Efficient Estimation in Partially Linear Single-Index Models for Longitudinal Data

Quan Cai and Suojin Wang

Dec 17, 2016

Abstract

In this paper, we consider the estimation of both the parameters and the nonparametric link function in partially linear single-index models for longitudinal data which may be unbalanced. In particular, a new three-stage approach is proposed to estimate the nonparametric link function using marginal kernel regression and the parametric components with generalized estimating equations. The resulting estimators properly account for the within-subject correlation. We show that the parameter estimators are asymptotically semiparametric efficient. We also show that the asymptotic variance of the link function estimator is 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. The finite sample performance of the proposed method is demonstrated by simulation studies. In addition, two real data examples are analyzed to illustrate the methodology.

Keywords: generalized estimating equation; kernel method; longitudinal data; partially linear single-index models; semiparametric efficiency.
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 Liang & Zeger (1986), Zeger & Liang (1986) and Diggle et al. (2002), 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., Carroll et al. (1997), Lin & Carroll (2000, 2001, 2006) and Zeger & Diggle (1994). One of the applications of semiparametric models is the partially linear single-index model

\[ Y_i = X_i \beta + \phi(Z_i \theta) + e_i \]

for \( i = 1, \ldots, n \), where \( \beta \) and \( \theta \) are parameters and \( \phi(\cdot) \) is an unknown link function, all to be estimated. 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 this model for the independent and identically distributed case. In this case, Liang et al. (2010) obtained the semiparametrically efficient profile least-squares estimators of regression coefficients. They also employed the smoothly clipped absolute deviation penalty (SCAD) approach to simultaneously select variables and estimate regression coefficients. Chen & Thomas (2014) specifically calculated semiparametric information bounds for a partially linear single-index model using a simple method developed by Severini & Tripathi (2001). Hu et al. (2004) and Li et al. (2010) discussed how partially linear single-index models are applied to longitudinal data.

However, how to properly estimate the parameters and link function in partially linear single-index models 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 partially linear single-index models for both sparse and dense unbalanced longitudinal data. Hereafter the method is referred as SGEE. They pointed out that the convergence rate and the asymptotic variances of the proposed method’s estimators are substantially different for the sparse and dense longitudinal cases. However, their parameter estimators are generally not semiparametric efficient. For the estimation of the link 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 & 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 link function. Wang (2003) re-examined the profile-kernel GEE methods. She pointed out that when kernel weights are used to control the bias for link 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 link function estimation. As a result, the profile-kernel GEE method is not optimal regarding to the asymptotic variance of the link 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 paper we focus on the estimation efficiency of both the parameters and the unknown link function in partially linear single-index models with a general unbalanced sparse longitudinal data setting. Specifically, first we use the working independence (WI) kernel GEE for the link function estimation by fixing the parameters as known and use the WI least square estimation for the parameters by fixing the link 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 link 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 link function estimator is more efficient than the link 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 $X_i$, $Z_i$ or $e_i$.

The rest of the paper is organized as follows. In Section 2 we describe an estimation procedure to obtain new estimators for the parameters and the unknown link function in longitudinal partially linear single-index models. Section 3 provides some asymptotic results for the estimators where the asymptotic variance, asymptotic bias and convergence rates are presented. We show that the proposed link function estimator is more efficient than the working independence estimator. Moreover, both the parameter and link 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. Section 4 gives some finite sample simulations to compare the proposed method with some existing ones. In Section 5 we apply the proposed method in two real data examples. In Section 6 we conclude the paper and discuss some possible extensions for future research. Regularity conditions and some technical arguments are given in the Appendices A–E.

2 Methodology

Rewrite the longitudinal partially linear single-index model as follows:

$$Y_{ij} = X_{ij}^T \beta + \phi(Z_{ij}^T \theta) + e_{ij}$$

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 $\beta$ and $\theta$ of dimension $p$ and $q$ respectively are to be estimated. The univariate unknown link function $\phi(\cdot)$ is also to be estimated. The continuous response variable $Y_{ij}$ and covariates $(X_{ij}^T, Z_{ij}^T)$ are observed at time $t_{ij}$ and $e_{ij}$ is a random error. For simplicity, in this paper we assume that $Z_{ij}$ are continuous random variables and $X_{ij}$ 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 $\Sigma_i = \text{cov}(E_i)$ with $E_i^T = (e_{i1}, \ldots, e_{im_i})$. 
The main task is to estimate the true parameters \( \Theta_0 = (\beta_0^T, \theta_0^T)^T \) and the true unknown nonparametric link function \( \phi_0(\cdot) \) in Equation (1). To guarantee identifiability, we assume that the Frobenius norm of \( \theta_0 \) is 1 with the first element of \( \theta_0 \) being positive as in Liang et al. (2010). To simplify the notation, we denote \( N = \sum_{i=1}^n m_i \), \( Y_i = (Y_{i1}, \ldots, Y_{im_i})^T \), \( X_i = (X_{i1}, \ldots, X_{im_i})^T \), \( Z_i = (Z_{i1}, \ldots, Z_{im_i})^T \), \( \phi(Z_i \theta) = (\phi(Z_{i1}^T \theta), \ldots, \phi(Z_{im_i}^T \theta))^T \) and \( \Theta = (\beta^T, \theta^T)^T \). 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 \( G \), the domain of \( Z_{ij}^T \theta \) defined in Assumption 4. Given \( \Theta \), the one-step estimate of \( \phi(\cdot) \) and \( \phi^{(1)}(\cdot) \), denoted by WI kernel GEE, is obtained by solving the following estimating equation

\[
 n^{-\frac{1}{2}} \sum_{i=1}^n L'(u)^T K_h(u) \{ Y_i - X_i \beta - L'(u)a \} = 0, \tag{2}
\]

where \( L'(u) \) is an \( m_i \times 2 \) matrix with the \((j,k)^{th}\) element \((Z_{ij}^T \theta - u)^{k-1} \) and \( K_h(u) = \text{diag}\{K_h(Z_{ij}^T \theta - u)\} \) with the \((l,l)^{th}\) entry being \( K_h(Z_{il}^T \theta - u) \). Here \( h \) is a proper bandwidth to be discussed later. The obtained estimates are \( \tilde{a}(u, \Theta) = \{ \tilde{\phi}(u, \Theta), \tilde{\phi}^{(1)}(u, \Theta) \}^T \).

2. With the estimated link function, the one-step estimates of \( \Theta \) which ignore the within-subject correlation structure are obtained by minimizing

\[
 n^{-\frac{1}{2}} \sum_{i=1}^n \{ Y_i - X_i \beta - \tilde{\phi}(Z_i \theta, \Theta) \}^T \{ Y_i - X_i \beta - \tilde{\phi}(Z_i \theta, \Theta) \}. \tag{3}
\]

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

3. Let \( V_i \) be the estimated working covariance matrix for subject \( i \), \( i = 1, \ldots, n \). From the initial estimates, we get the residual term \( \tilde{e}_{ij} = Y_{ij} - X_{ij}^T \beta - \tilde{\phi}(Z_{ij}^T \theta) \) as an estimate of \( e_{ij} \). A semiparametric variance correlation decomposition approach is applied to estimate the true covariance \( \Sigma_i \).

4. Re-estimate \( \phi(\cdot) \). Given \( \Theta \), let \( M_{ij} = [1_j, 1_j(Z_{ij}^T \theta - u)] \) be the \( m_i \times 2 \) matrix, where \( 1_j \) denotes the indicator vector with \( j^{th} \) entry equal to 1, and 0 elsewhere. Define
\[
\{ \hat{\phi}(u, \Theta), \hat{\phi}^{(1)}(u, \Theta) \} = (b_1, b_2), \text{ where } (b_1, b_2) \text{ solves the kernel-weighted estimating equation}
\]
\[
n^{-\frac{1}{2}} \sum_{i=1}^{n} \sum_{j=1}^{m_i} K_h(Z_{ij}^T \theta - u) M_{ij}^T V_i^{-1} [Y_i - \mu^*\{u, X_i, Z_i, \Theta, \hat{\phi}_c(\cdot), b_1, b_2\}] = 0, \quad (4)
\]
where \(\mu^*\{u, X_i, Z_i, \Theta, \hat{\phi}_c(\cdot), b_1, b_2\}\) is an \(m_i \times 1\) vector with the \(k^{th}\) element being
\[
X_{ik}^T \beta + I(k = j)\{b_1 + b_2(Z_{ij}^T \theta - u)\} + I(k \neq j)\hat{\phi}_c(Z_{ik}^T \theta, \Theta)
\]
and \(h\) is another proper bandwidth. Here \(\hat{\phi}_c(\cdot)\) is the current estimate of \(\phi(\cdot)\) and \(I(\cdot)\) is the indicator function.

5. Re-estimate the parameters \(\Theta\) using the following GEE
\[
n^{-\frac{1}{2}} \sum_{i=1}^{n} \frac{\partial \{X_i \beta + \hat{\phi}(Z_i, \Theta)\}}{\partial \Theta}^T V_i^{-1} \{Y_i - X_i \beta - \hat{\phi}(Z_i, \Theta)\} = 0. \quad (5)
\]
The solutions of (5) are updated parameter estimates. We then assign the updated parameter estimates values to Step 4 for iteration. The final estimates of the parameters \(\hat{\Theta} = (\hat{\beta}^T, \hat{\theta}^T)^T\) and the link function estimate \(\hat{\phi}(u, \hat{\Theta}) = \hat{\phi}(u)\) 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 link function and the residuals. The resulting estimates are \(\sqrt{n}\)-consistent (Lin & 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 link 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 1 The unknown link function \(\phi_0(u)\) in the partially linear single-index model does not depend on parameters \(\Theta_0\). However, while in the estimation steps of the unknown link function, the estimated link function depends on the parameter estimates in the current step.
When the iteration stops, we have the final estimate for the link function \( \hat{\phi}(u, \hat{\Theta}) = \hat{\phi}(u) \). For clarity, we use \( \hat{\phi}(u, \Theta) \) to emphasize its dependence on \( \Theta \) in the estimation procedure and use \( \hat{\phi}(u) \) for simplicity when no confusion arises.

**Remark 2** Given the residual estimates \( \tilde{e}_{ij} \) from the initial estimation step, there are several ways to obtain the working covariance matrix estimates \( V_i \) for longitudinal data, such as Wu & Pourahmadi (2003). One of the covariance modeling techniques is based on variance-correlation decomposition. Some recent works include Fan et al. (2007), Fan & 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 Theoretical properties

In our theoretical development, we assume that the number of clusters \( n \) goes to \( \infty \) and the number of measurements \( m_i \) is bounded for \( i = 1, \ldots, n \). Wang et al. (2005) proposed the semiparametric efficient marginal estimators for the generalized partially linear models under the multivariate normal assumption. Their theoretical and empirical investigations were focused on the parameter estimation part. Inspired by Lin & Carroll (2001) and Wang et al. (2005), we first derive the semiparametric efficient score and information bound for partially linear single-index models 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 link 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 1–6 with some detailed proofs shown in Appendices B–E.

3.1 Theory for semiparametric efficiency

First we define the $L_2$ norm of a square-integrable function $f(u)$ on $G$ by $\{\int_G \{f(u)\}^2 du\}^{1/2}$, where $G$ is defined in Assumption 4 in Appendix A. Besides, we denote $(X_i, Z_i) = [X_i, \{\phi^{(1)}(Z_i\theta_0) \otimes 1_q^T \} \odot Z_i]$, where $\otimes$ is the Kronecker product and $\odot$ is the component-wise product. Using the notation in Appendix B, we have

$$\tilde{X}_{i,e} = X_i - \psi_{n, \beta}(Z_i\theta_0), \quad \tilde{Z}_{i,e} = Z_i - \psi_{n, \theta}(Z_i\theta_0),$$

where $\psi_{n, \beta}(Z_i\theta_0) = \{\psi_{n, \beta_1}(Z_i\theta_0), \ldots, \psi_{n, \beta_p}(Z_i\theta_0)\}$ is an $m_i \times p$ matrix with $\psi_{n, \beta_l}(\cdot) \in L_2$, $l = 1, \ldots, p$. It satisfies

$$\frac{1}{n} \sum_{i=1}^n E \left( \tilde{X}_{i,e}^T \Sigma_i^{-1} \kappa_n(Z_i\theta_0) \right) = 0$$

for all $\kappa_n(\cdot) \in L_2$. Here $\kappa_n^T(Z_i\theta_0) = \{\kappa_n(Z_{i1}\theta_0), \ldots, \kappa_n(Z_{im}\theta_0)\}$. Besides, $\psi_{n, \theta}(Z_i\theta_0) = \{\psi_{n, \theta_1}(Z_i\theta_0), \ldots, \psi_{n, \theta_q}(Z_i\theta_0)\}$ is an $m_i \times q$ matrix with $\psi_{n, \theta_l}(\cdot) \in L_2$, $l = 1, \ldots, q$. It also satisfies

$$\frac{1}{n} \sum_{i=1}^n E \left( \tilde{Z}_{i,e}^T \Sigma_i^{-1} \kappa_n(Z_i\theta_0) \right) = 0$$

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 $\Theta$ is (see Appendix B)

$$S_e = \sum_{i=1}^n (\tilde{X}_{i,e}, \tilde{Z}_{i,e})^T \Sigma_i^{-1} \{Y_i - X_i\beta_0 - \phi_0(Z_i\theta_0)\}.$$  

We now present our first theorem below. Its proof is given in Appendix B.

**THEOREM 1** If Assumptions 4–5 in Appendix A hold, the semiparametric information bound for estimating $\Theta$ at $\Theta_0$ is

$$R_e = \lim_{n \to \infty} \left\{ \frac{1}{n} E\{S_e S_e^T\} \right\}^{-1} = \lim_{n \to \infty} \left[ \frac{1}{n} \sum_{i=1}^n E\{ (\tilde{X}_{i,e}, \tilde{Z}_{i,e})^T \Sigma_i^{-1}(\tilde{X}_{i,e}, \tilde{Z}_{i,e}) \} \right]^{-1}. $$
From Kress et al. (1989), each element of $\psi_{n, \beta}(\cdot)$ and $\psi_{n, \theta}(\cdot)$ solves the Fredholm integral equation of the second kind, which is shown in Equations (24) and (26). In Appendix D we use these equations to show that the proposed parameter estimators reach the semiparametric information bound.

### 3.2 Theory for the parameter estimators

We now study the asymptotic distribution of the estimators $\hat{\Theta}$ and show that they reach the semiparametric information bound. Define

$$
\varphi_{n, \beta}(u, \hat{\Theta}) = -\frac{\partial \hat{\phi}(u, \Theta)}{\partial \beta} \bigg|_{\Theta = \hat{\Theta}} \rightarrow \varphi_{\beta}(u),
$$

$$
\varphi_{n, \theta}(u, \hat{\Theta}) = -\frac{\partial \hat{\phi}(u, \Theta)}{\partial \theta} \bigg|_{\Theta = \hat{\Theta}} \rightarrow \varphi_{\theta}(u)
$$

in probability as $n \to \infty$. Moreover, by Assumption 3 on sufficient smoothness of $\phi(\cdot)$, it is readily seen that the convergence is uniform over $u$ in the compact domain $G$. Further define

$$
\Omega_0 = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} E \left\{ (\tilde{X}_i, \tilde{Z}_i)^T V_i^{-1} (\tilde{X}_i, \tilde{Z}_i) \right\},
$$

$$
\Omega_1 = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} E \left\{ (\tilde{X}_i, \tilde{Z}_i)^T V_i^{-1} \Sigma_i V_i^{-1} (\tilde{X}_i, \tilde{Z}_i) \right\},
$$

where $(\tilde{X}_i, \tilde{Z}_i) = \{X_i - \varphi_{\beta}(Z_i, \theta_0), Z_i - \varphi_{\theta}(Z_i, \theta_0)\}$. Assume that $\Omega_0$ and $\Omega_1$ are non-negative definite matrices. Then we have the following theorem a proof of which is given in Appendix E.

**THEOREM 2** If Assumptions 1–6 in Appendix A hold, we have

$$
n^{1/2}(\hat{\Theta} - \Theta_0) \xrightarrow{D} N(0, \Omega_0^{-1} \Omega_1 \Omega_0^{-1})
$$

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

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 $V_i$, $\Omega_0^{-1} \Omega_1 \Omega_0^{-1} - \Omega_i^{-1}$ is a semi-positive definite matrix.
First denote $A \geq 0$ when $A$ is a semi-positive definite matrix. By the extended Cauchy-Schwarz inequality, if

$$
\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \geq 0,
$$
then $\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \geq 0$. Further denote $U_i = (\tilde{X}_i, \tilde{Z}_i)$. Then the problem of minimizing $\Omega^{-1}_0 \Omega_1 \Omega^{-1}_0$ is asymptotically equivalent to minimizing

$$
\left( \sum_{i=1}^n U_i^T V_i^{-1} U_i \right)^{-1} \left( \sum_{i=1}^n U_i^T V_i^{-1} \Sigma_i V_i^{-1} U_i \right) \left( \sum_{i=1}^n U_i^T V_i^{-1} U_i \right)^{-1}.
$$

Since

$$
\begin{bmatrix} \sum_{i=1}^n U_i^T V_i^{-1} \Sigma_i V_i^{-1} U_i \\ \sum_{i=1}^n U_i^T V_i^{-1} U_i \end{bmatrix} = \sum_{i=1}^n U_i^T \left( V_i^{-1} \Sigma_i V_i^{-1} \right) U_i \geq 0,
$$
we have

$$
\sum_{i=1}^n U_i^T V_i^{-1} \Sigma_i V_i^{-1} U_i - \left( \sum_{i=1}^n U_i^T V_i^{-1} U_i \right) \left( \sum_{i=1}^n U_i^T \Sigma_i^{-1} U_i \right)^{-1} \left( \sum_{i=1}^n U_i^T V_i^{-1} U_i \right) \geq 0,
$$
which leads to

$$
\left( \sum_{i=1}^n U_i^T V_i^{-1} U_i \right)^{-1} \left( \sum_{i=1}^n U_i^T V_i^{-1} \Sigma_i V_i^{-1} U_i \right) \left( \sum_{i=1}^n U_i^T V_i^{-1} U_i \right)^{-1} - \left( \sum_{i=1}^n U_i^T \Sigma_i^{-1} U_i \right)^{-1} \geq 0.
$$
The inequality becomes equality if and only if $V_i = c \Sigma_i$ for some $c > 0$. Without loss of generality, we set $c = 1$ here and in the proof of Theorem 5.

Moreover, the asymptotic covariance of the parameter estimators reaches the semiparametric information bound when $V_i = \Sigma_i$. We formally include this result in the following theorem.

**THEOREM 3** Under Assumptions 1–6 in Appendix A, if the covariance matrices are correctly specified, then we have

$$
\psi_{n, \beta}(u) \to \varphi_\beta(u) \quad \text{and} \quad \psi_{n, \theta}(u) \to \varphi_\theta(u)
$$
in probability as \( n \to \infty \) for every \( u \in \mathcal{G} \), where \( \varphi_{\beta}(u) \) and \( \varphi_{\theta}(u) \) are defined in Equation (11). Furthermore, the proposed parameter estimators reach the semiparametric information bound and are thus semiparametric efficient.

A proof of Theorem 3 is given in Appendix D. 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 \( X \) and \( Z \) given the single-index part. Therefore, the asymptotic properties in Theorem 3 generally do not hold for their parameter estimators. This implies that their parameter estimators are generally not semiparametric efficient.

### 3.3 Theory for the link function estimator

To establish the asymptotic distribution theory for the nonparametric link function estimator \( \hat{\phi}(u) \), we first define the following notation. Let \( f_{ij}(\cdot) \) be the density of \( Z_{ij}^\top \theta \) and \( c_{i,j} = \int u^iK^j(u)du \). Of course, as a special case we may assume a common density function, i.e., \( f_{ij}(\cdot) = f(\cdot) \). Furthermore, define

\[
Q_1(u) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{m_i} \sum_{j=1}^{m} v_{ij} f_{ij}(u), \quad Q_2(u) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{m_i} \sum_{j=1}^{m} \xi_{ij} f_{ij}(u),
\]

\[
Q_3(u) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{m_i} \sum_{j=1}^{m} \sigma_{ij} f_{ij}(u), \quad Q_4(u) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{m_i} \sum_{j=1}^{m} \sigma_{ij}^{-1} f_{ij}(u),
\]

where \( v_{ij}, \xi_{ij}, \sigma_{ij} \) and \( \sigma_{ij} \) are the \((j, j)\)th element of \( V_i^{-1}, V_i^{-1} \Sigma_i V_i^{-1}, \Sigma_i^{-1} \) and \( \Sigma_i \) respectively. These \( Q \) functions are simply the limits of some weighted averages. In the special case when \( V_i = \Sigma_i \) for all \( i \), \( Q_1(u) = Q_2(u) = Q_3(u) \).

We now present the following theorem whose proof is given in Appendix C.

**THEOREM 4** If Assumptions 1–6 in Appendix A hold, we have

\[
\sqrt{nh} \{ \hat{\phi}(u) - \phi(u) - c_{2,1} b(u) h^2 \} \xrightarrow{D} N(0, \sigma^2(u)),
\]

where \( b(u) \) satisfies

\[
b(u) + \int b(w) \eta(u, w)dw = \frac{1}{2} \phi^{(2)}_0(u)
\]
with
\[
\eta(u, w) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} v_{ij} f_{ijk}(u, w)
\]
and \(\sigma^2_\xi(u) = c_{0,2}Q_2(u)/Q_1^2(u)\).

In the following we show that the asymptotic variance of \(\hat{\phi}(u)\) is minimized if and only if \(V_i = \Sigma_i, i = 1, \ldots, n\). Define \(F_i^{1/2}(u)\) as a \(m_i \times m_i\) diagonal matrix with the \((j, j)^{th}\) element being \(f_{ij}^{1/2}(u)\) and \(F_i^{-1/2}(u)\) as the inverse matrix of \(F_i^{1/2}(u)\). Further define \(G_i^{-1}(u) = F_i^{1/2}(u)V_i^{-1}F_i^{-1/2}(u)\) and \(H_i(u) = F_i^{-1/2}(u)\Sigma_iF_i^{-1/2}(u)\). Then the problem of minimizing \(\sigma^2_\xi(u)\) is asymptotically equivalent to minimizing
\[
\frac{\sum_{i=1}^{n} \text{tr}\{G_i^{-1}(u)H_i(u)G_i^{-1}(u)\}}{\left[\sum_{i=1}^{n} \text{tr}\{G_i^{-1}(u)\}\right]^2},
\]
where \(\text{tr}(A)\) is the trace of a matrix \(A\). The extended Cauchy-Schwarz inequality (Magnus & Neudecker (1999), p. 227, Theorem 2) shows that
\[
\{\text{tr}(A^T B)\}^2 \leq \{\text{tr}(A^T A)\}\{\text{tr}(B^T B)\}
\]
for any square matrices \(A\) and \(B\) with equality if and only if \(A = cB\) for some constant \(c\).

Now let \(A\) and \(B\) be two \(N \times N\) block diagonal matrices
\[
A = \text{diag}(A_1, \ldots, A_n) \quad \text{and} \quad B = \text{diag}(B_1, \ldots, B_n),
\]
where \(A_i = H_i^{-1/2}(u)\) and \(B_i = H_i^{1/2}(u)G_i^{-1}(u)\) for \(i = 1, \ldots, n\). Then we have
\[
\left[\sum_{i=1}^{n} \text{tr}\{G_i^{-1}(u)\}\right]^2 \leq \sum_{i=1}^{n} \text{tr}\{H_i^{-1}(u)\} \sum_{i=1}^{n} \text{tr}\{G_i^{-1}(u)H_i(u)G_i^{-1}(u)\}.
\]
It is equivalent to
\[
\frac{\sum_{i=1}^{n} \text{tr}\{G_i^{-1}(u)H_i(u)G_i^{-1}(u)\}}{\left[\sum_{i=1}^{n} \text{tr}\{G_i^{-1}(u)\}\right]^2} \geq \frac{1}{\sum_{i=1}^{n} \text{tr}\{H_i^{-1}(u)\}}.
\]
Without loss of generality, let \(c = 1\). Then the equality holds if and only if \(A = B\). It leads to \(H_i(u) = G_i(u)\), which is equivalent to \(V_i = \Sigma_i\) for \(i = 1, \ldots, n\). The result is formally presented in the following theorem.
THEOREM 5 If Assumptions 1–6 in Appendix A hold, the asymptotic variance of \( \hat{\phi}(u) \) is minimized if and only if \( V_i = \Sigma_i \) i.e., when the working covariance matrices are correctly specified. In this case, the estimated link function \( \hat{\phi}(u) \) has the following asymptotic normality property

\[
\sqrt{nh}\{\hat{\phi}(u) - \phi_0(u) - c_{2,1}b(u)h^2\} \overset{D}{\to} N(0, \sigma^2_\phi(u)),
\]

where \( \sigma^2_\phi(u) = c_{0,2}Q^{-1}_3(u) \).

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

\[
\mathrm{var}(\tilde{\phi}(u)) = \frac{1}{nh}c_{0,2}Q^{-1}_4(u).
\]

Comparing (15) and (16), the SGEE estimator and the proposed estimator of the link 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 \( \sigma^2_{ij} \geq \frac{1}{\sigma_{i,ij}} \) for all \( i, j \). We consider \( j = 1 \) without loss of generality. Writing \( \Sigma_i \) in the block matrix form, we have

\[
\Sigma_i = \begin{pmatrix}
\Sigma_{i,11} & \Sigma_{i,12} \\
\Sigma_{i,12}^T & \Sigma_{i,22}
\end{pmatrix} \succeq 0.
\]

By the extended Cauchy-Schwartz inequality, \( \sigma_{i,11}^2 = (\sigma_{i,11} - \Sigma_{i,12}\Sigma_{i,22}^{-1}\Sigma_{i,12}^T)^{-1} \geq (\sigma_{i,11})^{-1} \). Therefore, when the working covariance matrices are correctly specified, the proposed link 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 6 If Assumptions 1–6 in Appendix A hold, when the working covariance matrices are correctly specified, the asymptotic variance of the proposed estimator \( \hat{\phi}(u) \) has a variance less than or equal to that of the SGEE estimator for the link function.
Remark 3 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.

4 Simulation studies

4.1 Simulation setup

In our simulation studies, we considered the partially linear single index model in (1). The true parameter settings are similar to Chen et al. (2015). Parameters $\beta$ and $\theta$ are of dimensions 2 and 3 with true values $\beta_0 = (2, 1)^T$ and $\theta_0 = (2/3, 1/3, 2/3)^T$. The covariates $X$ and $Z$ are jointly generated from multivariate normal distribution with mean zero, standard deviation 1 and correlation 0.1. The true link 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, which corresponds to an average time dimension of $\bar{m}_i = 10$. 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 $e_i(t_j) = e_{ij}$, $\text{cor}(e_i(t_1), e_i(t_2)) = \gamma^{|t_2-t_1|}, t_1 \neq t_2$. 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 $e_{ij}$ are generated from a Gaussian process with mean 0, variance function $\text{var}\{e_i(t)\} = 0.25 \exp(t/12)$. 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 \( \text{cor}(e_i(t_1), e_i(t_2)) = \rho |t_2-t_1|, t_1 \neq t_2 \), but we model it as an AR(1) correlation structure. The residuals are generated with \( \text{var}\{e_i(t)\} = 0.25 \exp(t/12) \).

### 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 link 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 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 link function estimation \( \hat{\phi}(\cdot) \) for \( \phi_0(\cdot) \) is evaluated by averaged mean squared error (AMSE) evaluated at the observed data points:

\[
\text{AMSE} = \sum_{b=1}^{B} \text{MSE}_b, \quad \text{MSE}_b = N^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left\{ \hat{\phi}_b(Z_{ij}^T \hat{\theta}) - \phi_0(Z_{ij}^T \hat{\theta}) \right\}^2.
\]

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 $(\beta^T, \theta^T)$. 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 4: 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 paper, 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 1–4 with SE, SWSE and AMSE values being in percentage. In Table 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 link function are relatively close to those for the case when the true covariance is assumed. In Tables 2–4 we see that both SGEE and SMGEE outperform the WI estimators in all simulation settings. SMGEE is clearly more efficient than SGEE regarding the link function estimation and is also overall more efficient than SGEE in the parameter estimation in these settings.

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 maturity firms from the period 1980 to 1989.
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. After examining the scatterplot matrix for all variables, we applied model (1) to the data with $Z_{ij} = (Z_{1ij}, Z_{2ij}, Z_{3ij})^T$ and $\theta = (\theta_1, \theta_2, \theta_3)^T$.

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 5. Furthermore, the estimated link function $\hat{\phi}(\cdot)$ evaluated at $Z_{ij}^T\hat{\theta}$ and its pointwise 95% bootstrap confidence band are shown in Figure 1 produced by WI, SGEE and SMGEE respectively. The parameter estimates in Table 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. By comparing the estimated link function in Figure 1, the decreasing patterns are similar for all three methods.

From the linear component estimates in Table 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 & 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 link function estimate. Since the general trend for the link function is decreasing in Figure 1 by SMGEE, together with the sign and magnitude of the parameter 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.

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 $Z_1 =$ 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 (??) is fitted to the data with $i = 1, \ldots, 328$ and $j = 1, \ldots, 10$, $\beta = (\beta_1, \beta_2)^T$ and $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)^T$.

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 6. Moreover, the estimated link function evaluated at $Z_{ij}^T \hat{\theta}$ and its pointwise 95% bootstrap confidence band are shown in Figure 2. They are obtained with WI, SGEE and SMGEE respectively. Table 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 three link function estimates in Figure 2, 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 nonparametric link 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 nonparametric link 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.

Table 1: Performance comparisons of the estimates of the parameters and link function under different covariance estimation methods while all other estimation steps are the same. The partially linear single-index model with nonstationary residual variance and AR(1) correlation (γ = 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.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Models</th>
<th>β_1</th>
<th>β_2</th>
<th>θ_1</th>
<th>θ_2</th>
<th>θ_3</th>
<th>φ(·)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SE</td>
<td>SWSE</td>
<td>SE</td>
<td>SWSE</td>
<td>SE</td>
<td>SWSE</td>
</tr>
<tr>
<td>n = 30</td>
<td>GEE-TC</td>
<td>2.68</td>
<td>2.72</td>
<td>2.65</td>
<td>2.66</td>
<td>1.75</td>
<td>1.76</td>
</tr>
<tr>
<td>m = 10</td>
<td>GEE-NC</td>
<td>2.80</td>
<td>2.84</td>
<td>2.73</td>
<td>2.76</td>
<td>1.77</td>
<td>1.80</td>
</tr>
<tr>
<td></td>
<td>SMGEE</td>
<td>2.73</td>
<td>2.80</td>
<td>2.71</td>
<td>2.76</td>
<td>1.74</td>
<td>1.75</td>
</tr>
</tbody>
</table>

6 Concluding remarks

In this paper, we proposed a three-stage procedure to estimate the parameters and the unknown link function in partially linear single-index models under the general sparse longitudinal setting. The parameter estimators have been shown to be semiparametric efficient. Furthermore, the link function estimator is not only more efficient than the working independence kernel estimator of Chen et al. (2015), but also achieves the minimum asymptotic
Table 2: Performance comparisons of the estimates of the parameters and link function for estimation methods WI, SGEE and SMGEE. The data were generated using the partially linear single-index model 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.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Models</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$\theta_3$</th>
<th>$\phi(\cdot)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SE</td>
<td>SWSE</td>
<td>SE</td>
<td>SWSE</td>
<td>SE</td>
<td>SWSE</td>
</tr>
<tr>
<td>$n = 30$</td>
<td>WI</td>
<td>3.82</td>
<td>3.89</td>
<td>4.12</td>
<td>4.14</td>
<td>2.72</td>
<td>2.75</td>
</tr>
<tr>
<td></td>
<td>SGEE</td>
<td>2.77</td>
<td>2.79</td>
<td>2.69</td>
<td>2.71</td>
<td>1.73</td>
<td>1.72</td>
</tr>
<tr>
<td></td>
<td>SMGEE</td>
<td>2.65</td>
<td>2.67</td>
<td>2.65</td>
<td>2.70</td>
<td>1.73</td>
<td>1.73</td>
</tr>
<tr>
<td>$n = 50$</td>
<td>WI</td>
<td>3.06</td>
<td>3.11</td>
<td>2.71</td>
<td>2.79</td>
<td>1.86</td>
<td>1.92</td>
</tr>
<tr>
<td></td>
<td>SGEE</td>
<td>2.61</td>
<td>2.62</td>
<td>2.10</td>
<td>2.13</td>
<td>1.17</td>
<td>1.19</td>
</tr>
<tr>
<td></td>
<td>SMGEE</td>
<td>2.54</td>
<td>2.56</td>
<td>2.10</td>
<td>2.14</td>
<td>1.15</td>
<td>1.16</td>
</tr>
<tr>
<td>$n = 100$</td>
<td>WI</td>
<td>2.19</td>
<td>2.21</td>
<td>2.29</td>
<td>2.30</td>
<td>1.61</td>
<td>1.65</td>
</tr>
<tr>
<td></td>
<td>SGEE</td>
<td>1.58</td>
<td>1.58</td>
<td>1.60</td>
<td>1.50</td>
<td>0.87</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>SMGEE</td>
<td>1.53</td>
<td>1.55</td>
<td>1.58</td>
<td>1.50</td>
<td>0.87</td>
<td>0.89</td>
</tr>
</tbody>
</table>

variance among a class of estimators when the covariance matrices are correctly specified. These analytic results are supported by our empirical studies. 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 link function estimators. Another possible topic is variable selection for partially linear single-index models for longitudinal data since it is sometimes difficult to keep or discard some covariates in practice. Furthermore, for the retained covariates, there is still a problem of distinguishing them from the linear part to the single index part. All this is worth investigating in future research.

References

Table 3: Performance comparisons of the estimates of the parameters and link function for estimation methods WI, SGEE and SMGEE. The data were generated using the partially linear single-index model 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.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Models</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$\theta_3$</th>
<th>$\phi(\cdot)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SE</td>
<td>SWSE</td>
<td>SE</td>
<td>SWSE</td>
<td>SE</td>
<td>SWSE</td>
</tr>
<tr>
<td>$n = 30$</td>
<td>WI</td>
<td>3.98</td>
<td>4.02</td>
<td>4.17</td>
<td>4.20</td>
<td>2.84</td>
<td>2.86</td>
</tr>
<tr>
<td>$\tilde{m} = 10$</td>
<td>SGEE</td>
<td>2.87</td>
<td>2.95</td>
<td>2.73</td>
<td>2.81</td>
<td>1.78</td>
<td>1.82</td>
</tr>
<tr>
<td></td>
<td>SMGEE</td>
<td>2.73</td>
<td>2.80</td>
<td>2.71</td>
<td>2.76</td>
<td>1.74</td>
<td>1.75</td>
</tr>
<tr>
<td>$n = 50$</td>
<td>WI</td>
<td>3.04</td>
<td>3.08</td>
<td>2.79</td>
<td>2.83</td>
<td>1.88</td>
<td>1.92</td>
</tr>
<tr>
<td>$\tilde{m} = 10$</td>
<td>SGEE</td>
<td>2.67</td>
<td>2.70</td>
<td>2.15</td>
<td>2.20</td>
<td>1.20</td>
<td>1.23</td>
</tr>
<tr>
<td></td>
<td>SMGEE</td>
<td>2.61</td>
<td>2.56</td>
<td>2.11</td>
<td>2.13</td>
<td>1.19</td>
<td>1.20</td>
</tr>
<tr>
<td>$n = 100$</td>
<td>WI</td>
<td>2.21</td>
<td>2.23</td>
<td>2.33</td>
<td>2.34</td>
<td>1.68</td>
<td>1.69</td>
</tr>
<tr>
<td>$\tilde{m} = 10$</td>
<td>SGEE</td>
<td>1.60</td>
<td>1.66</td>
<td>1.50</td>
<td>1.52</td>
<td>0.91</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>SMGEE</td>
<td>1.56</td>
<td>1.59</td>
<td>1.47</td>
<td>1.50</td>
<td>0.91</td>
<td>0.92</td>
</tr>
</tbody>
</table>


Table 4: Performance comparisons of the estimates of the parameters and link function for estimation methods WI, SGEE and SMGEE. The data were generated using the partially linear single-index model 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.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Models</th>
<th>$\beta_1$ SE</th>
<th>$\beta_1$ SWSE</th>
<th>$\beta_2$ SE</th>
<th>$\beta_2$ SWSE</th>
<th>$\theta_1$ SE</th>
<th>$\theta_1$ SWSE</th>
<th>$\theta_2$ SE</th>
<th>$\theta_2$ SWSE</th>
<th>$\theta_3$ SWSE</th>
<th>$\phi(\cdot)$ AMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 30$</td>
<td>WI</td>
<td>4.20</td>
<td>4.24</td>
<td>4.26</td>
<td>4.29</td>
<td>3.44</td>
<td>3.53</td>
<td>3.93</td>
<td>4.09</td>
<td>2.95</td>
<td>2.95</td>
</tr>
<tr>
<td></td>
<td>SGEE</td>
<td>3.01</td>
<td>3.06</td>
<td>2.89</td>
<td>2.85</td>
<td>2.06</td>
<td>2.14</td>
<td>2.70</td>
<td>2.81</td>
<td>2.28</td>
<td>2.35</td>
</tr>
<tr>
<td></td>
<td>SMGEE</td>
<td>2.97</td>
<td>3.05</td>
<td>2.84</td>
<td>2.75</td>
<td>2.02</td>
<td>1.99</td>
<td>2.69</td>
<td>2.77</td>
<td>2.31</td>
<td>2.36</td>
</tr>
<tr>
<td>$n = 50$</td>
<td>WI</td>
<td>3.15</td>
<td>3.13</td>
<td>2.95</td>
<td>3.04</td>
<td>2.75</td>
<td>2.79</td>
<td>2.83</td>
<td>2.90</td>
<td>2.36</td>
<td>2.46</td>
</tr>
<tr>
<td></td>
<td>SGEE</td>
<td>2.83</td>
<td>2.85</td>
<td>2.20</td>
<td>2.28</td>
<td>1.62</td>
<td>1.70</td>
<td>1.40</td>
<td>1.48</td>
<td>2.04</td>
<td>2.10</td>
</tr>
<tr>
<td></td>
<td>SMGEE</td>
<td>2.80</td>
<td>2.79</td>
<td>2.21</td>
<td>2.26</td>
<td>1.63</td>
<td>1.68</td>
<td>1.42</td>
<td>1.48</td>
<td>2.07</td>
<td>2.11</td>
</tr>
<tr>
<td>$n = 100$</td>
<td>WI</td>
<td>2.66</td>
<td>2.69</td>
<td>2.56</td>
<td>2.64</td>
<td>1.85</td>
<td>1.93</td>
<td>2.08</td>
<td>2.11</td>
<td>1.88</td>
<td>1.84</td>
</tr>
<tr>
<td></td>
<td>SGEE</td>
<td>1.92</td>
<td>2.01</td>
<td>1.70</td>
<td>1.72</td>
<td>1.11</td>
<td>1.14</td>
<td>1.68</td>
<td>1.72</td>
<td>1.34</td>
<td>1.36</td>
</tr>
<tr>
<td></td>
<td>SMGEE</td>
<td>1.94</td>
<td>1.98</td>
<td>1.68</td>
<td>1.70</td>
<td>1.07</td>
<td>1.12</td>
<td>1.70</td>
<td>1.72</td>
<td>1.33</td>
<td>1.35</td>
</tr>
</tbody>
</table>


Table 5: Parameter estimates by WI, SGEE and SMGEE and their standard errors for the CD4 data.

<table>
<thead>
<tr>
<th>Model Estimates</th>
<th>WI Estimates</th>
<th>SE</th>
<th>SGEE Estimates</th>
<th>SE</th>
<th>SMGEE Estimates</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>β</td>
<td>0.044</td>
<td>0.058</td>
<td>0.050</td>
<td>0.055</td>
<td>0.046</td>
<td>0.052</td>
</tr>
<tr>
<td>θ₁</td>
<td>0.880</td>
<td>0.036</td>
<td>0.842</td>
<td>0.033</td>
<td>0.827</td>
<td>0.033</td>
</tr>
<tr>
<td>θ₂</td>
<td>0.065</td>
<td>0.013</td>
<td>0.061</td>
<td>0.009</td>
<td>0.057</td>
<td>0.008</td>
</tr>
<tr>
<td>θ₃</td>
<td>−0.470</td>
<td>0.041</td>
<td>−0.536</td>
<td>0.040</td>
<td>−0.559</td>
<td>0.040</td>
</tr>
</tbody>
</table>


Figure 1: Estimated link 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.


Table 6: Parameter estimates by WI, SGEE and SMGEE and their standard errors for the debt maturity data.

<table>
<thead>
<tr>
<th>Model Estimates</th>
<th>WI Estimate</th>
<th>SE</th>
<th>SGEE Estimate</th>
<th>SE</th>
<th>SMSEE Estimate</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>$-0.357$</td>
<td>0.037</td>
<td>$-0.348$</td>
<td>0.030</td>
<td>$-0.346$</td>
<td>0.027</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>$-0.136$</td>
<td>0.046</td>
<td>$-0.115$</td>
<td>0.045</td>
<td>$-0.121$</td>
<td>0.044</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>0.843</td>
<td>0.009</td>
<td>0.796</td>
<td>0.009</td>
<td>0.782</td>
<td>0.007</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>0.535</td>
<td>0.097</td>
<td>0.596</td>
<td>0.095</td>
<td>0.612</td>
<td>0.089</td>
</tr>
<tr>
<td>$\theta_3$</td>
<td>$-0.022$</td>
<td>0.017</td>
<td>0.048</td>
<td>0.014</td>
<td>0.064</td>
<td>0.012</td>
</tr>
<tr>
<td>$\theta_4$</td>
<td>$-0.050$</td>
<td>0.020</td>
<td>$-0.098$</td>
<td>0.016</td>
<td>$-0.102$</td>
<td>0.016</td>
</tr>
</tbody>
</table>


### A Assumptions

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

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

**ASSUMPTION 3** $\phi(\cdot)$ has continuous derivatives up to the second order.

**ASSUMPTION 4** For $i = 1, \ldots, n$, $j = 1, \ldots, m_i$, $Z_{ij}$ is bounded with a compact support. The density of $Z_{ij}^T \theta$, denoted by $f_{ij}(u)$, is twice continuously differentiable and positive for all $u \in \mathcal{G}$ with $\mathcal{G} = \{ u = Z_{ij}^T \theta : Z_{ij} \in \mathcal{Z}, \theta \in \Xi \}$. Here $\Xi$ is the a compact parameter space for $\theta$ and $\mathcal{Z}$ is a compact support for $Z_{ij}$. In addition, the joint density of $(Z_{ij}^T \theta, Z_{ik}^T \theta)$ has first partial derivatives.
Figure 2: Estimated link 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.

**Assumption 5** \( E(X_{ij}|Z_{ij}^T\theta = u) \) and \( E(Z_{ij}|Z_{ij}^T\theta = u) \) are smooth functions of \( u \) with continuous derivatives up to the second order. In addition, \( \sup_{u \in G} E(||X_{ij}||^2|Z_{ij}^T\theta = u) \) and \( \sup_{u \in G} E(||Z_{ij}||^2|Z_{ij}^T\theta = u) \) are bounded for all \( i = 1, \ldots, n, \ j = 1, \ldots, m_i \).
Assumption 6 \( h \to 0, \, nh^8 \to 0 \) and \( nh/\log(1/h) \to \infty \).

Assumption 1 lists some regularity conditions for the kernel function. Assumption 2 is imposed for the consistency and asymptotic normality of our estimators. Assumption 3 is the smoothness restriction for the unknown link function. Assumption 4 ensures that the denominator of the kernel estimator for the link function in Steps 1 and 4 in Section 2 is meaningful and that some relevant asymptotic expansions are valid. Assumption 5 is a commonly used moments condition for predictors in partially linear single-index models. Assumption 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 6 are sufficient for the properties obtained in this paper.

**B Proof of Theorem 1 on semiparametric information bound**

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 \( M = M_{\alpha, \phi} \), where \( \alpha \) is the finite dimensional parameters and \( \phi \) is the nonparametric part with infinite dimension. Let \( S_\alpha \) be the score function with respect to \( \alpha \) in submodel \( M_\alpha \) which is \( M \) with the true function \( \phi_0 \) given. Besides, let \( \dot{M}_\phi \) be the tangent space for submodel \( M_\phi \) which is model \( M \) evaluated at the true parameters values \( \alpha = \alpha_0 \). Consider \( S_\alpha \) as an element in the Hilbert space and \( \dot{M}_\phi \) as a subset of the same Hilbert space with inner product \( E(\eta_1^T \eta_2) \), where \( \eta_1 \) and \( \eta_2 \) are two elements in \( \dot{M}_\phi \). Then the residual from the projection of \( S_\alpha \) on \( \dot{M}_\phi \) exists and there is a unique vector \( S_e \) satisfying

\[
S_\alpha - S_e \in \dot{M}_\phi \quad \text{and} \quad E(S_e^T w) = 0 \quad \text{for all} \quad w \in \dot{M}_\phi. \tag{17}
\]

If the likelihood function is regular with score function \( S_\alpha \) and \( E(S_e S_e^T) \) is nonsingular, then the semiparametric information bound is \( R_e = \{E(S_e S_e^T)\}^{-1} \) and the semiparametric efficient score is \( S_e \).

Now we denote the longitudinal partially linear single-index model in Equation (1) by \( M \). This is a special case of the general framework above. Model \( M \) has three unknown parts: \( \Theta_0, \phi_0(\cdot) \) and the joint distribution of \( (Y_{ij}, X_{ij}, Z_{ij}^T \theta_0) \) for \( i = 1, \ldots, n, \, j = 1, \ldots, m_i \). To derive the efficient score for \( \Theta \), consider three submodels of model \( M \):

- \( M_1 \): Model \( M \) with only \( \Theta_0 \) unknown;
$M_2$: Model $M$ with only $\phi_0(\cdot)$ unknown;

$M_3$: Model $M$ with both $\Theta_0$ and $\phi_0(\cdot)$ known.

Let $S_{\Theta}$ 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 (17), the semiparametric efficient score for model $M$ is

$$S_e = S_{\Theta} - \Pi(S_{\Theta}|M_2 + \dot{M}_3)$$

$$= S_{\Theta} - \Pi(S_{\Theta}|M_3) - \Pi(S_{\Theta}|\Pi_{M_2^\perp} \dot{M}_2),$$

where $M^\perp$ is the perpendicular space of model $M$, $\Pi(S_{\Theta}|P)$ is the projection of score function $S_{\Theta}$ on space $P$ and $\Pi_{P^T}$ is the projection of space $T$ on space $P$.

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 (17) $S_{\Theta} - \Pi(S_{\Theta}|\dot{M}_3)$ is the efficient score for $\Theta$ in model $M_4$. According to Lemma A4 in Huang et al. (2007),

$$\Pi(S_{\Theta}|\dot{M}_3) = \sum_{i=1}^{n} (X_i, Z_i)^T \Sigma_i^{-1} \{Y_i - X_i\beta_0 - \phi_0(Z_i, \theta_0)\}. $$

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

$$\Pi_{M_2^\perp} \dot{M}_2 = \left\{ \sum_{i=1}^{n} \kappa_n^T(Z_i, \theta_0) \Sigma_i^{-1} \{Y_i - X_i\beta_0 - \phi_0(Z_i, \theta_0)\}, \kappa_n(\cdot) \in L_2 \right\}. $$

Projecting the score function $S_{\Theta}$ on space $\Pi_{M_2^\perp} \dot{M}_2$ and putting the above two equations back to (18), we have

$$S_e = \sum_{i=1}^{n} (\tilde{X}_{i,e}, \tilde{Z}_{i,e})^T \Sigma_i^{-1} \{Y_i - X_i\beta_0 - \phi_0(Z_i, \theta_0)\},$$

where $\tilde{X}_{i,e} = X_i - \psi_{n, \beta}(Z_i, \theta_0)$, $\tilde{Z}_{i,e} = Z_i - \psi_{n, \theta}(Z_i, \theta_0)$ for some unique $\psi_{n, \beta}(\cdot) \in L_2$ and $\psi_{n, \theta}(\cdot) \in L_2$ by (17).

According to (17), $S_e$ is orthogonal to any member in $\dot{M}_2 + \dot{M}_3$. Besides, $\Pi_{M_2^\perp} \dot{M}_2$ is a subset of $\dot{M}_2 + \dot{M}_3$. Therefore, (17) implies Equations (7) and (8) for all $\kappa_n(\cdot) \in L_2$. The existence and uniqueness of the semiparametric score function $S_e$ for model $M$ are guaranteed by (17). Therefore, we have the semiparametric efficient score function as shown in Equation (9). Finally, we have the semiparametric information bound $R_e$ given in Equation (10), completing the proof of Theorem 1.
C Proof of Theorem 4 on link function estimation

First denote \( \mu_{ij} = E(Y_{ij}) = X_{ij}^T \beta_0 + \phi_0(Z_{ij}^T \theta_0) \). From Equation (2) we have

\[
\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} K_h(Z_{ij}^T \theta - u) \left[ (Z_{ij}^T \theta - u) \{ Y_{ij} - X_{ij}^T \beta - \tilde{\phi}(u) - \phi^{(1)}(u)(Z_{ij}^T \theta - u) \} \right] = 0.
\]

Some lengthy but standard calculations show that, when the parameters are evaluated at the true values, we have

\[
\tilde{\phi}(u, \Theta_0) - \phi_0(u) = \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} f_{ij}(u) \right\}^{-1} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} K_h(Z_{ij}^T \theta_0 - u)(Y_{ij} - \mu_{ij})
+ \frac{1}{2} c_{2,1} b^0(u) h^2 + o_p\{h^2 + (nh)^{-1/2}\}.
\]

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

\[
a_n(u) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} v_{ij}^{1j} f_{ij}(u),
\]

\[
J_n(u_1, u_2) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} v_{ij}^{jk} E \{ a_n^{-1}(Z_{ik}^T \theta_0) \} f_{ijk}(u_1, u_2),
\]

\[
b^e_n(u) = b^0(u) - a_n^{-1}(u) \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} f_{ij}(u) \sum_{k \neq j} v_{ij}^{jk} b^e_n(Z_{ik}^T \theta_0),
\]

where \( b^0(u) = b^0_n(u) = \phi_0^{(2)}(u) \) and \( f_{ijk} \) is the joint density of \((Z_{ij}^T \theta, Z_{ik}^T \theta)\). For the first step updated estimating equation from Equation (4), we have

\[
\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} K_h(Z_{ij}^T \theta - u) \left[ (Z_{ij}^T \theta - u) \{ Y_{ij} - X_{ij}^T \beta - \tilde{\phi}(u) - \phi^{(1)}(u)(Z_{ij}^T \theta - u) \} \right]
+ \sum_{k \neq j} \left( Z_{ij}^T \theta - u \right) v_{ij}^{jk} \{ Y_{ik} - X_{ik}^T \beta - \tilde{\phi}(Z_{ik}^T \theta) \} \right] = 0.
\]

Similarly to (19), the one-step update of \( \tilde{\phi} \), defined as \( \tilde{\phi}^1 \), has the following asymptotic expansion

\[
\tilde{\phi}^1(u, \Theta_0) - \phi_0(u) = a_n^{-1}(u)(B_{1n} + B_{2n} + B_{3n}) + o_p\{h^2 + (nh)^{-1/2}\},
\]

(20)
where

\[ B_{1n} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} K_h(Z_{ij}^T \theta_0 - u)v_{ij} \left\{ Y_{ij} - X_{ij}^T \beta_0 - \phi_0(u) - \phi_0^{(1)}(u)(Z_{ij}^T \theta_0 - u) \right\}, \]

\[ B_{2n} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} K_h(Z_{ij}^T \theta_0 - u) \left\{ \sum_{k \neq j} v_{ik}^j (Y_{ik} - \mu_{ik}) \right\}, \]

\[ B_{3n} = -\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} K_h(Z_{ij}^T \theta_0 - u) \left[ \sum_{k \neq j} v_{ik}^j \left\{ \tilde{\phi}(Z_{ik}^T \theta_0) - \phi_0(Z_{ik}^T \theta_0) \right\} \right]. \]

By plugging (19) into $B_{3n}$, we have

\[ \tilde{\phi}^1(u, \Theta_0) - \phi_0(u) = D_{1n}(u) + D_{2n}(u) + \frac{1}{2} c_{2,1} b_n(u) h^2 + o_p \{ h^2 + (nh)^{-1/2} \}, \] (21)

where

\[ D_{1n}(u) = a_n^{-1}(u) \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} K_h(Z_{ij}^T \theta_0 - u) \left\{ \sum_{k=1}^{m_i} v_{ik}^j (Y_{ik} - \mu_{ik}) \right\}, \]

\[ D_{2n}(u) = -a_n^{-1}(u) \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} v_{ij} \sum_{k \neq j} J_n(u, Z_{ij}^T \theta_0)(Y_{ij} - \mu_{ij}). \]

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

\[ \tilde{\phi}^r(u, \Theta_0) - \phi_0(u) = D_{1n} + E_{1n}^r + E_{2n}^r + \frac{1}{2} c_{2,1} b_n(u) h^2 + o_p \{ h^2 + (nh)^{-1/2} \}, \] (22)

where

\[ E_{1n}^r = a_n^{-1}(u) \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} J_{n,1}^r(u, Z_{ij}^T \theta_0) \left\{ \sum_{k=1}^{m_i} v_{ik}^j (Y_{ik} - \mu_{ik}) \right\}, \]

\[ E_{2n}^r = a_n^{-1}(u) \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} v_{ij} J_{n,2}^r(u, Z_{ij}^T \theta_0)(Y_{ij} - \mu_{ij}). \]

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, $\tilde{\phi}(u) - \phi_0(u)$ has a form similar to Equation (22) 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:

\[ b_n(u) = \phi_0^{(2)}(u) - a_n^{-1}(u) \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k \neq j} v_{ik}^j E \{ b_n(Z_{ik}^T \theta_0) \} f_{ij}(u), \]

\[ J_{n,1}(u_1, u_2) = -J_n(u_1, u_2) + H_n(J_{n,1}; u_1, u_2), \]

\[ J_{n,2}(u_1, u_2) = H_n(J_{n,2}; u_1, u_2). \] (23)
Since $E(E_{1n}^r) = E(E_{2n}^r) = 0$ and the variances of $E_{1n}$ and $E_{2n}$ are of order $O(n^{-1}) = o((nh)^{-1})$, we have $E_{1n}^r = E_{2n}^r = o_p((nh)^{-1/2})$. Thus, Equation (22) can be simplified as

$$\hat{\phi}(u, \Theta_0) - \phi_0(u) = D_{1n} + \frac{1}{2}c_{2,1}b_n(u)h^2 + o_p\{h^2 + (nh)^{-1/2}\}.$$ 

Therefore, we have

$$\hat{\phi}(u) - \phi_0(u) = a_n^{-1}(u) \frac{1}{n} \sum_{i=1}^{m_n} \sum_{j=1}^{m_r} K_h(Z_{ij}^T \theta_0 - u) \{ \sum_{k=1}^{m_r} v_{ik}^2 (Y_{ik} - \mu_{ik}) \}
+ c_{2,1} \left[ \frac{h^2 \phi_n^{(2)}(u)}{2} - h^2 a_n^{-1}(u) \frac{1}{n} \sum_{i=1}^{m_n} \sum_{j=1}^{m_r} \sum_{k \neq j} v_{ij}^2 E\{b(Z_{ik}^T \theta_0)\} f_{ij}(u) \right]
+ o_p\{h^2 + (nh)^{-1/2}\},$$

where $b(\cdot)$ is defined in Theorem 4. Some standard calculations show that the asymptotic bias for $\hat{\phi}(u)$ is $c_{2,1}b(u)h^2$, where $b(u)$ satisfies

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

with $\eta(u, w)$ given in Theorem 4. The asymptotic variance of $\sqrt{nh}\{\hat{\phi}(u) - \phi_0(u)\}$ is readily seen to be $\sigma_n^2(u) = c_{0,2}Q_2(u)/Q_1^2(u)$. In fact, the one-step update $\hat{\phi}_1(u)$ also has the same asymptotic variance of $\sigma_n^2(u)$. That is, further iteration steps beyond the first update do not change the asymptotic variance.

It has been shown in Theorem 5 that the asymptotic variance of $\hat{\phi}(u)$ is minimized when $V_i = \Sigma_i$. Therefore, when the covariance matrices are correctly specified, the asymptotic variance of $\hat{\phi}(u)$ is $\sigma_n^2(u)$ as shown in Theorem 5.

### D Proof of Theorem 3 on semiparametric efficiency of the parameter estimators

Denote the $l$th element of $X_{ij}$ and $Z_{ij}$ by $X_{ijl}$ and $Z_{ijl}$, respectively. Similarly to Wang et al. (2005), by (7) we can obtain that $\psi_{n,\beta_l}(\cdot)$ solves the Fredholm integral equation of the second kind:

$$\psi_{n,\beta_l}(u_1) = r_{n,\beta_l}(u_1) - \int W_n(u_1, u_2)\psi_{n,\beta_l}(u_2)du_2, \quad (24)$$

where

$$W_n(u_1, u_2) = \frac{1}{n} \sum_{i=1}^{m_n} \sum_{j=1}^{m_r} \sum_{k \neq j} \sigma_{ij}^2 f_{ijk}(u_2, u_1)$$

$$- \frac{1}{n} \sum_{i=1}^{m_n} \sum_{j=1}^{m_r} \sigma_{ij} f_{ij}(u_1) \quad (25)$$
and

\[ r_{n,\beta_i}(u) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{\sigma_i^j K_h(Z_{ij}^T \beta - u)\{Y_{ij} - X_{ij}^T \beta - \hat{\phi}(u, \Theta) - \hat{\phi}(1)(u, \Theta)(Z_{ij}^T \beta - u)\}}{\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sigma_i^j f_{ij}(u)} \]

for \( l = 1, \ldots, p \). Similarly, from Equation (8) we can obtain that \( \psi_{n,\theta_i}(\cdot) \) solves the Fredholm integral equation of the second kind:

\[ \psi_{n,\theta_i}(u_1) = r_{n,\theta_i}(u_1) - \int W_n(u_1, u_2) \psi_{n,\theta_i}(u_2)dw_2, \quad (26) \]

where

\[ r_{n,\theta_i}(u) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{\sigma_i^j E(Z_{ijkl})f_{ij}(u)}{\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sigma_i^j f_{ij}(u)} \]

for \( l = 1, \ldots, q \). If we take the limit of Equations (24) and (26) as \( n \to \infty \) we see that \( \psi_{\beta}(u) = \lim_{n \to \infty} \psi_{n,\beta}(u) \) and \( \psi_{\theta}(u) = \lim_{n \to \infty} \psi_{n,\theta}(u) \) 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 \( \varphi_{\beta}(u) \) and \( \varphi_{\theta}(u) \) satisfy the Fredholm integral equations of the second kind as same as \( \psi_{\beta}(u) \) and \( \psi_{\theta}(u) \), respectively. By Equation (4), when \( V_i = \Sigma_i \) for \( i = 1, \ldots, n \), we have

\[ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sigma_i^j K_h(Z_{ij}^T \theta - u)\{Y_{ij} - X_{ij}^T \beta - \hat{\phi}(u, \Theta) - \hat{\phi}(1)(u, \Theta)(Z_{ij}^T \theta - u)\} + \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k\neq j} \sigma_i^j K_h(Z_{ik}^T \theta - u)\{Y_{ik} - X_{ik}^T \beta - \hat{\phi}(Z_{ik}^T \theta, \Theta)\} = 0. \]

(27)

Taking derivative with respect to \( \theta \) on both sides and evaluating at \( \theta = \theta_0 \), we get

\[ L_1 + L_2 + L_3 + L_4 = 0, \]

where

\[ L_1 = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sigma_i^j K_h^{-1}(Z_{ij}^T \theta_0 - u)Z_{ij} \]

\[ \{Y_{ij} - X_{ij}^T \beta_0 - \hat{\phi}(u, \Theta_0) - \hat{\phi}(1)(u, \Theta_0)(Z_{ij}^T \theta_0 - u)\}, \]

\[ L_2 = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k\neq j} \sigma_i^j K_h^{-1}(Z_{ij}^T \theta_0 - u)Z_{ij} \]

\[ \{Y_{ij} - X_{ij}^T \beta_0 - \hat{\phi}(u, \Theta_0) - \hat{\phi}(1)(u, \Theta_0)(Z_{ij}^T \theta_0 - u)\}, \]
From the GEE in Equation (5), the parameter estimates satisfy the equation

\[ L_3 = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sigma_{ij}^2 K_h(\textbf{Z}_{ij}^T \theta_0 - u) \left\{ \varphi_{n,\theta}(u, \Theta_0) - \frac{\partial \hat{\phi}(1)}{\partial \Theta_0}(\textbf{Z}_{ij}^T \theta_0 - u) \right\} \]

\[ - \left\{ \varphi_{n,\theta}(\textbf{Z}_{ij}^T \theta_0, \Theta_0) - \frac{\partial \hat{\phi}(1)}{\partial \Theta_0}(\textbf{Z}_{ij}^T \theta_0, \Theta_0) \right\} \textbf{Z}_{ij} \]

With Assumptions 1–6, when \( n \to \infty \) and taking expectations, some standard calculations lead to \( L_1 = o_p(1), L_2 = o_p(1) \) and \( L_3 = o_p(1) \). Then we get

\[ \varphi_{\theta}(u) \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sigma_{ij}^2 f_{ij}(u) = \left( \frac{1}{n} \right) \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sigma_{ij}^2 E \left\{ \textbf{Z}_{ik} \phi_{\theta}(1) \right\} f_{ij}(u) \]

\[ - \left( \frac{1}{n} \right) \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k \neq j} \int \sigma_{ij}^2 f_{ijk}(\textbf{Z}_{ik}^T \theta_0, u) \varphi_{\theta}(\textbf{Z}_{ik}^T \theta_0) d(\textbf{Z}_{ik}^T \theta_0) + o_p(1) \]

as \( n \to \infty \). By taking the limit of both (26) 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 \( \beta \) of both sides of Equation (27) and following similar steps, we can obtain Equation (24) as \( \psi_{n,\beta} \); \( r_{n,\beta} \) and \( W_n \) are replaced by \( \varphi_{\beta}, r_{\beta} = \lim_{n \to \infty} r_{n,\beta}, \) and \( W = \lim_{n \to \infty} W_n \) respectively when \( n \to \infty \). The uniqueness is guaranteed by the projection method in (17). Therefore, the parameter estimators reach the semiparametric information bound, completing the proof of Theorem 3.

E Proof of Theorem 2 on parameter estimation

We prove this theorem after proving Theorems 3 and 4 because they are used in this proof. To derive the asymptotic properties for the parameter estimators, the uniform consistency of the link function estimator is needed. The only difference between the uniform consistency and pointwise consistency of the link 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 \( \Delta_i \{ \beta, \phi(\textbf{Z}_i, \Theta) \}^T = \partial \{ \textbf{X}_i \beta + \phi(\textbf{Z}_i, \Theta) \}^T / \partial \Theta \).

From the GEE in Equation (5), the parameter estimates satisfy the equation

\[ n^{-1/2} \sum_{i=1}^{n} \Delta_i \{ \beta, \phi(\textbf{Z}_i, \Theta) \}^T \textbf{V}_i^{-1} \{ \textbf{Y}_i - \textbf{X}_i \beta - \phi(\textbf{Z}_i, \Theta) \} = 0. \]
By expending the equation above at $\Theta = \Theta_0$ and after some further algebra, we have the following equation

$$n^{-1/2} \sum_{i=1}^n \Delta_i \{\beta_0, \hat{\phi}(Z_i, \theta_0)\}^T V_i^{-1} \left[ Y_i - X_i \beta_0 - \hat{\phi}(Z_i, \theta_0) - \Delta_i \{\beta_0, \hat{\phi}(Z_i, \theta_0)\} (\hat{\Theta} - \Theta_0) \right] + o_p(1) = 0,$$

where we recall that $\hat{\phi}(Z_i, \theta_0) = \hat{\phi}(Z_i, \hat{\Theta})$. Thus we have

$$n^{-1/2} \sum_{i=1}^n \Delta_i \{\beta_0, \hat{\phi}(Z_i, \theta_0)\}^T V_i^{-1} \Delta_i \{\beta_0, \hat{\phi}(Z_i, \theta_0)\} (\hat{\Theta} - \Theta_0)$$

$$= n^{-1/2} \sum_{i=1}^n \Delta_i \{\beta_0, \hat{\phi}(Z_i, \theta_0)\}^T V_i^{-1} \left[ Y_i - X_i \beta_0 - \phi_0(Z_i, \theta_0) \right]$$

$$- \{\hat{\phi}(Z_i, \theta_0) - \phi_0(Z_i, \theta_0)\} + o_p(1).$$

Then it is readily seen that

$$\frac{1}{n} \sum_{i=1}^n (\tilde{X}_i, \tilde{Z}_i)^T V_i^{-1} (\tilde{X}_i, \tilde{Z}_i) \{n^{1/2}(\hat{\Theta} - \Theta_0)\}$$

$$= n^{-1/2} \sum_{i=1}^n (\tilde{X}_i, \tilde{Z}_i)^T V_i^{-1} \left[ Y_i - X_i \beta_0 - \phi_0(Z_i, \theta_0) \right]$$

$$- \{\hat{\phi}(Z_i, \theta_0) - \phi_0(Z_i, \theta_0)\} + o_p(1). \tag{28}$$

By Equations (22) and (23) in the proof of Theorem 4 and a second order bias expansion, we have

$$n^{-1/2} \sum_{i=1}^n (\tilde{X}_i, \tilde{Z}_i)^T V_i^{-1} \{\hat{\phi}(Z_i, \theta_0) - \phi_0(Z_i, \theta_0)\}$$

$$= n^{-1/2} \frac{h^2}{2} \left[ \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} v_{ij}^k (\tilde{X}_{ij}, \tilde{Z}_{ij})^T \left\{ b(Z_{ik}^T \theta_0) + h b_1(Z_{ik}^T \theta_0) + O_p(h^2) \right\} \right]$$

$$+ \left( n^{-1/2} \sum_{i=1}^n \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} v_{ij}^k (\tilde{X}_{ij}, \tilde{Z}_{ij})^T \left\{ a_{n-1}(Z_{ik}^T \theta_0) \frac{1}{n} \sum_{i'=1}^n \sum_{j'=1}^{m} v_{i'i'}^{j'} \right\} \right)$$

$$\times \left\{ K_h(Z_{ij}, \theta_0) - Z_{ik}^T \theta_0 \right\} \sum_{l=1}^{m_i} v_{i'i'}^{j'} (Y_{i'l} - \mu_{i'l})$$

$$+ J_{n,2}(Z_{ik}^T \theta_0, Z_{ij}^T \theta_0) (Y_{ij} - \mu_{ij})$$

$$+ J_{n,1}(Z_{ik}^T \theta_0, Z_{ij}^T \theta_0) \sum_{l=1}^{m_i} v_{i'i'}^{j'} (Y_{i'l} - \mu_{i'l}) \right\} + o_p(1)$$
\[ T_1 + T_2 = o_p(1), \]

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

\[
T_{21} = n^{-\frac{1}{2}} \sum_{i'=1}^{n} \sum_{j'=1}^{m_i} v_{i'j'} \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} K_h(Z_{i'j'} - Z_{ik} \theta_0)(\tilde{X}_{ij}, \tilde{Z}_{ij})^T v_{i'k} \right\} \left\{ \frac{m_i}{n} \sum_{l=1}^{m_i} v_{i'l} \left( Y_{i'l} - \mu_{i'l} \right) \right\}.
\]

We now note that it is easy to see that when working covariances \( V_i \) are used in place of \( \Sigma_i \), Equations (7) and (8) are asymptotically equivalent to the following equation:

\[
\frac{1}{n} \sum_{i=1}^{n} E \{ (\tilde{X}_i, \tilde{Z}_i)^T V_i^{-1} g_n(D_i) \odot f_i(D_i) | D_i = d_i \} = 0
\]

for any function \( g_n(\cdot) \in L_2 \), where \( D_i = Z_i \theta, d_i = (d_{i1}, \ldots, d_{imi})^T \) and \( f_i(d_i) = (f_{i1}(d_{i1}), \ldots, f_{imi}(d_{imi}))^T \) 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

\[
\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{k=1}^{m_i} E \{ (\tilde{X}_{ij}, \tilde{Z}_{ij})^T v_{i'k} a_n^{-1}(D_{ik}) | D_{ik} = d \} f_{ik}(d) | d = Z_{i'j}' \theta_0
\]

with \( D_{ik} = Z_{ik}^T \theta_0 \), which is 0 in probability by (29). 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 1–6, \( T_1 + T_2 = o_p(1) \). It follows from Equation (28) that

\[
n^{1/2}(\hat{\Theta} - \Theta_0) = \Omega_0^{-1} n^{-\frac{1}{2}} \sum_{i=1}^{n} (\tilde{X}_i, \tilde{Z}_i)^T V_i^{-1} \{ Y_i - X_i \beta_0 - \phi_0(Z_i \theta_0) \} + o_p(1).
\]

This directly leads to Theorem 2.