the three methods have different standard errors. Similarly to the results for the CD4 data, the SMGEE method leads to the smallest standard errors of the parameter estimates. Comparing the estimated single-index function in Figure 3.3, the patterns are different with the narrowest confidence band for the SMGEE method in the middle range where $1 < u < 5$.

From the linear component of the fitted model by SMGEE, we see that the firms debt maturity index is negatively related to both low bond and high bond, and low bond has a more significant negative effect on the debt maturity. The estimated single-index function by SMGEE increases sharply in the beginning and at the end, and is steady in the middle range. In general, the trend for the single-index function is increasing. Together with the sign and magnitude of parameter estimates, the firms debt maturity index is positively related to the leverage and assets maturities, slightly positively related to the scaled market value, but is negatively related to the VAR index.

3.6 Concluding Remarks

In this chapter, we proposed a three-stage procedure to estimate the parameters and the unknown single-index function in partially linear single-index models under the general sparse longitudinal setting. The parameter estimators have been shown to be semiparamet-

Table 3.6: Parameter estimates by WI, SGEE and SMGEE and their standard errors for the debt maturity data.

Model	WI		SGEE		SMSEE	
Estimates	Estimate	SE	Estimate	SE	Estimate	SE
β_1	−0.357	0.037	−0.348	0.030	−0.346	0.027
β_2	−0.136	0.046	−0.115	0.045	−0.121	0.044
θ_1	0.843	0.009	0.796	0.009	0.782	0.007
θ_2	0.535	0.097	0.596	0.095	0.612	0.089
θ_3	−0.022	0.017	0.048	0.014	0.064	0.012
θ_4	−0.050	0.020	−0.098	0.016	−0.102	0.016

rically efficient. Furthermore, the single-index function estimator is not only more efficient than the working independence estimator of Chen et al. (2015), but also achieves the minimum asymptotic variance among a class of estimators when the covariance matrices are correctly specified. These analytic results are supported by our empirical studies.

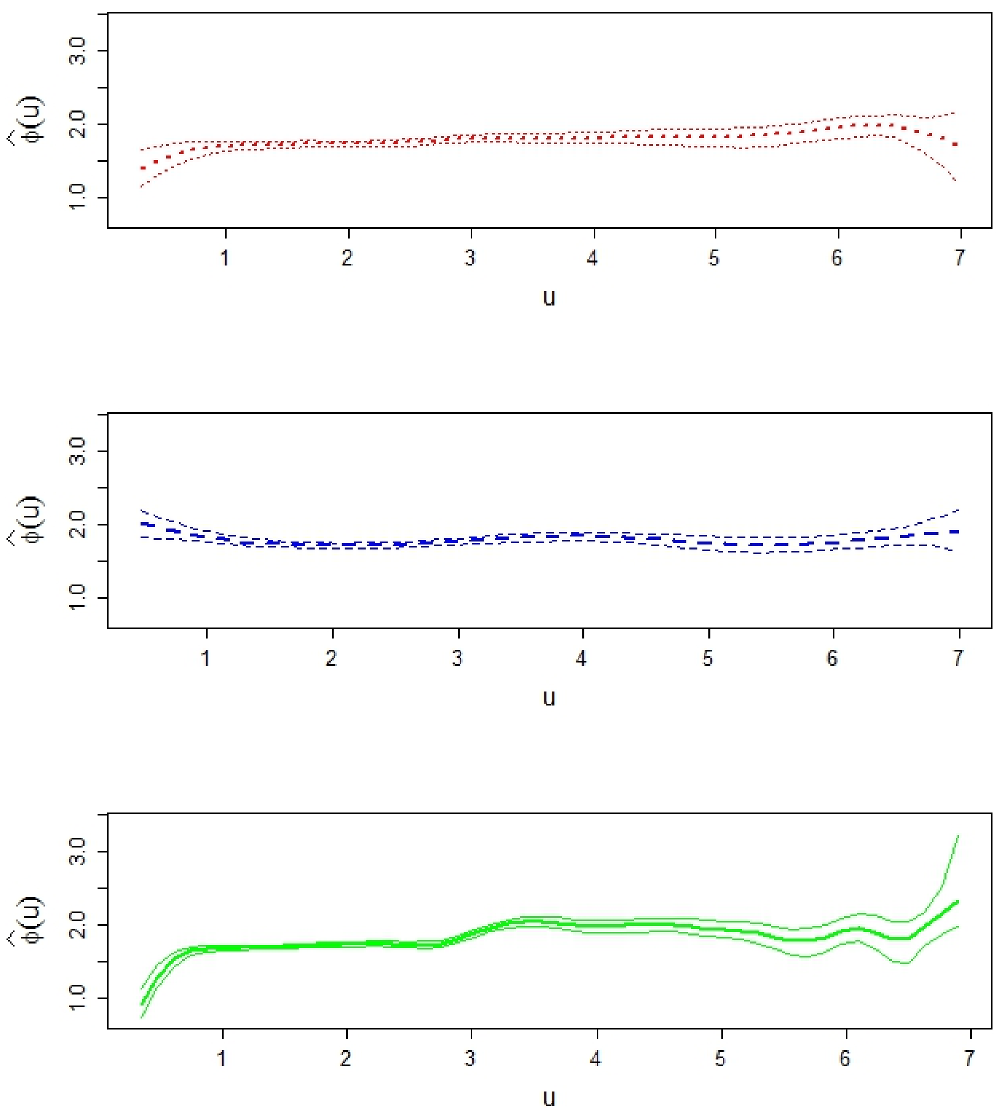

Figure 3.3: Estimated single-index function and its 95% bootstrap confidence band for the debt maturity data by WI, SGEE and SMGEE. The red dotted curves are for WI. The blue dashed curves are for SGEE. The green solid curves are for SMGEE.

4. GENERALIZED PARTIALLY LINEAR SINGLE-INDEX MODELS FOR LONGITUDINAL DATA

We study generalized partially linear single-index models for longitudinal data in this chapter. We propose a method to efficiently estimate both the parameters and the nonparametric single-index function in generalized partially linear single-index models when subjects are observed or measured over time. This is an extension of the method developed in Chapter 3. The proposed estimation approach is more flexible and more general in that we can model both categorical response and transformation-necessary response such as heavy-tailed variable with multiple covariates, especially when some covariates are parametrically correlated with the response and the others are nonparametrically correlated with the response. With minimal assumptions, we show that the semiparametric information bound is reached for the parameter estimators. We also show that the asymptotic variance of the single-index function estimator is generally less than that of existing estimators. Furthermore, we provide Monte Carlo simulation results and an empirical data analysis that support our new method.

4.1 Introduction

Semiparametric models are flexible in statistical modeling with their advantages of incorporating both the parametric and nonparametric components. It is popular in economics, biomedical science and many other research fields. The generalized partially linear single-index model is one of semiparametric models proposed by Carroll et al. (1997). Suppose we have an univariate response variable Y and possibly multi-dimensional covariates $\mathbf{X}$ and $\mathbf{W}$. GPLSIM has the form

$$\begin{aligned} &\mathrm{E}(Y|\mathbf{X}=\mathbf{x},\mathbf{W}=\mathbf{w})=\mu(\mathbf{x},\mathbf{w}), \\ &g^{-1}\{\mu(\mathbf{x},\mathbf{w})\}=\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}+\gamma(\mathbf{w}^{\mathrm{T}}\boldsymbol{\alpha}), \end{aligned} \tag{4.1}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are possibly multi-dimensional parameters associated with predictors $\mathbf{X}$ and $\mathbf{W}$ respectively, $\mathbf{x}$ and $\mathbf{w}$ are the realizations of $\mathbf{X}$ and $\mathbf{W}$ respectively and $\gamma(\cdot)$ is an unknown function, referred to as the single-index function hereafter. Besides, $g^{-1}(\cdot)$ is assumed to be a known monotonic and differentiable link function. The goal is to estimate the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ and the single-index function $\gamma(\cdot)$ in (4.1). GPLSIM is the generalized model of several types of models. When the single-index function is the identity function, it becomes the generalized linear model (Nelder and Baker (2004)). When there are no covariates $\mathbf{X}$, it becomes the generalized single-index model (Ichimura (1993)). When $\mathbf{W}$ is one-dimensional, it becomes the generalized partially linear model (Chen and Shiau (1994)). When the link function $g^{-1}(\cdot)$ is the identity function and Y is continuous, it becomes the partially linear single-index model (Carroll et al. (1997), Chen et al. (2015)). Therefore, efficient estimation of GLSIM is of major interest in that it can unify the estimation of several important models above and has broad applications.

Partially linear single-index models has been studied extensively for independent and identically distributed (i.i.d.) data. In the i.i.d. case, researchers have proposed a variety of methods for the mean function estimation. Carroll et al. (1997) proposed GPLSIM and the maximum quasi-likelihood method to estimate the parameters as well as the single-index function. Yu and Ruppert (2002) proposed the penalized spline estimation approach for PLSIM. Zhu and Xue (2006) provided the empirical likelihood confidence regions for parameters of this model. To extend the work of Liang and Zeger (1986) on generalized linear models for longitudinal data, Liang et al. (2010) proposed profile estimation method for the parameters and single-index function for PLSIM and obtained the semiparametric efficient parameter estimators. Xia et al. (1999) and Xia and Härdle (2006) studied the theoretical properties of estimators and extended the conditions for estimating PLSIM.

However, there are limited research on GPLSIM for dependent or correlated data, especially in the case of longitudinal/clustered data. There are some related literature discussing the problem of mean function estimation in longitudinal data. Lin and Carroll (2000, 2001) studied nonparametric function estimation for clustered data under various settings. They showed that for the profile-kernel GEE method, the estimated parameters in partially linear models are not semiparametric efficient and the most efficient estimators are obtained by ignoring the dependence and undersmoothing the nonparametric function. Wang (2003) proposed the marginal kernel generalized estimation equation (GEE) method for the mean function in nonparametric regression to account for within-subject correlation. Wang et al. (2005) extended this method to generalized partially linear models and showed that the parameter estimators reach the semiparametric efficiency bound under some mild conditions. Huang et al. (2007) proposed spline-based additive models for partially linear models and Cheng et al. (2014) extended the results to generalized partially linear additive models. They also showed that the parameter estimators are semiparametric efficient. Li et al. (2010) proposed a bias-corrected block empirical likelihood inference for partially linear single-index models and provided confidence regions for the parameters. Chen et al. (2015) proposed a unified semiparametric GEE analysis for partially linear single-index models for both sparse and dense longitudinal data. In Chapter 3 we discussed how to efficiently estimate the parameters and the single-index function in partially linear single-index models for longitudinal data. The main objective of Chapter 4 is to extend the work in Chapter 3 to GPLSIM and propose efficient estimators in a generalized approach.

The remainder of the chapter is organized as follows. In Section 4.2, the main problem is defined and the estimation method is introduced. In Section 4.3, the main theoretical results as well as the detailed proofs are presented. Simulation studies are performed in Section 4.4 to demonstrate the methodology and a real data analysis is given in Section 4.5 to apply the proposed method. Some concluding remarks are given in Section 4.6.

4.2 Methodology

One of the most common approaches to analyze longitudinal data is by assuming marginal models; see Pepe and Anderson (1994). Consider the marginal longitudinal generalized partially linear single-index model

$$\begin{aligned} \mathrm{E}(Y_{ij}|\mathbf{X}_i, \mathbf{W}_i) =& \mathrm{E}(Y_{ij}|\mathbf{X}_{ij}, \mathbf{W}_{ij}) = \mu_{ij}, \\ g^{-1}(\mu_{ij}) =& \mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \gamma(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}) \end{aligned} \tag{4.2}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m_i$. Besides, $g^{-1}(\cdot)$ is assumed to be a known monotonic and differentiable link function. This model extends model (4.1) for the i.i.d. case. Variable Y_{ij} indicates the univariate response of the j^{th} observation/measurement for subject i. Multivariate covariates $\mathbf{X}_{ij}$ and $\mathbf{W}_{ij}$ of dimensions r and s respectively are defined similarly. Joining together, $\mathbf{X}_i = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{im_i})^{\mathrm{T}}$ and $\mathbf{W}_i = (\mathbf{W}_{i1}, \ldots, \mathbf{W}_{im_i})^{\mathrm{T}}$ are covariates matrices for subject i. In this work we focus on the sparse case, i.e., $\max_{1\le i\le n}(m_i)$ is bounded as $n \to \infty$.

The parameters $\boldsymbol{\xi} = (\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\alpha}^{\mathrm{T}})^{\mathrm{T}}$ with dimensions r and s and the univariate single-index function $\gamma(\cdot)$ are to be estimated. To ensure identifiability, we make restrictions that $||\boldsymbol{\alpha}|| = 1$ and the first element of $\boldsymbol{\alpha}$ is positive. Let $K(\cdot)$ be a symmetric kernel density function defined in Condition **(C1)** in Section 4.3, $K_h(x) = h^{-1}K(x/h)$ with bandwidth h and $\mathcal{D}$ be the domain of $\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}$ defined in Condition **(C2)**. The algorithm for estimating the parameters and the single-index function has the following three stages.

(I) For $z \in \mathcal{D}$ and fixed $\boldsymbol{\xi}$, denote $\mathbf{a}_1(z, \boldsymbol{\xi}) = \{\widetilde{\gamma}(z, \boldsymbol{\xi}), \widetilde{\gamma}^{(1)}(z, \boldsymbol{\xi})\}^{\mathrm{T}}$ as the solution to

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}(1, \mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha} - z)^{\mathrm{T}}\mu^{(1)}(\boldsymbol{\xi}, \mathbf{a}_1, z)K_h(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha} - z) \\ &\{Y_{ij} - \mu(\boldsymbol{\xi}, \mathbf{a}_1, z)\} = 0, \end{aligned} \tag{4.3}$$

where $\mu(\boldsymbol{\xi}, \mathbf{a}_1, z)$ and $\mu^{(1)}(\boldsymbol{\xi}, \mathbf{a}_1, z)$ are $g(\cdot)$ and its first derivative evaluated at $\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + (1, \mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha} - z)^{\mathrm{T}}\mathbf{a}_1(z, \boldsymbol{\xi})$.

Given estimated $\widetilde{\gamma}(\cdot)$, let $\widetilde{\boldsymbol{\xi}} = (\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}, \widetilde{\boldsymbol{\alpha}}^{\mathrm{T}})^{\mathrm{T}}$ be the solution to

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\partial g\{\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \widetilde{\gamma}(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha})\}^{\mathrm{T}}}{\partial \boldsymbol{\xi}}\Big[Y_{ij} - g\big\{\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \widetilde{\gamma}(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha})\big\}\Big] = 0. \tag{4.4}$$

Iterating Stage (I) until convergence, we obtain the initial estimates $\widetilde{\boldsymbol{\xi}}$ and $\widetilde{\gamma}(z)$, $z \in \mathcal{D}$.

(II) Estimate the true covariance matrices $\mathbf{V}_i$ with the working covariance matrices $\mathbf{C}_i$, $i = 1, \ldots, n$. With the estimated mean function $\widehat{\mu}(\cdot)$ at observed values, the estimated working covariance matrices $\widehat{\mathbf{C}}_i$ can be decomposed into the estimated standard deviation matrices and the estimated correlation matrices components with the following equation:

$$\widehat{\mathbf{C}}_i = \mathrm{diag}(s_{i1}, \ldots, s_{im_i})\mathbf{R}_i(\widehat{\boldsymbol{\rho}})\mathrm{diag}(s_{i1}, \ldots, s_{im_i}),$$

where s_{ij} is the estimated standard deviation of Y_{ij} and $\mathbf{R}_i(\widehat{\boldsymbol{\rho}})$ is the working correlation matrix with estimated parameters $\widehat{\boldsymbol{\rho}}$ for subject i, $i = 1, \ldots, n$.

(III) Denote $\mathbf{a}_2^{\mathrm{T}} = (a_1, a_2) = \{\widehat{\gamma}(z, \boldsymbol{\xi}), \widehat{\gamma}^{(1)}(z, \boldsymbol{\xi})\}$ and $\mathbf{G}_{ij} = [\mathbf{k}_j, \mathbf{k}_j(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha} - z)]$ which is an $m_i \times 2$ matrix, where $\mathbf{k}_j$ denotes the indicator vector with j^{th} entry equal to

1, and 0 elsewhere, $j = 1, \ldots, m_i$. Further denote $\gamma_{[m]}(\cdot)$ as the estimate of $\gamma(\cdot)$ in the m^{th} step. Then in the $(m+1)^{th}$ estimation step, $\mathbf{a}_2$ solves the kernel-weighted estimating equation

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha} - z)\mu^{(1)}(\boldsymbol{\xi}, \mathbf{a}_2, z)\mathbf{G}_{ij}^{\mathrm{T}}\widehat{\mathbf{C}}_i^{-1} \\ &\Big[\mathbf{Y}_i - \boldsymbol{\mu}^*\big\{z, \mathbf{X}_i, \mathbf{W}_i, \boldsymbol{\xi}, \gamma_{[m]}(\cdot)\big\}\Big] = 0,\end{aligned} \tag{4.5}$$

where $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{im_i})^{\mathrm{T}}$ and the k^{th} element of the $m_i \times 1$ vector $\boldsymbol{\mu}^*\{z, \mathbf{X}_i, \mathbf{W}_i,$ $\boldsymbol{\xi}, \gamma_{[m]}(\cdot)\}$ is

$$g\left[\mathbf{X}_{ik}^{\mathrm{T}}\boldsymbol{\beta} + I(k=j)\{a_1 + a_2(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha} - z)\} + I(k \neq j)\gamma_{[m]}(\mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha})\right]$$

for $k = 1, \ldots, m_i$. Here $\mathbf{I}(\cdot)$ is the indicator function.

Given estimated $\widehat{\gamma}(z, \widehat{\boldsymbol{\xi}})$, the parameters $\boldsymbol{\xi} = (\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\alpha}^{\mathrm{T}})^{\mathrm{T}}$ are estimated by solving the GEE

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial g\{\mathbf{X}_i\boldsymbol{\beta} + \widehat{\gamma}(\mathbf{W}_i\boldsymbol{\alpha})\}^{\mathrm{T}}}{\partial\boldsymbol{\xi}}\widehat{\mathbf{C}}_i^{-1}\Big[\mathbf{Y}_i - g\big\{\mathbf{X}_i\boldsymbol{\beta} + \widehat{\gamma}(\mathbf{W}_i\boldsymbol{\alpha})\big\}\Big] = 0, \tag{4.6}$$

where $\widehat{\gamma}(\mathbf{W}_i\boldsymbol{\alpha}) = (\widehat{\gamma}(\mathbf{W}_{i1}^{\mathrm{T}}\boldsymbol{\alpha}), \ldots, \widehat{\gamma}(\mathbf{W}_{im_i}^{\mathrm{T}}\boldsymbol{\alpha}))^{\mathrm{T}}$. We obtain the final parameter estimate $\widehat{\boldsymbol{\xi}}$ and the single-index function estimate $\widehat{\gamma}(z, \widehat{\boldsymbol{\xi}}) = \widehat{\gamma}(z)$ by iterating Stage (III).

Note that at Stages (I) and (III), we can use the Newton-Raphson algorithm to update the estimates for $\boldsymbol{\xi}$. Moreover, for the within-subject covariances estimation at Stage (II), we have different approaches for each type of link function in generalized models. For example, for the identity link function and the working correlation matrix AR(1), we can estimate the variance component with local linear approximation and estimate the parameter in AR(1) with the quasi-likelihood method or with the minimum generalized variance method; see Fan and Wu (2008). For binary response data with logit link function, the standard deviation is estimated by $\widehat{s}_{ij} = \widehat{\mu}_{ij}(1-\widehat{\mu}_{ij})$. Assuming the working correlation matrix is compound symmetry with parameter ρ, we estimate the parameter ρ by

$$\widehat{\rho} = \frac{2}{\sum_{i=1}^{n} m_i(m_i-1)} \sum_{i=1}^{n}\sum_{j=2}^{m_i}\sum_{j'=1}^{j-1} \frac{(Y_{ij}-\widehat{\mu}_{ij})(Y_{ij'}-\widehat{\mu}_{ij'})}{\{\widehat{\mu}_{ij}(1-\widehat{\mu}_{ij})\widehat{\mu}_{ij'}(1-\widehat{\mu}_{ij'})\}^{1/2}}.$$

For other types of generalized models, the working covariance matrices can be estimated in a similar way; see Prentice (1988) and Chan (2014).

4.3 Theoretical Properties

In this section, we explore the asymptotic properties of the parameter estimators and the single-index function estimator obtained in Section 4.2. First of all, we list the necessary conditions for the proposed method. Under the given conditions, we first analyze the theoretical results for the parameter estimators. We derive that under some mild conditions the parameter estimators are asymptotically consistent and normal. The asymptotic variance is also specified. Furthermore, we obtain the asymptotic properties for the single-index function estimator. From now on we assume that the true parameters in (4.2) are $\boldsymbol{\xi}_0 = (\boldsymbol{\beta}_0^{\mathrm{T}}, \boldsymbol{\alpha}_0^{\mathrm{T}})^{\mathrm{T}}$ and the true single-index function is $\gamma_0(\cdot)$.

In our theoretical development, we assume the following conditions.

(C1) Kernel function $K(\cdot)$ is symmetric and bounded with a compact support, and has a continuous first derivative, denoted by $K^{(1)}(\cdot)$. Furthermore, when $n \to \infty$, $h \to 0$, $nh^8 \to 0$ and $\log(1/h)/(nh) \to 0$.

(C2) The density of $\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}$, denoted by $p_{ij}(z)$, is positive and has second continuous derivatives in the domain $\mathcal{D} = \{z = \mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha} : \mathbf{W}_{ij} \in \mathcal{W}, \boldsymbol{\alpha} \in \boldsymbol{\Theta}\}$ where $\mathcal{W}$ is a compact support for $\mathbf{W}_{ij}$ and $\boldsymbol{\Theta}$ is a compact parameter space for $\boldsymbol{\alpha}$. In addition, the joint density of $(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}, \mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha})$, denoted by $p_{ijk}(z_1, z_2)$, $z_1, z_2 \in \mathcal{D}$ has first partial derivatives. $i = 1, \ldots, n$, $j = 1, \ldots, m_i$ and $k = 1, \ldots, m_i$. Furthermore, $\sup_{z\in\mathcal{D}} \mathrm{E}(||\mathbf{X}_{ij}||^2|\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha} = z)$ and $\sup_{z\in\mathcal{D}} \mathrm{E}(||\mathbf{W}_{ij}||^2|\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha} = z)$ are bounded for all $i = 1, \ldots, n$, $j = 1, \ldots, m_i$.

(C3) The true covariance matrices $\mathbf{V}_i$ and the estimated working covariance matrices $\widehat{\mathrm{C}}_i$ are non-negative definite matrices for all $i = 1, \ldots, n$.

(C4) $\gamma(\cdot)$ and $g(\cdot)$ have second continuous derivatives.

In Condition **(C1)**, some common regularity and smoothness conditions are given for the kernel function. Besides, we make some mild conditions on bandwidths to allow the asymptotic results of the parameter estimators and the single-index function estimator to be valid. In Condition **(C2)**, we assume some conditions on the domain and the compact support for random variables. We also need the second moments to exist to make the proposed parameter and the single-index function estimators to be consistent and asymptotically normal. Condition **(C3)** guarantees that both the true covariance matrices and the estimated working covariance matrices are invertible, since the inverse of the matrices are used when estimating the parameters in generalized estimating equations at Stage (III). The smoothness conditions in Condition **(C4)** for the single-index function and the link function are required to guarantee that the theoretical results are well founded and the estimated single-index function is smooth.

4.3.1 Semiparametric Information Bound

We now derive the semiparametric efficiency bound for parameter estimators in GPLSIM with the projection method first proposed by Bickel et al. (1993). First we define model (4.2) by $\mathcal{B}$ and four submodels of $\mathcal{B}$ as follows:

$\mathcal{B}_1$: Model $\mathcal{B}$ with only $\boldsymbol{\xi}_0$ unknown;

$\mathcal{B}_2$: Model $\mathcal{B}$ with only $\gamma_0(\cdot)$ unknown;

$\mathcal{B}_3$: Model $\mathcal{B}$ with both $\boldsymbol{\xi}_0$ and $\gamma_0(\cdot)$ known;

$\mathcal{B}_4$: Model $\mathcal{B}$ with only $\gamma_0(\cdot)$ known.

In order to obtain the semiparametric efficient score for model $\mathcal{B}$, we further denote $\mathbf{S}_{\boldsymbol{\xi}}$ as the score function in submodel $\mathcal{B}_1$ and denote $T_{\mathcal{B}_l}$ as the tangent space for submodel $\mathcal{B}_l$ for $l = 2, 3$. With the projection method, we have the following equation for the semiparametric efficient score $\mathbf{S}_{\mathrm{s}}$ of model $\mathcal{B}$:

$$\mathbf{S}_{\mathrm{s}} = \mathbf{S}_{\boldsymbol{\xi}} - \text{proj}(\mathbf{S}_{\boldsymbol{\xi}}|T_{\mathcal{B}_3}) - \text{proj}(\mathbf{S}_{\boldsymbol{\xi}}|\text{proj}_{T_{\mathcal{B}_3}^{\perp}} T_{\mathcal{B}_2}), \tag{4.7}$$

where $\mathcal{B}^{\perp}$ is the perpendicular space of model $\mathcal{B}$, proj$(\mathbf{S}_{\boldsymbol{\xi}}|R)$ is the projection of score function $\mathbf{S}_{\boldsymbol{\xi}}$ on space R and proj$_{S1}S2$ is the projection of space $S2$ on space $S1$.

From the definition of the submodels above, $\mathcal{B}_1$ is a subspace of $\mathcal{B}_4$ associated with the finite dimensional component and $\mathcal{B}_3$ s a subspace of $\mathcal{B}_4$ associated with the infinite dimensional component. By applying the projection method, $\mathbf{S}_{\boldsymbol{\xi}} - \text{proj}(\mathbf{S}_{\boldsymbol{\xi}}|T_{\mathcal{B}_3})$ is the efficient score for $\boldsymbol{\xi}_0$ in model $\mathcal{B}_4$. Lemma A4 in Huang et al. (2007) indicates that

$$\text{proj}(\mathbf{S}_{\boldsymbol{\xi}}|T_{\mathcal{B}_3}^{\perp}) = \sum_{i=1}^{n} (\mathbf{X}_i, \mathbf{W}_i^*)^{\mathrm{T}} \boldsymbol{\Delta}_i \mathbf{V}_i^{-1} \Big[\mathbf{Y}_i - g\big\{\mathbf{X}_i \boldsymbol{\beta}_0 + \gamma_0(\mathbf{W}_i \boldsymbol{\alpha}_0)\big\}\Big], \tag{4.8}$$

and when considering the parametric submodels of $\mathcal{B}_2$, we have the following result:

$$\text{proj}_{T_{\mathcal{B}_3}^{\perp}} T_{\mathcal{B}_2} = \Bigg\{ \sum_{i=1}^{n} \tau_n^{\mathrm{T}}(\mathbf{W}_i \boldsymbol{\alpha}_0) \boldsymbol{\Delta}_i \mathbf{V}_i^{-1} \Big[\mathbf{Y}_i - g\big\{\mathbf{X}_i \boldsymbol{\beta}_0 + \gamma_0(\mathbf{W}_i \boldsymbol{\alpha}_0)\big\}\Big], \\ \tau_n(\cdot) \in L_2 \Bigg\}. \tag{4.9}$$

Here $\tau_n(\cdot)$ is any integrable function in the L_2 norm. With the result in (4.8) and (4.9) and then referring back to (4.7), we have

$$\mathbf{S}_{\mathrm{s}} = \sum_{i=1}^{n} (\widetilde{\mathbf{X}}_{i,\mathrm{s}}, \widetilde{\mathbf{W}}_{i,\mathrm{s}})^{\mathrm{T}} \boldsymbol{\Delta}_i \mathbf{V}_i^{-1} \Big[\mathbf{Y}_i - g\big\{\mathbf{X}_i \boldsymbol{\beta}_0 + \gamma_0(\mathbf{W}_i \boldsymbol{\alpha}_0)\big\}\Big],$$

where $\widetilde{\mathbf{X}}_{i,\mathrm{s}} = \mathbf{X}_i - \tau_{n,\boldsymbol{\beta}}(\mathbf{W}_i\boldsymbol{\alpha}_0)$, $\widetilde{\mathbf{W}}_{i,\mathrm{s}} = \mathbf{W}_i^* - \tau_{n,\boldsymbol{\alpha}}(\mathbf{W}_i\boldsymbol{\alpha}_0)$ for some functions $\tau_{n,\boldsymbol{\beta}}(\cdot)$ and $\tau_{n,\boldsymbol{\alpha}}(\cdot)$ in the L_2 norm. The requirement for $\tau_{n,\boldsymbol{\beta}}(\cdot)$ and $\tau_{n,\boldsymbol{\alpha}}(\cdot)$ is that $\mathbf{S}_\mathrm{s}$ should be orthogonal to $\mathrm{proj}_{T_{\mathcal{B}_3}^{\perp}} T_{\mathcal{B}_2}$. As a result,

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left\{\widetilde{\mathbf{X}}_{i,\mathrm{s}}^{\mathrm{T}} \boldsymbol{\Delta}_i \mathbf{V}_i^{-1} \boldsymbol{\Delta}_i \tau_n(\mathbf{W}_i\boldsymbol{\alpha}_0)\right\} = 0, \tag{4.10}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left\{\widetilde{\mathbf{W}}_{i,\mathrm{s}}^{\mathrm{T}} \boldsymbol{\Delta}_i \mathbf{V}_i^{-1} \boldsymbol{\Delta}_i \tau_n(\mathbf{W}_i\boldsymbol{\alpha}_0)\right\} = 0, \tag{4.11}$$

where $\tau_n^{\mathrm{T}}(\mathbf{W}_i\boldsymbol{\alpha}_0) = \{\tau_n(\mathbf{W}_{i1}^{\mathrm{T}}\boldsymbol{\alpha}_0), \ldots, \tau_n(\mathbf{W}_{im_i}^{\mathrm{T}}\boldsymbol{\alpha}_0)\}$ and for all $\tau_n(\cdot)$ in the L_2 norm. The existence and uniqueness of the semiparametric score function $\mathbf{S}_\mathrm{s}$ for model $\mathcal{B}$ are guaranteed by the projection method.

Therefore, we have the semiparametric efficient score function of $\boldsymbol{\xi}$ and the semiparametric information bound $\mathbf{U}(\boldsymbol{\xi}_0)$ summarized in the following theorem:

Theorem 4.3.1. *Assume that the conditions in Conditions **(C2)–(C4)** hold. Then the semiparametric efficient score function of $\boldsymbol{\xi}$ for GPLSIM in model (4.2) is*

$$\mathbf{S}_\mathrm{s} = \sum_{i=1}^{n} (\widetilde{\mathbf{X}}_{i,\mathrm{s}}, \widetilde{\mathbf{W}}_{i,\mathrm{s}})^{\mathrm{T}} \boldsymbol{\Delta}_i \mathbf{V}_i^{-1} \Big[\mathbf{Y}_i - g\big\{\mathbf{X}_i\boldsymbol{\beta}_0 + \gamma_0(\mathbf{W}_i\boldsymbol{\alpha}_0)\big\}\Big].$$

Therefore, we have the semiparametric information bound $\mathbf{U}(\boldsymbol{\xi}_0)$ by

$$\mathbf{U}(\boldsymbol{\xi}_0) = \{\lim_{n\to\infty} \frac{1}{n} \mathrm{E}(\mathbf{S}_\mathrm{s}\mathbf{S}_\mathrm{s}^{\mathrm{T}})\}^{-1} = \Big[\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\big\{(\widetilde{\mathbf{X}}_{i,\mathrm{s}}, \widetilde{\mathbf{W}}_{i,\mathrm{s}})^{\mathrm{T}} \boldsymbol{\Delta}_i \mathbf{V}_i^{-1} \boldsymbol{\Delta}_i (\widetilde{\mathbf{X}}_{i,\mathrm{s}}, \widetilde{\mathbf{W}}_{i,\mathrm{s}})\big\}\Big]^{-1}.$$

4.3.2 Theory for the Single-Index Function Estimator

Let $k_0 = \int z^2 K(z)dz$, $k_1 = \int K^2(z)dz$. Furthermore, let c_i^{jj} and $\eta_{i,jj}$ be the $(j,j)^{th}$ element of $\mathbf{C}_i^{-1}$ and $\mathbf{C}_i^{-1}\mathbf{V}_i\mathbf{C}_i^{-1}$ respectively. The following theorem gives the asymptotic properties of the single-index function estimator.

Theorem 4.3.2. *Assume that the regularity and smoothness conditions in Conditions **(C1)**–**(C4)** hold. Then for $z \in \mathcal{D}$ we have*

$$\sqrt{nh}\{\widehat{\gamma}(z) - \gamma_0(z) - k_0 b(z)h^2\} \xrightarrow{D} N(0, \sigma_\gamma^2(z)), \tag{4.12}$$

where

$$\sigma_\gamma^2(z) = k_1 \frac{\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^{m_i} \mathrm{E}\{(\mu_{ij}^{(1)})^2 | \mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha} = z\}\eta_{i,jj} p_{ij}(z)}{\left[\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^{m_i} \mathrm{E}\left\{(\mu_{ij}^{(1)})^2 | \mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha} = z\right\} c_i^{jj} p_{ij}(z)\right]^2},$$

and $b(z)$ satisfies

$$b(z) + \int b(v)\omega(z,v)dv = \frac{1}{2}\gamma_0^{(2)}(z)$$

with

$$\omega(z,v) = \frac{\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k\neq j} \mathrm{E}(\mu_{ij}^{(1)} c_i^{jk} \mu_{ik}^{(1)} | \mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha} = z) p_{ijk}(z,v)}{\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^{m_i} \mathrm{E}\{(\mu_{ij}^{(1)})^2 c_i^{jj} | \mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha} = v\} p_{ij}(v)}.$$

The proof of Theorem 4.3.2 is given in Proof 4.3.2 below.

Proof. From Equation (4.3) with the second-order Taylor expansion, given that the parameters are evaluated at true values, we have

$$\begin{aligned}\widetilde{\gamma}(z,\boldsymbol{\xi}_0)-\gamma_0(z)=&\{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mathrm{E}(\Delta_{ijj}^2|\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0=z)p_{ij}(z)\}^{-1}\\&\times\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mu_{ij}^{(1)}K_h(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0-z)(Y_{ij}-\mu_{ij})\\&+\frac{1}{2}c_0\gamma_0^{(2)}(z)h^2+o_p\{h^2+(nh)^{-1/2}\}.\end{aligned}\tag{4.13}$$

Similarly to (4.13), from Equation (4.5) at Stage (III), we have the one-step update $\widehat{\gamma}_{(1)}$ of $\widetilde{\gamma}$ as follows:

$$\widehat{\gamma}_{(1)}(z,\boldsymbol{\xi}_0)-\gamma_0(z)=e_n^{-1}(z)(H_{1n}+H_{2n})+o_p\{h^2+(nh)^{-1/2}\},\tag{4.14}$$

where

$$\begin{aligned}e_n(z)=&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mathrm{E}(c_i^{jj}\Delta_{ijj}^2|\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0=z\}p_{ij}(z),\\H_{1n}=&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}K_h(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0-z)\mu_{ij}^{(1)}c_i^{jj}\Big[Y_{ij}-g\big\{\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0-\gamma_0(z)\\&-\gamma_0^{(1)}(z)(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0-z)\big\}\Big]\\&+\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}K_h(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0-z)\mu_{ij}^{(1)}\Big\{\sum_{k\neq j}c_i^{jk}(Y_{ik}-\mu_{ik})\Big\},\\H_{2n}=&-\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}K_h(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0-z)\mu_{ij}^{(1)}\Big[\sum_{k\neq j}c_i^{jk}\{\widetilde{\gamma}(\mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0)-\gamma_0(\mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0)\}\Big].\end{aligned}$$

Using (4.13) in H_{2n} leads to the following re-expression of (4.14):

$$\widetilde{\gamma}_{(1)}(z,\boldsymbol{\xi}_0)-\gamma_0(z)=J_{1n}(z)+J_{2n}(z)+\frac{1}{2}k_0b_{n,1}(z)h^2+o_p\{h^2+(nh)^{-1/2}\},\tag{4.15}$$

where

$$
\begin{aligned}
J_{1n}(z) =& \; e_n^{-1}(z)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0 - z)\mu_{ij}^{(1)}\{\sum_{k=1}^{m_i} c_i^{jk}(Y_{ik}-\mu_{ik})\}, \\
J_{2n}(z) =& \; -e_n^{-1}(z)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mu_{ij}^{(1)}c_i^{jj}L_n(z,\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0)(Y_{ij}-\mu_{ij}), \\
L_n(z_1,z_2) =& \; \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k\neq j}\mathrm{E}\Big\{\Delta_{ijj}c_i^{jk}\Delta_{ikk}e_n^{-1}(\mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0)|\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0 = z_1, \mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0 = z_2\Big\} \\
& \times p_{ijk}(z_1,z_2), \\
b_{n,0}(z) =& \; \gamma_0^{(2)}(z), \\
b_{n,t}(z) =& \; \gamma_0^{(2)}(z) - e_n^{-1}(z)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k\neq j}\mathrm{E}\{\Delta_{ijj}c_i^{jk}\Delta_{ikk}b_{n,t-1}(\mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0)|\mathbf{W}_{il}^{\mathrm{T}}\boldsymbol{\alpha}_0 = z\} \\
& \times p_{ij}(z),
\end{aligned}
$$

for $t = 1, 2, \ldots$. For the second iteration step, by applying the result in (4.15), we have an equation similar to (4.14) by replacing $\widetilde{\gamma}(\mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0) - \gamma_0(z)$ in H_{2n} with $\widehat{\gamma}_{(1)}(\mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0) - \gamma_0(z)$ which leads to the following equation:

$$
\widehat{\gamma}_{(2)}(z,\boldsymbol{\xi}_0) - \gamma_0(z) = e_n^{-1}(z)(H_{1n} + H_{3n}) + o_p\{h^2 + (nh)^{-1/2}\},
$$

where

$$
\begin{aligned}
H_{3n} =& \; -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0 - z)\mu_{ij}^{(1)}\Big[\sum_{k\neq j} c_i^{jk}\{J_{1n}(\mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0) + J_{2n}(\mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0) \\
& + \frac{1}{2}k_0 b_{n,1}(\mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0)h^2\}\Big].
\end{aligned}
$$

We can obtain the asymptotic bias term as

$$
\begin{aligned}
&\frac{1}{2}k_0h^2\Big[\gamma_0^{(2)}(z) - e_n^{-1}(z)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k\neq j}\mathrm{E}\big\{\Delta_{jj}c_i^{jk}\Delta_{kk}b_{n,1}(\mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0)\big\}p_{ij}(z)\Big]\\
=&\frac{1}{2}k_0b_{n,2}(z).
\end{aligned}
$$

Now we extend the above results to the t^{th} $(t \geq 2)$ iteration step. First define

$$
\begin{aligned}
A_n(L_n; z_1, z_2) =& -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k\neq j}\mathrm{E}\Big\{\Delta_{jj}c_i^{jk}\Delta_{kk}e_n^{-1}(\mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0)L_n(\mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0, z_2)\Big\}p_{ij}(z_1),\\
Q_{1n,t} =& e_n^{-1}(z)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mu_{ij}^{(1)}G_{n,1}^{t}(z, \mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0)\Big\{\sum_{k=1}^{m_i}c_i^{jk}(Y_{ik}-\mu_{ik})\Big\},\\
Q_{2n,t} =& e_n^{-1}(z)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mu_{ij}^{(1)}c_i^{jj}G_{n,2}^{t}(z, \mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0)(Y_{ij}-\mu_{ij}),
\end{aligned}
$$

where $G_{n,1}^1(z_1, z_2) = 0$, $G_{n,1}^t(z_1, z_2) = -L_n(z_1, z_2) + A_n(G_{n,1}^{l-1}; z_1, z_2)$, $G_{n,2}^1(z_1, z_2) = -L_n(z_1, z_2)$, and $G_{n,2}^t(z_1, z_2) = A_n(G_{n,2}^{t-1}; z_1, z_2)$. Now we have the following expansion:

$$
\widehat{\gamma}_{(t)}(z, \boldsymbol{\xi}_0) - \gamma_0(z) = J_{1n} + Q_{1n,t} + Q_{2n,t} + \frac{1}{2}k_0b_{n,t}(z)h^2 + o_p\{h^2 + (nh)^{-1/2}\}. \quad (4.16)
$$

At convergence, we replace $b_{n,t}$, $G_{n,1}^t$, $G_{n,2}^t$ with their limits b_n, $G_{n,1}$ and $G_{n,2}$, So $\widehat{\gamma}(z) - \gamma_0(z)$ satisfies the following equation:

$$
\begin{aligned}
b_n(z) =& \gamma_0^{(2)}(z) - e_n^{-1}(z)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k\neq j}\mathrm{E}\Big\{\Delta_{jj}c_i^{jk}\Delta_{kk}b_n(\mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0)\Big\}p_{ij}(z),\\
G_{n,1}(z_1, z_2) =& -L_n(z_1, z_2) + A_n(G_{n,1}; z_1, z_2),\\
G_{n,2}(z_1, z_2) =& A_n(G_{n,2}; z_1, z_2).
\end{aligned} \quad (4.17)
$$

Since $\mathrm{E}(Q_{1n,t}) = \mathrm{E}(Q_{2n,t}) = 0$ and the variances of $Q_{1n,t}$ and $Q_{2n,t}$ are of order $O(n^{-1}) =$

$o\{(nh)^{-1}\}$, we have $Q_{1n,t} = Q_{2n,t} = o_p\{(nh)^{-1/2}\}$. Thus, Equation (4.16) can be simplified as

$$\widehat{\gamma}_{(t)}(z, \boldsymbol{\xi}_0) - \gamma_0(z) = J_{1n}(z) + \frac{1}{2} k_0 b_{n,t}(z) h^2 + o_p\{h^2 + (nh)^{-1/2}\}.$$

It is now trivial to see that

$$\begin{aligned}
\widehat{\gamma}(z) - \gamma_0(z) = e_n^{-1}(z) \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} K_h(\mathbf{W}_{ij}^{\mathrm{T}} \boldsymbol{\alpha}_0 - z) \mu_{ij}^{(1)} \Big\{ \sum_{k=1}^{m_i} c_i^{jk} (Y_{ik} - \mu_{ik}) \Big\} \\
+ k_0 \Big[\frac{h^2 \gamma_0^{(2)}(z)}{2} - h^2 e_n^{-1}(z) \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k \neq j} \mathrm{E}\big\{ \Delta_{jj} c_i^{jk} \Delta_{kk} b(\mathbf{W}_{ik}^{\mathrm{T}} \boldsymbol{\alpha}_0) \big\} p_{ij}(z) \Big] \\
+ o_p\{h^2 + (nh)^{-1/2}\},
\end{aligned}$$

which leads to the result in Theorem 4.3.2. □

Similarly to the theoretical results for the parameter estimators, by applying the extended Cauchy-Schwarz inequality and some calculations, we can derive that the asymptotic variance of $\widehat{\gamma}(z)$ is minimized when the covariance matrices are correctly specified. Furthermore, when the covariance matrices are correctly specified, the asymptotic variance of $\widehat{\gamma}(z)$ is generally smaller than the asymptotic variance of $\widetilde{\gamma}(z)$, the working independence estimator for the single-index function. The detailed proof is similar to the proof in Chapter 3.

4.3.3 Theory for the Parameter Estimators

To obtain theoretical results for the parameter estimators, we have the following notation. When $n \to \infty$, we have

$$\lambda_{n,\boldsymbol{\beta}}(z, \widehat{\boldsymbol{\xi}}) = -\frac{\partial \widehat{\gamma}(z, \boldsymbol{\xi})}{\partial \boldsymbol{\beta}^{\mathrm{T}}}|_{\boldsymbol{\xi}=\widehat{\boldsymbol{\xi}}} \xrightarrow{P} \lambda_{\boldsymbol{\beta}}(z), \;\; \lambda_{n,\boldsymbol{\alpha}}(z, \widehat{\boldsymbol{\xi}}) = -\frac{\partial \widehat{\gamma}(z, \boldsymbol{\xi})}{\partial \boldsymbol{\alpha}^{\mathrm{T}}}|_{\boldsymbol{\xi}=\widehat{\boldsymbol{\xi}}} \xrightarrow{P} \lambda_{\boldsymbol{\alpha}}(z), \tag{4.18}$$

where $\xrightarrow{P}$ indicates convergence in probability. Further define

$$\begin{aligned}
\boldsymbol{\Sigma}_0 &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}\left\{(\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{W}}_i)^{\mathrm{T}} \boldsymbol{\Delta}_i \mathbf{C}_i^{-1} \boldsymbol{\Delta}_i (\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{W}}_i)\right\}, \\
\boldsymbol{\Sigma}_1 &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}\left\{(\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{W}}_i)^{\mathrm{T}} \boldsymbol{\Delta}_i \mathbf{C}_i^{-1} \mathbf{V}_i \mathbf{C}_i^{-1} \boldsymbol{\Delta}_i (\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{W}}_i)\right\},
\end{aligned} \tag{4.19}$$

where $(\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{W}}_i) = \{\mathbf{X}_i - \lambda_{\boldsymbol{\beta}}(\mathbf{W}_i\boldsymbol{\alpha}_0), \mathbf{W}_i^* - \lambda_{\boldsymbol{\alpha}}(\mathbf{W}_i\boldsymbol{\alpha}_0)\}$ and $\boldsymbol{\Delta}_i$ is the $m_i \times m_i$ diagonal matrix with its $(j,j)^{th}$ element Δ_{ijj} being $g^{(1)}\{\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0 + \gamma_0(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0)\}$. Here $g^{(1)}(\cdot)$ indicates the first derivative of $g(\cdot)$ and $\mathbf{W}_i^* = \{\gamma_0^{(1)}(\mathbf{W}_i\boldsymbol{\alpha}_0) \otimes \mathbf{1}_s^{\mathrm{T}}\} \circ \mathbf{Z}_i$, where $\mathbf{1}_s$ is a vector of 1 of dimension s, $\otimes$ is the Kronecker product and $\circ$ is the component-wise product. Assume that $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$ are non-negative definite matrices. Then we have the following theorem.

Theorem 4.3.3. *Assume that the regularity and smoothness conditions in Conditions **(C1)**–**(C4)** are satisfied. Then*

$$n^{1/2}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{-1}), \tag{4.20}$$

where $\xrightarrow{D}$ *denotes convergence in distribution.*

The proof of Theorem 4.3.3 is given below.

Proof. In the proof of the asymptotic properties of the parameter estimators, it is required that the single-index function estimator is uniformly consistent. As a result, the convergence rate of the single-index function estimator changes from $o_p\{h^2 + (nh)^{-1/2}\}$ to $o_p(h^2 + \{\log(n)/nh\}^{1/2})$.

From (4.6), we have

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \frac{\partial g\{\mathbf{X}_i\boldsymbol{\beta} + \widehat{\gamma}(\mathbf{W}_i\boldsymbol{\alpha})\}^{\mathrm{T}}}{\partial \boldsymbol{\xi}} |_{\boldsymbol{\xi}=\widehat{\boldsymbol{\xi}}} \mathbf{C}_i^{-1} \Big[\mathbf{Y}_i - g\big\{\mathbf{X}_i\widehat{\boldsymbol{\beta}} + \widehat{\gamma}(\mathbf{W}_i\widehat{\boldsymbol{\alpha}})\big\}\Big] = 0. \quad (4.21)$$

Now we apply the first-order Taylor expansion to $g\big\{\mathbf{X}_i\widehat{\boldsymbol{\beta}} + \widehat{\gamma}(\mathbf{W}_i\widehat{\boldsymbol{\alpha}})\big\}$ at $g\big\{\mathbf{X}_i\boldsymbol{\beta}_0 + \gamma_0(\mathbf{W}_i\boldsymbol{\alpha}_0)\big\}$. We have

$$\begin{aligned} & g\big\{\mathbf{X}_i\widehat{\boldsymbol{\beta}} + \widehat{\gamma}(\mathbf{W}_i\widehat{\boldsymbol{\alpha}})\big\} \\ &= g\big\{\mathbf{X}_i\boldsymbol{\beta}_0 + \gamma_0(\mathbf{W}_i\boldsymbol{\alpha}_0)\big\} + g^{(1)}\big\{\mathbf{X}_i\boldsymbol{\beta}_0 + \gamma_0(\mathbf{W}_i\boldsymbol{\alpha}_0)\big\} \circ (\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{W}}_i)(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0) \\ &\quad + g^{(1)}\big\{\mathbf{X}_i\boldsymbol{\beta}_0 + \gamma_0(\mathbf{W}_i\boldsymbol{\alpha}_0)\big\} \circ \big\{\widehat{\gamma}(\mathbf{W}_i\boldsymbol{\alpha}_0) - \gamma_0(\mathbf{W}_i\boldsymbol{\alpha}_0)\big\} + o_p(n^{-1/2}). \quad (4.22) \end{aligned}$$

Then by plugging (4.22) into (4.21), we obtain that

$$\begin{aligned} & n^{-\frac{1}{2}} \sum_{i=1}^{n} (\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{W}}_i)^{\mathrm{T}} g^{(1)}\big\{\mathbf{X}_i\boldsymbol{\beta}_0 + \gamma_0(\mathbf{W}_i\boldsymbol{\alpha}_0)\big\} \circ \mathbf{C}_i^{-1} \\ & \Big[\mathbf{Y}_i - g\big\{\mathbf{X}_i\boldsymbol{\beta}_0 + \gamma_0(\mathbf{W}_i\boldsymbol{\alpha}_0)\big\} - g^{(1)}\big\{\mathbf{X}_i\boldsymbol{\beta}_0 + \gamma_0(\mathbf{W}_i\boldsymbol{\alpha}_0)\big\} \circ (\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{W}}_i)(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0) \\ & - g^{(1)}\big\{\mathbf{X}_i\boldsymbol{\beta}_0 + \gamma_0(\mathbf{W}_i\boldsymbol{\alpha}_0)\big\} \circ \big\{\widehat{\gamma}(\mathbf{W}_i\boldsymbol{\alpha}_0) - \gamma_0(\mathbf{W}_i\boldsymbol{\alpha}_0)\big\}\Big] = o_p(1). \end{aligned}$$

Then it is easy to derive that

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^{n} (\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{W}}_i)^{\mathrm{T}} \boldsymbol{\Delta}_i \mathbf{C}_i^{-1} \boldsymbol{\Delta}_i (\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{W}}_i) \Big\{ n^{1/2}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0) \Big\} \\ &= n^{-\frac{1}{2}} \sum_{i=1}^{n} (\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{W}}_i)^{\mathrm{T}} \boldsymbol{\Delta}_i \mathbf{C}_i^{-1} \Big[\mathbf{Y}_i - g\big\{\mathbf{X}_i\boldsymbol{\beta}_0 + \gamma_0(\mathbf{W}_i\boldsymbol{\alpha}_0)\big\} \end{aligned}$$

$$- g^{(1)}\{\mathbf{X}_i\boldsymbol{\beta}_0 + \gamma_0(\mathbf{W}_i\boldsymbol{\alpha}_0)\} \circ \{\widehat{\gamma}(\mathbf{W}_i\boldsymbol{\alpha}_0) - \gamma_0(\mathbf{W}_i\boldsymbol{\alpha}_0)\}\Big] + o_p(1). \tag{4.23}$$

Then by Equations (4.16) and (4.17) and a second-order bias expansion, it is readily seen that

$$\begin{aligned} L &= n^{-\frac{1}{2}}\sum_{i=1}^{n}(\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{W}}_i)^{\mathrm{T}}\boldsymbol{\Delta}_i\mathbf{C}_i^{-1}g^{(1)}\{\mathbf{X}_i\boldsymbol{\beta}_0 + \gamma_0(\mathbf{W}_i\boldsymbol{\alpha}_0)\} \circ \{\widehat{\gamma}(\mathbf{W}_i\boldsymbol{\alpha}_0) - \gamma_0(\mathbf{W}_i\boldsymbol{\alpha}_0)\} \\ &= L_1 + L_2, \end{aligned}$$

where

$$L_1 = n^{-\frac{1}{2}}\frac{h^2}{2}\Big[\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}\mu_{ij}^{(1)}c_i^{jk}\mu_{ik}^{(1)}(\widetilde{\mathbf{X}}_{ij}, \widetilde{\mathbf{W}}_{ij})\{b(\mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0) + hb_1(\mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0) + O_p(h^2)\}\Big],$$

and

$$\begin{aligned} L_2 = n^{-\frac{1}{2}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}\mu_{ij}^{(1)}c_i^{jk}\mu_{ik}^{(1)}(\widetilde{\mathbf{X}}_{ij}, \widetilde{\mathbf{W}}_{ij})\Bigg(&e_n^{-1}(\mathbf{W}_{il}^{\mathrm{T}}\boldsymbol{\alpha})n^{-1}\sum_{i'=1}^{n}\sum_{j'=1}^{m_i}\mu_{i'j'}^{(1)} \\ &\times\Big[K_h(\mathbf{W}_{i'j'}^{\mathrm{T}}\boldsymbol{\alpha} - \mathbf{W}_{il}^{\mathrm{T}}\boldsymbol{\alpha})\{\sum_{l=1}^{m_i}(c_{i'})^{j'l}(Y_{i'l} - \mu_{i'l})\} \\ &+ c_{i'}^{j'j'}G_{n,2}(\mathbf{W}_{il}^{\mathrm{T}}\boldsymbol{\alpha}, \mathbf{W}_{i'j'}^{\mathrm{T}}\boldsymbol{\alpha})(Y_{i'j'} - \mu_{i'j'}) \\ &+ G_{n,1}(\mathbf{W}_{il}^{\mathrm{T}}\boldsymbol{\alpha}, \mathbf{W}_{i'j'}^{\mathrm{T}}\boldsymbol{\alpha})\{\sum_{l=1}^{m_i}c_{i'}^{j'l}(Y_{i'l} - \mu_{i'l})\}\Big]\Bigg). \end{aligned}$$

Here $b(\cdot)$ is the first-order and $b_1(\cdot)$ is the second-order bias expansion of $\widehat{\gamma}(z)$.

When working covariances $\mathbf{C}_i$ are used for $\mathbf{V}_i$, Equations (4.10) and (4.11) are asymptotically equivalent to the following equation:

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\Big\{(\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{W}}_i)^{\mathrm{T}}\boldsymbol{\Delta}_i\mathbf{C}_i^{-1}\boldsymbol{\Delta}_iq_n(\mathbf{Z}_i) \circ \mathbf{f}_i(\mathbf{Z}_i)|\mathbf{Z}_i = \mathbf{z}_i\Big\} = 0 \tag{4.24}$$

for any function $q_n(\cdot) \in L_2$, where $\mathbf{Z}_i = \mathbf{W}_i\boldsymbol{\alpha}$, $\mathbf{z}_i = (z_{11}, \ldots, z_{im_i})^{\mathrm{T}}$ and $\mathbf{f}_i(\mathbf{z}_i) = (p_{i1}(z_{i1}), \ldots, p_{im_i}(z_{im_i}))^{\mathrm{T}}$ for $z_{ij} \in \mathcal{D}$, $j = 1, \ldots, m_i$. We can then use arguments similar to that in Section A.4 of Wang et al. (2005) to obtain that $L_1 = o_p(1)$ if $nh^8 \to 0$. As for L_2, rewrite it as $L_2 = L_{21} + L_{22} + L_{23}$. Here L_{21} is asymptotically equivalent to

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}\mathrm{E}\big\{(\widetilde{\mathbf{X}}_{ij}, \widetilde{\mathbf{W}}_{ij})^{\mathrm{T}}\mu_{ij}^{(1)}c_i^{jk}\mu_{ik}^{(1)}e_n^{-1}(Z_{ik})|Z_{ik} = z\big\}p_{ik}(z)|_{z=\mathbf{W}_{i'j'}^{\mathrm{T}}\boldsymbol{\alpha}_0}$$

with $Z_{ik} = \mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0$, which converges to 0 in probability by (4.24). Similarly, it can be shown by the Central Limit Theorem that $L_{22} = o_p(1)$ and $L_{23} = o_p(1)$.

Consequently, from (4.23) we have

$$n^{\frac{1}{2}}(\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0) = \boldsymbol{\Sigma}_0^{-1}n^{-\frac{1}{2}}\sum_{i=1}^{n}(\widetilde{\mathbf{X}}_{ij}, \widetilde{\mathbf{W}}_{ij})^{\mathrm{T}}\boldsymbol{\Delta}_i\mathbf{C}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i) + o_p(1).$$

By the Central Limit Theorem, Theorem 4.3.3 has been shown. □

From Theorem 1, we know that the parameter estimators are asymptotic normal, consistent and with convergence rate $O_p(n^{-1/2})$. Furthermore, when the covariance matrices are correctly specified, the asymptotic covariance of the parameter estimators can be simplified from $\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{-1}$ to

$$\boldsymbol{\Sigma}_2 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left\{(\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{W}}_i)^{\mathrm{T}}\boldsymbol{\Delta}_i\mathbf{V}_i^{-1}\boldsymbol{\Delta}_i(\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{W}}_i)\right\}.$$

Moreover, by the extended Cauchy-Schwarz inequality, we obtain that the asymptotic covariance of the parameter estimators is minimized when the covariance matrices are correctly specified in the sense that for all estimated working covariances $\mathbf{C}_i$, $\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\Sigma}_2$ is a non-negative definite matrix. We omit the proof here since it is similar to the proof in Chapter 3.

4.3.4 Semiparametric Efficiency of the Proposed Parameter Estimators

To show that the proposed parameter estimators reach semiparametric information bound, it is equivalent to show that when $n \to \infty$, the limit of $\tau_{n,\boldsymbol{\beta}}(z)$ is $\lambda_{\boldsymbol{\beta}}(z)$ and the limit of $\tau_{n,\boldsymbol{\alpha}}(z)$ is $\lambda_{\boldsymbol{\alpha}}(z)$. Actually, the asymptotic covariance of the parameter estimators reaches the semiparametric information bound when $\mathbf{C}_i = \mathbf{V}_i$ for $i = 1, \dots, n$. We formally state this result in the following theorem.

Theorem 4.3.4. *Assume that the conditions in Conditions* **(C1)–(C4)** *are satisfied and assume that the covariance matrices are correctly specified. Then the centering part of the asymptotic covariance of the proposed parameter estimators are asymptotically the same as the centering part of the asymptotic covariance of semiparametric efficient estimators. That is, for $z \in \mathcal{D}$, when $n \to \infty$ we have*

$$\tau_{n,\boldsymbol{\beta}}(z) \xrightarrow{P} \lambda_{\boldsymbol{\beta}}(z) \quad \text{and} \quad \tau_{n,\boldsymbol{\alpha}}(z) \xrightarrow{P} \lambda_{\boldsymbol{\alpha}}(z)$$

Therefore, the proposed parameter estimators are semiparametrically efficient.

The proof of Theorem 4.3.4 is given in Proof 4.3.4.

Proof. From (4.10) we obtain that $\kappa_{n,\beta_l}(\cdot)$ satisfies the following equation:

$$\tau_{n,\beta_l}(z_1) = q_{n,\beta_l}(z_1) - \int Q_n(z_1, z_2)\tau_{n,\beta_l}(z_2)dz_2, \tag{4.25}$$

where

$$Q_n(z_1, z_2) = \frac{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k\neq j} \mathrm{E}(\Delta_{ijj}c_i^{jk}\Delta_{ikk}|\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha} = z_1, \mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha} = z_2)p_{ijk}(z_2, z_1)}{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} \mathrm{E}(\Delta_{ijj}^2 c_i^{jj}|\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha} = z_1)p_{ij}(z_1)} \tag{4.26}$$

and

$$q_{n,\beta_l}(z)=\frac{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}\mathrm{E}(\Delta_{ijj}c_i^{jk}\Delta_{ikk}|\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}=z)p_{ij}(z)}{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mathrm{E}(\Delta_{ijj}^2c_i^{jj}|\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}=z)p_{ij}(z)}$$

for $l=1,\ldots,r$. When taking the limit of (4.25) as $n\to\infty$, $\tau_{\boldsymbol{\beta}}(z)=\lim_{n\to\infty}\tau_{n,\boldsymbol{\beta}}(z)$ satisfies the Fredholm integral equations of the second kind in the limiting form. Similarly we have that $\tau_{n,\alpha_l}(\cdot)$ satisfies the following equation:

$$\tau_{n,\alpha_l}(z_1)=q_{n,\alpha_l}(z_1)-\int Q_n(z_1,z_2)\tau_{n,\alpha_l}(z_2)dz_2, \tag{4.27}$$

where

$$q_{n,\alpha_l}(z)=\frac{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}\mathrm{E}(\Delta_{ijj}c_i^{jk}\Delta_{ikk}W_{ijl}^*|\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}=z)p_{ij}(z)}{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mathrm{E}(\Delta_{ijj}^2c_i^{jj}|\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}=z)p_{ij}(z)}$$

for $l=1,\ldots,s$ and W_{ijl}^* is the l^{th} element of $\mathbf{W}_{ij}^*$. Moreover, as $n\to\infty$, $\tau_{\boldsymbol{\alpha}}(z)=\lim_{n\to\infty}\tau_{n,\boldsymbol{\alpha}}(z)$ satisfies the Fredholm integral equations of the second kind as well.

To show that the asymptotic covariance of the proposed parameter estimators reach the semiparametric information bound, it is equivalent to prove that $\lambda_{\boldsymbol{\beta}}(z)$ and $\lambda_{\boldsymbol{\alpha}}(z)$ satisfy the Fredholm integral equations of the second kind in (4.25) and (4.27). In the following, we only show that $\lambda_{\boldsymbol{\alpha}}(z)$ satisfies (4.27) in the limiting form. The proof of $\lambda_{\boldsymbol{\beta}}(z)$ satisfying (4.25) is similar.

When $\mathbf{C}_i=\mathbf{V}_i$ for $i=1,\ldots,n$, Equation (4.5) can be expressed by

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}K_h(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}-z)\mu_{ij}^{(1)}(\boldsymbol{\xi},\mathbf{a}_1,z)c_i^{jj}\Big[Y_{ij}-g\big\{\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}+\widehat{\gamma}(z,\boldsymbol{\xi})+\widehat{\gamma}^{(1)}(z,\boldsymbol{\xi})$$
$$\times(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}-z)\big\}\Big]+\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k\neq j}K_h(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}-z)\mu_{ij}^{(1)}(\boldsymbol{\xi},\mathbf{a}_1,z)c_i^{jk}$$

$$\times\Big[Y_{ik} - g\big\{\mathbf{X}_{ik}^{\mathrm{T}}\boldsymbol{\beta} + \widehat{\gamma}(\mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}, \boldsymbol{\xi})\big\}\Big] = 0. \tag{4.28}$$

Now we take the partial derivative with respect to $\boldsymbol{\alpha}$ on both sides of (4.28) and then evaluate the parameters at the true values, obtaining the following equation:

$$M_{1n} + M_{2n} + M_{3n} + M_{4n} + M_{5n} + M_{6n} = 0, \tag{4.29}$$

where

$$\begin{aligned}
M_{1n} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{1}{h}K_h^{(1)}(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0 - z)\mathbf{W}_{ij}\mu_{ij}^{(1)}(\boldsymbol{\xi}_0, \mathbf{a}_1, z)c_i^{jj} \\
&\qquad\times\Big[Y_{ij} - g\big\{\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0 + \widehat{\gamma}(z, \boldsymbol{\xi}_0) + \widehat{\gamma}^{(1)}(z, \boldsymbol{\xi}_0)(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0 - z)\big\}\Big], \\
M_{2n} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k\neq j}\frac{1}{h}K_h^{(1)}(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0 - z)\mathbf{W}_{ij}\mu_{ij}^{(1)}(\boldsymbol{\xi}_0, \mathbf{a}_1, z)c_i^{jk} \\
&\qquad\times\Big[Y_{ik} - g\big\{\mathbf{X}_{ik}^{\mathrm{T}}\boldsymbol{\beta}_0 + \widehat{\gamma}(\mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0, \boldsymbol{\xi}_0)\big\}\Big], \\
M_{3n} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}K_h(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0 - z)\mu_{ij}^{(2)}(\boldsymbol{\xi}_0, \mathbf{a}_1, z)\{-\lambda_{n,\boldsymbol{\alpha}}(z, \boldsymbol{\xi}_0) \\
&\qquad+\frac{\partial\widehat{\gamma}^{(1)}(z, \boldsymbol{\xi}_0)}{\partial\boldsymbol{\xi}_0}(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0 - z) + \widehat{\gamma}^{(1)}(z, \boldsymbol{\xi}_0)\mathbf{W}_{ij}\}c_i^{jj} \\
&\qquad\Big[Y_{ij} - g\big\{\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0 + \widehat{\gamma}(z, \boldsymbol{\xi}_0) + \widehat{\gamma}^{(1)}(z, \boldsymbol{\xi}_0)(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0 - z)\big\}\Big], \\
M_{4n} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k\neq j}K_h(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0 - z)\mu_{ij}^{(2)}(\boldsymbol{\xi}_0, \mathbf{a}_1, z)\{-\lambda_{n,\boldsymbol{\alpha}}(z, \boldsymbol{\xi}_0) \\
&\qquad+\frac{\partial\widehat{\gamma}^{(1)}(z, \boldsymbol{\xi}_0)}{\partial\boldsymbol{\xi}_0}(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0 - z) + \widehat{\gamma}^{(1)}(z, \boldsymbol{\xi}_0)\mathbf{W}_{ij}\}c_i^{jk} \\
&\qquad\Big[Y_{ik} - g\big\{\mathbf{X}_{ik}^{\mathrm{T}}\boldsymbol{\beta}_0 + \widehat{\gamma}(\mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0, \boldsymbol{\xi}_0)\big\}\Big], \\
M_{5n} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}K_h(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0 - z)\{\mu_{ij}^{(1)}(\boldsymbol{\xi}_0, \mathbf{a}_1, z)\}^2 c_i^{jj}\{-\lambda_{n,\boldsymbol{\alpha}}(z, \boldsymbol{\xi}_0) \\
&\qquad+\frac{\partial\widehat{\gamma}^{(1)}(z, \boldsymbol{\xi}_0)}{\partial\boldsymbol{\xi}_0}(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0 - z) + \widehat{\gamma}^{(1)}(z, \boldsymbol{\xi}_0)\mathbf{W}_{ij}\},
\end{aligned}$$

$$M_{6n} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k\neq j} K_h(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0 - z)\mu_{ij}^{(1)}(\boldsymbol{\xi}_0, \mathbf{a}_1, z)c_i^{jk}\mu_{ik}^{(1)}(\boldsymbol{\xi}_0, \mathbf{a}_1, z)$$
$$\times\{\lambda_{n,\boldsymbol{\alpha}}(\mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0, \boldsymbol{\xi}_0) - \widehat{\gamma}^{(1)}(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0, \boldsymbol{\xi}_0)\mathbf{W}_{ik}\}.$$

Under Conditions **(C1)–(C4)**, with some lengthy but standard calculations it is seen that $M_{1n} = o_p(1)$, $M_{2n} = o_p(1)$, $M_{3n} = o_p(1)$ and $M_{4n} = o_p(1)$ when $n \to \infty$. For M_{5n}, we separate it by $M_{5n} = M_{5n1} + M_{5n2} + M_{5n3}$. When taking expectation with respect to $\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0$, we get $M_{5n2} = o_p(1)$ because of the symmetry of the kernel function. For M_{5n1} and M_{5n3}, we have

$$\mathrm{E}(M_{5n1}) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mathrm{E}(\Delta_{ijj}^2|\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0 = z)c_i^{jj}p_{ij}(z)\lambda_{n,\boldsymbol{\alpha}}(z, \boldsymbol{\xi}_0) + o_p(1),$$
$$\mathrm{E}(M_{5n3}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mathrm{E}(\Delta_{ijj}^2\mathbf{W}_{ij}|\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0 = z)c_i^{jj}p_{ij}(z)\widehat{\gamma}^{(1)}(z, \boldsymbol{\xi}_0) + o_p(1).$$

Similarly, after we separate M_{6n} by $M_{6n} = M_{6n1} + M_{6n2}$ and take the expectation with respect to $\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0$, we have

$$\mathrm{E}(M_{6n1}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k\neq j}\mathrm{E}\Big\{\Delta_{ijj}\Delta_{ikk}\widehat{\gamma}^{(1)}(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0, \boldsymbol{\xi}_0)\mathbf{W}_{ik}|\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0 = z\Big\}c_i^{jk}p_{ij}(z) + o_p(1),$$
$$\mathrm{E}(M_{6n2}) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k\neq j}\int \mathrm{E}(\Delta_{ijj}\Delta_{ikk}|\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0 = z)c_i^{jk}\lambda_{n,\boldsymbol{\alpha}}(\mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0, \boldsymbol{\xi}_0)$$
$$\times\, p_{ijk}(z, \mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0)d(\mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0) + o_p(1).$$

When $n \to \infty$, (4.29) becomes

$$\lambda_{\boldsymbol{\alpha}}(z)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mathrm{E}(\Delta_{ijj}^2|\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0 = z)c_i^{jj}p_{ij}(z)$$
$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}\mathrm{E}\{\Delta_{ijj}\Delta_{ikk}\gamma_0^{(1)}(\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0)\mathbf{W}_{ik}|\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0 = z\}c_i^{jk}p_{ij}(z)$$

$$-\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\sum_{k\neq j}\int \mathrm{E}(\Delta_{ijj}\Delta_{ikk}|\mathbf{W}_{ij}^{\mathrm{T}}\boldsymbol{\alpha}_0 = z)c_i^{jk}p_{ijk}(\mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0, z)\lambda_{\boldsymbol{\alpha}}(\mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0)$$

$$d(\mathbf{W}_{ik}^{\mathrm{T}}\boldsymbol{\alpha}_0) + o_p(1). \tag{4.30}$$

By taking the limit of (4.27) and (4.30), the limiting equations are the same for (4.27) and (4.30). As a consequence, $\lambda_{\boldsymbol{\alpha}}(z)$ and $\tau_{\boldsymbol{\alpha}}(z)$ are asymptotically equivalent. Similarly, after taking the partial derivative with respect to $\boldsymbol{\beta}$ on both sides of Equation (4.28) and evaluating each term when $n \to \infty$, $\lambda_{\boldsymbol{\beta}}(z)$ and $\tau_{\boldsymbol{\beta}}(z)$ have the same limiting equation as (4.25). Therefore, $\lambda_{\boldsymbol{\beta}}(z)$ and $\tau_{\boldsymbol{\beta}}(z)$ are asymptotically equivalent as well. In conclusion, the parameter estimators are semiparametrically efficient. □

4.4 Monte Carlo Simulations

In this section, we conduct simulation studies in various settings to evaluate the performance of the proposed method, denoted by SEE (semiparametric efficient estimators), in estimating the parameters and the single-index function, in comparison with two existing methods which have not fully taken into consideration the within-subject correlations. The first method is totally working independence, denoted by WI. It applies the profile least square method for the parameters estimation and the local linear approximation for the single-index function estimation. The second method to be compared with is the method of Chen et al. (2015) denoted by SGEE. With the initial estimators and working covariances estimation, SGEE incorporates within-subject correlation in GEE for the parameter estimation, while the local linear approximation is used for the single-index estimation. Furthermore, we examine the sensitivity of the covariance estimation by comparing the SEE estimation results with true covariances and estimated covariances.

In the first simulation study, we consider binary partially linear single-index model in (4.2) with the link function by logit function $g^{-1}(p) = \log\{(p/(1-p)\}$. As in Chen et al. (2015), parameters $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ have dimensions 2 and 3. Let true values $\boldsymbol{\beta}_0 = (2, 1)^{\mathrm{T}}$, $\boldsymbol{\alpha}_0 = (2/3, 1/3, 2/3)^{\mathrm{T}}$ and the single index function $\gamma_0(x) = 0.5 \sin(x)$. The covariates $\mathbf{X}$ and $\mathbf{W}$ are jointly generated from the multivariate normal distribution with mean zero, standard deviation 1 and correlation 0.1. The observation times t_{ij} are generated in the same way as in Fan and Wu (2008), i.e., for each subject, there is a set of time points $\{1, 2, \ldots, 12\}$ with each time point having a chance of 0.2 to be missed. Each simulated observation time is the sum of a non-missing observation and a randomly generated value from the uniform $[0, 1]$ distribution. The binary response variable is generated in the same way as in Chen and Zhou (2011) and Leisch et al. (1998) which can be easily implemented by the "bindata" package in R. For the within-subject correlation structure, we use compound symmetry with $\rho = 0.75$ so that for $Y_i(t_j) = Y_{ij}$, $\mathrm{cor}(Y_i(t_1), Y_i(t_2)) = \rho, t_1 \neq t_2$. The estimated variance term is calculated with the estimated mean.

In the second simulation study, we consider Poisson partially linear single-index model with the log link function. The parameters $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, covariates $\mathbf{X}$ and $\mathbf{W}$, the single-index function as well as the observation times t_{ij} are generated in the same way as in the first simulation study. The correlated Poisson response is generated by the Trivariate Reduction method; see Mardia (1970) and Xu and Rahman (2004). This method is implemented in the "corcounts" package in R. We use the compound symmetry correlation function with $\rho = 0.60$. The estimated variance term is calculated with the estimated mean. In optimization steps, simulated annealing method is used ("GenSA" package in R) to find the global optimal points. The use of simulated annealing method reduces the impact of initial values selection.

In our simulations, the average time dimension $\overline{m}_i$ is about 10. The number of subjects was set to be $n = 30$, 50 and 100. The leave-one-subject-out cross validation is applied to select the bandwidths. The simulation repetition is 200 due to heavy computational costs.

The simulation results for simulations 1 and 2 are shown in Tables 4.1 and 4.2. In each table, we compare the estimation accuracy for both the parameter estimators and the single-index function estimator for WI, SGEE and SEE. Specifically, we provide the Monte Carlo standard errors (SE) and empirical asymptotic standard errors – "sandwich" standard errors (SWSE) for the parameter estimators, and the averaged mean squared errors (AMSE) for the single-index function estimator. From both tables, we see that for the parameter estimators, SE and SWSE for SEE are generally smaller than that for SGEE and WI. For the single-index function estimator, AMSE for SEE are generally smaller than that for SGEE and WI. The results indicate that in these simulation settings, SEE has better estimation performance than WI and SGEE, which supports our theoretical findings.

In order to measure the sensitivity of within-subject covariance estimation to the parameter and single-index function estimation, we evaluated the proposed method in simulation 1 with the estimated covariances and when the true covariances are assumed. The comparison is limited to a particular case of $n = 30$ because of the highly computational complexity. Table 4.3 gives the simulation results of the proposed method with the true covariances plugged in (Tru-Cov) and with the estimated covariances (Est-Cov). Similarly to the format of Tables 4.1 and 4.2, we provide SE and SWSE for the parameter estimators and AMSE for the single-index estimator. While the method with the true covariances plugged in has smaller SE and SWSE for the parameter estimators and smaller AMSE for the single-index function estimator, the differences are generally as expected. Therefore, it seems that the estimation accuracy is not very sensitive to the covariances estimation in the proposed method with the given sample size even when the sample size is not large.

4.5 An Example of Real Data Analysis

We apply the proposed method to the Indonesian Children's Health Study (ICHS) data to analyze the problem of vitamin A deficiency in preschool children in the Aceh Province of Indonesia. This longitudinal dataset was analyzed by Zeger and Liang (1991) with

Table 4.1: Empirical comparisons of three methods WI, SGEE and SEE in the case of binary partially linear single-index model with compound symmetry correlation ($\rho = 0.75$). SE = Monte Carlo standard error, SWSE = empirical asymptotic standard error and AMSE = averaged mean squared error. All the values are in percentage.

n	Methods	β_1		β_2		α_1		α_2		α_3		$\gamma(\cdot)$
		SE	SWSE	SE	SWSE	SE	SWSE	SE	SWSE	SE	SWSE	AMSE
30	WI	3.79	3.81	4.39	4.30	2.78	2.76	3.25	3.21	1.91	1.92	1.70
	SGEE	3.71	3.69	4.27	4.20	2.75	2.68	2.87	2.82	1.88	1.91	1.68
	SEE	3.55	3.49	4.20	4.16	2.69	2.65	2.86	2.80	1.81	1.80	1.58
50	WI	3.08	3.13	3.42	3.40	2.29	2.31	2.47	2.44	1.44	1.45	1.36
	SGEE	2.88	2.88	3.42	3.38	2.27	2.24	2.34	2.40	1.42	1.42	1.30
	SEE	2.84	2.82	3.35	3.36	2.15	2.20	2.32	2.34	1.32	1.29	1.20
100	WI	1.35	1.37	1.68	1.65	0.92	0.93	1.17	1.14	0.89	0.87	0.73
	SGEE	1.29	1.25	1.61	1.60	0.92	0.90	1.13	1.14	0.81	0.84	0.65
	SEE	1.21	1.23	1.56	1.59	0.92	0.86	1.09	1.12	0.78	0.79	0.60

generalized linear models. In this dataset, the binary response variable $Y_{ij} = 1$ indicates that the i^{th} child suffers from respiratory infection at the j^{th} visit, and 0 otherwise. The predictor variable is 'Xerop' which indicates the presence (1) or absence (0) of xerophthalmia. It is an ocular symptom of the chronic vitamin A deficiency. Other covariates are 'Time' which indicates the time passed by; 'Age', which, by centering at 36, indicates the baseline of children's age in months; 'Height', which, by centering at 90%, represents the percent of the National Center for Health Statistics (NCHS) standard. 'Height' indicates the long-term nutritional status. Chowdhury and Sinha (2015) applied GPLSIM to analyze ICHS data with different estimating equations which are specifically for binary responses. However, the dataset they used has a different sample size. In the data that we are analyzing, the sample size $n = 275$ and the observation time m_i ranges from 1 to 6. While in Chowdhury and Sinha (2015) the sample size $n = 137$ and $m_i = m = 4$.

Assume that the marginal mean response $p_{ij} = \mathrm{E}(Y_{ij})$ and the covariates has the fol-

Table 4.2: Empirical comparisons of three methods WI, SGEE and SEE in the case of Poisson partially linear single-index model with compound symmetry correlation ($\rho = 0.60$). SE = Monte Carlo standard error, SWSE = empirical asymptotic standard error and AMSE = averaged mean squared error. All the values are in percentage.

n	Methods	β_1		β_2		α_1		α_2		α_3		$\gamma(\cdot)$
		SE	SWSE	SE	SWSE	SE	SWSE	SE	SWSE	SE	SWSE	AMSE
30	WI	4.10	4.03	4.80	4.88	3.58	3.47	3.01	3.14	3.09	3.03	2.72
	SGEE	4.09	4.07	4.06	4.03	3.34	3.37	2.83	2.77	2.38	2.32	2.37
	SEE	3.45	3.38	3.81	3.71	3.11	3.19	2.53	2.51	2.59	2.56	2.29
50	WI	3.20	3.16	3.81	3.75	2.85	2.91	2.43	2.48	2.47	2.44	2.19
	SGEE	3.15	3.15	3.30	3.26	2.57	2.59	2.31	2.24	2.22	2.19	1.94
	SEE	2.72	2.65	2.99	3.96	2.40	2.41	2.03	2.12	2.01	2.06	1.80
100	WI	2.27	2.23	2.73	2.76	2.00	2.08	1.68	1.71	1.70	1.67	1.56
	SGEE	2.21	2.18	2.36	2.36	1.86	1.87	1.52	1.49	1.48	1.53	1.38
	SEE	1.91	1.91	2.12	2.11	1.80	1.77	1.40	1.38	1.44	1.47	1.25

Table 4.3: Empirical comparisons of SEE under true (Tru-Cov) and estimated (Est-Cov) covariance estimation methods in the case of binary partially linear single-index model with compound symmetry correlation ($\rho = 0.75$). GEE-TC and SEE are estimation methods with the true covariance and semiparametrically estimated covariance (proposed) respectively. SE = Monte Carlo standard error, SWSE = empirical asymptotic standard error and AMSE = averaged mean squared error. All the values are in percentage.

n	Methods	β_1		β_2		α_1		α_2		α_3		$\gamma(\cdot)$
		SE	SWSE	SE	SWSE	SE	SWSE	SE	SWSE	SE	SWSE	AMSE
30	Tru-Cov	3.10	3.04	4.43	4.41	2.15	2.19	2.23	2.16	1.81	1.74	1.40
	Est-Cov	3.55	3.49	4.20	4.16	2.69	2.65	2.86	2.80	1.81	1.80	1.58

lowing form

$$\text{logit}(p_{ij}) = \beta_0 + \beta_1 \text{Xerop}_{ij} + \gamma(\alpha_1 \text{Time}_{ij} + \alpha_2 \text{Age}_{ij} + \alpha_3 \text{Height}_{ij}).$$

The compound symmetry working correlation structure is also assumed. The parameter estimates by WI, SGEE and SEE are given in Table 4.4. The SEE estimates are slightly different from others and have the smallest estimated standard errors, while the WI estimates have the largest estimated standard errors. Furthermore, the coefficient of covariate

Age is seen to be significant by SEE. However, it is not significant by WI and SGEE. The single-index function estimates evaluated at the observed values after smoothing with their 95% bootstrap confidence bands by WI, SGEE and SEE are shown in Figure 4.1. There is a general increasing trend by SEE. However, this trend is not obvious by WI and SGEE. Furthermore, comparing with WI and SGEE, the confidence band for the single-index function by SEE is generally narrower. Table 4.4 shows that there is significant evidence that the presence of xerophthalmia is positively related to respiratory infection. That is, the presence of xerophthalmia is more likely for children to have vitamin A deficiency. Moreover, from Table 4.4 and Figure 4.1 we conclude that in general it is statistically significant that older and taller children are less likely to get vitamin A deficiency. However, measuring time is not a statistically significant factor for children's vitamin A deficiency. These conclusions are generally in line with that of Zeger and Liang (1991) and Chowdhury and Sinha (2015) obtained with different approaches.

Table 4.4: Parameter estimates and their standard errors for the ICHS data by WI, SGEE and SEE.

Model	WI		SGEE		SEE	
Estimates	Estimate	SE	Estimate	SE	Estimate	SE
β_0	−2.220	0.624	−2.133	0.615	−2.537	0.604
β_1	1.775	0.530	1.690	0.526	1.701	0.509
α_1	0.541	0.587	0.455	0.551	0.748	0.530
α_2	0.041	0.041	0.021	0.038	−0.118	0.031
α_3	−0.840	0.195	−0.890	0.196	−0.653	0.192

4.6 Conclusions

A semiparametrically efficient estimation method for the longitudinal generalized partially linear single-index model was proposed in this chapter, which extends the results

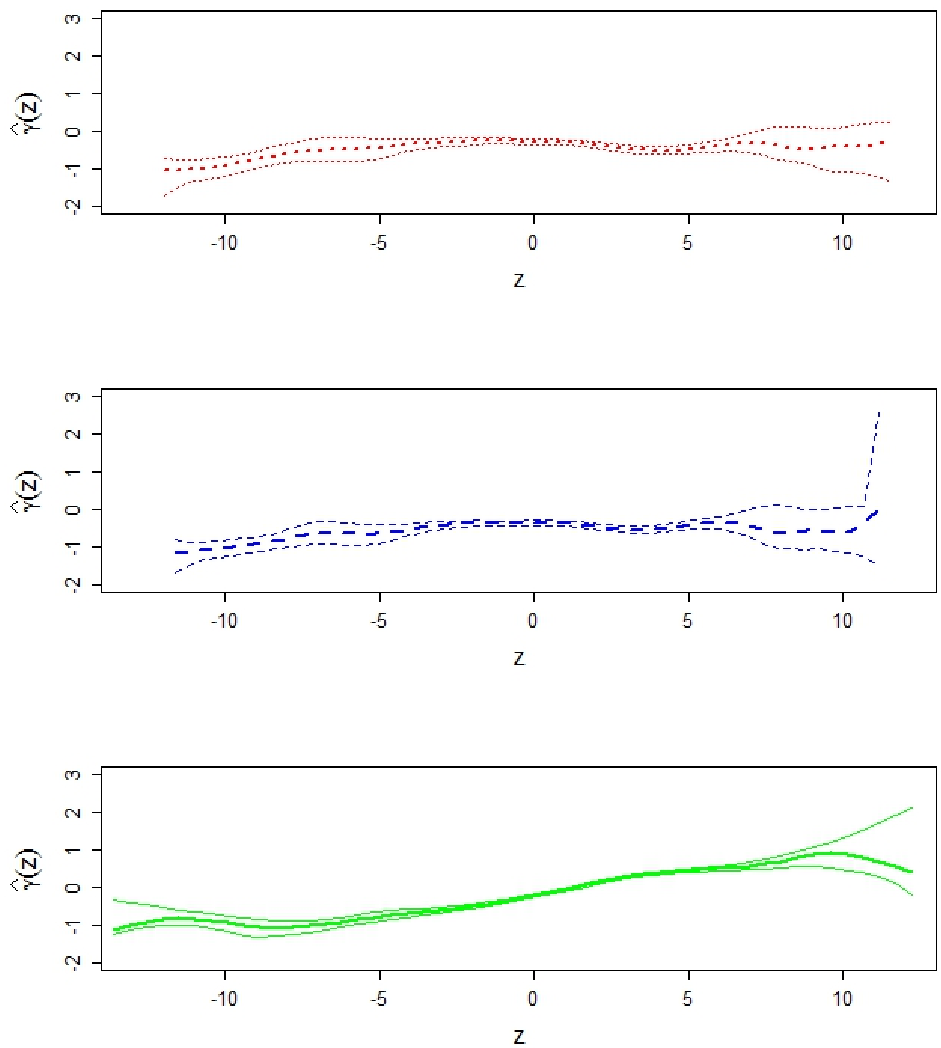

Figure 4.1: Estimated single-index function and its 95% bootstrap confidence band for the ICHS data by three estimation methods. The WI estimated single-index function and its 95% bootstrap confidence band are shown with red dotted curves. The SGEE estimated single-index function and its 95% bootstrap confidence band are shown with blue dot-dashed curves. The SEE estimated single-index function and its 95% bootstrap confidence band are shown with green solid curves.

for the estimation of the longitudinal partially linear single-index model to the generalized models. Asymptotic properties of the parameter estimators and the single-index function estimator were derived. The estimated parameters were shown to be semiparametrically efficient and the single-index function estimator generally has a smaller variance when compared with that of existing methods. Simulation studies and real data analysis were performed that support the methodology and theoretical results.

5. SUMMARIES AND FURTHER STUDIES

5.1 Summary

In this dissertation, we investigated efficient estimation methods in partially linear single-index models and generalized partially linear single-index models for longitudinal data. In these models, one or more covariates can be modeled parametrically and more than one covariate should be modeled nonparametrically with the response variable. The nonparametric component is indexed by a single-index function.

The parameters in both the parametric component and the within single-index function are estimated when the single-index function is fixed. Meanwhile, the single-index function is estimated when all parameters are fixed. The estimation procedure is processed in an iterative way. By taking into consideration the within-subject correlation properly, the parameter estimators can reach semiparametric efficient bound when the covariances are correctly specified. We also studied the asymptotic properties of the nonparametric single-index function. We presented the asymptotic convergence rate, bias and variance for the estimator. We compared the asymptotic results with those for existing methods and showed that the asymptotic variance is not only more efficient, but also minimized when within-subject covariances are correctly specified.

Simulation studies under various settings were conducted. The simulation results support our theoretical development. Two real data analyses using PLSIM and one data analysis using GPLSIM were performed to demonstrate the proposed methods.

5.2 Further Studies

There are several directions of further studies on this topic. Further research problems include studying possibly superior within-subject covariance estimation and direct precision matrix estimation since the efficiency and convergence rates of the estimated covariance or precision matrix can affect the finite sample performance of parameter and single-index function estimators. Another possible topic is variable selection for (generalized) partially linear single-index models for longitudinal data since it is sometimes difficult to decide whether to keep or discard some covariates in practice. Furthermore, for the retained covariates, there is still a problem of distinguishing them from the parametric component to the nonparametric component. All this is worth investigating in future research.

REFERENCES

Bickel, P. J., Klaassen, C. A., Ritov, Y., and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. John Hopkins University Press, Baltimore.

Buja, A., Hastie, T., and Tibshirani, R. (1989). Linear smoothers and additive models. *The Annals of Statistics*, 17:453–510.

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.

Chan, J. S. (2014). Analysis of correlation structures using generalized estimating equation approach for longitudinal binary data. *Journal of Data Science*, 12:293–305.

Chen, B. and Zhou, X.-H. (2011). Doubly robust estimates for binary longitudinal data analysis with missing response and missing covariates. *Biometrics*, 67:830–842.

Chen, H. and Shiau, J.-J. H. (1994). Data-driven efficient estimators for a partially linear model. *The Annals of Statistics*, 22:211–237.

Chen, J., Li, D., Liang, H., and Wang, S. (2015). Semiparametric GEE analysis in partially linear single-index models for longitudinal data. *The Annals of Statistics*, 43:1682–1715.

Chen, K. and Jin, Z. (2006). Partial linear regression models for clustered data. *Journal of the American Statistical Association*, 101:195–204.

Chen, T. and Parker, T. (2014). Semiparametric efficiency for partially linear single-index regression models. *Journal of Multivariate Analysis*, 130:376–386.

Cheng, G., Zhou, L., and Huang, J. Z. (2014). Efficient semiparametric estimation in generalized partially linear additive models for longitudinal/clustered data. *Bernoulli*, 20:141–163.

Chowdhury, S. K. and Sinha, S. K. (2015). Semiparametric marginal models for binary longitudinal data. *International Journal of Statistics and Probability*, 4:107–120.

Diggle, P., Heagerty, P., Liang, K.-Y., and Zeger, S. (2013). *Analysis of Longitudinal Data*. Oxford University Press, Oxford.

Fan, J., Huang, T., and Li, R. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of the American Statistical Association*, 102:632–641.

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

Fan, J. and Wu, Y. (2008). Semiparametric estimation of covariance matrices for longitudinal data. *Journal of the American Statistical Association*, 103:1520–1533.

Grace, Y. Y., He, W., and Liang, H. (2009a). Analysis of correlated binary data under partially linear single-index logistic models. *Journal of Multivariate Analysis*, 100:278–290.

Grace, Y. Y., He, W., and Liang, H. (2009b). Semiparametric marginal and association regression methods for clustered binary data. *Annals of the Institute of Statistical Mathematics*, 100:278–290.

Hardle, W., Hall, P., and Ichimura, H. (1993). Optimal smoothing in single-index models. *The Annals of Statistics*, 21:157–178.

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.

Hu, Z., Wang, N., and Carroll, R. J. (2004). Profile-kernel versus backfitting in the partially linear models for longitudinal/clustered data. *Biometrika*, 91:251–262.

Huang, J. Z., Chen, M., Maadooliat, M., and Pourahmadi, M. (2012). A cautionary note on generalized linear models for covariance of unbalanced longitudinal data. *Journal of*

Statistical Planning and Inference, 142:743–751.

Huang, J. Z., Zhang, L., and Zhou, L. (2007). Efficient estimation in marginal partially linear models for longitudinal/clustered data using splines. *Scandinavian Journal of Statistics*, 34:451–477.

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

Kim, S. and Zhao, Z. (2012). Unified inference for sparse and dense longitudinal models. *Biometrika*, 100:203–212.

Kress, R., Maz'ya, V., and Kozlov, V. (1989). *Linear Integral Equations*. Springer, Verlag.

Lai, P., Li, G., and Lian, H. (2013). Quadratic inference functions for partially linear single-index models with longitudinal data. *Journal of Multivariate Analysis*, 118:115–127.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38:963–974.

Leisch, F., Weingessel, A., and Hornik, K. (1998). On the generation of correlated artificial binary data. Technical report, WU Vienna University of Economics and Business.

Li, G., Zhu, L., Xue, L., and Feng, S. (2010). Empirical likelihood inference in partially linear single-index models for longitudinal data. *Journal of Multivariate Analysis*, 101:718–732.

Li, Y. (2011). Efficient semiparametric regression for longitudinal data with nonparametric covariance estimation. *Biometrika*, 98:1–16.

Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics*, 38:3321–3351.

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.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 1:13–22.

Lin, D. Y. and Wei, L.-J. (1989). The robust inference for the cox proportional hazards model. *Journal of the American statistical Association*, 84:1074–1078.

Lin, W. and Kulasekera, K. (2007). Identifiability of single-index models and additive-index models. *Biometrika*, 94:496–501.

Lin, X. and Carroll, R. J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association*, 95:520–534.

Lin, X. and Carroll, R. J. (2001). Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association*, 96:1045–1056.

Lipsitz, S. R., Laird, N. M., and Harrington, D. P. (1991). Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. *Biometrika*, 78:153–160.

Ma, S., Liang, H., and Tsai, C.-L. (2014). Partially linear single index models for repeated measurements. *Journal of Multivariate Analysis*, 130:354–375.

Magnus, J. R. and Neudecker, H. (1995). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, New York.

Mardia, K. V. (1970). *Families of Bivariate Distributions*. Griffin, London.

Nelder, J. A. and Baker, R. J. (2004). *Generalized Linear Models*. John Wiley & Sons, New York.

Pepe, M. and Anderson, G. L. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics-Simulation and Computation*, 23:939–951.

Ping, T., Lin, Y., and Liugen, X. (2010). Asymptotic properties of estimators in partially linear single-index model for longitudinal data. *Acta Mathematica Scientia*, 30:677–687.

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika*, 86:677–690.

Pourahmadi, M. (2000). Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika*, 87:425–435.

Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics*, 44:1033–1048.

Qu, A. and Li, R. (2006). Quadratic inference functions for varying-coefficient models with longitudinal data. *Biometrics*, 62:379–391.

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

Severini, T. A. and Staniswalis, J. G. (1994). Quasi-likelihood estimation in semiparametric models. *Journal of the American Statistical Association*, 89:501–511.

Severini, T. A. and Tripathi, G. (2001). A simplified approach to computing efficiency bounds in semiparametric models. *Journal of Econometrics*, 102:23–66.

Uppal, S., Verma, S., and Dhot, P. (2003). Normal values of CD4 and CD8 lymphocyte subsets in healthy indian adults and the effects of sex, age, ethnicity, and smoking. *Cytometry Part B: Clinical Cytometry*, 52:32–36.

Van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

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

Wang, N., Carroll, R. J., and Lin, X. (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data. *Journal of the American Statistical Association*, 100:147–157.

Wang, X., Zhang, J., Yu, L., and Yin, G. (2015). Generalized partially linear single-index model for zero-inflated count data. *Statistics in Medicine*, 34:876–886.

Wu, W. B. and Pourahmadi, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika*, 90:831–844.

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

Xia, Y., Tong, H., and Li, W. (1999). On extended partially linear single-index models. *Biometrika*, 86:831–842.

Xu, H. and Rahman, S. (2004). A generalized dimension-reduction method for multi-dimensional integration in stochastic mechanics. *International Journal for Numerical Methods in Engineering*, 61:1992–2019.

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

Zeger, S. L. and Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42:121–130.

Zeger, S. L. and Liang, K.-Y. (1991). Feedback models for discrete and continuous time series. *Statistica Sinica*, 1:51–64.

Zeger, S. L., Liang, K.-Y., and Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, 44:1049–1060.

Zeger, S. L., Liang, K.-Y., and Self, S. G. (1985). The analysis of binary longitudinal data with time-independent covariates. *Biometrika*, 72:31–38.

Zhu, L. and Xue, L. (2006). Empirical likelihood confidence regions in a partially linear single-index model. *Journal of the Royal Statistical Society: Series B*, 68:549–570.