The Annals of Statistics
2015, Vol. 43, No. 4, 1682–1715
DOI: 10.1214/15-AOS1320

SEMIPARAMETRIC GEE ANALYSIS IN PARTIALLY LINEAR SINGLE-INDEX MODELS FOR LONGITUDINAL DATA

BY JIA CHEN*, DEGUI LI*, HUA LIANG[†,1] AND SUOJIN WANG[‡,2]

*University of York**, *George Washington University*[†] *and Texas A&M University*[‡]

In this article, we study a partially linear single-index model for longitudinal data under a general framework which includes both the sparse and dense longitudinal data cases. A semiparametric estimation method based on a combination of the local linear smoothing and generalized estimation equations (GEE) is introduced to estimate the two parameter vectors as well as the unknown link function. Under some mild conditions, we derive the asymptotic properties of the proposed parametric and nonparametric estimators in different scenarios, from which we find that the convergence rates and asymptotic variances of the proposed estimators for sparse longitudinal data would be substantially different from those for dense longitudinal data. We also discuss the estimation of the covariance (or weight) matrices involved in the semiparametric GEE method. Furthermore, we provide some numerical studies including Monte Carlo simulation and an empirical application to illustrate our methodology and theory.

1. Introduction. Consider a semiparametric partially linear single-index model defined by

$$Y(t) = \mathbf{Z}^\top(t)\boldsymbol{\beta} + \eta\big(\mathbf{X}^\top(t)\boldsymbol{\theta}\big) + e(t), \qquad t \in \mathcal{T}, \tag{1.1}$$

where $\mathcal{T}$ is a bounded time interval, $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are two unknown vectors of parameters with dimensions d and p, respectively, $\eta(\cdot)$ is an unknown link function, $Y(t)$ is a scalar stochastic process, $\mathbf{Z}(t)$ and $\mathbf{X}(t)$ are covariates with dimensions d and p, respectively, and $e(t)$ is the random error process. For the case of independent and identically distributed (i.i.d.) or weakly dependent time series data, there has been extensive literature on statistical inference of model (1.1) since its introduction by Carroll et al. (1997). Several different approaches have been proposed to estimate the unknown parameters and link function involved; see, for example, Xia, Tong and Li (1999), Yu and Ruppert (2002), Xia and Härdle (2006), Wang et al. (2010) and Ma and Zhu (2013). A recent paper by Liang et al. (2010) further

Received May 2014; revised February 2015.

[1]Supported in part by NSF Grants DMS-14-40121 and DMS-14-18042 and by Award Number 11228103, made by National Natural Science Foundation of China.

[2]Supported in part by Award Number KUS-CI-016-04, made by King Abdullah University of Science and Technology (KAUST).

MSC2010 subject classifications. 62G09, 62H99, 62G99.

Key words and phrases. Efficiency, GEE, local linear smoothing, longitudinal data, semiparametric estimation, single-index models.

developed semiparametric techniques for the variable selection and model specification testing issues in the context of model (1.1).

In this paper, we are interested in studying partially linear single-index model (1.1) in the context of longitudinal data which arise frequently in many fields of research, such as biology, climatology, economics and epidemiology, and thus have attracted considerable attention in the literature in recent years. Various parametric models and methods have been studied in depth for longitudinal data; see Diggle et al. (2002) and the references therein. However, the parametric models may be misspecified in practice, and the misspecification may lead to inconsistent estimates and incorrect conclusions being drawn. Hence, to circumvent this issue, in recent years, there has been a large literature on how to relax the parametric assumptions on longitudinal data models and many nonparametric, and semiparametric models have thus been investigated; see, for example, Lin and Ying (2001), He, Zhu and Fung (2002), Fan and Li (2004), Wang, Carroll and Lin (2005), Lin and Carroll (2006), Wu and Zhang (2006), Li and Hsing (2010), Jiang and Wang (2011) and Yao and Li (2013).

Suppose that we have a random sample with n subjects from model (1.1). For the ith subject, $i = 1, \dots, n$, the response variable $Y_i(t)$ and the covariates $\{\mathbf{Z}_i(t), \mathbf{X}_i(t)\}$ are collected at random time points t_{ij}, $j = 1, \dots, m_i$, which are distributed in a bounded time interval $\mathcal{T}$ according to the probability density function $f_T(t)$. Here m_i is the total number of observations for the ith subject. To accommodate such longitudinal data, model (1.1) is written in the following framework:

$$Y_i(t_{ij}) = \mathbf{Z}_i^\top(t_{ij})\boldsymbol{\beta} + \eta\big(\mathbf{X}_i^\top(t_{ij})\boldsymbol{\theta}\big) + e_i(t_{ij}) \tag{1.2}$$

for $i = 1, \dots, n$ and $j = 1, \dots, m_i$. When m_i varies across the subjects, the longitudinal data set under investigation is unbalanced. Several nonparametric and semiparametric models can be viewed as special cases of model (1.2). For instance, when $\boldsymbol{\beta} = \mathbf{0}$, model (1.2) reduces to the single-index longitudinal data model [Jiang and Wang (2011), Chen, Gao and Li (2013a)]; when $p = 1$ and $\boldsymbol{\theta} = 1$, model (1.2) reduces to the partially linear longitudinal data model [Fan and Li (2004)]. To avoid confusion, we let $\boldsymbol{\beta}_0$ and $\boldsymbol{\theta}_0$ be the true values of the two parameter vectors. For identifiability reasons, $\boldsymbol{\theta}_0$ is assumed to be a unit vector with the first nonzero element being positive. Furthermore, we allow that there exists certain within-subject correlation structure for $e_i(t_{ij})$, which makes the model assumption more realistic but the development of estimation methodology more challenging.

To estimate the parameters $\boldsymbol{\beta}_0$, $\boldsymbol{\theta}_0$ as well as the link function $\eta(\cdot)$ in model (1.2), we first apply the local linear approximation to the unknown link function, and then introduce a profile weighted least squares approach to estimate the two parameter vectors based on the technique of generalized estimation equations (GEE). Under some mild conditions, we derive the asymptotic properties of the developed parametric and nonparametric estimators in different scenarios. Our framework is flexible in that m_i can either be bounded or tend to infinity.

Thus both the dense and sparse longitudinal data cases can be included. Dense longitudinal data means that there exists a sequence of positive numbers M_n such that $\min_i m_i \geq M_n$, and $M_n \to \infty$ as $n \to \infty$ [see, e.g., Hall, Müller and Wang (2006) and Zhang and Chen (2007)], whereas sparse longitudinal data means that there exists a positive constant M_* such that $\max_i m_i \leq M_*$; see, for example, Yao, Müller and Wang (2005), Wang, Qian and Carroll (2010). We show that the convergence rates and asymptotic variances of our semiparametric estimators in the sparse case are substantially different from those in the dense case. Furthermore, we show that the proposed semiparametric GEE (SGEE)-based estimators are asymptotically more efficient than the profile unweighted least squares (PULS) estimators, when the weights in the SGEE method are chosen as the inverse of the covariance matrix of the errors. We also introduce a semiparametric approach to estimate the covariance matrices (or weights) involved in the SGEE method, which is based on a variance–correlation decomposition and consists of two steps: first, estimate the conditional variance function using a robust nonparametric method that accommodates heavy-tailed errors, and second, estimate the parameters in the correlation matrix. A simulation study and a real data analysis are provided to illustrate our methodology and theory.

The rest of the paper is organized as follows. In Section 2, we introduce the SGEE methodology for estimating $\boldsymbol{\beta}_0$, $\boldsymbol{\theta}_0$ and $\eta(\cdot)$. Section 3 establishes the large sample theory for the proposed parametric and nonparametric estimators and gives some related discussions. Section 4 discusses how to determine the weight matrices in the estimation equations. Section 5 gives some numerical examples to investigate the finite sample performance of the proposed approach. Section 6 concludes the paper. Technical assumptions are given in Appendix A. The proofs of the main results are given in Appendix B. Some auxiliary lemmas and their proofs are provided in the supplementary material [Chen et al. (2015)].

2. Estimation methodology. Various semiparametric estimation approaches have been proposed to estimate model (1.1) in the case of i.i.d. observations (or weakly dependent time series data). See, for example, Carroll et al. (1997) and Liang et al. (2010) for the profile likelihood method, Yu and Ruppert (2002) and Wang et al. (2010) for the "remove-one-component" technique using penalized spline and local linear smoothing, respectively, and Xia and Härdle (2006) for the minimum average variance estimation approach. However, there is limited literature on partially linear single-index models for longitudinal data because of the more complicated structures involved. Recently, Chen, Gao and Li (2013b) studied a partially linear single-index longitudinal data model with individual effects. To remove the individual effects and derive consistent semiparametric estimators, they had to limit their discussions to the dense and balanced longitudinal data case. Ma, Liang and Tsai (2014) considered a partially linear single-index longitudinal data model by using polynomial splines to approximate the unknown link function, but their discussion was limited to the sparse and balanced longitudinal data case. In contrast, as mentioned in Section 1, our framework includes both the sparse and dense longitudinal data cases. Meanwhile, observations are allowed to be collected at irregular and subject specific time points. All this provides much wider applicability of our framework. Furthermore, to improve the efficiency of the semiparametric estimation, we develop a new profile weighted least squares approach to estimate the parameters $\boldsymbol{\beta}_0$, $\boldsymbol{\theta}_0$ as well as the link function $\eta_0(\cdot)$.

To simplify the presentation, let

$$\begin{aligned} \mathbf{Y}_i &= \big(Y_i(t_{i1}), \ldots, Y_i(t_{im_i})\big)^\top, \qquad \mathbf{X}_i = \big(\mathbf{X}_i(t_{i1}), \ldots, \mathbf{X}_i(t_{im_i})\big)^\top, \\ \mathbf{Z}_i &= \big(\mathbf{Z}_i(t_{i1}), \ldots, \mathbf{Z}_i(t_{im_i})\big)^\top, \qquad \mathbf{e}_i = \big(e_i(t_{i1}), \ldots, e_i(t_{im_i})\big)^\top, \\ \boldsymbol{\eta}(\mathbf{X}_i, \boldsymbol{\theta}) &= \big(\eta(\mathbf{X}_i^\top(t_{i1})\boldsymbol{\theta}), \ldots, \eta(\mathbf{X}_i^\top(t_{im_i})\boldsymbol{\theta})\big)^\top. \end{aligned}$$

With the above notation, model (1.2) can then be re-written as

$$\mathbf{Y}_i = \mathbf{Z}_i\boldsymbol{\beta}_0 + \boldsymbol{\eta}(\mathbf{X}_i, \boldsymbol{\theta}_0) + \mathbf{e}_i. \tag{2.1}$$

We further let $\mathbb{Y} = (\mathbf{Y}_1^\top, \ldots, \mathbf{Y}_n^\top)^\top$, $\mathbb{Z} = (\mathbf{Z}_1^\top, \ldots, \mathbf{Z}_n^\top)^\top$, $\mathbb{E} = (\mathbf{e}_1^\top, \ldots, \mathbf{e}_n^\top)^\top$, $\boldsymbol{\eta}(\mathbb{X}, \boldsymbol{\theta}) = (\boldsymbol{\eta}^\top(\mathbf{X}_1, \boldsymbol{\theta}), \ldots, \boldsymbol{\eta}^\top(\mathbf{X}_n, \boldsymbol{\theta}))^\top$. Then model (2.1) is equivalent to

$$\mathbb{Y} = \mathbb{Z}\boldsymbol{\beta}_0 + \boldsymbol{\eta}(\mathbb{X}, \boldsymbol{\theta}_0) + \mathbb{E}. \tag{2.2}$$

Our estimation procedure is based on the profile likelihood method, which is commonly used in semiparametric estimation; see, for example, Carroll et al. (1997), Fan and Huang (2005) and Fan, Huang and Li (2007). Let $Y_{ij} = Y_i(t_{ij})$, $\mathbf{Z}_{ij} = \mathbf{Z}_i(t_{ij})$ and $\mathbf{X}_{ij} = \mathbf{X}_i(t_{ij})$. For given $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, we can estimate $\eta(\cdot)$ and its derivative $\dot{\eta}(\cdot)$ at point u by minimizing the following loss function:

$$\begin{aligned} &L_n(a, b|\boldsymbol{\beta}, \boldsymbol{\theta}) \\ &\quad = \sum_{i=1}^n \left\{ \frac{w_i}{h} \sum_{j=1}^{m_i} [Y_{ij} - \mathbf{Z}_{ij}^\top\boldsymbol{\beta} - a - b(\mathbf{X}_{ij}^\top\boldsymbol{\theta} - u)]^2 K\left(\frac{\mathbf{X}_{ij}^\top\boldsymbol{\theta} - u}{h}\right) \right\}, \end{aligned} \tag{2.3}$$

where $K(\cdot)$ is a kernel function, h is a bandwidth and w_i, $i = 1, \ldots, n$, are some weights. It is well known that the local linear smoothing has advantages over the Nadaraya–Watson kernel method, such as higher asymptotic efficiency, design adaption and automatic boundary correction [Fan and Gijbels (1996)]. Following the existing literature such as Wu and Zhang (2006), the weights w_i can be specified by two schemes: $w_i = 1/T_n$ (type 1) and $w_i = 1/(nm_i)$ (type 2), where $T_n = \sum_{i=1}^n m_i$. The type 1 weight scheme corresponds to an equal weight for each observation, while the type 2 scheme corresponds to an equal weight within each subject. As discussed in Huang, Wu and Zhou (2002) and Wu and Zhang (2006), the type 2 scheme may be appropriate if the number of observations varies across subjects. As the longitudinal data under investigation in this paper are allowed to be unbalanced, we use $w_i = 1/(nm_i)$, which was also used by Li and Hsing (2010) and Kim and Zhao (2013). We denote

$$(2.4)\qquad \bigl(\widehat{\eta}(u|\boldsymbol{\beta},\boldsymbol{\theta}), \widehat{\dot{\eta}}(u|\boldsymbol{\beta},\boldsymbol{\theta})\bigr)^\top = \arg\min_{a,b} L_n(a,b|\boldsymbol{\beta},\boldsymbol{\theta}).$$

By some elementary calculations [see, e.g., Fan and Gijbels (1996)], we have

$$(2.5)\qquad \widehat{\eta}(u|\boldsymbol{\beta},\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathbf{s}_i(u|\boldsymbol{\theta})(\mathbf{Y}_i - \mathbf{Z}_i\boldsymbol{\beta})$$

for given $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, where

$$(2.6)\qquad \begin{aligned} \mathbf{s}_i(u|\boldsymbol{\theta}) &= (1,0)\left[\sum_{i=1}^{n} \overline{\mathbf{X}}_i^\top(u|\boldsymbol{\theta})\mathbf{K}_i(u|\boldsymbol{\theta})\overline{\mathbf{X}}_i(u|\boldsymbol{\theta})\right]^{-1} \overline{\mathbf{X}}_i^\top(u|\boldsymbol{\theta})\mathbf{K}_i(u|\boldsymbol{\theta}),\\ \overline{\mathbf{X}}_i(u|\boldsymbol{\theta}) &= \bigl(\overline{\mathbf{X}}_{i1}(u|\boldsymbol{\theta}),\dots,\overline{\mathbf{X}}_{im_i}(u|\boldsymbol{\theta})\bigr)^\top,\\ \overline{\mathbf{X}}_{ij}(u|\boldsymbol{\theta}) &= \bigl(1, \mathbf{X}_{ij}^\top\boldsymbol{\theta} - u\bigr)^\top,\\ \mathbf{K}_i(u|\boldsymbol{\theta}) &= \operatorname{diag}\left(w_i K\left(\frac{\mathbf{X}_{i1}^\top\boldsymbol{\theta}-u}{h}\right),\dots, w_i K\left(\frac{\mathbf{X}_{im_i}^\top\boldsymbol{\theta}-u}{h}\right)\right). \end{aligned}$$

Based on the profile least squares approach with the first-stage local linear smoothing, we can construct estimators of the parameters $\boldsymbol{\beta}_0$ and $\boldsymbol{\theta}_0$. We start with the PULS method which ignores the possible within-subject correlation structure. Define the PULS loss function by

$$(2.7)\qquad \begin{aligned} Q_{n0}(\boldsymbol{\beta},\boldsymbol{\theta}) &= \sum_{i=1}^{n}\bigl[\mathbf{Y}_i - \mathbf{Z}_i\boldsymbol{\beta} - \widehat{\boldsymbol{\eta}}(\mathbf{X}_i|\boldsymbol{\beta},\boldsymbol{\theta})\bigr]^\top\bigl[\mathbf{Y}_i - \mathbf{Z}_i\boldsymbol{\beta} - \widehat{\boldsymbol{\eta}}(\mathbf{X}_i|\boldsymbol{\beta},\boldsymbol{\theta})\bigr]\\ &= \bigl[\mathbb{Y} - \mathbb{Z}\boldsymbol{\beta} - \widehat{\boldsymbol{\eta}}(\mathbb{X}|\boldsymbol{\beta},\boldsymbol{\theta})\bigr]^\top\bigl[\mathbb{Y} - \mathbb{Z}\boldsymbol{\beta} - \widehat{\boldsymbol{\eta}}(\mathbb{X}|\boldsymbol{\beta},\boldsymbol{\theta})\bigr], \end{aligned}$$

where, for given $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\eta}}(\mathbf{X}_i|\boldsymbol{\beta},\boldsymbol{\theta})$ and $\widehat{\boldsymbol{\eta}}(\mathbb{X}|\boldsymbol{\beta},\boldsymbol{\theta})$ are the local linear estimators of the vectors $\boldsymbol{\eta}(\mathbf{X}_i,\boldsymbol{\theta})$ and $\boldsymbol{\eta}(\mathbb{X},\boldsymbol{\theta})$, respectively; that is, each element of $\widehat{\boldsymbol{\eta}}(\mathbf{X}_i|\boldsymbol{\beta},\boldsymbol{\theta})$ and $\widehat{\boldsymbol{\eta}}(\mathbb{X}|\boldsymbol{\beta},\boldsymbol{\theta})$ is defined as in (2.5). The PULS estimators of $\boldsymbol{\beta}_0$ and $\boldsymbol{\theta}_0$ are obtained by minimizing the loss function $Q_{n0}(\boldsymbol{\beta},\boldsymbol{\theta})$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ and normalizing the minimizer $\boldsymbol{\theta}$. We denote the resulting estimators by $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\theta}}$, respectively.

Although it is easy to verify that both $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\theta}}$ are consistent, they are not efficient as the within-subject correlation structure is not taken into account. Hence, to improve the efficiency of the parametric estimators, we next introduce a GEE-based method to estimate the parameters $\boldsymbol{\beta}_0$ and $\boldsymbol{\theta}_0$. Existing literature on GEE-based method in longitudinal data analysis includes Liang and Zeger (1986), Xie and Yang (2003) and Wang (2011). Let $\mathbb{W} = \operatorname{diag}\{\mathbf{W}_1,\dots,\mathbf{W}_n\}$, where $\mathbf{W}_i = \mathbf{R}_i^{-1}$ and $\mathbf{R}_i$ is an $m_i \times m_i$ working covariance matrix whose estimation will be discussed in Section 4. Define

$$\boldsymbol{\rho}_{\mathbf{Z}}(\mathbf{X}_i, \boldsymbol{\theta}) = \big(\rho_{\mathbf{Z}}(\mathbf{X}_{i1}^\top \boldsymbol{\theta}|\boldsymbol{\theta}), \ldots, \rho_{\mathbf{Z}}(\mathbf{X}_{im_i}^\top \boldsymbol{\theta}|\boldsymbol{\theta})\big)^\top, \qquad \rho_{\mathbf{Z}}(u|\boldsymbol{\theta}) = \mathrm{E}\big[\mathbf{Z}_{ij}|\mathbf{X}_{ij}^\top \boldsymbol{\theta} = u\big],$$

$$\boldsymbol{\rho}_{\mathbf{X}}(\mathbf{X}_i, \boldsymbol{\theta}) = \big(\rho_{\mathbf{X}}(\mathbf{X}_{i1}^\top \boldsymbol{\theta}|\boldsymbol{\theta}), \ldots, \rho_{\mathbf{X}}(\mathbf{X}_{im_i}^\top \boldsymbol{\theta}|\boldsymbol{\theta})\big)^\top, \qquad \rho_{\mathbf{X}}(u|\boldsymbol{\theta}) = \mathrm{E}\big[\mathbf{X}_{ij}|\mathbf{X}_{ij}^\top \boldsymbol{\theta} = u\big],$$

$$\boldsymbol{\Lambda}_i(\boldsymbol{\theta}) = \big(\mathbf{Z}_i - \boldsymbol{\rho}_{\mathbf{Z}}(\mathbf{X}_i, \boldsymbol{\theta}), \big[\dot{\boldsymbol{\eta}}(\mathbf{X}_i, \boldsymbol{\theta}) \otimes \mathbf{1}_p^\top\big] \odot \big[\mathbf{X}_i - \boldsymbol{\rho}_{\mathbf{X}}(\mathbf{X}_i, \boldsymbol{\theta})\big]\big),$$

where $\dot{\boldsymbol{\eta}}(\mathbf{X}_i, \boldsymbol{\theta})$ is a column vector with its elements being the derivatives of $\eta(\cdot)$ at points $\mathbf{X}_{ij}^\top \boldsymbol{\theta}$, $j = 1, \ldots, m_i$, $\mathbf{1}_p$ is a p-dimensional vector of ones, $\otimes$ is the Kronecker product and $\odot$ denotes the componentwise product. The construction of the parametric estimators is based on solving the following equation with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$:

$$\sum_{i=1}^{n} \widehat{\boldsymbol{\Lambda}}_i^\top(\boldsymbol{\theta}) \mathbf{W}_i \big[\mathbf{Y}_i - \mathbf{Z}_i \boldsymbol{\beta} - \widehat{\boldsymbol{\eta}}(\mathbf{X}_i|\boldsymbol{\beta}, \boldsymbol{\theta})\big] = \mathbf{0}, \tag{2.8}$$

where $\widehat{\boldsymbol{\Lambda}}_i(\boldsymbol{\theta})$ is an estimator of $\boldsymbol{\Lambda}_i(\boldsymbol{\theta})$ with $\boldsymbol{\rho}_{\mathbf{Z}}(\mathbf{X}_i, \boldsymbol{\theta})$, $\boldsymbol{\rho}_{\mathbf{X}}(\mathbf{X}_i, \boldsymbol{\theta})$ and $\dot{\boldsymbol{\eta}}(\mathbf{X}_i, \boldsymbol{\theta})$ replaced by their corresponding local linear estimated values. Let $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}_1$ be the solutions to the estimation equations in (2.8), and let the SGEE-based estimator of $\boldsymbol{\theta}_0$ be defined as $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_1 / \|\widehat{\boldsymbol{\theta}}_1\|$, where $\|\cdot\|$ is the Euclidean norm. Note that the solutions to the equations in (2.8) generally do not have a closed form. In the numerical studies, we use the trust-region dogleg algorithm within the Matlab command "fsolve" to obtain the solutions to (2.8). Corollary 1 below shows that the SGEE-based estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$ are generally asymptotically more efficient than the PULS estimators $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\theta}}$, when the weights are chosen appropriately.

Replacing $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ in $\widehat{\eta}(\cdot)$ by $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$, respectively, we obtain the local linear estimator of the link function $\eta(\cdot)$ at u as

$$\widehat{\eta}(u) = \widehat{\eta}(u|\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \mathbf{s}_i(u|\widehat{\boldsymbol{\theta}})(\mathbf{Y}_i - \mathbf{Z}_i \widehat{\boldsymbol{\beta}}). \tag{2.9}$$

In Section 3 below, we will give the large sample properties of the estimators proposed above, and in Section 4, we will discuss how to choose the working covariance matrix $\mathbf{R}_i$.

3. Theoretical properties. Before establishing the large sample theory for the proposed parametric and nonparametric estimators, we introduce some notation. Let $\mathbf{B}_0$ be a $p \times (p-1)$ matrix such that $\mathbf{M} = (\boldsymbol{\theta}_0, \mathbf{B}_0)$ is a $p \times p$ orthogonal matrix, and define

$$\mathbf{I}(\mathbf{B}_0) = \begin{pmatrix} \mathbf{I}_d & \mathbf{O}_{d\times(p-1)} \\ \mathbf{O}_{p\times d} & \mathbf{B}_0 \end{pmatrix},$$

where $\mathbf{I}_k$ is a $k \times k$ identity matrix and $\mathbf{O}_{k\times l}$ is a $k \times l$ null matrix. Let $\boldsymbol{\Lambda}_i = \boldsymbol{\Lambda}_i(\boldsymbol{\theta}_0)$, and assume that there exist two positive semi-definite matrices $\boldsymbol{\Omega}_0$ and $\boldsymbol{\Omega}_1$ as well as a sequence of numbers ω_n such that $\omega_n \to \infty$,

$$\frac{1}{\omega_n}\sum_{i=1}^{n}\boldsymbol{\Lambda}_i^\top \mathbf{W}_i \boldsymbol{\Lambda}_i \xrightarrow{P} \boldsymbol{\Omega}_0, \tag{3.1}$$

$$\frac{1}{\omega_n}\sum_{i=1}^{n}\mathrm{E}\big[\boldsymbol{\Lambda}_i^\top \mathbf{W}_i \mathbf{e}_i \mathbf{e}_i^\top \mathbf{W}_i \boldsymbol{\Lambda}_i\big] \to \boldsymbol{\Omega}_1, \tag{3.2}$$

$$\max_{1\le i\le n}\mathrm{E}\big[\boldsymbol{\Lambda}_i^\top \mathbf{W}_i \mathbf{e}_i \mathbf{e}_i^\top \mathbf{W}_i \boldsymbol{\Lambda}_i\big] = o(\omega_n), \tag{3.3}$$

as $n \to \infty$, and $\mathbf{I}^\top(\mathbf{B}_0)\boldsymbol{\Omega}_0\mathbf{I}(\mathbf{B}_0)$ is positive definite. Conditions (3.2) and (3.3) ensure that the Lindeberg–Feller condition can be satisfied, and thus the classical central limit theorem for independent sequence [Petrov (1995)] is applicable. It is not difficult to verify the assumption in (3.3) for the dense and sparse longitudinal data. In particular, (3.3) excludes the case where the term $\boldsymbol{\Lambda}_i^\top \mathbf{W}_i \mathbf{e}_i$ from one or a few subjects dominates those from the others. For the latter case, it may be possible to derive the consistency of the proposed parametric estimation, but the proof of the asymptotic normality would be difficult. Let $\boldsymbol{\Omega}_0^+$ be the Moore–Penrose inverse matrix of $\boldsymbol{\Omega}_0$, which is defined as $\boldsymbol{\Omega}_0^+ = \mathbf{I}(\mathbf{B}_0)[\mathbf{I}^\top(\mathbf{B}_0)\boldsymbol{\Omega}_0\mathbf{I}(\mathbf{B}_0)]^{-1}\mathbf{I}^\top(\mathbf{B}_0)$. We next give the asymptotic distribution theory for the SGEE-based estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$.

THEOREM 1. *Suppose that Assumptions 1–5 in Appendix A and (3.1)–(3.3) are satisfied. Then we have*

$$\omega_n^{1/2}\begin{pmatrix}\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\\ \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\end{pmatrix} \xrightarrow{d} \mathrm{N}\big(\mathbf{0}, \boldsymbol{\Omega}_0^+\boldsymbol{\Omega}_1\boldsymbol{\Omega}_0^+\big) \tag{3.4}$$

as $n \to \infty$.

REMARK 1. Theorem 1 establishes the asymptotically normal distribution theory for $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$ with convergence rate $\omega_n^{1/2}$. This ω_n is linked to h through n in a certain way. Specifically, the condition $\omega_n h^6 \to 0$ in Assumption 5 needs to be satisfied to ensure that the bias term of the parametric estimation is asymptotically negligible. The specific forms of ω_n, $\boldsymbol{\Omega}_0$ and $\boldsymbol{\Omega}_1$ can be derived for some particular cases, for instance, when longitudinal data are balanced, that is, $m_i \equiv m$, $\omega_n = nm$. Furthermore, assume that the covariates and the error are i.i.d. with $\mathrm{E}[e_i^2(t_{ij})] \equiv \sigma_e^2$, $e_i(t_{ij})$ is independent of the covariates and $\mathbf{W}_i$, $i = 1,\dots,n$, are $m \times m$ identity matrices. Then we can show that

$$\boldsymbol{\Omega}_0 = \begin{pmatrix}\Omega_0(1) & \Omega_0(2)\\ \Omega_0^\top(2) & \Omega_0(3)\end{pmatrix} \quad\text{and}\quad \boldsymbol{\Omega}_1 = \sigma_e^2\begin{pmatrix}\Omega_0(1) & \Omega_0(2)\\ \Omega_0^\top(2) & \Omega_0(3)\end{pmatrix},$$

where

$$\Omega_0(1) = \mathrm{E}\{[\mathbf{Z}(t) - \boldsymbol{\rho}_{\mathbf{Z}}(\mathbf{X}^\top(t)\boldsymbol{\theta}_0|\boldsymbol{\theta}_0)][\mathbf{Z}(t) - \boldsymbol{\rho}_{\mathbf{Z}}(\mathbf{X}^\top(t)\boldsymbol{\theta}_0|\boldsymbol{\theta}_0)]^\top\},$$

$$\Omega_0(2) = \mathrm{E}\{\dot{\eta}(\mathbf{X}^\top(t)\boldsymbol{\theta}_0)[\mathbf{Z}(t) - \rho_{\mathbf{Z}}(\mathbf{X}^\top(t)\boldsymbol{\theta}_0|\boldsymbol{\theta}_0)][\mathbf{X}(t) - \rho_{\mathbf{X}}(\mathbf{X}^\top(t)\boldsymbol{\theta}_0|\boldsymbol{\theta}_0)]^\top\},$$

$$\Omega_0(3) = \mathrm{E}\{[\dot{\eta}(\mathbf{X}^\top(t)\boldsymbol{\theta}_0)]^2[\mathbf{X}(t) - \rho_{\mathbf{X}}(\mathbf{X}^\top(t)\boldsymbol{\theta}_0|\boldsymbol{\theta}_0)][\mathbf{X}(t) - \rho_{\mathbf{X}}(\mathbf{X}^\top(t)\boldsymbol{\theta}_0|\boldsymbol{\theta}_0)]^\top\}.$$

Hence $\boldsymbol{\Omega}_0^+\boldsymbol{\Omega}_1\boldsymbol{\Omega}_0^+$ reduces to $\sigma_e^2\boldsymbol{\Omega}_0^+$.

In Theorem 1 above, we only require $n \to \infty$. As mentioned in Section 1, both the sparse and dense longitudinal data cases can be included in a unified framework. For the sparse longitudinal data case when m_i is bounded by a certain positive constant, we can take $\omega_n = n$ and prove that (3.4) holds. For the dense longitudinal data case where $\min_i m_i \geq M_n$ with $M_n \to \infty$, under some regularity conditions we may prove (3.4) with $w_n = \sum_{i=1}^n m_i$. As more observations are available in the dense longitudinal data case and the order for the total number of the observations is higher than n, the convergence rate for the parametric estimators is faster than the well-known root-n rate in the sparse longitudinal data case.

Using Theorem 1, we can obtain the following corollary.

COROLLARY 1. *Suppose that the weights $\mathbf{W}_i$ in (2.8) are chosen as the inverse of the conditional covariance matrix of $\mathbf{e}_i$, and the conditions of Theorem 1 are satisfied. Then the SGEE-based estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$ are asymptotically more efficient than the PULS estimators $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\theta}}$ defined in Section 2.*

REMARK 2. In the proof of the above corollary, we show that the asymptotic covariance matrix of the PULS estimators $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\theta}}$ (after appropriate normalization) minus that of the SGEE-based estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$ is positive semi-definite, although the two estimation methods have the same convergence rates. That is, under the conditions assumed in Theorem 1, the limit matrix of $\omega_n[\mathrm{Var}(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\theta}}) - \mathrm{Var}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}})]$ is positive semi-definite. For the case of independent observations, a recent paper by Luo, Li and Yin (2014) discussed the efficient bound for the semiparametric estimation in single-index models. Following their idea, we conjecture that modification of our estimation procedure may be needed to obtain the efficient estimation in the partially linear single-index longitudinal data models. We will study this issue in our future research.

To establish the asymptotic distribution theory for the nonparametric estimator $\widehat{\eta}(u)$ under a unified framework, we assume that there exist a sequence $\varphi_n(h)$ and a constant $0 < \sigma_*^2 < \infty$ such that

$$\varphi_n(h) = o(\omega_n), \qquad \varphi_n(h) \max_{1 \leq i \leq n} \mathrm{E}[\mathbf{s}_i(u|\boldsymbol{\theta}_0)\mathbf{e}_i\mathbf{e}_i^\top\mathbf{s}_i^\top(u|\boldsymbol{\theta}_0)] = o(1) \tag{3.5}$$

and

$$\varphi_n(h)\sum_{i=1}^{n}\mathrm{E}\big[\mathbf{s}_i(u|\boldsymbol{\theta}_0)\mathbf{e}_i\mathbf{e}_i^\top\mathbf{s}_i^\top(u|\boldsymbol{\theta}_0)\big]\to\sigma_*^2. \tag{3.6}$$

The first restriction in (3.5) is imposed to ensure that the parametric convergence rates are faster than the nonparametric convergence rates, and the second restriction in (3.5) and the condition in (3.6) are imposed for the derivation of the asymptotic variance of the local linear estimator $\widehat{\eta}(u)$ and the satisfaction of the Lindeberg–Feller condition. The specific forms of $\varphi_n(h)$ and σ_*^2 will be discussed in Remark 3 below. Let $\mu_j=\int v^j K(v)\,dv$ for $j=0,1,2$ and $\ddot{\eta}_0(\cdot)$ be the second-order derivative of $\eta_0(\cdot)$.

THEOREM 2. *Suppose that the conditions of Theorem 1, (3.5) and (3.6) are satisfied. Then we have*

$$\varphi_n^{1/2}(h)\big[\widehat{\eta}(u)-\eta_0(u)-b_\eta(u)h^2\big]\xrightarrow{d}\mathrm{N}\big(0,\sigma_*^2\big), \tag{3.7}$$

where $b_\eta(u)=\ddot{\eta}_0(u)\mu_2/2$.

REMARK 3. Theorem 2 provides the asymptotically normal distribution theory for the nonparametric estimator $\widehat{\eta}(u)$ with a convergence rate $O_P(\varphi_n^{-1/2}(h)+h^2)$. The forms of $\varphi_n(h)$ and σ_*^2 in Theorem 2 depend on the type of the longitudinal data under study, that is, whether it is sparse or dense. We can derive their specific forms for some particular cases. Consider, for example, the case where $e_i(t_{ij})=v_i+\varepsilon_{ij}$, in which ε_{ij} are i.i.d. across both i and j with $\mathrm{E}[\varepsilon_{ij}]=0$ and $\mathrm{E}[\varepsilon_{ij}^2]=\sigma_\varepsilon^2$, and $\{v_i\}$ is an i.i.d. sequence of random variables with $\mathrm{E}[v_i]=0$ and $\mathrm{E}[v_i^2]=\sigma_v^2$ and is independent of $\{\varepsilon_{ij}\}$. In this case, we note that

$$\begin{aligned}
&\mathrm{E}\biggl\{\biggl[\sum_{j=1}^{m_i}K\biggl(\frac{\mathbf{X}_{ij}^\top\boldsymbol{\theta}_0-u}{h}\biggr)e_{ij}\biggr]^2\biggr\}\\
&\quad=\mathrm{E}\biggl\{\biggl[\sum_{j=1}^{m_i}K\biggl(\frac{\mathbf{X}_{ij}^\top\boldsymbol{\theta}_0-u}{h}\biggr)(v_i+\varepsilon_{ij})\biggr]^2\biggr\}\\
&\quad=\sum_{j=1}^{m_i}\mathrm{E}\biggl[K^2\biggl(\frac{\mathbf{X}_{ij}^\top\boldsymbol{\theta}_0-u}{h}\biggr)(v_i+\varepsilon_{ij})^2\biggr]\\
&\qquad+\sum_{j_1\neq j_2}\mathrm{E}\biggl[K\biggl(\frac{\mathbf{X}_{ij_1}^\top\boldsymbol{\theta}_0-u}{h}\biggr)K\biggl(\frac{\mathbf{X}_{ij_2}^\top\boldsymbol{\theta}_0-u}{h}\biggr)(v_i+\varepsilon_{ij_1})(v_i+\varepsilon_{ij_2})\biggr]\\
&\quad\sim m_ih\nu_0f_{\boldsymbol{\theta}_0}(u)\big(\sigma_v^2+\sigma_\varepsilon^2\big)+m_i(m_i-1)h^2\mu_0^2f_{\boldsymbol{\theta}_0}^2(u)\sigma_v^2,
\end{aligned}$$

where $\nu_0 = \int K^2(v)\,dv$ and $f_{\boldsymbol{\theta}_0}(\cdot)$ is the probability density function of $\mathbf{X}_{ij}^\top \boldsymbol{\theta}_0$.

For the sparse longitudinal data case, $m_i(m_i-1)h^2\mu_0^2 f_{\boldsymbol{\theta}_0}^2(u)\sigma_v^2$ is dominated by $m_i h\nu_0 f_{\boldsymbol{\theta}_0}(u)(\sigma_v^2+\sigma_\varepsilon^2)$, as m_i is bounded and $h \to 0$. Then, by Lemma 1 in the supplementary document [Chen et al. (2015)] and some elementary calculations, we can prove that

$$\begin{aligned}\sum_{i=1}^n \mathrm{E}\big[\mathbf{s}_i(u|\boldsymbol{\theta}_0)\mathbf{e}_i\mathbf{e}_i^\top \mathbf{s}_i^\top(u|\boldsymbol{\theta}_0)\big] &\sim \frac{1}{(nh)^2}\sum_{i=1}^n \frac{m_i h\nu_0(\sigma_v^2+\sigma_\varepsilon^2)}{m_i^2 f_{\boldsymbol{\theta}_0}(u)} \\ &\sim \frac{\nu_0(\sigma_v^2+\sigma_\varepsilon^2)}{n^2 h f_{\boldsymbol{\theta}_0}(u)}\sum_{i=1}^n \frac{1}{m_i}.\end{aligned} \tag{3.8}$$

Hence, in this case, we can take $\varphi_n(h) = (n^2h)(\sum_{i=1}^n 1/m_i)^{-1}$ which has the same order as nh, and $\sigma_*^2 = \nu_0(\sigma_v^2+\sigma_\varepsilon^2)/f_{\boldsymbol{\theta}_0}(u)$. This result is similar to Theorem 1(i) in Kim and Zhao (2013).

For the dense longitudinal data case, $m_i h\nu_0 f_{\boldsymbol{\theta}_0}(u)(\sigma_v^2+\sigma_\varepsilon^2)$ is dominated by $m_i(m_i-1)h^2\mu_0^2 f_{\boldsymbol{\theta}_0}^2(u)\sigma_v^2$ if we assume that $m_i h \to \infty$. Then, again by Lemma 1 in the supplementary material [Chen et al. (2015)], we can prove that

$$\begin{aligned}\sum_{i=1}^n \mathrm{E}\big[\mathbf{s}_i(u|\boldsymbol{\theta}_0)\mathbf{e}_i\mathbf{e}_i^\top \mathbf{s}_i^\top(u|\boldsymbol{\theta}_0)\big] &\sim \frac{1}{(nh)^2}\sum_{i=1}^n \frac{m_i(m_i-1)h^2\mu_0^2\sigma_v^2}{m_i^2} \\ &\sim \frac{\mu_0^2\sigma_v^2}{n}.\end{aligned}$$

Hence, in this case, we can take $\varphi_n(h) = n$ and $\sigma_*^2 = \mu_0^2\sigma_v^2$, which are analogous to those in Theorem 1(ii) of Kim and Zhao (2013) and quite different from those in the sparse longitudinal data case.

4. Estimation of covariance matrices. Estimation of the weight or working covariance matrices, which are involved in the SGEE (2.8), is critical to improving the efficiency of the proposed semiparametric estimators. However, the unbalanced longitudinal data structure, which can be either sparse or dense, makes such covariance matrix estimation very challenging, and some existing estimation methods based on balanced data [such as Wang (2011)] cannot be directly used here. In this section, we introduce a semiparametric estimation approach that is applicable to both sparse and dense unbalanced longitudinal data. This approach is based on a variance–correlation decomposition, and the estimation of the working covariance matrices then consists of two steps: first, estimate the conditional variance function using a robust nonparametric method that accommodates heavy-tailed errors, and second, estimate the parameters in the correlation matrix. For recent developments on the study of the covariance structure in longitudinal data analysis, we refer to Fan and Wu (2008), Zhang, Leng and Tang (2015) and the references therein.

For each $1 \le i \le n$, let $\mathbf{R}_i$ be the covariance matrix of $\mathbf{e}_i$ and

$$\boldsymbol{\Sigma}_i = \operatorname{diag}\{\sigma^2(t_{i1}), \dots, \sigma^2(t_{im_i})\}$$

with $\sigma^2(t_{ij}) = \mathrm{E}[e_i^2(t_{ij})|t_{ij}] = \mathrm{E}[e_i^2(t_{ij})|t_{ij}, \mathbf{X}_i(t_{ij}), \mathbf{Z}_i(t_{ij})]$ for $j = 1, \dots, m_i$, and $\mathbf{C}_i$ be the correlation matrix of $\mathbf{e}_i$. Assume that there exists a q-dimensional parameter vector $\boldsymbol{\phi}$ such that $\mathbf{C}_i = \mathbf{C}_i(\boldsymbol{\phi})$ where $\mathbf{C}_i(\cdot)$, $1 \le i \le n$, are pre-specified. By the variance–correlation decomposition, we have

$$\mathbf{R}_i = \boldsymbol{\Sigma}_i^{1/2} \mathbf{C}_i(\boldsymbol{\phi}) \boldsymbol{\Sigma}_i^{1/2}. \tag{4.1}$$

The above semiparametric covariance structure has been studied in some of the existing literature [see, e.g., Fan, Huang and Li (2007) and Fan and Wu (2008)] and provides a flexible framework to capture the error covariance structure, especially when the dimension of $\boldsymbol{\phi}$ is large. For example, it is satisfied when $e_i(t_{ij})$ has the AR(1) or ARMA(1, 1) dependence structure for each i; see, for example, the simulated example in Section 5.1. When $e_i(t_{ij}) = \sigma(t_{ij})(v_i + \varepsilon_{ij})$ in which v_i and ε_{ij} satisfy the conditions discussed in Remark 3 and $\sigma_\varepsilon^2 + \sigma_v^2 = 1$, we can also show that the semiparametric covariance structure is satisfied with $\boldsymbol{\phi}$ being σ_ε^2 or σ_v^2. Some existing papers such as Wu and Pourahmadi (2003) suggest the use of a nonparametric smoothing method to estimate the covariance matrix. However, they usually need to assume that the longitudinal data are balanced or nearly balanced, which would be violated when the data are collected at irregular and possibly subject-specific time points. Yao, Müller and Wang (2005) proposed the approach of functional data analysis to estimate the covariance structure for sparse and irregularly-spaced longitudinal data. However, some substantial modification may be needed to extend the method of Yao, Müller and Wang (2005) to our framework, which includes both the sparse and dense longitudinal data.

In the present paper, we first estimate the conditional variance function $\sigma^2(\cdot)$ in the diagonal matrix $\boldsymbol{\Sigma}_i$ by using a nonparametric method. In recent years, there has been a rich literature on the study of nonparametric conditional variance estimation; see, for example, Fan and Yao (1998), Yu and Jones (2004), Fan, Huang and Li (2007) and Leng and Tang (2011). However, when the errors are heavy-tailed, which is not uncommon in economic and financial data analysis, most of these existing methods may not perform well. This motivates us to devise an estimation method that is robust to heavy-tailed errors. Let $r(t_{ij}) = [Y_{ij} - \mathbf{Z}_{ij}^\top \boldsymbol{\beta}_0 - \eta(\mathbf{X}_{ij}^\top \boldsymbol{\theta}_0)]^2$. We can then find a random variable $\xi(t_{ij})$ so that $r(t_{ij}) = \sigma^2(t_{ij})\xi^2(t_{ij})$ and $\mathrm{E}[\xi^2(t_{ij})|t_{ij}] = 1$ with probability one. By applying the log-transformation [see Peng and Yao (2003) and Chen, Cheng and Peng (2009) for the application of this transformation in time series analysis] to $r(t_{ij})$, we have

$$\log r(t_{ij}) = \log[\tau\sigma^2(t_{ij})] + \log[\tau^{-1}\xi^2(t_{ij})] \equiv \sigma_\diamond^2(t_{ij}) + \xi_\diamond(t_{ij}), \tag{4.2}$$

where τ is a positive constant such that $\mathrm{E}[\xi_\diamond(t_{ij})] = \mathrm{E}\{\log[\tau^{-1}\xi^2(t_{ij})]\} = 0$. Here, $\xi_\diamond(t_{ij})$ could be viewed as an error term in model (4.2). As $r_{ij} \equiv r(t_{ij})$ are unobservable, we replace them with

$$\widehat{r}_{ij} = \bigl[Y_{ij} - \mathbf{Z}_{ij}^\top \widetilde{\boldsymbol{\beta}} - \widehat{\eta}(\mathbf{X}_{ij}^\top \widetilde{\boldsymbol{\theta}} | \widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\theta}})\bigr]^2,$$

where $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\theta}}$ are the PULS estimators of $\boldsymbol{\beta}_0$ and $\boldsymbol{\theta}_0$, respectively. In order to estimate $\sigma_\diamond^2(t)$, we define

$$\widetilde{L}_n(a,b) = \sum_{i=1}^n \left\{ \frac{w_i}{h_1} \sum_{j=1}^{m_i} [\log(\widehat{r}_{ij} + \zeta_n) - a - b(t_{ij} - t)]^2 K_1\left(\frac{t_{ij}-t}{h_1}\right)\right\}, \tag{4.3}$$

where $K_1(\cdot)$ is a kernel function, h_1 is a bandwidth satisfying Assumption 9 in Appendix A, $w_i = 1/(nm_i)$ as in Section 2 and $\zeta_n \to 0$ as $n \to \infty$. Throughout this paper, we set $\zeta_n = 1/T_n$, where $T_n = \sum_{i=1}^n m_i$. The ζ_n is added in $\log(\widehat{r}_{ij} + \zeta_n)$ to avoid the occurrence of invalid $\log 0$ as $\zeta_n > 0$ for any n. Such a modification would not affect the asymptotic distribution of the conditional variance estimation under certain mild restrictions. Then $\sigma_\diamond^2(t)$ can be estimated as

$$\widehat{\sigma}_\diamond^2(t) = \widehat{a} \qquad \text{where } (\widehat{a}, \widehat{b})^\top = \arg\min_{a,b} \widetilde{L}_n(a,b). \tag{4.4}$$

On the other hand, noting that $\frac{\exp\{\sigma_\diamond^2(t_{ij})\}}{\tau}\xi^2(t_{ij}) = r_{ij}$ and $\mathrm{E}[\xi^2(t_{ij})] = 1$, the constant τ can be estimated by

$$\widehat{\tau} = \left[\frac{1}{T_n}\sum_{i=1}^n \sum_{j=1}^{m_i} \widehat{r}_{ij} \exp\{-\widehat{\sigma}_\diamond^2(t_{ij})\}\right]^{-1}. \tag{4.5}$$

We then estimate $\sigma^2(t)$ by

$$\widehat{\sigma}^2(t) = \frac{\exp\{\widehat{\sigma}_\diamond^2(t)\}}{\widehat{\tau}}. \tag{4.6}$$

It is easy to see that thus defined estimator $\widehat{\sigma}^2(t)$ is always positive.

Suppose that there exists a sequence $\varphi_{n\diamond}(h_1)$ which depends on h_1, and a constant $0 < \sigma_\diamond^2 < \infty$ such that

$$\begin{gathered} \varphi_{n\diamond}(h_1) = o(\omega_n), \\ \frac{\varphi_{n\diamond}(h_1)}{h_1^2} \max_{1\le i\le n} w_i^2 \mathrm{E}\left[\sum_{j=1}^{m_i} \xi_\diamond(t_{ij}) K_1\left(\frac{t_{ij}-t}{h_1}\right)\right]^2 = o(1) \end{gathered} \tag{4.7}$$

and

$$\frac{\varphi_{n\diamond}(h_1)}{f_T(t) h_1^2} \mathrm{E}\left[\sum_{i=1}^n w_i \sum_{j=1}^{m_i} \xi_\diamond(t_{ij}) K_1\left(\frac{t_{ij}-t}{h_1}\right)\right]^2 \to \sigma_\diamond^2, \tag{4.8}$$

which are similar to those in (3.5) and (3.6), where $f_T(\cdot)$ is the density function of the observation times t_{ij}. Define

$$b_{\sigma 1}(t) = \frac{\exp\{\sigma_\diamond^2(t)\}}{2\tau} \ddot{\sigma}_\diamond^2(t) \int v^2 K_1(v)\,dv,$$
$$b_{\sigma 2}(t) = \frac{\exp\{\sigma_\diamond^2(t)\}}{2\tau} \mathrm{E}\big[\ddot{\sigma}_\diamond^2(t_{ij})\big] \int v^2 K_1(v)\,dv,$$

where $\ddot{\sigma}_\diamond^2(\cdot)$ is the second-order derivative of $\sigma_\diamond^2(\cdot)$. We then establish the asymptotic distribution of $\widehat{\sigma}^2(t)$ in the following theorem, whose proof is given in the supplementary material [Chen et al. (2015)].

THEOREM 3. *Suppose the conditions in Theorems 1 and 2, Assumptions 6–9 in Appendix A, (4.7) and (4.8) are satisfied. Then we have*

$$\varphi_{n\diamond}^{1/2}(h_1)\{\widehat{\sigma}^2(t) - \sigma^2(t) - \big[b_{\sigma 1}(t) - b_{\sigma 2}(t)\big]h_1^2\} \xrightarrow{d} \mathrm{N}\bigg(0, \frac{\sigma^4(t)}{f_T(t)}\sigma_\diamond^2\bigg). \tag{4.9}$$

REMARK 4. Theorem 3 can be seen as an extension of Theorem 1 in Chen, Cheng and Peng (2009) from the time series case to the longitudinal data case. The longitudinal data framework in this paper is more flexible and includes both sparse and dense data types. If $\xi_\diamond(t_{ij}) = v_i^\diamond + \varepsilon_{ij}^\diamond$, where $\varepsilon_{ij}^\diamond$ are i.i.d. across both i and j with $\mathrm{E}[\varepsilon_{ij}^\diamond] = 0$ and $\mathrm{E}[(\varepsilon_{ij}^\diamond)^2] < \infty$, and $\{v_i^\diamond\}$ is an i.i.d. sequence of random variables with $\mathrm{E}[v_i^\diamond] = 0$ and $\mathrm{E}[(v_i^\diamond)^2] < \infty$ and is independent of $\{\varepsilon_{ij}^\diamond\}$, following the discussion in Remark 3, we can again show that the form of $\varphi_{n\diamond}(h_1)$ depends on the type of the longitudinal data, and thus the nonparametric conditional variance estimation has different convergence rates for sparse and dense data.

We next discuss how to obtain the optimal value of the parameter vector $\boldsymbol{\phi}$. Construct the residuals $\widetilde{\mathbf{e}}_i = \mathbf{Y}_i - \mathbf{Z}_i\widetilde{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\eta}}(\mathbf{X}_i, \widetilde{\boldsymbol{\theta}})$, where $\widetilde{\boldsymbol{\eta}}(\mathbf{X}_i, \widetilde{\boldsymbol{\theta}})$ is defined in the same way as $\boldsymbol{\eta}(\mathbf{X}_i, \boldsymbol{\theta})$ but with $\eta(\cdot)$ and $\boldsymbol{\theta}$ replaced by $\widetilde{\eta}(\cdot) \equiv \widehat{\eta}(\cdot|\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\theta}})$ and $\widetilde{\boldsymbol{\theta}}$, respectively. Let $\widetilde{\boldsymbol{\Lambda}}_i \equiv \widehat{\boldsymbol{\Lambda}}_i(\widetilde{\boldsymbol{\theta}})$, $\widehat{\boldsymbol{\Sigma}}_i = \mathrm{diag}\{\widehat{\sigma}^2(t_{i1}), \ldots, \widehat{\sigma}^2(t_{im_i})\}$, and define $\mathbf{R}_i^*(\boldsymbol{\phi}) = \widehat{\boldsymbol{\Sigma}}_i^{1/2}\mathbf{C}_i(\boldsymbol{\phi})\widehat{\boldsymbol{\Sigma}}_i^{1/2}$. Motivated by equations (3.1) and (3.2), we construct

$$\boldsymbol{\Omega}_0^*(\boldsymbol{\phi}) = \sum_{i=1}^n \widetilde{\boldsymbol{\Lambda}}_i^\top \big[\mathbf{R}_i^*(\boldsymbol{\phi})\big]^{-1}\widetilde{\boldsymbol{\Lambda}}_i \tag{4.10}$$

and

$$\boldsymbol{\Omega}_1^*(\boldsymbol{\phi}) = \sum_{i=1}^n \widetilde{\boldsymbol{\Lambda}}_i^\top \big[\mathbf{R}_i^*(\boldsymbol{\phi})\big]^{-1}\widetilde{\mathbf{e}}_i\widetilde{\mathbf{e}}_i^\top \big[\mathbf{R}_i^*(\boldsymbol{\phi})\big]^{-1}\widetilde{\boldsymbol{\Lambda}}_i. \tag{4.11}$$

By Theorem 1, the sandwich formula estimate $[\boldsymbol{\Omega}_0^*(\boldsymbol{\phi})]^+\boldsymbol{\Omega}_1^*(\boldsymbol{\phi})[\boldsymbol{\Omega}_0^*(\boldsymbol{\phi})]^+$ is asymptotically proportional to the asymptotic covariance of the proposed SGEE estimators when the inverse of $\mathbf{R}_i^*(\boldsymbol{\phi})$ is chosen as the weight matrix. The optimal value of $\boldsymbol{\phi}$, denoted by $\widehat{\boldsymbol{\phi}}$, can be chosen to minimize the determinant $|[\boldsymbol{\Omega}_0^*(\boldsymbol{\phi})]^+\boldsymbol{\Omega}_1^*(\boldsymbol{\phi})[\boldsymbol{\Omega}_0^*(\boldsymbol{\phi})]^+|$. Such a method is called the minimum generalized variance method [Fan, Huang and Li (2007)]. With the chosen $\widehat{\boldsymbol{\phi}}$, we can estimate the covariance matrices by

$$\mathbf{R}_i(\widehat{\boldsymbol{\phi}}) = \widehat{\boldsymbol{\Sigma}}_i^{1/2}\mathbf{C}_i(\widehat{\boldsymbol{\phi}})\widehat{\boldsymbol{\Sigma}}_i^{1/2}, \tag{4.12}$$

whose inverse will be used as the weight matrices in the SGEE method.

5. Numerical studies. In this section, we first study the finite sample performance of the proposed SGEE estimators through Monte Carlo simulation, and then give an empirical application of the proposed model and methodology.

5.1. *Simulation studies.* We investigate both sparse and dense longitudinal data cases with an average time dimension $\overline{m}$ of 10 for the sparse data and 30 for the dense data. We use two types of within-subject correlation structure, AR(1) and ARMA(1, 1), in the error terms $e_i(t_{ij})$. We investigate the finite sample performance of the proposed estimators under both correct specification and misspecification of the correlation structure in the construction of the covariance matrix estimator proposed in Section 4. For the misspecified case, we fit an AR(1) correlation structure while the true underlying structure is ARMA(1, 1) and examine the robustness of the estimators.

Simulated data are generated from model (1.2) with two-dimensional $\mathbf{Z}_i(t_{ij})$ and three-dimensional $\mathbf{X}_i(t_{ij})$, and

$$\boldsymbol{\beta}_0 = (2, 1)^\top, \qquad \boldsymbol{\theta}_0 = (2, 1, 2)^\top/3 \quad \text{and} \quad \eta(u) = 0.5\exp(u).$$

The covariates $(\mathbf{Z}_i^\top(t_{ij}), \mathbf{X}_i^\top(t_{ij}))^\top$ are generated independently from a five-dimensional Gaussian distribution with mean $\mathbf{0}$, variance 1 and pairwise correlation 0.1. The observation times t_{ij} are generated in the same way as in Fan, Huang and Li (2007): for each subject, $\{0, 1, 2, \dots, T\}$ is a set of scheduled times, and each scheduled time from 1 to T has a 0.2 probability of being skipped; each actual observation time is a perturbation of a nonskipped scheduled time; that is, a uniform [0, 1] random number is added to the nonskipped scheduled time. Here T is set to be 12 or 36, which corresponds to an average time dimension of $\overline{m} = 10$ or $\overline{m} = 30$, respectively. For each i, the error terms $e_i(t_{ij})$ are generated from a Gaussian process with mean 0, variance function

$$\operatorname{var}[e(t)] = \sigma^2(t) = 0.25\exp(t/12) \tag{5.1}$$

and serial correlation structure

$$\operatorname{cor}(e(t), e(s)) = \begin{cases} 1, & t = s, \\ \gamma\rho^{|t-s|}, & t \neq s. \end{cases} \tag{5.2}$$

Note that (5.2) corresponds to an ARMA$(1, 1)$ correlation structure and reduces to an AR(1) correlation structure when $\gamma = 1$. The number of subjects, n, is taken to be 30 or 50. The values for γ and ρ are $(\gamma, \rho) = (0.85, 0.9)$ in the ARMA$(1, 1)$ correlation structure and $(\gamma, \rho) = (1, 0.9)$ in the AR(1) structure.

For each combination of $\overline{m}$, n, and the correlation structure, the number of simulation replications is 200. For the selection of the bandwidth, however, due to the running time limitation, we first run a leave-one-unit-out (i.e., leave out observations from one subject at a time) cross-validation (CV) to choose the optimal bandwidths from 20 replications. We then use the average of the optimal bandwidths from these 20 replications as the bandwidth in the 200 replications of the simulation study. For the SGEE method, we choose the weight matrix as the inverse of the estimated within-subject covariance matrix as constructed in (4.12) of Section 4. We first study the performance of the proposed estimators in the case where the correlation structure in the estimation of the covariance matrix is correctly specified, and then investigate the robustness of the estimators to the misspecification of the correlation structure. The bias, calculated as the average of the estimates from the 200 replications minus the true parameter values, the standard deviation (SD), calculated as the sample standard deviation of the 200 estimates and the median absolute deviation (MAD), calculated as the median absolute deviation of the 200 estimates are reported in Tables 1 and 2. Table 1 gives the results obtained under the correct specification of an underlying within-subject AR(1) correlation structure in $e_i(t_{ij})$, and Table 2 gives those obtained under the correct specification of an underlying ARMA$(1, 1)$ structure in $e_i(t_{ij})$. For comparison, we also report the results from the PULS estimation. The results in Tables 1 and 2 show that the SGEE estimates are comparable with the PULS estimates in terms of bias and are more efficient than the PULS estimates, which supports the asymptotic theory developed in Section 3. In Figures 1 and 2, we plot the local linear estimated link function from a typical realization together with the real curve for each combination of n and $\overline{m}$.

To investigate the robustness of the SGEE and PULS estimators to correlation structure misspecification, we also carry out a simulation study in which an AR(1) correlation structure is used in the covariance matrix estimation detailed in Section 4, when the true underlying correlation structure is ARMA$(1, 1)$. Table 3 reports the results under this misspecification. The table shows that in the presence of correlation structure misspecification, SGEE still produces more efficient parameter estimates than PULS.

We also include a simulated example where the covariates in $\mathbf{Z}$ follow discrete distributions. The same model as above is used except that the covariates $\mathbf{X}_i^\top(t_{ij})$ are drawn independently from a three-dimensional Gaussian distribution with mean $\mathbf{0}$, variance 1 and pairwise correlation 0.1, and $\mathbf{Z}_i^\top(t_{ij})$ are independently drawn from a binomial distribution with success probability 0.5. The errors $e_i(t_{ij})$ are generated with the AR(1) serial correlation structure of $(\gamma, \rho) = (1, 0.9)$. The simulation results for this example are presented in Table 4. The same finding as

TABLE 1
Performance of parameter estimation methods under correct specification of an underlying AR(1) *correlation structure*

	n		**30**			**50**		
$\overline{m}$	**Parameters**	**Methods**	**Bias**	**SD**	**MAD**	**Bias**	**SD**	**MAD**
10	β_1	PULS	0.0048	0.0402	0.0288	−0.0030	0.0308	0.0195
		SGEE	−0.0026	0.0508	0.0081	−0.0016	0.0259	0.0074
	β_2	PULS	−0.0024	0.0409	0.0243	0.0049	0.0267	0.0180
		SGEE	−0.0018	0.0298	0.0110	0.0033	0.0310	0.0077
	θ_1	PULS	−0.0049	0.0299	0.0180	−0.0009	0.0197	0.0134
		SGEE	−0.0013	0.0164	0.0083	−0.0002	0.0118	0.0046
	θ_2	PULS	0.0011	0.0380	0.0229	−0.0016	0.0237	0.0161
		SGEE	0.0026	0.0188	0.0100	0.0006	0.0108	0.0067
	θ_3	PULS	0.0018	0.0314	0.0188	0.0006	0.0203	0.0147
		SGEE	−0.0007	0.0182	0.0090	−0.0004	0.0088	0.0052
30	β_1	PULS	0.0003	0.0408	0.0277	0.0016	0.0328	0.0222
		SGEE	−0.0081	0.1134	0.0106	0.0007	0.0108	0.0083
	β_2	PULS	−0.0020	0.0425	0.0317	0.0005	0.0351	0.0202
		SGEE	−0.0017	0.0420	0.0096	−0.0064	0.0152	0.0079
	θ_1	PULS	0.0020	0.0315	0.0213	−0.0020	0.0244	0.0182
		SGEE	−0.0008	0.0247	0.0075	0.0001	0.0148	0.0064
	θ_2	PULS	−0.0035	0.0340	0.0240	−0.0083	0.0278	0.0163
		SGEE	−0.0027	0.0242	0.0090	−0.0013	0.0104	0.0066
	θ_3	PULS	−0.0027	0.0321	0.0185	0.0045	0.0267	0.0169
		SGEE	0.0009	0.0230	0.0074	0.0001	0.0162	0.0068

above can be obtained. Some additional results, that is, those on the average angles between the estimated and the true parameter vectors, are given in Appendix D of the supplementary material [Chen et al. (2015)].

5.2. *Real data analysis*. We next illustrate the partially linear single-index model and the proposed SGEE estimation method through an empirical example which explores the relationship between lung function and air pollution. There is voluminous literature studying the effects of air pollution on people's health. For a review of the literature, the reader is referred to Pope, Bates and Raizenne (1995). Many studies have found association between air pollution and health problems such as increased respiratory symptoms, decreased lung function, increased hospitalizations or hospital visits for respiratory and cardiovascular diseases and increased respiratory morbidity [Dockery et al. (1989), Kinney et al. (1989), Pope (1991), Braun-Fahrländer et al. (1992), Lipfert and Hammerstrom (1992)]. While earlier research often used time series or cross-sectional data to evaluate the health

TABLE 2
Performance of parameter estimation methods under correct specification of an underlying ARMA(1, 1) *correlation structure*

	n		**30**			**50**		
$\overline{m}$	**Parameters**	**Methods**	**Bias**	**SD**	**MAD**	**Bias**	**SD**	**MAD**
10	β_1	PULS	−0.0029	0.0400	0.0280	0.0006	0.0322	0.0221
		SGEE	−0.0025	0.0244	0.0155	0.0000	0.0193	0.0124
	β_2	PULS	0.0032	0.0386	0.0282	−0.0045	0.0299	0.0205
		SGEE	0.0009	0.0249	0.0171	0.0001	0.0212	0.0126
	θ_1	PULS	−0.0004	0.0267	0.0181	−0.0003	0.0188	0.0126
		SGEE	−0.0002	0.0161	0.0104	0.0006	0.0146	0.0073
	θ_2	PULS	−0.0047	0.0343	0.0209	0.0005	0.0223	0.0156
		SGEE	−0.0031	0.0192	0.0113	−0.0002	0.0145	0.0087
	θ_3	PULS	0.0008	0.0253	0.0158	−0.0009	0.0201	0.0121
		SGEE	0.0011	0.0148	0.0102	−0.0009	0.0146	0.0074
30	β_1	PULS	−0.0026	0.0450	0.0296	−0.0016	0.0374	0.0273
		SGEE	0.0005	0.0214	0.0138	0.0015	0.0288	0.0105
	β_2	PULS	−0.0013	0.0461	0.0291	0.0035	0.0361	0.0252
		SGEE	0.0040	0.0335	0.0147	0.0014	0.0152	0.0104
	θ_1	PULS	−0.0014	0.0296	0.0192	−0.0010	0.0207	0.0159
		SGEE	−0.0005	0.0166	0.0095	0.0006	0.0092	0.0063
	θ_2	PULS	−0.0050	0.0355	0.0231	0.0011	0.0229	0.0173
		SGEE	−0.0037	0.0371	0.0120	−0.0003	0.0116	0.0072
	θ_3	PULS	0.0017	0.0279	0.0186	−0.0006	0.0215	0.0154
		SGEE	0.0009	0.0181	0.0095	−0.0007	0.0100	0.0070

effects of air pollution, recent advances in longitudinal data analysis techniques offer greater opportunities for studying this problem. In this paper, we will examine whether air pollution has a significant adverse effect on lung function, and, if so, to what extent. The use of the partially linear single-index model and the SGEE method would provide greater modeling flexibility than linear models and allow the within-subject correlation to be adequately taken into account. We will use a longitudinal data set obtained from a study where a total of 971 4th-grade children aged between 8 and 14 years (at their first visit to the hospital/clinic) were followed over 10 years. For each yearly visit of the children to the hospital/clinic, records on their forced expiratory volume (FEV), asthma symptom at visit (ASSPM, 1 for those with symptoms and 0 for those without), asthmatic status (ASS, 1 for asthma patient and 0 for nonasthma patient), gender (G, 1 for males and 0 for females), race (R, 1 for nonwhites and 0 for whites), age (A), height (H), BMI and respiratory infection at visit (RINF, 1 for those with infection and 0 for those without) were taken. Together with the measurements from the children, the mean levels of

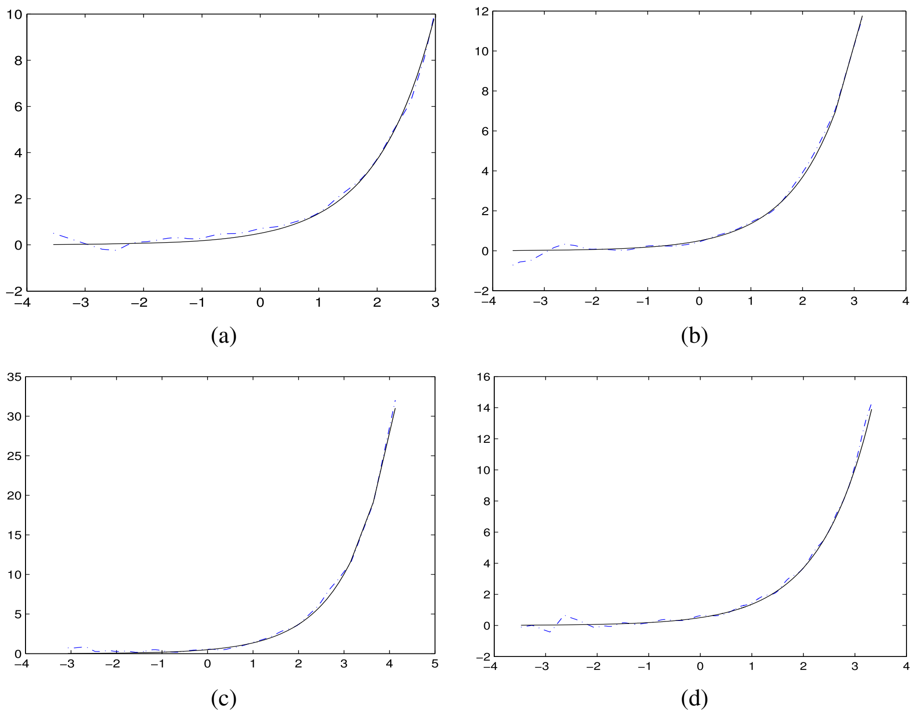

FIG. 1. *Estimated link function* (*dot-dashed line*), *together with the true link function* (*solid line*), *from a typical realization of model* (1.2) *with* AR(1) *correlation structure for each combination of* n *and* $\overline{m}$: (a) $n = 30$, $\overline{m} = 10$; (b) $n = 50$, $\overline{m} = 10$; (c) $n = 30$, $\overline{m} = 30$; (d) $n = 50$, $\overline{m} = 30$.

ozone and NO_2 in the month prior to the visit were also recorded. Due to dropout or other reasons, the majority of children had 4 to 5 years of records, and the total number of observations in the data set is 3809.

As in many other studies, the FEV will be used as a measure of lung function, and its log-transformed values, log(FEV), will be used as the response values in our model. Our main interest is to determine whether higher levels of ozone and NO_2 would lead to decrements in lung function. To account for the effects of other confounding factors, we include all other recorded variables. As age and height exhibit strong co-linearity (with a correlation of 0.78), we will only use height in the study. In fitting the partially linear single-index model to the data, all the continuous variables (i.e., FEV, H, BMI, OZONE and NO_2) are log-transformed, and the log(BMI), log(OZONE) and log(NO_2) are included in the single-index part. The log(H) and all the binary variables are included in the linear part of the model.

The scatter plots of the response variable against the continuous regressors are shown in Figure 3, and the box plots of the response against the binary regressors are given in Figure 4. We use an ARMA(1, 1) within-subject correlation structure in the estimation of the covariance matrix for the proposed SGEE method. The

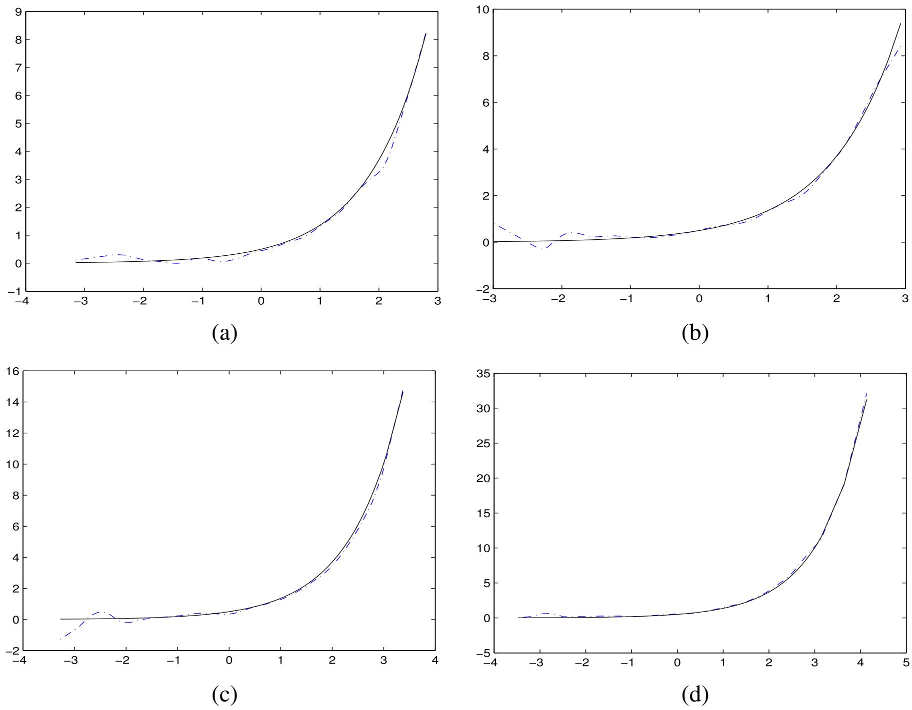

FIG. 2. *Estimated link function* (*dot-dashed line*), *together with the true link function* (*solid line*), *from a typical realization of model* (1.2) *with* ARMA(1, 1) *correlation structure for each combination of* n *and* $\overline{m}$: (a) $n = 30$, $\overline{m} = 10$; (b) $n = 50$, $\overline{m} = 10$; (c) $n = 30$, $\overline{m} = 30$; (d) $n = 50$, $\overline{m} = 30$.

resulting estimated model is as follows:

$$\begin{aligned}
&\log(\mathrm{FEV})\\
&\approx \underset{(0.0041)}{0.0325} * \mathrm{G} - \underset{(0.0080)}{0.0111} * \mathrm{ASS} - \underset{(0.0059)}{0.0671} * \mathrm{R}\\
&\qquad - \underset{(0.0085)}{0.0047} * \mathrm{ASSPM} - \underset{(0.0043)}{0.0068} * \mathrm{RINF} + \underset{(0.0307)}{2.3206} * \log(\mathrm{H}),\\
&\qquad + \widehat{\eta}\big[\underset{(0.0560)}{0.9929} * \log(\mathrm{BMI}) - \underset{(0.0127)}{0.0924} * \log(\mathrm{OZONE}) - \underset{(0.0125),}{0.0753} * \log(\mathrm{NO_2})\big]
\end{aligned}$$

where the numbers in the parentheses under the estimated coefficients are their respective estimated standard errors. The estimated link function and its 95% point-wise confidence intervals are plotted in Figure 5.

From Figure 5, it can be seen that the estimated link function is overall increasing. The 95% point-wise confidence intervals show that a linear functional form for the unknown link function would be rejected, and thus the partially liner single-index model might be more appropriate than the traditional linear regression

TABLE 3
Performance of parameter estimation methods under misspecification of an underlying ARMA(1, 1) *correlation structure*

	n		30			50		
$\overline{m}$	**Parameters**	**Methods**	**Bias**	**SD**	**MAD**	**Bias**	**SD**	**MAD**
10	β_1	PULS	0.0072	0.0410	0.0357	−0.0038	0.0299	0.0201
		SGEE	−0.0054	0.0261	0.0210	−0.0055	0.0211	0.0147
	β_2	PULS	0.0068	0.0336	0.0256	0.0037	0.0290	0.0163
		SGEE	0.0025	0.0267	0.0157	0.0023	0.0190	0.0136
	θ_1	PULS	0.0037	0.0166	0.0114	0.0061	0.0157	0.0096
		SGEE	0.0033	0.0144	0.0122	0.0016	0.0163	0.0081
	θ_2	PULS	−0.0092	0.0303	0.0184	−0.0084	0.0224	0.0174
		SGEE	−0.0007	0.0198	0.0144	−0.0045	0.0203	0.0130
	θ_3	PULS	−0.0005	0.0229	0.0158	−0.0028	0.0160	0.0111
		SGEE	−0.0035	0.0141	0.0094	0.0000	0.0134	0.0092
30	β_1	PULS	0.0066	0.0403	0.0259	−0.0221	0.0502	0.0252
		SGEE	0.0093	0.0144	0.0087	0.0001	0.0165	0.0118
	β_2	PULS	−0.0138	0.0435	0.0353	0.0107	0.0312	0.0233
		SGEE	−0.0017	0.0268	0.0096	0.0035	0.0170	0.0096
	θ_1	PULS	0.0027	0.0252	0.0165	0.0020	0.0181	0.0067
		SGEE	0.0054	0.0136	0.0078	0.0019	0.0096	0.0098
	θ_2	PULS	−0.0063	0.0265	0.0245	0.0021	0.0315	0.0273
		SGEE	0.0009	0.0198	0.0118	0.0046	0.0136	0.0094
	θ_3	PULS	−0.0011	0.0285	0.0258	−0.0042	0.0217	0.0136
		SGEE	−0.0065	0.0178	0.0137	−0.0046	0.0120	0.0084

model. Meanwhile, it can be seen from the above estimated model that height and BMI are significant positive factors in accounting for lung function. Taller children and children with larger BMI tend to have higher FEV. Furthermore, male and white children have, on average, higher FEV than female or nonwhite children. Furthermore, both OZONE and NO_2 in the single-index component have negative effects on children's lung function, as the estimated coefficients for OZONE and NO_2 are negative, and the estimated link function is increasing. Although these negative effects are relatively small in magnitude compared to the effect of BMI, they are statistically significant. This means that higher levels of ozone and NO_2 tend to lead to reduced lung function as represented by lower values of FEV.

6. Conclusions and discussions. In this paper, we study a partially linear single-index modeling structure for possibly unbalanced longitudinal data in a general framework, which includes both the sparse and dense longitudinal data

TABLE 4
Performance of parameter estimation methods under correct specification of an underlying AR(1) *correlation structure when the covariates in* $\mathbf{Z}$ *are discrete*

	n		30			50		
$\overline{m}$	**Parameters**	**Methods**	**Bias**	**SD**	**MAD**	**Bias**	**SD**	**MAD**
10	β_1	PULS	0.0215	0.0530	0.0404	0.0018	0.0646	0.0472
		SGEE	0.0228	0.0511	0.0208	0.0037	0.0298	0.0138
	β_2	PULS	−0.0309	0.0858	0.0735	0.0193	0.0526	0.0498
		SGEE	0.0024	0.0313	0.0193	0.0074	0.0339	0.0274
	θ_1	PULS	−0.0012	0.0185	0.0090	−0.0116	0.0201	0.0175
		SGEE	−0.0060	0.0157	0.0082	0.0020	0.0086	0.0066
	θ_2	PULS	−0.0020	0.0263	0.0232	0.0138	0.0229	0.0172
		SGEE	0.0122	0.0241	0.0143	−0.0004	0.0087	0.0063
	θ_3	PULS	0.0012	0.0206	0.0075	0.0036	0.0153	0.0132
		SGEE	−0.0008	0.0078	0.0048	−0.0020	0.0070	0.0034
30	β_1	PULS	0.0075	0.0427	0.0222	0.0108	0.0723	0.0513
		SGEE	0.0061	0.0284	0.0233	0.0033	0.0226	0.0175
	β_2	PULS	−0.0143	0.0768	0.0401	0.0023	0.0681	0.0417
		SGEE	0.0116	0.0275	0.0125	−0.0039	0.0259	0.0196
	θ_1	PULS	−0.0159	0.0310	0.0252	0.0031	0.0218	0.0168
		SGEE	−0.0030	0.0083	0.0045	0.0015	0.0098	0.0064
	θ_2	PULS	−0.0026	0.0192	0.0112	0.0048	0.0252	0.0200
		SGEE	0.0040	0.0200	0.0133	0.0002	0.0115	0.0084
	θ_3	PULS	0.0151	0.0331	0.0308	−0.0067	0.0228	0.0150
		SGEE	0.0006	0.0133	0.0083	−0.0018	0.0103	0.0064

cases. An SGEE method with the first-stage local linear smoothing is introduced to estimate the two parameter vectors as well as the unspecified link function. In Theorems 1 and 2, we derive the asymptotic properties of the proposed parametric and nonparametric estimators in different scenarios, from which we find that the convergence rates and asymptotic variances of the resulting estimators in the sparse longitudinal data case could be substantially different from those in the dense longitudinal data. In Section 4, we propose a semiparametric method to estimate the error covariance matrices which are involved in the estimation equations. The conditional variance function is estimated by using the log-transformed local linear method, and the parameters in the correlation matrices are estimated by the minimum generalized variance method. In particular, if the correlation matrices are correctly specified, as is stated in Corollary 1, the SGEE-based estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$ are generally asymptotically more efficient than the corresponding PULS estimators $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\theta}}$ in the sense that the asymptotic covariance matrix of the SGEE estimators minus that of the PULS estimators is negative semi-definite. Both the

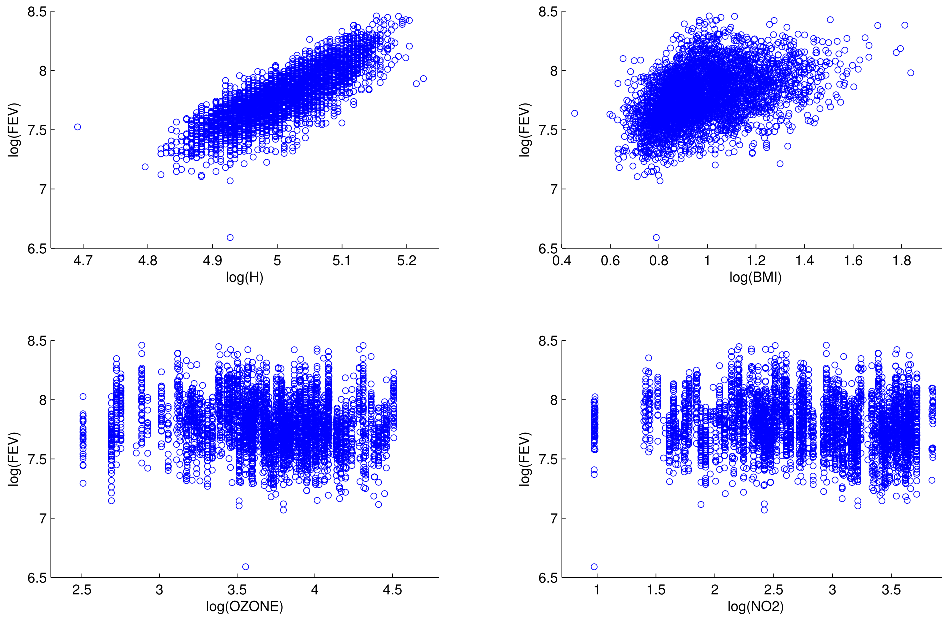

FIG. 3. *The scatter plots of the response variable* log(FEV) *against the continuous regressors*, *that is*, (*clockwise from top left*) log(H), log(BMI), log(NO_2), log(OZONE).

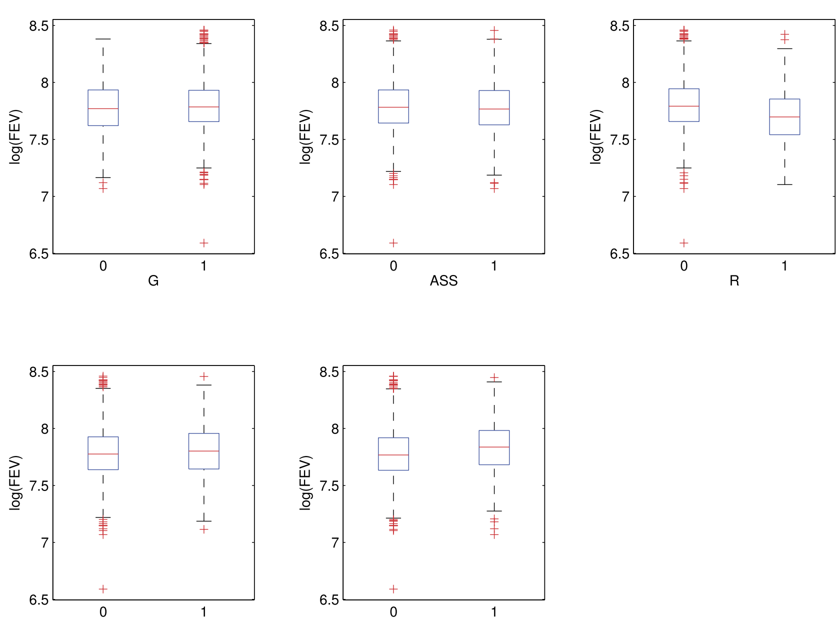

FIG. 4. *The box plots of the response variable* log(FEV) *against the binary regressors*, *that is*, (*clockwise from top left*) *G*, *ASS*, *R*, *RINF*, *ASSPM*.

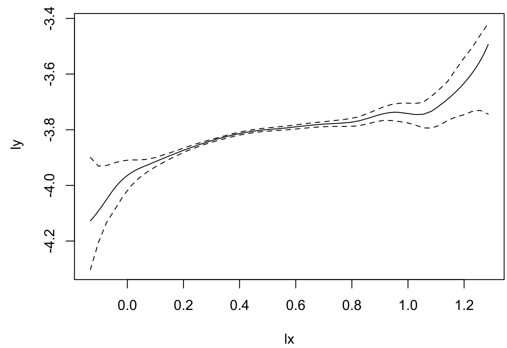

FIG. 5. *The estimated link function and its* 95% *point-wise confidence intervals.*

simulation study and empirical data analysis in Section 5 show that the proposed methods work well in the finite samples.

Recently, Yao and Li (2013) developed a new nonparametric regression function estimation method for a longitudinal regression model. This method takes into account the within-subject correlation information and thus generally improves the asymptotic estimation efficiency. It would also be interesting to incorporate the within-subject correlation information in the local linear estimation of the unknown link function in this paper and to examine both theoretical and empirical performance of the resulting estimator. We will leave this issue for future research. Another possible future topic is to extend the semiparametric techniques of variable selection and specification testing proposed by Liang et al. (2010) from the i.i.d. case to the general longitudinal data case discussed in the present paper.

APPENDIX A: REGULARITY CONDITIONS

To establish the asymptotic properties of the SGEE estimators proposed in Section 2, we introduce the following regularity conditions, although some of them might not be the weakest possible.

ASSUMPTION 1. The kernel function $K(\cdot)$ is a bounded and symmetric probability density function with compact support. Furthermore, the kernel function has a continuous first-order derivative function denoted by $\dot{K}(\cdot)$.

ASSUMPTION 2. (i) The errors $e_{ij} \equiv e_i(t_{ij})$, $1 \le i \le n$, $1 \le j \le m_i$, are independent across i; that is, $\mathbf{e}_i$ defined in Section 2, $1 \le i \le n$, are mutually independent.

(ii) The covariates $\mathbf{X}_{ij}$ and $\mathbf{Z}_{ij}$, $1 \le i \le n$, $1 \le j \le m_i$, are i.i.d. random vectors.

(iii) The errors e_{ij} are independent of the covariates $\mathbf{Z}_{ij}$ and $\mathbf{X}_{ij}$, and for each i, e_{ij}, $1 \le j \le m_i$, may be correlated with each other. Furthermore, $\mathrm{E}[e_{ij}] = 0$, $0 <$

$\mathrm{E}[e_{ij}^2] < \infty$ and $\mathrm{E}[|e_{ij}|^{2+\delta}] < \infty$ for some $\delta > 0$. The largest eigenvalues of $\mathbf{W}_i$ and $\mathbf{W}_i \mathrm{E}[\mathbf{e}_i \mathbf{e}_i]\mathbf{W}_i$ are bounded for any i.

ASSUMPTION 3. (i) The density function $f_{\boldsymbol{\theta}}(\cdot)$ of $\mathbf{X}_{ij}^\top \boldsymbol{\theta}$ is positive and has a continuous second-order derivative in $\mathcal{U} = \{\mathbf{x}^\top \boldsymbol{\theta} : \mathbf{x} \in \mathcal{X}, \boldsymbol{\theta} \in \Theta\}$, where Θ is a compact parameter space for $\boldsymbol{\theta}$ and $\mathcal{X}$ is a compact support of $\mathbf{X}_{ij}$.

(ii) The function $\rho_{\mathbf{Z}}(u|\boldsymbol{\theta}) = \mathrm{E}[\mathbf{Z}_{ij}|\mathbf{X}_{ij}^\top \boldsymbol{\theta} = u]$ has a bounded and continuous second-order derivative (with respect to u) for any $\boldsymbol{\theta} \in \Theta$, and $\mathrm{E}[\|\mathbf{Z}_{ij}\|^{2+\delta}] < \infty$, where δ was defined in Assumption 2(iii).

ASSUMPTION 4. The link function $\eta(\cdot)$ has continuous derivatives up to the second order.

ASSUMPTION 5. The bandwidth h satisfies

$$\omega_n h^6 \to 0, \qquad \frac{n^2 h^2}{N_n(h) \log n} \to \infty, \qquad \frac{T_n^{2/(2+\delta)} \log n}{h^2 N_n(h)} = o(1), \tag{A.1}$$

where $N_n(h) = \sum_{i=1}^n 1/(m_i h)$, $T_n = \sum_{i=1}^n m_i$ and δ was defined in Assumption 2(iii). Furthermore, $\max_{1 \le i \le n}(m_i^4 + m_i^3 h^{-1}) = o(w_n)$.

REMARK 5. Assumption 1 imposes some mild restrictions on the kernel functions, which have been used in the existing literature in i.i.d. and weakly dependent time series cases; see, for example, Fan and Gijbels (1996) and Gao (2007). The compact support restriction on the kernel functions can be removed if we impose certain restrictions on the tail of the kernel function. In Assumption 2(i), the longitudinal data under investigation is assumed to be independent across subjects i, which is not uncommon in longitudinal data analysis; see, for example, Wu and Zhang (2006) and Zhang, Fan and Sun (2009). Assumption 2(ii) is imposed to simplify the presentation of the asymptotic results. However, we may replace Assumption 2(ii) with the conditions that the covariates $\mathbf{X}_{ij}$ and $\mathbf{Z}_{ij}$ are i.i.d. across i and identically distributed across j, and in the case of dense longitudinal data, it is further satisfied that for $\kappa = 0, 1, 2, \ldots,$

$$\mathrm{Var}\left[\frac{1}{m_i} \sum_{j=1}^{m_i} \frac{U_{ij}}{h} \left(\frac{\mathbf{X}_{ij}^\top \boldsymbol{\theta} - u}{h}\right)^{\kappa} K\left(\frac{\mathbf{X}_{ij}^\top \boldsymbol{\theta} - u}{h}\right)\right] \le C(m_i h)^{-1} \tag{A.2}$$

uniformly for $u \in \mathcal{U}$ and $\boldsymbol{\theta} \in \Theta$, where U_{ij} can be 1, $\mathbf{Z}_{ij} B_1(\mathbf{Z}_{ij})$, or $\mathbf{X}_{ij} B_2(\mathbf{X}_{ij})$, $B_1(\cdot)$ and $B_2(\cdot)$ are two bounded functions, and C is a positive constant which is independent of i. When $\mathbf{X}_{ij}$ and $\mathbf{Z}_{ij}$ are stationary and α-mixing dependent across j for the case of dense longitudinal data, it is easy to validate the high-level condition (A.2). In Assumption 2(iii), we allow the error terms to have certain within-subject correlation, which makes the model assumptions more realistic. Assumption 3 gives some commonly-used conditions in partially linear single-index models; see Xia and Härdle (2006) and Chen, Gao and Li (2013b), for example. Assumption 4 is a mild smoothness condition on the link function imposed for the application of the local linear fitting. Assumption 5 gives a set of restrictions on the bandwidth h, which is involved in the estimation of the link function. Note that the bandwidth conditions in Assumption 5 imply that the milder bandwidth conditions in (C.1) of Lemma 1 in the supplemental material [Chen et al. (2015)] are satisfied. Hence we can use Lemma 1 to prove our main theoretical results.

We next give some regularity conditions, which are needed to derive the asymptotic property of the nonparametric conditional variance estimators in Section 4.

ASSUMPTION 6. The kernel function $K_1(\cdot)$ is a continuous and symmetric probability density function with compact support.

ASSUMPTION 7. The observation times, t_{ij}, are i.i.d. and have a continuous and positive probability density function $f_T(t)$, which has a compact support $\mathcal{T}$. The density function of $\xi^2(t_{ij})$ is continuous and bounded. Let $\delta > 2$, which strengthens the moment conditions in Assumptions 2 and 3.

ASSUMPTION 8. The conditional variance function $\sigma^2(\cdot)$ has a continuous second-order derivative and satisfies $\inf_{t\in\mathcal{T}}\sigma^2(t) > 0$. Let $\dot\sigma^2(\cdot)$ and $\ddot\sigma^2(\cdot)$ be its first-order and second-order derivative functions, respectively.

ASSUMPTION 9. The bandwidth h_1 satisfies

$$h_1 \to 0, \qquad \frac{T_n^{2/(2+\delta/2)}\log n}{h_1^2 N_n(h_1)} = o(1), \tag{A.3}$$

where $N_n(h_1) = \sum_{i=1}^n 1/(m_i h_1)$.

REMARK 6. Assumption 7 imposes a mild condition on the observation times [see, e.g., Jiang and Wang (2011)] and strengthens the moment conditions on e_{ij} and $\mathbf{Z}_{ij}$. However, such moment conditions are not uncommon in the asymptotic theory for nonparametric conditional variance estimation [Chen, Cheng and Peng (2009)]. Since the local linear smoothing technique is applied, a certain smoothness condition has to be assumed on $\sigma^2(\cdot)$, as is done in Assumption 8. Assumption 9 gives some mild restrictions on the bandwidth h_1, which is used in the estimation of the conditional variance function.

APPENDIX B: PROOFS OF THE MAIN RESULTS

In this appendix, we provide the detailed proofs of the main results given in Section 3.

B.1. Proof of Theorem 1. By the definition of the weighted local linear estimators in (2.4) and (2.5), we have

$$
\begin{aligned}
\widehat{\eta}(u|\boldsymbol{\beta},\boldsymbol{\theta}) - \eta(u) &= \sum_{i=1}^{n} \mathbf{s}_i(u|\boldsymbol{\theta})(\mathbf{Y}_i - \mathbf{Z}_i\boldsymbol{\beta}) - \eta(u) \\
&= \sum_{i=1}^{n} \mathbf{s}_i(u|\boldsymbol{\theta})\mathbf{e}_i + \sum_{i=1}^{n} \mathbf{s}_i(u|\boldsymbol{\theta})\mathbf{Z}_i(\boldsymbol{\beta}_0 - \boldsymbol{\beta}) \\
&\quad + \sum_{i=1}^{n} \mathbf{s}_i(u|\boldsymbol{\theta})\big[\boldsymbol{\eta}(\mathbf{X}_i,\boldsymbol{\theta}_0) - \boldsymbol{\eta}(\mathbf{X}_i,\boldsymbol{\theta})\big] \\
&\quad + \sum_{i=1}^{n} \mathbf{s}_i(u|\boldsymbol{\theta})\boldsymbol{\eta}(\mathbf{X}_i,\boldsymbol{\theta}) - \eta(u) \\
&\equiv I_{n1} + I_{n2} + I_{n3} + I_{n4}.
\end{aligned} \tag{B.1}
$$

For I_{n1}, note that by a first-order Taylor expansion of $K(\cdot)$, we have, for $i = 1, \ldots, n$ and $j = 1, \ldots, m_i$,

$$
K\left(\frac{\mathbf{X}_{ij}^{\top}\boldsymbol{\theta} - u}{h}\right) = K\left(\frac{\mathbf{X}_{ij}^{\top}\boldsymbol{\theta}_0 - u}{h}\right) + \dot{K}\left(\frac{\mathbf{X}_{ij}^{\top}\boldsymbol{\theta}_* - u}{h}\right)\frac{\mathbf{X}_{ij}^{\top}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)}{h},
$$

where $\dot{K}(\cdot)$ is the first-order derivative of $K(\cdot)$ and $\boldsymbol{\theta}_* = \boldsymbol{\theta}_0 + \lambda_*(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$, $0 < \lambda_* < 1$. Hence, by some standard calculations and the assumption that $n^2h^2/\{N_n(h)\log n\} \to \infty$, we have

$$
\begin{aligned}
I_{n1} &= \sum_{i=1}^{n} \mathbf{s}_i(u|\boldsymbol{\theta}_0)\mathbf{e}_i + \sum_{i=1}^{n} [\mathbf{s}_i(u|\boldsymbol{\theta}) - \mathbf{s}_i(u|\boldsymbol{\theta}_0)]\mathbf{e}_i \\
&= \sum_{i=1}^{n} \mathbf{s}_i(u|\boldsymbol{\theta}_0)\mathbf{e}_i + O_P\left(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \cdot \frac{\sqrt{N_n(h)\log n}}{nh}\right) \\
&= \sum_{i=1}^{n} \mathbf{s}_i(u|\boldsymbol{\theta}_0)\mathbf{e}_i + o_P(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|)
\end{aligned} \tag{B.2}
$$

for any $u \in \mathcal{U}$ and $\boldsymbol{\theta} \in \Theta$.

By Lemma 2 in the supplementary material [Chen et al. (2015)], we can prove that

$$
I_{n2} = -\rho_{\mathbf{Z}}^{\top}(u)(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + O_P(\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2 + \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2) \tag{B.3}
$$

for any $u \in \mathcal{U}$, where $\rho_{\mathbf{Z}}(u) \equiv \rho_{\mathbf{Z}}(u|\boldsymbol{\theta}_0) = \mathrm{E}[\mathbf{Z}_{ij}|\mathbf{X}_{ij}^{\top}\boldsymbol{\theta}_0 = u]$.

Note that

$$
\eta(\mathbf{X}_{ij}^{\top}\boldsymbol{\theta}) - \eta(\mathbf{X}_{ij}^{\top}\boldsymbol{\theta}_0) = \dot{\eta}(\mathbf{X}_{ij}^{\top}\boldsymbol{\theta}_0)\mathbf{X}_{ij}^{\top}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + O_P(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2),
$$

which, together with Lemma 3 in the supplementary material [Chen et al. (2015)], leads to

$$I_{n3} = -\dot{\eta}(u)\rho_{\mathbf{X}}^{\top}(u)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + O_P(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2) \tag{B.4}$$

for any $u \in \mathcal{U}$, where $\rho_{\mathbf{X}}(u) \equiv \rho_{\mathbf{X}}(u|\boldsymbol{\theta}_0) = \mathrm{E}[\mathbf{X}_{ij}|\mathbf{X}_{ij}^{\top}\boldsymbol{\theta}_0 = u]$.

By a second-order Taylor expansion of $\eta(\cdot)$ and the first-order Taylor expansion of $K(\cdot)$ used to handle I_{n1}, we can prove that, for any $u \in \mathcal{U}$, we have

$$I_{n4} = \tfrac{1}{2}\mu_2\ddot{\eta}(u)h^2[1 + O_P(h)] + o_P(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|). \tag{B.5}$$

Recall that $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}_1$ are the solutions to the equations in (2.8). By (B.1)–(B.5), we can prove that, uniformly for $i = 1, \dots, n$ and $j = 1, \dots, m_i$,

$$\begin{aligned}
&\widehat{\eta}(\mathbf{X}_{ij}^{\top}\widehat{\boldsymbol{\theta}}_1|\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}_1) - \eta(\mathbf{X}_{ij}^{\top}\boldsymbol{\theta}_0)\\
&\quad= \widehat{\eta}(\mathbf{X}_{ij}^{\top}\widehat{\boldsymbol{\theta}}_1|\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}_1) - \widehat{\eta}(\mathbf{X}_{ij}^{\top}\boldsymbol{\theta}_0|\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}_1) + \widehat{\eta}(\mathbf{X}_{ij}^{\top}\boldsymbol{\theta}_0|\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}_1) - \eta(\mathbf{X}_{ij}^{\top}\boldsymbol{\theta}_0)\\
&\quad= \widehat{\dot{\eta}}(\mathbf{X}_{ij}^{\top}\boldsymbol{\theta}_0|\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}_1)\mathbf{X}_{ij}^{\top}(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0) + \widehat{\eta}(\mathbf{X}_{ij}^{\top}\boldsymbol{\theta}_0|\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}_1) - \eta(\mathbf{X}_{ij}^{\top}\boldsymbol{\theta}_0)\\
&\qquad + O_P(\|\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0\|^2)\\
&\quad= \dot{\eta}(\mathbf{X}_{ij}^{\top}\boldsymbol{\theta}_0)[\mathbf{X}_{ij} - \rho_{\mathbf{X}}(\mathbf{X}_{ij}^{\top}\boldsymbol{\theta}_0)]^{\top}(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0)(1 + o_P(1))\\
&\qquad + \sum_{k=1}^{n}\mathbf{s}_k(\mathbf{X}_{ij}^{\top}\boldsymbol{\theta}_0)\mathbf{e}_k - \rho_{\mathbf{Z}}^{\top}(\mathbf{X}_{ij}^{\top}\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)(1 + o_P(1))\\
&\qquad + \tfrac{1}{2}\mu_2\ddot{\eta}(\mathbf{X}_{ij}^{\top}\boldsymbol{\theta}_0)h^2 + O_P(h^3) + O_P(\|\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0\|^2 + \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2),
\end{aligned} \tag{B.6}$$

where $\mathbf{s}_k(\mathbf{X}_{ij}^{\top}\boldsymbol{\theta}_0) \equiv \mathbf{s}_k(\mathbf{X}_{ij}^{\top}\boldsymbol{\theta}_0|\boldsymbol{\theta}_0)$.

By the definitions of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}_1$ [see (2.8) in Section 2], we have

$$\sum_{i=1}^{n}\widehat{\boldsymbol{\Lambda}}_i^{\top}(\widehat{\boldsymbol{\theta}}_1)\mathbf{W}_i[\mathbf{Y}_i - \mathbf{Z}_i\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\eta}}(\mathbf{X}_i|\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}_1)] = \mathbf{0}. \tag{B.7}$$

By the uniform consistency results for the local linear estimators (such as Lemmas 2 and 3 in the supplementary material [Chen et al. (2015)]), we can approximate $\widehat{\boldsymbol{\Lambda}}_i(\widehat{\boldsymbol{\theta}}_1)$ in (B.7) by $\boldsymbol{\Lambda}_i = \boldsymbol{\Lambda}_i(\boldsymbol{\theta}_0)$ when deriving the asymptotic distribution theory. Then we have

$$\begin{aligned}
\mathbf{0} &= \sum_{i=1}^{n}\widehat{\boldsymbol{\Lambda}}_i^{\top}(\widehat{\boldsymbol{\theta}}_1)\mathbf{W}_i[\mathbf{Y}_i - \mathbf{Z}_i\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\eta}}(\mathbf{X}_i|\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}_1)]\\
&= \sum_{i=1}^{n}\boldsymbol{\Lambda}_i^{\top}\mathbf{W}_i[\mathbf{Y}_i - \mathbf{Z}_i\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\eta}}(\mathbf{X}_i|\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}_1)]\\
&\quad + \sum_{i=1}^{n}(\widehat{\boldsymbol{\Lambda}}_i(\widehat{\boldsymbol{\theta}}_1) - \boldsymbol{\Lambda}_i)^{\top}\mathbf{W}_i[\mathbf{Y}_i - \mathbf{Z}_i\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\eta}}(\mathbf{X}_i|\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}_1)]\\
&\overset{P}{\sim} \sum_{i=1}^{n}\boldsymbol{\Lambda}_i^{\top}\mathbf{W}_i[\mathbf{Y}_i - \mathbf{Z}_i\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\eta}}(\mathbf{X}_i|\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}_1)][1 + O_P(\|\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0\|)],
\end{aligned} \tag{B.8}$$

where and below $a_n \stackrel{P}{\sim} b_n$ denotes $a_n = b_n(1 + o_P(1))$. Furthermore, note that

$$\mathbf{Y}_i - \mathbf{Z}_i\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\eta}}(\mathbf{X}_i|\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}_1) = \mathbf{e}_i - \mathbf{Z}_i(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) - [\widehat{\boldsymbol{\eta}}(\mathbf{X}_i|\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}_1) - \boldsymbol{\eta}(\mathbf{X}_i, \boldsymbol{\theta}_0)],$$

which, together with (B.6) and the bandwidth condition $\omega_n h^6 = o(1)$, implies that

$$\begin{aligned}
\sum_{i=1}^n &\boldsymbol{\Lambda}_i^\top \mathbf{W}_i[\mathbf{Y}_i - \mathbf{Z}_i\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\eta}}(\mathbf{X}_i|\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}_1)] \\
&= \sum_{i=1}^n \boldsymbol{\Lambda}_i^\top \mathbf{W}_i\mathbf{e}_i - \sum_{i=1}^n \boldsymbol{\Lambda}_i^\top \mathbf{W}_i\mathbf{Z}_i(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
&\quad - \sum_{i=1}^n \boldsymbol{\Lambda}_i^\top \mathbf{W}_i[\widehat{\boldsymbol{\eta}}(\mathbf{X}_i|\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}_1) - \boldsymbol{\eta}(\mathbf{X}_i, \boldsymbol{\theta}_0)] \\
&= -\sum_{i=1}^n \boldsymbol{\Lambda}_i^\top \mathbf{W}_i[\mathbf{Z}_i - \boldsymbol{\rho}_{\mathbf{Z}}(\mathbf{X}_i, \boldsymbol{\theta}_0)](\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)(1 + o_P(1)) \\
&\quad - \sum_{i=1}^n \boldsymbol{\Lambda}_i^\top \mathbf{W}_i\{[\dot{\boldsymbol{\eta}}(\mathbf{X}_i, \boldsymbol{\theta}_0) \otimes \mathbf{1}_p^\top] \odot [\mathbf{X}_i - \boldsymbol{\rho}_{\mathbf{X}}(\mathbf{X}_i, \boldsymbol{\theta}_0)]\} \\
&\qquad \times (\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0)(1 + o_P(1)) \\
&\quad + \sum_{i=1}^n \boldsymbol{\Lambda}_i^\top \mathbf{W}_i\left[\mathbf{e}_i - \sum_{k=1}^n \mathbf{s}_k(\mathbf{X}_i, \boldsymbol{\theta}_0)\mathbf{e}_k\right] \\
&\quad + O_P(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2 + \|\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0\|^2),
\end{aligned} \tag{B.9}$$

where $\mathbf{s}_k(\mathbf{X}_i, \boldsymbol{\theta}_0) = [\mathbf{s}_k^\top(\mathbf{X}_{i1}^\top\boldsymbol{\theta}_0), \ldots, \mathbf{s}_k^\top(\mathbf{X}_{im_i}^\top\boldsymbol{\theta}_0)]^\top$, $\boldsymbol{\rho}_{\mathbf{Z}}(\mathbf{X}_i, \boldsymbol{\theta}_0)$ and $\boldsymbol{\rho}_{\mathbf{X}}(\mathbf{X}_i, \boldsymbol{\theta}_0)$ were defined in Section 2. Following the standard proof in the existing literature [see, e.g., Ichimura (1993), Chen, Gao and Li (2013b)], we can show the weak consistency of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}_1$. Note that

$$\begin{aligned}
\sum_{i=1}^n &\boldsymbol{\Lambda}_i^\top \mathbf{W}_i\boldsymbol{\Lambda}_i \begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0 \end{pmatrix} \\
&= \sum_{i=1}^n \boldsymbol{\Lambda}_i^\top \mathbf{W}_i\{[\dot{\boldsymbol{\eta}}(\mathbf{X}_i, \boldsymbol{\theta}_0) \otimes \mathbf{1}_p^\top] \odot [\mathbf{X}_i - \boldsymbol{\rho}_{\mathbf{X}}(\mathbf{X}_i, \boldsymbol{\theta}_0)]\}(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0) \\
&\quad + \sum_{i=1}^n \boldsymbol{\Lambda}_i^\top \mathbf{W}_i[\mathbf{Z}_i - \boldsymbol{\rho}_{\mathbf{Z}}(\mathbf{X}_i, \boldsymbol{\theta}_0)](\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)
\end{aligned}$$

and

$$\sum_{i=1}^n \boldsymbol{\Lambda}_i^\top \mathbf{W}_i\left[\sum_{k=1}^n \mathbf{s}_k(\mathbf{X}_i, \boldsymbol{\theta}_0)\mathbf{e}_k\right] = o_P(\omega_n^{1/2}),$$

which, together with (B.8) and (B.9), lead to

$$\left[\sum_{i=1}^{n} \boldsymbol{\Lambda}_i^{\top} \mathbf{W}_i \boldsymbol{\Lambda}_i\right] \begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0 \end{pmatrix} \overset{P}{\sim} \sum_{i=1}^{n} \boldsymbol{\Lambda}_i^{\top} \mathbf{W}_i \mathbf{e}_i. \tag{B.10}$$

Define $\mathbf{I}(\boldsymbol{\theta}_0, \mathbf{B}_0) = \operatorname{diag}\{\mathbf{I}_d, \mathbf{M}\}, \mathbf{O}(\boldsymbol{\theta}_0) = \begin{pmatrix} \mathbf{O}_{d\times d} & \mathbf{O}_{d\times 1} \\ \mathbf{O}_{p\times d} & \boldsymbol{\theta}_0 \end{pmatrix}$, where $\mathbf{M} = (\boldsymbol{\theta}_0, \mathbf{B}_0)$ was defined in Section 3. It is easy to find that

$$\mathbf{I}_{d+p} = \mathbf{I}(\boldsymbol{\theta}_0, \mathbf{B}_0)\mathbf{I}^{\top}(\boldsymbol{\theta}_0, \mathbf{B}_0) = \mathbf{O}(\boldsymbol{\theta}_0)\mathbf{O}^{\top}(\boldsymbol{\theta}_0) + \mathbf{I}(\mathbf{B}_0)\mathbf{I}^{\top}(\mathbf{B}_0). \tag{B.11}$$

By the identification condition on $\boldsymbol{\theta}_0$, we may show that

$$\begin{aligned} \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 &= \frac{\widehat{\boldsymbol{\theta}}_1}{\|\widehat{\boldsymbol{\theta}}_1\|} - \frac{\boldsymbol{\theta}_0}{\|\boldsymbol{\theta}_0\|} = \frac{\widehat{\boldsymbol{\theta}}_1}{\|\widehat{\boldsymbol{\theta}}_1\|} - \frac{\boldsymbol{\theta}_0}{\|\widehat{\boldsymbol{\theta}}_1\|} + \frac{\boldsymbol{\theta}_0}{\|\widehat{\boldsymbol{\theta}}_1\|} - \frac{\boldsymbol{\theta}_0}{\|\boldsymbol{\theta}_0\|} \\ &\overset{P}{\sim} \frac{\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0}{\|\boldsymbol{\theta}_0\|} - \boldsymbol{\theta}_0\boldsymbol{\theta}_0^{\top} \frac{\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0}{\|\boldsymbol{\theta}_0\|} = (\mathbf{I}_p - \boldsymbol{\theta}_0\boldsymbol{\theta}_0^{\top})(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0), \end{aligned}$$

which implies that $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = \mathbf{B}_0\mathbf{B}_0^{\top}(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0)$ and

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \end{pmatrix} = \mathbf{I}(\mathbf{B}_0)\mathbf{I}^{\top}(\mathbf{B}_0) \begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0 \end{pmatrix}. \tag{B.12}$$

By (B.10), (B.11) and using the fact that $\boldsymbol{\Lambda}_i \mathbf{O}(\boldsymbol{\theta}_0) = \mathbf{0}$, we have

$$\mathbf{I}^{\top}(\mathbf{B}_0)\left[\sum_{i=1}^{n} \boldsymbol{\Lambda}_i^{\top} \mathbf{W}_i \boldsymbol{\Lambda}_i\right]\mathbf{I}(\mathbf{B}_0)\mathbf{I}^{\top}(\mathbf{B}_0) \begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0 \end{pmatrix} \overset{P}{\sim} \mathbf{I}^{\top}(\mathbf{B}_0)\left[\sum_{i=1}^{n} \boldsymbol{\Lambda}_i^{\top} \mathbf{W}_i \mathbf{e}_i\right],$$

which, together with (B.12), implies that

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \end{pmatrix} \overset{P}{\sim} \mathbf{I}(\mathbf{B}_0)\left\{\mathbf{I}^{\top}(\mathbf{B}_0)\left[\sum_{i=1}^{n} \boldsymbol{\Lambda}_i^{\top} \mathbf{W}_i \boldsymbol{\Lambda}_i\right]\mathbf{I}(\mathbf{B}_0)\right\}^{-1} \mathbf{I}^{\top}(\mathbf{B}_0)\left[\sum_{i=1}^{n} \boldsymbol{\Lambda}_i^{\top} \mathbf{W}_i \mathbf{e}_i\right].$$

Thus, by (3.1)–(3.3), the definition of the Moore–Penrose inverse and the classical central limit theorem for independent sequence, we can show that (3.4) in Theorem 1 holds.

B.2. Proof of Corollary 1. By Theorem 1, the PULS estimators $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\theta}}$ have the following asymptotic normal distribution:

$$\omega_n^{1/2} \begin{pmatrix} \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \end{pmatrix} \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega}_{0*}^{+}\boldsymbol{\Omega}_{1*}\boldsymbol{\Omega}_{0*}^{+}), \tag{B.13}$$

where $\boldsymbol{\Omega}_{0*}$ and $\boldsymbol{\Omega}_{1*}$ are two matrices such that

$$\frac{1}{\omega_n}\sum_{i=1}^{n} \boldsymbol{\Lambda}_i^{\top} \boldsymbol{\Lambda}_i \xrightarrow{P} \boldsymbol{\Omega}_{0*}, \qquad \frac{1}{\omega_n}\sum_{i=1}^{n} \mathrm{E}[\boldsymbol{\Lambda}_i^{\top} \mathbf{V}_i \boldsymbol{\Lambda}_i] \to \boldsymbol{\Omega}_{1*},$$

and $\mathbf{V}_i$ is the conditional covariance matrix of $\mathbf{e}_i$.

On the other hand, when the weights $\mathbf{W}_i$, $i = 1, \ldots, n$, are chosen as the inverse of $\mathbf{V}_i$, by Theorem 1, we have

$$\omega_n^{1/2} \begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \end{pmatrix} \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega}_*^+), \tag{B.14}$$

where $\boldsymbol{\Omega}_*$ is a positive semi-definite matrix such that

$$\frac{1}{\omega_n} \sum_{i=1}^n \mathrm{E}[\boldsymbol{\Lambda}_i^\top \mathbf{V}_i^{-1} \boldsymbol{\Lambda}_i] \to \boldsymbol{\Omega}_*.$$

In order to prove Corollary 1, by (B.13) and (B.14), we need only to show $\boldsymbol{\Omega}_{0*}^+ \boldsymbol{\Omega}_{1*} \boldsymbol{\Omega}_{0*}^+ - \boldsymbol{\Omega}_*^+$ is positive semi-definite. Letting $\boldsymbol{\Theta}_i = \boldsymbol{\Omega}_{0*}^+ \boldsymbol{\Lambda}_i \mathbf{V}_i^{1/2} - \boldsymbol{\Omega}_*^+ \boldsymbol{\Lambda}_i \mathbf{V}_i^{-1/2}$, we have

$$\begin{aligned} \boldsymbol{\Theta}_i \boldsymbol{\Theta}_i^\top &= (\boldsymbol{\Omega}_{0*}^+ \boldsymbol{\Lambda}_i \mathbf{V}_i^{1/2} - \boldsymbol{\Omega}_*^+ \boldsymbol{\Lambda}_i \mathbf{V}_i^{-1/2})(\boldsymbol{\Omega}_{0*}^+ \boldsymbol{\Lambda}_i \mathbf{V}_i^{1/2} - \boldsymbol{\Omega}_*^+ \boldsymbol{\Lambda}_i \mathbf{V}_i^{-1/2})^\top \\ &= \boldsymbol{\Omega}_{0*}^+ \boldsymbol{\Lambda}_i \mathbf{V}_i \boldsymbol{\Lambda}_i \boldsymbol{\Omega}_{0*}^+ - \boldsymbol{\Omega}_{0*}^+ \boldsymbol{\Lambda}_i \boldsymbol{\Lambda}_i \boldsymbol{\Omega}_*^+ - \boldsymbol{\Omega}_*^+ \boldsymbol{\Lambda}_i \boldsymbol{\Lambda}_i \boldsymbol{\Omega}_{0*}^+ + \boldsymbol{\Omega}_*^+ \boldsymbol{\Lambda}_i \mathbf{V}_i^{-1} \boldsymbol{\Lambda}_i \boldsymbol{\Omega}_*^+, \end{aligned}$$

which indicates that

$$\frac{1}{\omega_n} \sum_{i=1}^n \mathrm{E}[\boldsymbol{\Theta}_i \boldsymbol{\Theta}_i^\top] \to \boldsymbol{\Omega}_{0*}^+ \boldsymbol{\Omega}_{1*} \boldsymbol{\Omega}_{0*}^+ - \boldsymbol{\Omega}_*^+. \tag{B.15}$$

As $\mathrm{E}[\boldsymbol{\Theta}_i \boldsymbol{\Theta}_i^\top]$ is positive semi-definite, by (B.15) we know that $\boldsymbol{\Omega}_{0*}^+ \boldsymbol{\Omega}_{1*} \boldsymbol{\Omega}_{0*}^+ - \boldsymbol{\Omega}_*^+$ is also positive semi-definite. Hence the proof of Corollary 1 is complete.

B.3. Proof of Theorem 2. Note that

$$\begin{aligned} \widehat{\eta}(u) - \eta(u) &= \sum_{i=1}^n \mathbf{s}_i(u|\widehat{\boldsymbol{\theta}})(\mathbf{Y}_i - \mathbf{Z}_i^\top \widehat{\boldsymbol{\beta}}) - \eta(u) \\ &= \sum_{i=1}^n \mathbf{s}_i(u|\widehat{\boldsymbol{\theta}}) \mathbf{e}_i + \left[\sum_{i=1}^n \mathbf{s}_i(u|\widehat{\boldsymbol{\theta}}) \boldsymbol{\eta}(\mathbf{X}_i, \boldsymbol{\theta}_0) - \eta(u) \right] \\ &\quad + \sum_{i=1}^n \mathbf{s}_i(u|\widehat{\boldsymbol{\theta}}) \mathbf{Z}_i^\top (\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}) \\ &\equiv I_{n1,*} + I_{n2,*} + I_{n3,*}. \end{aligned} \tag{B.16}$$

By Assumption 1, we have

$$K\left(\frac{\mathbf{X}_{ij}^\top \widehat{\boldsymbol{\theta}} - u}{h}\right) = K\left(\frac{\mathbf{X}_{ij}^\top \boldsymbol{\theta}_0 - u}{h}\right) + \dot{K}\left(\frac{\mathbf{X}_{ij}^\top \boldsymbol{\theta}_\diamond - u}{h}\right) \frac{\mathbf{X}_{ij}^\top (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)}{h}, \tag{B.17}$$

where $\boldsymbol{\theta}_\diamond = \boldsymbol{\theta}_0 + \lambda_\diamond (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ for some $0 < \lambda_\diamond < 1$. By Theorem 1, we have

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| + \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| = O_P(\omega_n^{-1/2}). \tag{B.18}$$

It follows from (B.17), (B.18) and (3.5) that

$$
\begin{aligned}
I_{n3,*} &= \sum_{i=1}^{n} \mathbf{s}_i(u|\boldsymbol{\theta}_0)\mathbf{Z}_i^\top(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}) + \sum_{i=1}^{n}[\mathbf{s}_i(u|\widehat{\boldsymbol{\theta}}) - \mathbf{s}_i(u|\boldsymbol{\theta}_0)]\mathbf{Z}_i^\top(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}) \\
&= O_P(\omega_n^{-1/2}) + O_P(\omega_n^{-1}) \qquad \text{(B.19)} \\
&= o_P(\varphi_n^{-1/2}(h)).
\end{aligned}
$$

Similar to the proof of (B.5), we can show that

$$
I_{n2,*} = \tfrac{1}{2}\ddot{\eta}(u)\mu_2 h^2(1 + o_P(1)). \tag{B.20}
$$

For $I_{n1,*}$, note that by (B.17) and (B.18), we can show that $\sum_{i=1}^{n} \mathbf{s}_i(u|\boldsymbol{\theta}_0)\mathbf{e}_i$ is the leading term of $I_{n1,*}$. Letting $z_i(\boldsymbol{\theta}_0) = \mathbf{s}_i(u|\boldsymbol{\theta}_0)\mathbf{e}_i$ and by Assumption 2, it is easy to check that $\{z_i(\boldsymbol{\theta}_0) : i \geq 1\}$ is a sequence of independent random variables. By Assumption 2(iii), we have $\mathrm{E}[z_i(\boldsymbol{\theta}_0)] = 0$. By (3.5), (3.6) and the central limit theorem, it can be readily seen that

$$
\varphi_n^{1/2}(h) I_{n1,*} \xrightarrow{d} \mathrm{N}(0, \sigma_*^2). \tag{B.21}
$$

In view of (B.16), (B.19)–(B.21), the proof of Theorem 2 is complete.

Acknowledgements. The authors wish to thank the Co-editor, the Associate Editor and two referees for their valuable comments and suggestions, which substantially improved an earlier version of the paper.

SUPPLEMENTARY MATERIAL

Supplement to "Semiparametric GEE analysis in partially linear single-index models for longitudinal data" (DOI: 10.1214/15-AOS1320SUPP; .pdf). The supplement gives the proof of Theorem 3 and some technical lemmas that were used to prove the main results in Appendix B. It also includes some additional results of our simulation studies described in Section 5.

REFERENCES

BRAUN-FAHRLÄNDER, C., ACKERMANN-LIEBRICH, U., SCHWARTZ, J., GNEHM, H. P., RUTISHAUSER, M. and WANNER, H. U. (1992). Air pollution and respiratory symptoms in preschool children. *Am. Rev. Respir. Dis.* **145** 42–47.

CARROLL, R. J., FAN, J., GIJBELS, I. and WAND, M. P. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.* **92** 477–489. MR1467842

CHEN, L.-H., CHENG, M.-Y. and PENG, L. (2009). Conditional variance estimation in heteroscedastic regression models. *J. Statist. Plann. Inference* **139** 236–245. MR2474001

CHEN, J., GAO, J. and LI, D. (2013a). Estimation in single-index panel data models with heterogeneous link functions. *Econometric Rev.* **32** 928–955. MR3041108

CHEN, J., GAO, J. and LI, D. (2013b). Estimation in partially linear single-index panel data models with fixed effects. *J. Bus. Econom. Statist.* **31** 315–330. MR3173684

CHEN, J., LI, D., LIANG, H. and WANG, S. (2015). Supplement to "Semiparametric GEE analysis in partially linear single-index models for longitudinal data." DOI:10.1214/15-AOS1320SUPP.

DIGGLE, P. J., HEAGERTY, P. J., LIANG, K.-Y. and ZEGER, S. L. (2002). *Analysis of Longitudinal Data*, 2nd ed. Oxford Univ. Press, Oxford. MR2049007

DOCKERY, D. W., SPEIZER, F. E., STRAM, D. O., WARE, J. H., SPENGLER, J. D. and FERRIS, B. G. JR. (1989). Effects of inhalable particles on respiratory health of children. *Am. Rev. Respir. Dis.* **139** 587–594.

FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London. MR1383587

FAN, J. and HUANG, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli* **11** 1031–1057. MR2189080

FAN, J., HUANG, T. and LI, R. (2007). Analysis of longitudinal data with semiparametric estimation of convariance function. *J. Amer. Statist. Assoc.* **102** 632–641. MR2370857

FAN, J. and LI, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *J. Amer. Statist. Assoc.* **99** 710–723. MR2090905

FAN, J. and WU, Y. (2008). Semiparametric estimation of covariance matrixes for longitudinal data. *J. Amer. Statist. Assoc.* **103** 1520–1533. MR2504201

FAN, J. and YAO, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* **85** 645–660. MR1665822

GAO, J. (2007). *Nonlinear Time Series: Semiparametric and Nonparametric Methods*. Chapman & Hall/CRC, Boca Raton, FL. MR2297190

HALL, P., MÜLLER, H.-G. and WANG, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34** 1493–1517. MR2278365

HE, X., ZHU, Z.-Y. and FUNG, W.-K. (2002). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika* **89** 579–590. MR1929164

HUANG, J. Z., WU, C. O. and ZHOU, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* **89** 111–128. MR1888349

ICHIMURA, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J. Econometrics* **58** 71–120. MR1230981

JIANG, C.-R. and WANG, J.-L. (2011). Functional single index models for longitudinal data. *Ann. Statist.* **39** 362–388. MR2797850

KIM, S. and ZHAO, Z. (2013). Unified inference for sparse and dense longitudinal models. *Biometrika* **100** 203–212. MR3034333

KINNEY, P. L., WARE, J. H., SPENGLER, J. D., DOCKERY, D. W., SPEIZER, F. E. and FERRIS, B. G. JR. (1989). Short-term pulmonary function change in association with ozone levels. *Am. Rev. Respir. Dis.* **139** 56–61.

LENG, C. and TANG, C. Y. (2011). Improving variance function estimation in semiparametric longitudinal data analysis. *Canad. J. Statist.* **39** 656–670. MR2860832

LI, Y. and HSING, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Statist.* **38** 3321–3351. MR2766854

LIANG, K. Y. and ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73** 13–22. MR0836430

LIANG, H., LIU, X., LI, R. and TSAI, C.-L. (2010). Estimation and testing for partially linear single-index models. *Ann. Statist.* **38** 3811–3836. MR2766869

LIN, X. and CARROLL, R. J. (2006). Semiparametric estimation in general repeated measures problems. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 69–88. MR2212575

LIN, D. Y. and YING, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *J. Amer. Statist. Assoc.* **96** 103–126. MR1952726

LIPFERT, F. W. and HAMMERSTROM, T. (1992). Temporal patterns in air pollution and hospital admissions. *Environ. Res.* **59** 374–399.

LUO, W., LI, B. and YIN, X. (2014). On efficient dimension reduction with respect to a statistical functional of interest. *Ann. Statist.* **42** 382–412. MR3189490

MA, S., LIANG, H. and TSAI, C.-L. (2014). Partially linear single index models for repeated measurements. *J. Multivariate Anal.* **130** 354–375. MR3229543

MA, Y. and ZHU, L. (2013). Doubly robust and efficient estimators for heteroscedastic partially linear single-index models allowing high dimensional covariates. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 305–322. MR3021389

PENG, L. and YAO, Q. (2003). Least absolute deviations estimation for ARCH and GARCH models. *Biometrika* **90** 967–975. MR2024770

PETROV, V. V. (1995). *Limit Theorems of Probability Theory: Sequences of Independent Random Variables, Oxford Science Publications*. Oxford Univ. Press, New York. MR1353441

POPE, C. A. III (1991). Respiratory hospital admissions associated with $\mathrm{PM}_1 0$ pollution in utah, salt lake, and cache valleys. *Archives of Environmental Health: An International Journal* **46** 90–97.

POPE, C. A. III, BATES, D. V. and RAIZENNE, M. E. (1995). Health effects of particulate air pollution: Time for reassessment? *Environ. Health Perspect.* **103** 472–480.

WANG, L. (2011). GEE analysis of clustered binary data with diverging number of covariates. *Ann. Statist.* **39** 389–417. MR2797851

WANG, N., CARROLL, R. J. and LIN, X. (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data. *J. Amer. Statist. Assoc.* **100** 147–157. MR2156825

WANG, S., QIAN, L. and CARROLL, R. J. (2010). Generalized empirical likelihood methods for analyzing longitudinal data. *Biometrika* **97** 79–93. MR2594418

WANG, J.-L., XUE, L., ZHU, L. and CHONG, Y. S. (2010). Estimation for a partial-linear single-index model. *Ann. Statist.* **38** 246–274. MR2589322

WU, W. B. and POURAHMADI, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika* **90** 831–844. MR2024760

WU, H. and ZHANG, J.-T. (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis*. Wiley, Hoboken, NJ. MR2216899

XIA, Y. and HÄRDLE, W. (2006). Semi-parametric estimation of partially linear single-index models. *J. Multivariate Anal.* **97** 1162–1184. MR2276153

XIA, Y., TONG, H. and LI, W. K. (1999). On extended partially linear single-index models. *Biometrika* **86** 831–842. MR1741980

XIE, M. and YANG, Y. (2003). Asymptotics for generalized estimating equations with large cluster sizes. *Ann. Statist.* **31** 310–347. MR1962509

YAO, W. and LI, R. (2013). New local estimation procedure for a non-parametric regression function for longitudinal data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 123–138. MR3008274

YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. MR2160561

YU, K. and JONES, M. C. (2004). Likelihood-based local linear estimation of the conditional variance function. *J. Amer. Statist. Assoc.* **99** 139–144. MR2054293

YU, Y. and RUPPERT, D. (2002). Penalized spline estimation for partially linear single-index models. *J. Amer. Statist. Assoc.* **97** 1042–1054. MR1951258

ZHANG, J.-T. and CHEN, J. (2007). Statistical inferences for functional data. *Ann. Statist.* **35** 1052–1079. MR2341698

ZHANG, W., FAN, J. and SUN, Y. (2009). A semiparametric model for cluster data. *Ann. Statist.* **37** 2377–2408. MR2543696

ZHANG, W., LENG, C. and TANG, C. Y. (2015). A joint modelling approach for longitudinal studies. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 219–238. MR3299406

J. CHEN
DEPARTMENT OF ECONOMICS AND RELATED STUDIES
UNIVERSITY OF YORK
HESLINGTON WEST CAMPUS
YORK, YO10 5DD
UNITED KINGDOM
E-MAIL: jia.chen@york.ac.uk

D. LI
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF YORK
HESLINGTON WEST CAMPUS
YORK, YO10 5DD
UNITED KINGDOM
E-MAIL: degui.li@york.ac.uk

H. LIANG
DEPARTMENT OF STATISTICS
GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DISTRICT OF COLUMBIA 20052
USA
E-MAIL: hliang@gwu.edu

S. WANG
DEPARTMENT OF STATISTICS
TEXAS A&M UNIVERSITY
COLLEGE STATION, TEXAS 77843
USA
E-MAIL: sjwang@stat.tamu.edu