UNIVERSITY *of York*

Discussion Papers in Economics

No. 14/26

Semiparametric GEE Analysis in Partially Linear Single-Index Models for Longitudinal Data

Jia Chen, Degui Li, Hua Liang and Suojin Wang

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

Semiparametric GEE Analysis in Partially Linear Single-Index Models for Longitudinal Data

Jia Chen* Degui Li† Hua Liang‡ Suojin Wang§

December 1, 2014

Abstract

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 the 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 to illustrate our methodology and theory.

Keywords: GEE, local linear smoothing, longitudinal data, semiparametric estimation, single-index models.

JEL Classifications: C14, C13, C33

*Department of Economics and Related Studies, University of York, Heslington, York, YO10 5DD, UK. Email address: jia.chen@york.ac.uk

†Department of Mathematics, University of York, Heslington, York, YO10 5DD, UK. Email address: degui.li@york.ac.uk

‡Department of Statistics, George Washington University, Washington, D.C., 20052, US. Email address: hliang@gwu.edu

§Department of Statistics, Texas A&M University, College Station, TX, 77843, US. Email address: sjwang@stat.tamu.edu

1. Introduction

Consider a semiparametric partially linear single-index model defined by

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

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 *et al.* (1999), Yu and Ruppert (2002), Xia and Härdle (2006), Wang *et al.* (2010) and Ma and Zhu (2013). The recent paper by Liang *et al.* (2010) further 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 the above partially linear single-index model in the context of longitudinal data which arise frequently in many fields of research, such as biology, climatology, economics and epidemiology, and thus has 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, which may lead to inconsistent estimates and incorrect conclusions being drawn from the longitudinal data. Hence, to address this issue, in recent years, there has been a large literature on how to relax the parametric assumptions on the longitudinal data models and many nonparametric and semiparametric models have thus been investigated; see, for example, Lin and Ying (2001), Lin and Carroll (2001, 2006), He *et al.* (2002), Fan and Li (2004), Wang *et al.* (2005), Wu and Zhang (2006), Zhang *et al.* (2009), Li and Hsing (2010), and Jiang and Wang (2011).

Suppose that we have a random sample with n subjects from model (1.1). For the ith subject, $i = 1, \ldots, n$, the response variable $Y_i(t)$ and the covariates $\big\{\mathbf{Z}_i(t),\ \mathbf{X}_i(t)\big\}$ are collected at random time points t_{ij}, $j = 1, \ldots, m_i$, 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 $i = 1, \ldots, n$ and $j = 1, \ldots, m_i$. 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 *et al.*, 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 non-zero 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, for example, Hall *et al.*, 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 *et al.*, 2005; Wang *et al.*, 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 generally asymptotically more efficient than the profile unweighted least squares (PULS) estimators, when the weights in the SGEE method are chosen as the conditional covariance matrix of the errors given the covariates. 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 to estimate $\boldsymbol{\beta}_0$, $\boldsymbol{\theta}_0$ 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 as well as their proofs are provided in Appendix C.

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 "remove-one-component" technique using penalized spline and local linear smoothing, respectively; 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 *et al.* (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 *et al.* (2013) 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 the Introduction, 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 in 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 $\mathbf{Y}_i = \big(Y_i(t_{i1}), \ldots, Y_i(t_{im_i})\big)^\top$, $\mathbf{X}_i = \big(\mathbf{X}_i(t_{i1}), \ldots, \mathbf{X}_i(t_{im_i})\big)^\top$, $\mathbf{Z}_i = \big(\mathbf{Z}_i(t_{i1}), \ldots, \mathbf{Z}_i(t_{im_i})\big)^\top$, $\mathbf{e}_i = \big(e_i(t_{i1}), \ldots, e_i(t_{im_i})\big)^\top$, and $\boldsymbol{\eta}(\mathbf{X}_i, \boldsymbol{\theta}) = \big(\eta\big(\mathbf{X}_i^\top(t_{i1})\boldsymbol{\theta}\big), \ldots,$

$\eta\big(\mathbf{X}_i^\top(t_{im_i})\boldsymbol{\theta}\big)\big)^\top$. 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 $\mathbb{Y} = (\mathbf{Y}_1^\top, \ldots, \mathbf{Y}_n^\top)^\top$, $\mathbb{Z} = \big(\mathbf{Z}_1^\top, \ldots, \mathbf{Z}_n^\top\big)^\top$, $\mathbb{E} = (\mathbf{e}_1^\top, \ldots, \mathbf{e}_n^\top)^\top$, $\boldsymbol{\eta}(\mathbb{X}, \boldsymbol{\theta}) = \big(\boldsymbol{\eta}^\top(\mathbf{X}_1, \boldsymbol{\theta}), \ldots, \boldsymbol{\eta}^\top(\mathbf{X}_n, \boldsymbol{\theta})\big)^\top$. 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, Fan and Huang (2005) and Fan *et al.* (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 $\eta'(\cdot)$ at point u by minimizing the following loss function

$$L_n\big(a, b\big|\boldsymbol{\beta}, \boldsymbol{\theta}\big) = \sum_{i=1}^{n} \Big\{ w_i \sum_{j=1}^{m_i} \big[y_{ij} - \mathbf{Z}_{ij}^\top\boldsymbol{\beta} - a - b(\mathbf{X}_{ij}^\top\boldsymbol{\theta} - u)\big]^2 K\Big(\frac{\mathbf{X}_{ij}^\top\boldsymbol{\theta} - u}{h}\Big)\Big\}, \tag{2.3}$$

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). As in 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 *et al.* (2002) and Wu and Zhang (2006), the type 1 scheme might be a practical choice if the number of observations is relatively similar across the subjects, while the type 2 scheme may be appropriate otherwise. As the longitudinal data under investigation are allowed to be unbalanced, in this paper, we use $w_i = 1/(nm_i)$, which was also used by Li and Hsing (2010), and Kim and Zhao (2012). We denote

$$\big(\widehat{\eta}(u|\boldsymbol{\beta}, \boldsymbol{\theta}),\ \widehat{\eta}'(u|\boldsymbol{\beta}, \boldsymbol{\theta})\big)^\top = \arg\min_{a,b} L_n\big(a, b\big|\boldsymbol{\beta}, \boldsymbol{\theta}\big). \tag{2.4}$$

By some elementary calculations (see, for example, Fan and Gijbels, 1996), we have

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

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

$$
\begin{aligned}
\mathbf{s}_i(u|\boldsymbol{\theta}) &= (1,0)\Big[\sum_{i=1}^{n}\overline{\mathbf{X}}_i^{\top}(u|\boldsymbol{\theta})\mathbf{K}_i(u|\boldsymbol{\theta})\overline{\mathbf{X}}_i(u|\boldsymbol{\theta})\Big]^{-1}\overline{\mathbf{X}}_i^{\top}(u|\boldsymbol{\theta})\mathbf{K}_i(u|\boldsymbol{\theta}),\\
\overline{\mathbf{X}}_i(u|\boldsymbol{\theta}) &= \big(\overline{\mathbf{X}}_{i1}(u|\boldsymbol{\theta}),\ldots,\overline{\mathbf{X}}_{im_i}(u|\boldsymbol{\theta})\big)^{\top},\quad \overline{\mathbf{X}}_{ij}(u|\boldsymbol{\theta}) = \big(1,\mathbf{X}_{ij}^{\top}\boldsymbol{\theta}-u\big)^{\top},\\
\mathbf{K}_i(u|\boldsymbol{\theta}) &= \mathrm{diag}\Big(w_iK\Big(\frac{\mathbf{X}_{i1}^{\top}\boldsymbol{\theta}-u}{h}\Big),\ldots,w_iK\Big(\frac{\mathbf{X}_{im_i}^{\top}\boldsymbol{\theta}-u}{h}\Big)\Big).
\end{aligned}
$$

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 PLUS (profile unweighted least squares) method which ignores the possible within-subject correlation structure. Define the loss function by

$$
\begin{aligned}
Q_{n0}(\boldsymbol{\beta},\boldsymbol{\theta}) &= \sum_{i=1}^{n}\Big[\mathbf{Y}_i-\mathbf{Z}_i\boldsymbol{\beta}-\widehat{\boldsymbol{\eta}}(\mathbf{X}_i|\boldsymbol{\beta},\boldsymbol{\theta})\Big]^{\top}\Big[\mathbf{Y}_i-\mathbf{Z}_i\boldsymbol{\beta}-\widehat{\boldsymbol{\eta}}(\mathbf{X}_i|\boldsymbol{\beta},\boldsymbol{\theta})\Big]\\
&= \Big[\mathbb{Y}-\mathbb{Z}\boldsymbol{\beta}-\widehat{\boldsymbol{\eta}}(\mathbb{X}|\boldsymbol{\beta},\boldsymbol{\theta})\Big]^{\top}\Big[\mathbb{Y}-\mathbb{Z}\boldsymbol{\beta}-\widehat{\boldsymbol{\eta}}(\mathbb{X}|\boldsymbol{\beta},\boldsymbol{\theta})\Big], \qquad (2.6)
\end{aligned}
$$

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. The PULS estimators of $\boldsymbol{\beta}_0$ and $\boldsymbol{\theta}_0$ are obtained by minimizing $Q_{n0}(\boldsymbol{\beta},\boldsymbol{\theta})$, and we denote them 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 $\mathbb{W}=\mathrm{diag}\{\mathbf{W}_1,\ldots,\mathbf{W}_n\}$, 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

$$
\begin{aligned}
\boldsymbol{\rho}_{\mathbf{Z}}(\mathbf{X}_i,\boldsymbol{\theta}) &= \big(\rho_{\mathbf{Z}}(\mathbf{X}_{i1}^{\top}\boldsymbol{\theta}|\boldsymbol{\theta}),\ldots,\rho_{\mathbf{Z}}(\mathbf{X}_{im_i}^{\top}\boldsymbol{\theta}|\boldsymbol{\theta})\big)^{\top},\qquad \rho_{\mathbf{Z}}(u|\boldsymbol{\theta})=\mathrm{E}\big[\mathbf{Z}_{ij}|\mathbf{X}_{ij}^{\top}\boldsymbol{\theta}=u\big],\\
\boldsymbol{\rho}_{\mathbf{X}}(\mathbf{X}_i,\boldsymbol{\theta}) &= \big(\rho_{\mathbf{X}}(\mathbf{X}_{i1}^{\top}\boldsymbol{\theta}|\boldsymbol{\theta}),\ldots,\rho_{\mathbf{X}}(\mathbf{X}_{im_i}^{\top}\boldsymbol{\theta}|\boldsymbol{\theta})\big)^{\top},\qquad \rho_{\mathbf{X}}(u|\boldsymbol{\theta})=\mathrm{E}\big[\mathbf{X}_{ij}|\mathbf{X}_{ij}^{\top}\boldsymbol{\theta}=u\big],\\
\boldsymbol{\Lambda}_i(\boldsymbol{\theta}) &= \Big(\mathbf{Z}_i-\boldsymbol{\rho}_{\mathbf{Z}}(\mathbf{X}_i,\boldsymbol{\theta}),\ \big[\boldsymbol{\eta}'(\mathbf{X}_i,\boldsymbol{\theta})\otimes\mathbf{1}_p^{\top}\big]\odot\big[\mathbf{X}_i-\boldsymbol{\rho}_{\mathbf{X}}(\mathbf{X}_i,\boldsymbol{\theta})\big]\Big),
\end{aligned}
$$

where $\boldsymbol{\eta}'(\mathbf{X}_i,\boldsymbol{\theta})$ 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 the following equation:

$$
\sum_{i=1}^{n}\widehat{\boldsymbol{\Lambda}}_i^{\top}(\boldsymbol{\theta})\mathbf{W}_i\Big[\mathbf{Y}_i-\mathbf{Z}_i\boldsymbol{\beta}-\widehat{\boldsymbol{\eta}}(\mathbf{X}_i|\boldsymbol{\beta},\boldsymbol{\theta})\Big]=\mathbf{0}, \qquad (2.7)
$$

where $\widehat{\mathbf{\Lambda}}_i(\boldsymbol{\theta})$ is an estimator of $\mathbf{\Lambda}_i(\boldsymbol{\theta})$ with $\boldsymbol{\rho}_{\mathbf{Z}}(\mathbf{X}_i, \boldsymbol{\theta})$, $\boldsymbol{\rho}_{\mathbf{X}}(\mathbf{X}_i, \boldsymbol{\theta})$, and $\boldsymbol{\eta}'(\mathbf{X}_i, \boldsymbol{\theta})$ replaced by their corresponding local linear estimated values. Let $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$ be the solutions to the weighted estimation equations defined in (2.7). Corollary 3.1 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 by

$$\widehat{\eta}(u) = \widehat{\eta}(u|\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \mathbf{s}_i(u|\widehat{\boldsymbol{\theta}})\big(\mathbf{Y}_i - \mathbf{Z}_i\widehat{\boldsymbol{\beta}}\big). \tag{2.8}$$

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 notations. Let $\mathbf{\Lambda}_i = \mathbf{\Lambda}_i(\boldsymbol{\theta}_0)$, and assume that there exist two positive definite matrices $\mathbf{\Omega}_0$ and $\mathbf{\Omega}_1$ as well as a sequence ω_n such that $\omega_n \to \infty$,

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

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

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

as $n \to \infty$. The 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) would be applicable. Throughout the paper, we assume that the choice of $\mathbf{W}_i$ would not affect the form of ω_n. We first give the asymptotic distribution theory for the SGEE-based estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$.

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

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

as $n \to \infty$, where $\mathbf{A}^{+}$ is the Moore-Penrose inverse matrix of $\mathbf{A}$.

Remark 3.1. Theorem 3.1 establishes the asymptotically normal distribution theory for $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$ with convergence rate $\omega_n^{1/2}$, which is usually with the same asymptotic order as $T_n^{1/2}$ (T_n is the total number of observations). The specific forms of ω_n, $\mathbf{\Omega}_0$ and $\mathbf{\Omega}_1$ can be derived for some particular cases. For instance, when longitudinal data is balanced, i.e., $m_i \equiv m$, $\omega_n = nm$. Furthermore, if $\mathrm{E}[e_{ij}^2] \equiv \sigma_e^2$, and $\mathbf{W}_i$, $i = 1, \ldots, n$, are $m \times m$ identity matrices (i.e., e_{ij} are i.i.d.), where $e_{ij} = e_i(t_{ij})$ is independent of the covariates, then we can show that

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

where

$$\begin{aligned}
\Omega_0(1) &= \mathrm{E}\Big\{\big[\mathbf{Z}(t) - \boldsymbol{\rho}_{\mathbf{Z}}(\mathbf{X}(t)^\top\boldsymbol{\theta}_0|\boldsymbol{\theta}_0)\big]\big[\mathbf{Z}(t) - \boldsymbol{\rho}_{\mathbf{Z}}(\mathbf{X}(t)^\top\boldsymbol{\theta}_0|\boldsymbol{\theta}_0)\big]^\top\Big\}, \\
\Omega_0(2) &= \mathrm{E}\Big\{\eta'(\mathbf{X}(t)^\top\boldsymbol{\theta}_0)\big[\mathbf{Z}(t) - \rho_{\mathbf{Z}}(\mathbf{X}(t)^\top\boldsymbol{\theta}_0|\boldsymbol{\theta}_0)\big]\big[\mathbf{X}(t) - \rho_{\mathbf{X}}(\mathbf{X}(t)^\top\boldsymbol{\theta}_0|\boldsymbol{\theta}_0)\big]^\top\Big\}, \\
\Omega_0(3) &= \mathrm{E}\Big\{\big[\eta'(\mathbf{X}(t)^\top\boldsymbol{\theta}_0)\big]^2\big[\mathbf{X}(t) - \rho_{\mathbf{X}}(\mathbf{X}(t)^\top\boldsymbol{\theta}_0|\boldsymbol{\theta}_0)\big]\big[\mathbf{X}(t) - \rho_{\mathbf{X}}(\mathbf{X}(t)^\top\boldsymbol{\theta}_0|\boldsymbol{\theta}_0)\big]^\top\Big\}.
\end{aligned}$$

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

In Theorem 3.1 above, we only require $n \to \infty$. Thus, 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 certain positive constant, we can take $\omega_n = n$ and prove that (3.4) still holds. For the dense longitudinal data case where $\min_i m_i \geq M_n$ with $M_n \to \infty$, we assume that there exists $v(\cdot)$ such that

$$v(m_i) \to \infty, \quad \frac{\mathbf{\Lambda}_i^\top\mathbf{W}_i\mathbf{\Lambda}_i}{v(m_i)} \xrightarrow{P} \mathbf{\Omega}_0 \ \text{ and } \ \frac{\mathrm{E}\Big[\mathbf{\Lambda}_i^\top\mathbf{W}_i\mathbf{e}_i\mathbf{e}_i^\top\mathbf{W}_i\mathbf{\Lambda}_i\Big]}{v(m_i)} \to \mathbf{\Omega}_1, \quad \text{as} \ \ m_i \to \infty.$$

Letting $\omega_n = \sum_{i=1}^n v(m_i)$, we can prove (3.4). As more observations are available in the dense longitudinal data case, the convergence rate for the parametric estimators is faster than $O_P(\sqrt{n})$ in the sparse longitudinal data case.

The following corollary 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 $\mathbf{W}_i$ in (2.7) are chosen as the inverse of the conditional covariance matrix of $\mathbf{e}_i$ given $\mathbf{X}_i$ and $\mathbf{Z}_i$.

Corollary 3.1. *Suppose that the weights $\mathbf{W}_i$ in (2.7) are chosen as the inverse of the conditional covariance matrix of $\mathbf{e}_i$ given $\mathbf{X}_i$ and $\mathbf{Z}_i$, and the conditions of Theorem 3.1 are satisfied. Then, the SGEE-based estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$ are generally asymptotically more* *efficient than the PULS estimators $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\theta}}$ which minimize $Q_{n0}(\boldsymbol{\beta},\boldsymbol{\theta})$ in (2.6) with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$.*

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

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

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 3.2 below. Let $\mu_j=\int v^jK(v)dv$ for $j=0,1,2,\cdots$, and $\eta_0''(\cdot)$ be the second-order derivative of $\eta_0(\cdot)$.

Theorem 3.2. *Suppose that the conditions of Theorem 3.1, (3.5) and (3.6) are satisfied. Then, we have*

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

where $b_\eta(u)=\eta_0''(u)\mu_2/2$.

Remark 3.2. Theorem 3.2 provides the asymptotically normal distribution theory for the nonparametric estimator $\widehat{\eta}(u)$ with convergence rate $\varphi_n^{1/2}(h)$. The forms of $\varphi_n(h)$ and σ_*^2 can be specified for some particular cases. As an example, consider the case where $e_{ij}=v_i+\varepsilon_{ij}$, 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

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

where $\nu_j = \int v^j K^2(v)dv$, $j = 0, 1, 2$, 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 C.1 in Appendix C and some elementary calculations, we can prove that

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

Hence, in this case, we can take $\varphi_n(h) = (n^2h)\Big(\sum_{i=1}^n \frac{1}{m_i}\Big)^{-1}$ which has the same order as nh, and $\sigma_*^2 = \nu_0 f_{\boldsymbol{\theta}_0}(u)(\sigma_v^2+\sigma_\varepsilon^2)$. Such result is similar to Theorem 1 (i) in Kim and Zhao (2012).

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, by Lemma C.1 again, we can prove that

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

Hence, in this case, we can take $\varphi_n(h) = n$ and $\sigma_*^2 = \mu_0^2 f_{\boldsymbol{\theta}_0}^2(u)\sigma_v^2$, which are analogous to Theorem 1 (ii) in Kim and Zhao (2012) 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 (2.7) 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 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 each $1 \le i \le n$, let $\mathbf{R}_i$ be the covariance matrix of $\mathbf{e}_i$, $\boldsymbol{\Sigma}_i = \mathrm{diag}\big\{\sigma^2(t_{i1}), \ldots, \sigma^2(t_{im_i})\big\}$ with $\sigma^2(t_{ij}) = \mathrm{E}\big[e_i^2(t_{ij})\big|t_{ij}\big] = \mathrm{E}\big[e_i^2(t_{ij})\big|t_{ij}, \mathbf{X}_i(t_{ij}), \mathbf{Z}_i(t_{ij})\big]$ for $j = 1, \ldots, m_i$, 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

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

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, Ruppert *et al.* (1997), Fan and Yao (1998), Yu and Jones (2004) and Fan *et al.* (2007). However, when the errors are heavy-tailed, which is not uncommon is 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 $r(t_{ij}) = \big[Y_{ij} - \mathbf{Z}_{ij}^\top\boldsymbol{\beta}_0 - \eta(\mathbf{X}_{ij}^\top\boldsymbol{\theta}_0)\big]^2$, where $Y_{ij} = Y_i(t_{ij})$, $\mathbf{Z}_{ij} = \mathbf{Z}_i(t_{ij})$, and $\mathbf{X}_{ij} = \mathbf{X}_i(t_{ij})$. We can then find random variable $\xi(t_{ij})$ so that $r(t_{ij}) = \sigma^2(t_{ij})\xi^2(t_{ij})$ and $\mathrm{E}\big[\xi^2(t_{ij})|t_{ij}\big] = 1$ with probability 1. By applying the log-transformation (see Peng and Yao, 2003; Gao, 2007; and Chen *et al.*, 2009 for the application of this transformation in time series analysis) to $r(t_{ij})$, we have

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

where τ is a positive constant such that $\mathrm{E}[\xi_\diamond(t_{ij})] = \mathrm{E}\big\{\log\big[\tau^{-1}\xi^2(t_{ij})\big]\big\} = 0$. Here, $\xi_\diamond(t_{ij})$ could be viewed as an error term in the model (4.2). As $r_{ij} = r(t_{ij})$ are unobservable, we replace them with $\widehat{r}_{ij} = \big[Y_{ij} - \mathbf{Z}_{ij}^\top\widehat{\boldsymbol{\beta}} - \widehat{\eta}(\mathbf{X}_{ij}^\top\widehat{\boldsymbol{\theta}})\big]^2$. To estimate $\sigma_\diamond^2(t)$, we define

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

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 by

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

On the other hand, noting that

$$\frac{\exp\big\{\sigma_\diamond^2(t_{ij})\big\}}{\tau}\xi^2(t_{ij}) = r_{ij} \quad \text{and} \quad \mathrm{E}[\xi^2(t_{ij})] = 1,$$

the constant τ can be estimated by

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

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

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

and

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

which are similar to those in (3.5) and (3.6). Define

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

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 Appendix C.

Theorem 4.1. *Suppose the conditions in Theorems 3.1 and 3.2, Assumptions 6–9 in Appendix A, (4.7) and (4.8) are satisfied. Then, we have*

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

where $f_T(\cdot)$ is the density function of the observation times t_{ij}.

Remark 4.1. Theorem 4.1 can be seen as an extension of Theorem 1 in Chen *et al.* (2009) from the time series case to the longitudinal data case. The longitudinal data framework in this paper is quite flexible and includes both sparse and dense data types. Following the discussion in Remark 3.2, we can also show that under some mild conditions, the nonparametric conditional variance estimation would have different convergence rates for the two data types.

We next discuss how to obtain the optimal value of the parameter vector $\boldsymbol{\phi}$. Let $\widehat{\boldsymbol{\Sigma}}_i$ be the estimator of $\boldsymbol{\Sigma}_i$ with $\sigma^2(t_{ij})$ being replaced by $\widehat{\sigma}^2(t_{ij})$ which was defined in (4.6) and $\mathbf{R}_i^*(\boldsymbol{\phi}) = \widehat{\boldsymbol{\Sigma}}_i^{1/2}\mathbf{C}_i(\boldsymbol{\phi})\widehat{\boldsymbol{\Sigma}}_i^{1/2}$. Recall that $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\theta}}$ are the estimated values of the parameters $\boldsymbol{\beta}_0$ and $\boldsymbol{\theta}_0$, respectively, by taking $\mathbf{W}_i$ as the identity matrix in the estimation equations, and that they are consistent. With $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\theta}}$, we then construct the local linear estimator of the link function $\widetilde{\eta}(u) = \widehat{\eta}(u|\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\theta}})$, the residuals $\widetilde{\mathbf{e}}_i \equiv \mathbf{Y}_i - \mathbf{Z}_i\widetilde{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\eta}}(\mathbf{X}_i, \widetilde{\boldsymbol{\theta}})$, and $\widetilde{\boldsymbol{\Lambda}}_i \equiv \widehat{\boldsymbol{\Lambda}}_i(\widetilde{\boldsymbol{\theta}})$, 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 $\widetilde{\eta}(\cdot)$ and $\widetilde{\boldsymbol{\theta}}$, respectively. Motivated by equations (3.1) and (3.2), we construct

$$\boldsymbol{\Omega}_0^*(\boldsymbol{\phi}) = \sum_{i=1}^n \widetilde{\boldsymbol{\Lambda}}_i^\top \left[\mathbf{R}_i^*(\boldsymbol{\phi})\right]^{-1}\widetilde{\boldsymbol{\Lambda}}_i \quad \text{and} \quad \boldsymbol{\Omega}_1^*(\boldsymbol{\phi}) = \sum_{i=1}^n \widetilde{\boldsymbol{\Lambda}}_i^\top \left[\mathbf{R}_i^*(\boldsymbol{\phi})\right]^{-1}\widetilde{\mathbf{e}}_i\widetilde{\mathbf{e}}_i^\top \left[\mathbf{R}_i^*(\boldsymbol{\phi})\right]^{-1}\widetilde{\boldsymbol{\Lambda}}_i. \quad (4.10)$$

By Theorem 3.1, the sandwich formula estimate $\left[\boldsymbol{\Omega}_0^*(\boldsymbol{\phi})\right]^+\boldsymbol{\Omega}_1^*(\boldsymbol{\phi})\left[\boldsymbol{\Omega}_0^*(\boldsymbol{\phi})\right]^+$ is asymptotically proportional to the asymptotic covariance of $\left(\widehat{\boldsymbol{\beta}}^\top, \widehat{\boldsymbol{\theta}}^\top\right)^\top$. The optimal value of $\boldsymbol{\phi}$, denoted by $\widehat{\boldsymbol{\phi}}$, can be chosen to minimize the determinant $\left|\left[\boldsymbol{\Omega}_0^*(\boldsymbol{\phi})\right]^+\boldsymbol{\Omega}_1^*(\boldsymbol{\phi})\left[\boldsymbol{\Omega}_0^*(\boldsymbol{\phi})\right]^+\right|$. Such a method is called the minimum generalized variance method (Fan *et al.*, 2007). Then, we can choose the covariance matrices as $\mathbf{R}_i(\widehat{\boldsymbol{\phi}}) = \widehat{\boldsymbol{\Sigma}}_i^{1/2}\mathbf{C}_i(\widehat{\boldsymbol{\phi}})\widehat{\boldsymbol{\Sigma}}_i^{1/2}$.

5. Numerical studies

In this section, we first study the finite sample performance of the proposed SGEE estimator through Monte Carlo simulation, and then give an empirical application of the proposed model and methodology. In the simulation study, for comparison, we also report the performance of the PULS estimators which minimizes the loss function defined in (2.6).

5.1. Simulation study

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. The data are generated with one of the two types of within-subject correlation structure: AR(1) and ARMA(1,1), and with each type we investigate the robustness of the proposed estimator to misspecification of the correlation structure.

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

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

The covariates $(\mathbf{Z}_i^\top(t_{ij}), \mathbf{X}_i^\top(t_{ij}))^\top$ are generated independently from a five-dimensional normal distribution with mean $\mathbf{0}$, variance 1 and correlation 0.1. The observation times t_{ij} are generated in the same way as in Fan *et al.* (2007). For each subject, $\{0, 1, 2, \ldots, T\}$ 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 non-skipped scheduled time, i.e., a uniform [0, 1] random number is added to the non-skipped 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

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

and an ARMA(1,1) correlation structure

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

or an AR(1) correlation structure with $\gamma = 1$ in (5.2). 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 on one subject at a time) cross-validation (CV) to choose the optimal bandwidth in 20 replications. We then use the average of the optimal bandwidths from these 20 replications as the bandwidth for the following 200 replications. 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 5.1 and 5.2. Table 5.1 gives the results obtained under the correct specification of an underlying AR(1) correlation structure, and Table 5.2 gives those obtained under correct specification of an underlying ARMA(1,1) structure. The results show that the SGEE estimates are comparable with the corresponding PULS estimates in terms of bias and are more efficient than the PULS estimates, which supports the asymptotic theory developed in Section 3. The performance of both estimators improves as either time dimension or the number of the subjects increases.

Insert Table 5.1 here

Insert Table 5.2 here

In Figures 5.1 and 5.2, we also plot the local linear estimated link function from a typical realization together with the real curve for each combination of n and $\overline{m}$.

Insert Figure 5.1 here

Insert Figure 5.2 here

To study the robustness of the SGEE and PULS estimators to correlation structure misspecification, we fit an AR(1) working correlation structure in (2.7) when the true correlation structure is ARMA(1,1). Table 5.3 reports the results under this misspecification. The table shows that in the presence of correlation structure misspecification, the SGEE still produces more efficient estimates of the parameters than the PULS method.

Insert Table 5.3 here

5.2. Real data analysis

We next illustrate the partially linear single-index model and the proposed SGEE estimation method through an empirical example for exploring 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 Arden Pope III *et al.* (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-Fahrlander *et al.*, 1992, Lipfert and Hammerstrom, 1992). While earlier research often used time series or cross-sectional data to evaluate the health 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, by what extent. The use of the partially linear single-index model and the SGEE method would provide greater modelling 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 years and 14 years old (at their first visit to the hospital/clinic) were followed over 10 years. During 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 non-asthma patient), gender (G, 1 for males and 0 for females), race (R, 1 for non-whites 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 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 or 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. The 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 5.3, and the box plots of the response against the binary regressors are given in Figure 5.4. The estimated model is as follows

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

where the numbers in the parentheses under the estimated coefficients are their respective standard errors. The estimated link function and its 95% confidence band are plotted in Figure 5.5.

From Figure 5.5, it can be seen that the estimated link function is overall increasing. The 95% confidence bands show that the linear approximation for the unspecified link function would be rejected, and thus the partially liner single-index model might be more appropriate than the traditional linear regression 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 ($R = 0$ for whites and 1 for non-white) have, on average, higher lung function than female or non-white 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. And 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.

Insert Figure 5.3 here

Insert Figure 5.4 here

Insert Figure 5.5 here

6. Conclusions

In this paper, we study a partially linear single-index modelling structure for possible unbalanced longitudinal data under a general framework which includes both the sparse and dense longitudinal data 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 3.1 and 3.2, 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 also propose a semiparametric method to estimate the 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 3.1, 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 SGEE estimators have equal or smaller asymptotic variances.

Both the simulation study and empirical data analysis in Section 5 show that the proposed approaches work well in the finite sample case.

7. Acknowledgements

Liang's research was partially supported by NSF grants DMS-1007167 and DMS-1207444 and by Award Number 11228103, made by National Natural Science Foundation of China. Wang's research was partially supported by Award Number KUS-CI-016-04 made by King Abdullagh University of Science and Technology (KAUST).

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 the continuous first-order derivative function denoted by $K'(\cdot)$.

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 (i.e., $\mathbf{e}_i$, $1 \le i \le n$, are mutually independent, where $\mathbf{e}_i$ were defined in Section 2).

(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 uncorrelated with 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 $\delta > 0$.

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 the 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 $\mathrm{E}\big[\|\mathbf{Z}_{ij}\|^{2+\delta}\big] < \infty$, where δ was defined in *Assumption 2 (iii)* and $\|\cdot\|$ is the Euclidean norm.

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

Assumption 5. Let the bandwidth h satisfy

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

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).

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 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. Let the bandwidth h_1 satisfy

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

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

Remark A.1. *Assumptions 1 and 6* impose some mild restrictions on the kernel functions. These conditions have been used by existing literature in i.i.d. and weakly dependent time series cases (see, for example, Fan and Gijbels, 1996; Gao, 2007). The compact support restriction on the kernel functions can be removed if we impose certain restriction 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; Zhang *et al.*, 2009). *Assumption 2(ii)* is imposed to simplify the presentation of the asymptotic results, and it can be relaxed at the cost of more complicated forms for asymptotic variances of the proposed estimators.

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 *et al.* (2013b) for example. *Assumption 4* is a mild smoothness condition on the link function imposed for the application of the local linear fitting. *Assumptions 5 and 9* give a set of restrictions on the two bandwidths h and h_1, which are involved in the estimation of the link function and the conditional variance function, respectively. *Assumption 7* imposes a mild condition on the observation times (see, for example, 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 *et al.*, 2009). Since the local linear smoothing technique is applied, certain smoothness condition has to be assumed on $\sigma^2(\cdot)$, as is done in *Assumption 8*.

Appendix B: Proofs of the main results

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

Proof of Theorem 3.1. By the definition of the weighted local linear estimators in (2.4) and (2.5), we have

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

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

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

where $K'(\cdot)$ 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

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

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

By Lemma C.2 in Appendix C, we can prove that

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

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

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

which, together with Lemma C.3 in Appendix C, leads to

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

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

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

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

By (B.1)–(B.5), we can prove that, uniformly for $i = 1,\dots,n$ and $j = 1,\dots,m_i$,

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

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

By the definitions of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$ (see (2.7) in Section 2), we have

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

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

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

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

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

which together with (B.6), (B.8) and the bandwidth condition $\omega_n h^6 = o(1)$, we can show that

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

where $\mathbf{s}_k(\mathbf{X}_i,\boldsymbol{\theta}) = \big(\mathbf{s}_k^\top(\mathbf{X}_{i1}^\top\boldsymbol{\theta}|\boldsymbol{\theta}),\ldots,\mathbf{s}_k^\top(\mathbf{X}_{im_i}^\top\boldsymbol{\theta}|\boldsymbol{\theta})\big)^\top$, $\boldsymbol{\rho}_{\mathbf{Z}}(\mathbf{X}_i,\boldsymbol{\theta}_0)$ 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, for example, Ichimura, 1993; Chen *et al.*, 2013b), we can show the weak consistency of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$. Also note that

$$\begin{aligned}
\sum_{i=1}^n \boldsymbol{\Lambda}_i^\top\mathbf{W}_i\boldsymbol{\Lambda}_i\begin{pmatrix}\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\end{pmatrix} &= \sum_{i=1}^n \boldsymbol{\Lambda}_i^\top\mathbf{W}_i\big\{\big[\boldsymbol{\eta}'(\mathbf{X}_i,\boldsymbol{\theta}_0)\otimes\mathbf{1}_p^\top\big]\odot\big[\mathbf{X}_i - \boldsymbol{\rho}_{\mathbf{X}}(\mathbf{X}_i,\boldsymbol{\theta}_0)\big]\big\}\big(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big), \\
&\quad + \sum_{i=1}^n \boldsymbol{\Lambda}_i^\top\mathbf{W}_i\big[\mathbf{Z}_i - \boldsymbol{\rho}_{\mathbf{Z}}(\mathbf{X}_i,\boldsymbol{\theta}_0)\big]\big(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\big)
\end{aligned} \tag{B.10}$$

and

$$\sum_{i=1}^{n}\mathbf{\Lambda}_i^\top\mathbf{W}_i\Big[\sum_{k=1}^{n}\mathbf{s}_k(\mathbf{X}_i,\boldsymbol{\theta}_0)\mathbf{e}_k\Big]=o_P(\omega_n^{1/2}). \tag{B.11}$$

By (B.8)–(B.11) and the fact that $\sum_{i=1}^n\mathbf{\Lambda}_i^\top\mathbf{W}_i\mathbf{\Lambda}_i=O_P(\omega_n)$, we have

$$\begin{pmatrix}\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\\ \widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\end{pmatrix}\overset{P}{\sim}\Big[\sum_{i=1}^{n}\mathbf{\Lambda}_i^\top\mathbf{W}_i\mathbf{\Lambda}_i\Big]^+\Big[\sum_{i=1}^{n}\mathbf{\Lambda}_i^\top\mathbf{W}_i\mathbf{e}_i\Big]. \tag{B.12}$$

By (3.1)–(3.3), (B.12) and the classical central limit theorem for independent sequence, we can show that (3.4) in Theorem 3.1 holds. ■

Proof of Corollary 3.1. By Theorem 3.1, the PULS estimators $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\theta}}$ have the following asymptotic normal distribution:

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

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

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

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

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 3.1, we have

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

where $\mathbf{\Omega}_*$ is a positive definite matrix such that

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

By (B.13) and (B.14), to prove Corollary 3.1, we only need to show $\mathbf{\Omega}_{0*}^+\mathbf{\Omega}_{1*}\mathbf{\Omega}_{0*}^+-\mathbf{\Omega}_*^+$ is nonnegative definite. Letting $\mathbf{\Theta}_i=\mathbf{\Omega}_{0*}^+\mathbf{\Lambda}_i\mathbf{V}_i^{1/2}-\mathbf{\Omega}_*^+\mathbf{\Lambda}_i\mathbf{V}_i^{-1/2}$, we have

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

which indicates that

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

As $\mathrm{E}\big[\mathbf{\Theta}_i\mathbf{\Theta}_i^\top\big]$ is nonnegative definite, by (B.15), we can prove that $\mathbf{\Omega}_{0*}^{+}\mathbf{\Omega}_{1*}\mathbf{\Omega}_{0*}^{+} - \mathbf{\Omega}_{*}^{+}$ is also nonnegative definite. Hence, the proof of Corollary 3.1 is completed. ■

Proof of Theorem 3.2. Note that

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

By Assumption 1, we have

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

where $\boldsymbol{\theta}_\diamond=\boldsymbol{\theta}_0+\lambda_\diamond(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$ for some $0<\lambda_\diamond<1$. By Theorem 3.1, we have

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

From (B.17), (B.18) and (3.6), it follows that

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

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

$$I_{n2,*}=\frac{1}{2}\eta''(u)\mu_2h^2(1+o_P(1)). \tag{B.20}$$

We finally consider $I_{n1,*}$. By (B.17) and (B.18), we can show that $\sum_{i=1}^{n}\mathbf{s}_i(u\big|\boldsymbol{\theta}_0)\mathbf{e}_i$ is the leading term of $I_{n1,*}$. Letting $z_i(\boldsymbol{\theta}_0)=\mathbf{s}_i(u\big|\boldsymbol{\theta}_0)\mathbf{e}_i$ 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 3.2 is completed. ■

Appendix C: Some auxiliary lemmas and proof of Theorem 4.1

In this appendix, we give some technical lemmas which have been used to prove the main results in Appendix B, and the proof of Theorem 4.1 in Section 4. As in Appendix B, let C denote a generic positive constant whose value may change from line to line. Define

$$V_{ij}(u,\boldsymbol{\theta},\kappa)=\frac{1}{h}\Big(\frac{\mathbf{X}_{ij}^\top\boldsymbol{\theta}-u}{h}\Big)^\kappa K\Big(\frac{\mathbf{X}_{ij}^\top\boldsymbol{\theta}-u}{h}\Big),\ \kappa=0,1,2,\ldots,$$

for $i=1,\ldots,n$ and $j=1,\ldots,m_i$. We next give the uniform consistency results of the weighted nonparametric kernel-based estimators for the longitudinal data, which are of independent interest.

Lemma C.1. *Suppose that Assumptions 1, 2(ii) and 3(i) in Appendix A are satisfied and*

$$h\to 0,\quad \frac{n^2}{N_n(h)\log n}\to\infty,\quad \frac{\log n}{h^2N_n(h)}=O(1), \tag{C.1}$$

where $N_n(h)=\sum_{i=1}^n 1/m_ih$. *Then, we have, for any integer* $\kappa\ge 0$ *and as* $n\to\infty$,

$$\sup_{(u,\boldsymbol{\theta}^\top)^\top\in\mathcal{U}(\Theta)}\Big|\frac{1}{n}\sum_{i=1}^n\frac{1}{m_i}\sum_{j=1}^{m_i}V_{ij}(u,\boldsymbol{\theta},\kappa)-f_{\boldsymbol{\theta}}(u)\mu_\kappa\Big|=O_P\Big(h^{\tau_\kappa}+\frac{\sqrt{N_n(h)\log n}}{n}\Big), \tag{C.2}$$

where $\mathcal{U}(\Theta)=\big\{(u,\boldsymbol{\theta}^\top)^\top: u\in\mathcal{U},\ \boldsymbol{\theta}\in\Theta\big\}$, $\mathcal{U}$ *is defined in Assumption 3(i),* Θ *is a parameter space,* $\mu_\kappa=\int v^\kappa K(v)dv$, $\tau_\kappa=1$ *if* κ *is odd, and* $\tau_\kappa=2$ *if* κ *is even.*

Proof. For simplicity, let $\epsilon_n=\frac{\sqrt{N_n(h)\log n}}{n}$. To prove (C.2), it suffices to show that

$$\sup_{(u,\boldsymbol{\theta}^\top)^\top\in\mathcal{U}(\Theta)}\Big|\frac{1}{n}\sum_{i=1}^n\frac{1}{m_i}\sum_{j=1}^{m_i}\Big\{V_{ij}(u,\boldsymbol{\theta},\kappa)-\mathrm{E}[V_{ij}(u,\boldsymbol{\theta},\kappa)]\Big\}\Big|=O_P(\epsilon_n), \tag{C.3}$$

and

$$\sup_{(u,\boldsymbol{\theta}^\top)^\top\in\mathcal{U}(\Theta)}\Big|\mathrm{E}[V_{ij}(u,\boldsymbol{\theta},\kappa)]-f_{\boldsymbol{\theta}}(u)\mu_\kappa\Big|=O(h^{\tau_\kappa}). \tag{C.4}$$

By Assumptions 1, 2(ii) and 3(i) in Appendix A, we have

$$\begin{aligned}\mathrm{E}[V_{ij}(u,\boldsymbol{\theta},\kappa)] &= \frac{1}{h}\int\Big(\frac{u_1-u}{h}\Big)^\kappa K\Big(\frac{u_1-u}{h}\Big)f_{\boldsymbol{\theta}}(u_1)du_1\\ &= \int v^\kappa K(v)f_{\boldsymbol{\theta}}(u+hv)dv\\ &= f_{\boldsymbol{\theta}}(u)\mu_\kappa+f_{\boldsymbol{\theta}}'(u)\mu_{\kappa+1}h+O(h^2)\end{aligned}$$

uniformly for $(u,\boldsymbol{\theta}^\top)^\top\in\mathcal{U}(\Theta)$, which implies that (C.4) holds.

Let us now turn to the proof of (C.3). The main idea is to consider covering of the set $\mathcal{U}(\Theta)$ by a finite number of subsets $S(k)$, which are centered at $s_k^\top\equiv\big(u_k,\boldsymbol{\theta}_k^\top\big)$ with radius

$r = o(h^2)$. Letting $\mathcal{N}_n$ be the total number of such subsets, $S(k)$, $k = 1, 2, \ldots, \mathcal{N}_n$, then $\mathcal{N}_n = O(r^{-(p+1)})$. It is easy to show that

$$\begin{aligned}
&\sup_{(u,\boldsymbol{\theta}^\top)^\top \in \mathcal{U}(\Theta)} \Big|\frac{1}{n}\sum_{i=1}^n \frac{1}{m_i}\sum_{j=1}^{m_i} \Big\{V_{ij}(u,\boldsymbol{\theta},\kappa) - \mathrm{E}[V_{ij}(u,\boldsymbol{\theta},\kappa)]\Big\}\Big| \\
&= \max_{1\le k\le \mathcal{N}_n} \Big|\frac{1}{n}\sum_{i=1}^n \frac{1}{m_i}\sum_{j=1}^{m_i} \Big\{V_{ij}(s_k,\kappa) - \mathrm{E}[V_{ij}(s_k,\kappa)]\Big\}\Big| \\
&\quad + \max_{1\le k\le \mathcal{N}_n} \sup_{(u,\boldsymbol{\theta}^\top)^\top \in S(k)} \Big|\frac{1}{n}\sum_{i=1}^n \frac{1}{m_i}\sum_{j=1}^{m_i} \Big[V_{ij}(u,\boldsymbol{\theta},\kappa) - V_{ij}(s_k,\kappa)\Big]\Big| \\
&\quad + \max_{1\le k\le \mathcal{N}_n} \sup_{(u,\boldsymbol{\theta}^\top)^\top \in S(k)} \Big|\frac{1}{n}\sum_{i=1}^n \frac{1}{m_i}\sum_{j=1}^{m_i} \Big\{\mathrm{E}[V_{ij}(u,\boldsymbol{\theta},\kappa)] - \mathrm{E}[V_{ij}(s_k,\kappa)]\Big\}\Big| \\
&\equiv \Pi_{n1} + \Pi_{n2} + \Pi_{n3}, \qquad (C.5)
\end{aligned}$$

where $V_{ij}(s_k,\kappa) = V_{ij}(u_k,\boldsymbol{\theta}_k,\kappa)$.

Noting that $K(\cdot)$ is Lipschitz continuous by Assumption 1 and taking $r = C\epsilon_n h^2$ for some positive constant C, we have

$$\Pi_{n2} = O_P\Big(\frac{r}{h^2}\Big) = O_P(\epsilon_n), \quad \Pi_{n3} = O(\epsilon_n). \qquad (C.6)$$

For Π_{n1}, we apply the Bernstein inequality for i.i.d. random variables (see, for example, van der Vaart and Wellner, 1996) to obtain the convergence rate. Note that by Assumptions 1, 2(ii) and 3(i),

$$\frac{1}{m_i}\sum_{j=1}^{m_i} \Big|V_{ij}(s_k,\kappa) - \mathrm{E}[V_{ij}(s_k,\kappa)]\Big| \le \frac{C}{h} \quad \text{for some } C > 0, \qquad (C.7)$$

and

$$\mathrm{Var}\Big[\frac{1}{m_i}\sum_{i=1}^{m_i} V_{ij}(s_k,\kappa)\Big] = \frac{1}{m_i^2}\cdot \mathrm{Var}\Big[\sum_{i=1}^{m_i} V_{ij}(s_k,\kappa)\Big] \le \frac{C}{m_i h}, \qquad (C.8)$$

By (C.7), (C.8), Assumption 2(ii) and the Bernstein inequality, we have, for some sufficiently large positive constant C_ϵ,

$$\begin{aligned}
\mathrm{P}(\Pi_{n1} > C_\epsilon \epsilon_n) &\le \mathcal{N}_n \exp\Big\{\frac{-n^2 C_\epsilon^2 \epsilon_n^2}{\big(2CN_n(h) + \frac{2nC\epsilon_n}{3h}\big)}\Big\} \\
&\le \mathcal{N}_n \exp\Big\{\frac{-n^2 C_\epsilon^2 \epsilon_n^2}{C_\epsilon N_n(h)}\Big\} \\
&\le \mathcal{N}_n \exp\{-C_\epsilon \log n\} = o(1), \qquad (C.9)
\end{aligned}$$

which implies that

$$\Pi_{n1} = O_P(\epsilon_n). \qquad (C.10)$$

In view of (C.5), (C.6) and (C.10), we have shown (C.3), completing the proof of Lemma C.1. ■

Lemma C.2. *Suppose that Assumptions 3(ii) and 5, and the conditions in Lemma C.1 are satisfied. Then, we have*

$$\sup_{(u,\boldsymbol{\theta}^\top)^\top\in\mathcal{U}(\Theta)}\Big|\sum_{i=1}^n \mathbf{s}_i(u|\boldsymbol{\theta})\mathbf{Z}_i-\rho_{\mathbf{Z}}(u|\boldsymbol{\theta})\Big| = O_P\left(h^2+\epsilon_n\right), \tag{C.11}$$

where $\mathbf{s}_i(u|\boldsymbol{\theta})$ *was defined in Section 2,* $\rho_{\mathbf{Z}}(u|\boldsymbol{\theta}) = \mathrm{E}\big[\mathbf{Z}_{ij}|\mathbf{X}_{ij}^\top\boldsymbol{\theta}=u\big]$*,* $\epsilon_n = \frac{\sqrt{N_n(h)\log n}}{n}$ *and* $N_n(h)$ *was defined in Lemma C.1.*

Proof. Letting $H=\mathrm{diag}(1,h)$, then by Lemma C.1 we have

$$H^{-1}\Big[\frac{1}{n}\sum_{i=1}^n \overline{\mathbf{X}}_i^\top(u|\boldsymbol{\theta})K_i(u|\boldsymbol{\theta})\overline{\mathbf{X}}_i(u|\boldsymbol{\theta})\Big]H^{-1} = f_{\boldsymbol{\theta}}(u)\ \mathrm{diag}(1,\mu_2)+o_P(1). \tag{C.12}$$

uniformly for $(u,\boldsymbol{\theta}^\top)^\top\in\mathcal{U}(\Theta)$, where $\overline{\mathbf{X}}_i(u|\boldsymbol{\theta})$ and $K_i(u|\boldsymbol{\theta})$ were defined in Section 2.

We then use arguments similar to those in the proof of Lemma C.1 to show that

$$\sup_{(u,\boldsymbol{\theta}^\top)^\top\in\mathcal{U}(\Theta)}\Big|\frac{1}{n}\sum_{i=1}^n\frac{1}{m_i}\sum_{j=1}^{m_i}V_{ij,\mathbf{Z}}(u,\boldsymbol{\theta},\kappa)-f_{\boldsymbol{\theta}}(u)\mu_\kappa\rho_{\mathbf{Z}}(u|\boldsymbol{\theta})\Big| = O_P(h^{\tau_\kappa}+\epsilon_n), \tag{C.13}$$

where

$$V_{ij,\mathbf{Z}}(u,\boldsymbol{\theta},\kappa)=\frac{\mathbf{Z}_{ij}}{h}\Big(\frac{\mathbf{X}_{ij}^\top\boldsymbol{\theta}-u}{h}\Big)^\kappa K\Big(\frac{\mathbf{X}_{ij}^\top\boldsymbol{\theta}-u}{h}\Big),\quad \kappa=0,1,\ldots.$$

By the bandwidth condition in Assumption 5, we can choose a positive and slowly varying function $L(\cdot)$ such that as $n\to\infty$

$$L(n)\to\infty \quad\text{and}\quad \frac{L(n)T_n^{\frac{2}{2+\delta}}\log n}{h^2N_n(h)}=o(1).$$

Furthermore, let $l(\cdot)$ be any positive function such that

$$l(n)\to\infty,\ \ l(n)\ll L(n),\ \ \text{as}\ \ n\to\infty,$$

where $a_n\ll b_n$ means $a_n=o(b_n)$. To prove (C.13), we need only to show that

$$\sup_{(u,\boldsymbol{\theta}^\top)^\top\in\mathcal{U}(\Theta)}\frac{1}{n}\sum_{i=1}^n\frac{1}{m_i}\sum_{j=1}^{m_i}\Big\{V_{ij,\mathbf{Z}}(u,\boldsymbol{\theta},\kappa)-\mathrm{E}[V_{ij,\mathbf{Z}}(u,\boldsymbol{\theta},\kappa)]\Big\} = o_P\left(l(n)\epsilon_n\right) \tag{C.14}$$

and

$$\sup_{(u,\boldsymbol{\theta}^\top)^\top\in\mathcal{U}(\Theta)}\Big|\mathrm{E}[V_{ij,\mathbf{Z}}(u,\boldsymbol{\theta},\kappa)]-f_{\boldsymbol{\theta}}(u)\mu_\kappa\rho_{\mathbf{Z}}(u|\boldsymbol{\theta})\Big| = O_P\big(h^{\tau_\kappa}\big). \tag{C.15}$$

By Assumptions 1, 2(ii) and 3(ii), we have

$$
\begin{aligned}
\mathrm{E}[V_{ij,\mathbf{Z}}(u,\boldsymbol{\theta},\kappa)] &= \frac{1}{h}\int\Big(\frac{u_1-u}{h}\Big)^{\kappa}K\Big(\frac{u_1-u}{h}\Big)f_{\boldsymbol{\theta}}(u_1)\rho_{\mathbf{Z}}(u_1|\boldsymbol{\theta})du_1 \\
&= \int v^{\kappa}K(v)f_{\boldsymbol{\theta}}(u+hv)\rho_{\mathbf{Z}}(u+hv|\boldsymbol{\theta})dv \\
&= f_{\boldsymbol{\theta}}(u)\rho_{\mathbf{Z}}(u|\boldsymbol{\theta})\mu_{\kappa}+f'_{\boldsymbol{\theta}}(u|\boldsymbol{\theta})\rho_{\mathbf{Z}}(u|\boldsymbol{\theta})\mu_{\kappa+1}h \\
&\quad +f_{\boldsymbol{\theta}}(u)\rho'_{\mathbf{Z}}(u|\boldsymbol{\theta})\mu_{\kappa+1}h+O(h^2)
\end{aligned}
$$

uniformly in $(u,\boldsymbol{\theta}^\top)^\top\in\mathcal{U}(\Theta)$, which implies (C.15).

As in the proof of Lemma C.1, the main idea in proving (C.14) is to consider covering of the set $\mathcal{U}(\Theta)$ by a finite number of subsets $S(k)$ centered at s_k with radius $r=o(h^2)$. Letting s_k and $\mathcal{N}_n$ be defined as in the proof of Lemma C.1, it is easy to show that

$$
\begin{aligned}
&\sup_{(u,\boldsymbol{\theta}^\top)^\top\in\mathcal{U}(\Theta)}\Big|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j=1}^{m_i}\Big\{V_{ij,\mathbf{Z}}(u,\boldsymbol{\theta},\kappa)-\mathrm{E}[V_{ij,\mathbf{Z}}(u,\boldsymbol{\theta},\kappa)]\Big\}\Big| \\
&= \max_{1\le k\le\mathcal{N}_n}\Big|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j=1}^{m_i}\Big\{V_{ij,\mathbf{Z}}(s_k,\kappa)-\mathrm{E}[V_{ij,\mathbf{Z}}(s_k,\kappa)]\Big\}\Big| \\
&\quad+\max_{1\le k\le\mathcal{N}_n}\sup_{(u,\boldsymbol{\theta}^\top)^\top\in S(k)}\Big|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j=1}^{m_i}\Big[V_{ij,\mathbf{Z}}(u,\boldsymbol{\theta},\kappa)-V_{ij,\mathbf{Z}}(s_k,\kappa)\Big]\Big| \\
&\quad+\max_{1\le k\le\mathcal{N}_n}\sup_{(u,\boldsymbol{\theta}^\top)^\top\in S(k)}\Big|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j=1}^{m_i}\Big\{\mathrm{E}[V_{ij,\mathbf{Z}}(u,\boldsymbol{\theta},\kappa)]-\mathrm{E}[V_{ij,\mathbf{Z}}(s_k,\kappa)]\Big\}\Big| \\
&\equiv \Pi_{n4}+\Pi_{n5}+\Pi_{n6}, \qquad\qquad (C.16)
\end{aligned}
$$

where $V_{ij,\mathbf{Z}}(s_k,\kappa)=V_{ij,\mathbf{Z}}(u_k,\boldsymbol{\theta}_k,\kappa)$.

Similar to the proof of (C.6) as above, taking $r=O\left(\epsilon_n h^2\right)$, we have

$$
\Pi_{n5}+\Pi_{n6}=O_P\Big(\frac{r}{h^2}\Big)=O_P(\epsilon_n)=o_P(\epsilon_n l(n)). \qquad (C.17)
$$

We next obtain the convergence rate for Π_{n4}, which is slightly more complicated than its counterpart in the proof of Lemma C.1. As $\mathbf{Z}_{ij}$ may be unbounded, we apply a truncation method. For this purpose, we define

$$
\overline{V}_{ij,\mathbf{Z}}(k)=V_{ij,\mathbf{Z}}(s_k,\kappa)I\big\{\|\mathbf{Z}_{ij}\|\le T_n^{\frac{1}{2+\delta}}l(n)\big\}
$$

and

$$
\widetilde{V}_{ij,\mathbf{Z}}(k)=V_{ij,\mathbf{Z}}(s_k,\kappa)-\overline{V}_{ij,\mathbf{Z}}(k),
$$

where $I\{\cdot\}$ is an indicator function. It is easy to show that

$$\begin{aligned}\Pi_{n4} &\leq \max_{1\leq k\leq \mathcal{N}_n}\Big|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j=1}^{m_i}(\overline{V}_{ij,\mathbf{Z}}(k)-\mathrm{E}[\overline{V}_{ij,\mathbf{Z}}(k)])\Big| \\ &\quad+\max_{1\leq k\leq \mathcal{N}_n}\Big|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j=1}^{m_i}(\widetilde{V}_{ij,\mathbf{Z}}(k)-\mathrm{E}[\widetilde{V}_{ij,\mathbf{Z}}(k)])\Big| \\ &\equiv \Pi_{n4,1}+\Pi_{n4,2}. \end{aligned}\tag{C.18}$$

Note that for any $\eta>0$,

$$\begin{aligned}\mathrm{P}\Big(\Pi_{n4,2}>\eta\epsilon_n l(n)\Big) &\leq \mathrm{P}\Big(\max_{1\leq k\leq \mathcal{N}_n}\max_{1\leq i\leq n,1\leq j\leq m_i}|\widetilde{V}_{ij,\mathbf{Z}}(k)|>0\Big) \\ &\leq \sum_{i=1}^{n}\sum_{j=1}^{m_i}\mathrm{P}\Big(\|\mathbf{Z}_{ij}\|>T_n^{\frac{1}{2+\delta}}l(n)\Big) \\ &= O\big(l^{-(2+\delta)}(n)\big)=o(1),\end{aligned}$$

which leads to

$$\Pi_{n4,2}=o_P(\epsilon_n l(n)).\tag{C.19}$$

We then use the Bernstein inequality to deal with the convergence of $\Pi_{n4,1}$. Note that for any k, we have

$$\frac{1}{m_i}\sum_{j=1}^{m_i}\Big|\overline{V}_{ij,\mathbf{Z}}(k)-\mathrm{E}[\overline{V}_{ij,\mathbf{Z}}(k)]\Big|\leq\frac{CT_n^{\frac{1}{2+\delta}}l(n)}{h}\tag{C.20}$$

and

$$\mathrm{Var}\Big[\frac{1}{m_i}\sum_{j=1}^{m_i}\overline{V}_{ij,\mathbf{Z}}(k)\Big]\leq\frac{C}{m_i h},\tag{C.21}$$

where C is a positive constant which is independent of k.

By (C.20), (C.21), Assumptions 2(ii), 5 and the Bernstein inequality for i.i.d. random variables, we have, for any $\eta>0$,

$$\begin{aligned}\mathrm{P}\Big(\Pi_{n4,1}>\eta\epsilon_n l(n)\Big) &\leq \mathcal{N}_n\exp\Big\{\frac{-\eta^2n^2\epsilon_n^2l^2(n)}{2C\big[N_n(h)+nT_n^{\frac{1}{2+\delta}}l(n)h^{-1}\eta\epsilon_n\big]}\Big\} \\ &\leq \mathcal{N}_n\exp\{-Cl(n)\log(n)\} \\ &= O(\mathcal{N}_n n^{-Cl(n)})=o(1).\end{aligned}\tag{C.22}$$

Hence, we have

$$\Pi_{n4,1}=o_P(\epsilon_n l(n)).\tag{C.23}$$

By (C.16)–(C.19) and (C.23), we know that (C.14) holds, which, together with (C.15), implies that (C.13) holds. In view of (C.12) and (C.13) as well as the definition of $\mathbf{s}_i(u|\boldsymbol{\theta})$, (C.11) is readily seen. ■

Lemma C.3. *Let*

$$\widetilde{\boldsymbol{\eta}}(\mathbf{X}_i,\boldsymbol{\theta}) = \Big(\eta'(\mathbf{X}_{i1}^\top\boldsymbol{\theta})\mathbf{X}_{i1},\ldots,\eta'(\mathbf{X}_{im_i}^\top\boldsymbol{\theta})\mathbf{X}_{im_i}\Big)^\top,$$

and suppose that the conditions in Lemma C.2 are satisfied. Then we have

$$\sup_{(u,\boldsymbol{\theta}^\top)^\top\in\mathcal{U}(\Theta)}\Big|\sum_{i=1}^n \mathbf{s}_i(u|\boldsymbol{\theta})\widetilde{\boldsymbol{\eta}}(\mathbf{X}_i,\boldsymbol{\theta}) - \eta'(u)\rho_{\mathbf{X}}(u|\boldsymbol{\theta})\Big| = O_P\left(h^2+\epsilon_n\right), \tag{C.24}$$

where $\rho_{\mathbf{X}}(u|\boldsymbol{\theta}) = \mathrm{E}\big[\mathbf{X}_{ij}|\mathbf{X}_{ij}^\top\boldsymbol{\theta}=u\big]$.

Proof. The proof is similar to the proofs of Lemmas C.1 and C.2 given above. We thus omit the details. ■

We next give the proof of Theorem 4.1, whose main idea is analogous to the proof of Theorem 1 in Chen *et al.* (2009) in the time series context.

Proof of Theorem 4.1. Note that

$$\begin{aligned}
\widehat{\sigma}^2(t)-\sigma^2(t) &= \frac{\exp\{\widehat{\sigma}_\diamond^2(t)\}}{\widehat{\tau}} - \frac{\exp\{\sigma_\diamond^2(t)\}}{\tau} \\
&= \Big[\frac{\exp\{\widehat{\sigma}_\diamond^2(t)\}}{\widehat{\tau}} - \frac{\exp\{\widehat{\sigma}_\diamond^2(t)\}}{\tau}\Big] + \Big[\frac{\exp\{\widehat{\sigma}_\diamond^2(t)\}}{\tau} - \frac{\exp\{\sigma_\diamond^2(t)\}}{\tau}\Big] \\
&\equiv \Xi_{n1}+\Xi_{n2}.
\end{aligned} \tag{C.25}$$

We first consider Ξ_{n2}. By a first-order Taylor expansion and some standard techniques in local linear estimation, we can show that

$$\begin{aligned}
\Xi_{n2} &\overset{P}{\sim} \frac{\exp\{\sigma_\diamond^2(t)\}}{\tau}\big[\widehat{\sigma}_\diamond^2(t)-\sigma_\diamond^2(t)\big] \\
&\overset{P}{\sim} \frac{\exp\{\sigma_\diamond^2(t)\}}{\tau f_T(t)h_1}\Big\{\sum_{i=1}^n w_i\sum_{j=1}^{m_i}\big[\log(\widehat{r}_{ij}+\zeta_n)-\sigma_\diamond^2(t)-\dot{\sigma}_\diamond^2(t)(t_{ij}-t)\big]K_1\big(\frac{t_{ij}-t}{h_1}\big)\Big\} \\
&= \frac{\exp\{\sigma_\diamond^2(t)\}}{\tau f_T(t)h_1}\Big\{\sum_{i=1}^n w_i\sum_{j=1}^{m_i}\big[\sigma_\diamond^2(t_{ij})-\sigma_\diamond^2(t)-\dot{\sigma}_\diamond^2(t)(t_{ij}-t)\big]K_1\big(\frac{t_{ij}-t}{h_1}\big)\Big\} \\
&\quad+\frac{\exp\{\sigma_\diamond^2(t)\}}{\tau f_T(t)h_1}\Big\{\sum_{i=1}^n w_i\sum_{j=1}^{m_i}\big[\log(\widehat{r}_{ij}+\zeta_n)-\log(r_{ij})\big]K_1\big(\frac{t_{ij}-t}{h_1}\big)\Big\} \\
&\quad+\frac{\exp\{\sigma_\diamond^2(t)\}}{\tau f_T(t)h_1}\Big\{\sum_{i=1}^n w_i\sum_{j=1}^{m_i}\xi_\diamond(t_{ij})K_1\big(\frac{t_{ij}-t}{h_1}\big)\Big\} \\
&\equiv \Xi_{n2,1}+\Xi_{n2,2}+\Xi_{n2,3},
\end{aligned} \tag{C.26}$$

where $a_n \overset{P}{\sim} b_n$ denotes $a_n = b_n(1+o_P(1))$.

Noting that $\mathrm{E}[\xi_\diamond(t_{ij})] = 0$, by (4.7) and the central limit theorem, it is readily proven that

$$\varphi_{n\diamond}^{1/2}(h_1) \cdot \Xi_{n2,3} \xrightarrow{d} \mathrm{N}\big(0, \frac{\exp\{2\sigma_\diamond^2(t)\}}{\tau^2 f_T(t)} \sigma_\diamond^2\big), \tag{C.27}$$

where $\frac{\exp\{2\sigma_\diamond^2(t)\}}{\tau^2} = \sigma^4(t)$ by relevant definition in (4.2).

By Assumption 8 and a second-order Taylor expansion of $\sigma_\diamond^2(\cdot)$, we can show that

$$\Xi_{n2,1} = \frac{\exp\{\sigma_\diamond^2(t)\}}{2\tau} \ddot{\sigma}_\diamond^2(t) h_1^2 \int v^2 K_1(v) dv + o_P(h_1^2) = h_1^2 b_{\sigma 1}(t) + o_P(h_1^2). \tag{C.28}$$

We next prove that $\Xi_{n2,2}$ is asymptotically negligible (compared with $\Xi_{n2,3}$). Let $\chi_n = \log^{-2}(T_n)\varphi_{n\diamond}^{-1/2}(h_1)$ and $\xi(t_{ij})$ be defined as in Section 4 such that $r(t_{ij}) = \sigma^2(t_{ij})\xi^2(t_{ij})$ and $\mathrm{E}[\xi^2(t_{ij})|t_{ij}] = 1$ with probability 1. Note that

$$\begin{aligned}
\Xi_{n2,2} &= \frac{\exp\{\sigma_\diamond^2(t)\}}{\tau f_T(t) h_1} \Big\{ \sum_{i=1}^n w_i \sum_{j=1}^{m_i} \big[\log(\widehat{r}_{ij} + \zeta_n) - \log(r_{ij})\big] K_1\big(\frac{t_{ij} - t}{h_1}\big) I\{\xi^2(t_{ij}) \le \chi_n\} \Big\} \\
&\quad + \frac{\exp\{\sigma_\diamond^2(t)\}}{\tau f_T(t) h_1} \Big\{ \sum_{i=1}^n w_i \sum_{j=1}^{m_i} \big[\log(\widehat{r}_{ij} + \zeta_n) - \log(r_{ij})\big] K_1\big(\frac{t_{ij} - t}{h_1}\big) I\{\xi^2(t_{ij}) > \chi_n\} \Big\} \\
&\equiv \Xi_{n2,21} + \Xi_{n2,22}.
\end{aligned} \tag{C.29}$$

Recalling that $\zeta_n = 1/T_n$, by Assumption 7 and the definitions of $\widehat{r}_{ij}$ and r_{ij}, we have

$$\begin{aligned}
|\Xi_{n2,21}| &\le \frac{\exp\{\sigma_\diamond^2(t)\}}{\tau f_T(t) h_1} \Big\{ \log(T_n) \sum_{i=1}^n w_i \sum_{j=1}^{m_i} K_1\big(\frac{t_{ij} - t}{h_1}\big) I(\xi^2(t_{ij}) \le \chi_n) \Big\} \\
&\quad + \frac{\exp\{\sigma_\diamond^2(t)\}}{\tau f_T(t) h_1} \Big\{ \sum_{i=1}^n w_i \sum_{j=1}^{m_i} \big|\log(\sigma^2(t_{ij})\xi^2(t_{ij}))\big| K_1\big(\frac{t_{ij} - t}{h_1}\big) I(\xi^2(t_{ij}) \le \chi_n) \Big\} \\
&\le O_P(\chi_n \log T_n) + \frac{\exp\{\sigma_\diamond^2(t)\}}{\tau f_T(t) h_1} \Big\{ |\log \chi_n| \sum_{i=1}^n w_i \sum_{j=1}^{m_i} K_1\big(\frac{t_{ij} - t}{h_1}\big) I(\xi^2(t_{ij}) \le \chi_n) \Big\} \\
&= O_P(\chi_n \log T_n + \chi_n |\log \chi_n|) \\
&= o_P(\varphi_n^{-1/2}(h_1)).
\end{aligned} \tag{C.30}$$

In a way similar to Fan and Yao (1998) and Chen *et al.* (2009), we can show that $\Xi_{n2,22} = o_P(\varphi_n^{-1/2}(h_1))$, which, in combination with (C.30), implies

$$\Xi_{n2,2} = o_P(\varphi_n^{-1/2}(h_1)). \tag{C.31}$$

We next consider Ξ_{n1}. Following the proofs of Lemmas C.2 and C.3 above, we can similarly prove the uniform convergence rates for the local linear estimator of the link function.

Then, by Theorem 3.1 and (4.5), we can show that

$$
\begin{aligned}
\frac{1}{\widehat{\tau}} - \frac{1}{\tau} &\overset{P}{\sim} \Big[\frac{1}{T_n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} r_{ij}\big(\exp\{-\widehat{\sigma}_{\diamond}^2(t_{ij})\} - \exp\{-\sigma_{\diamond}^2(t_{ij})\}\big)\Big] \\
&\quad + \Big[\frac{1}{T_n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}(\widehat{r}_{ij} - r_{ij})\exp\{-\widehat{\sigma}_{\diamond}^2(t_{ij})\}\Big] \\
&= \frac{1}{T_n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} r_{ij}\exp\{-\sigma_{\diamond}^2(t_{ij})\}\big(\exp\{-\widehat{\sigma}_{\diamond}^2(t_{ij}) + \sigma_{\diamond}^2(t_{ij})\} - 1\big) + o_P(\varphi^{-1/2}(h_1) + h_1^2) \\
&= -\frac{1}{T_n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\xi^2(t_{ij})}{2\tau}\ddot{\sigma}_{\diamond}^2(t_{ij})h_1^2\int v^2 K_1(v)dv + o_P(\varphi^{-1/2}(h_1) + h_1^2) \\
&= -\frac{h_1^2}{2\tau}\mathrm{E}\big[\ddot{\sigma}_{\diamond}^2(t_{ij})\big]\mu_2 + o_P(\varphi^{-1/2}(h_1) + h_1^2),
\end{aligned}
$$

which implies that

$$
\Xi_{n1} = -\frac{\exp\{\sigma_{\diamond}^2(t)\}h_1^2}{2\tau}\mathrm{E}\big[\ddot{\sigma}_{\diamond}^2(t_{ij})\big]\mu_2 + o_P(\varphi^{-1/2}(h_1) + h_1^2) = -h_1^2 b_{\sigma 2}(t) + o_P(\varphi^{-1/2}(h_1) + h_1^2). \tag{C.32}
$$

In view of (C.25)–(C.32), we have completed the proof of Theorem 4.1. ■

References

[1] Braun-Fahrlander, C., Ackermann-Liebrich, U., Schwartz, J., Gnehm, H. P., Rutishauser, M. and Wanner H. U. (1992). Air pollution and respiratory symptoms in preschool children. *American Review of Respiratory Disease* **145**, 42–47.

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

[3] Chen, J., Gao, J. and Li, D. (2013a). Estimation in single-index panel data models with heterogeneous link function. *Econometric Reviews* **32**, 928–955.

[4] Chen, J., Gao, J. and Li, D. (2013b). Estimation in partially linear single-index panel data models with fixed effects. *Journal of Business and Economic Statistics* **31**, 315–330.

[5] Chen, L. H., Cheng, M. Y. and Peng, L. (2009). Conditional variance estimation in heteroscedastic regression models. *Journal of Statistical Planning and Inference* **139**, 236–245.

[6] Diggle, P. J., Heagerty, P., Liang, K. Y. and Zeger, S. L. (2002). *Analysis of Longitudinal Data*. Oxford University Press: London.

[7] 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. *American Review of Respiratory Disease* **139**, 587–594.

[8] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications*. Chapman and Hall: London.

[9] Fan, J. and Huang, T. (2005). Profile likelihood inferences on semiparametric varying coefficient partially linear models. *Bernoulli* **11**, 1031–1059.

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

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

[12] Fan, J. and Yao, Q. (1998). Efficient estimation of conditional variance function in stochastic regression. *Biometrika* **85**, 645–660.

[13] Gao, J. (2007). *Nonlinear Time Series: Semiparametric and Nonparametric Methods*. Chapman and Hall/CRC: London.

[14] Hall, P., Müller, H. G. and Wang, J. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Annals of Statistics* **34**, 1493–1517.

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

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

[17] Jiang, C. and Wang, J. (2011). Functional single-index models for longitudinal data. *Annals of Statistics* **39**, 362–388.

[18] Kim, S. and Zhao, Z. (2013). Unified inference for sparse and dense longitudinal models. *Biometrika* **100**, 203–212.

[19] 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. *American Review of Respiratory Disease* **139**, 56–61.

[20] Li, Y. and Hsing, T. (2010). Uniform convergence rate for nonparametric regression and principle component analysis in functional/longitudinal data. *Annals of Statistics* **38**, 3321–3351.

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

[22] Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalised linear models. *Biometrika* **73**, 12–22.

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

[24] Lipfert, F. W. and Hammerstrom, T. (1992). Temporal patterns in air pollution and hospital admissions. *Environmental Research* **59**, 374–399.

[25] Lin, X. and Carroll, R. J. (2001). Semiparametric regression for clustered data using generalised estimation equations. *Journal of the American Statistical Association* **96**, 1045–1056.

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

[27] Ma, S., Liang, H. and Tsai, C. L. (2013). Partially linear single-index models for repeated measurements. Working paper.

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

[29] Peng, L. and Yao, Q. (2003). Least absolute deviations estimation for ARCH and GARCH models. *Biometrika* **90**, 967–975.

[30] Petrov, V. V. (1995). *Limit Theorems of Probability Theory-Sequence of Independent Random Variables.* Oxford Science Publications: Oxford.

[31] Pope, C. A. III. (1991). Respiratory hospital admissions associated with PM_10 pollution in Utah, Salt Lake, and Cache Valleys. *Archives of Environmental Health: An International Journal* **46**, 90–97.

[32] Pope, C. A. III., Bates, D. V. and Raizenne, M. E. (1995). Health effects of particulate air pollution: time for reassessment? *Environmental Health Perspectives* **103**, 472–480.

[33] Ruppert, D., Wand, M. P., Holst, U. and Hössjer, O. (1997). Local polynomial variance function estimation. *Technometrics* **39**, 262–273.

[34] van der Vaart, A. W. and Wellner, J. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics.* Springer.

[35] Wang, L. (2011). GEE analysis of clustered binary data with diverging number of covariates. *Annals of Statistics* **39**, 389–417.

[36] Wang, J., Xue, L., Zhu, L. and Chong, Y. (2010). Estimation for a partially-linear single-index model. *Annals of Statistics* **38**, 246–274.

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

[38] Wang, S., Qian, L. and Carroll, R. J. (2010). Generalized empirical likelihood methods for analyzing longitudinal data. *Biometrika* **97**, 79–93.

[39] Wu, H. and Zhang, J. (2006) *Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modeling Approaches.* Wiley Series in Probability and Statistics.

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

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

[42] Xie, M. and Yang, Y. (2003). Asymptotics for generalized estimating equations with large cluster sizes. *Annals of Statistics* **31**, 310–347.

[43] Yao, F., Müller, H. G. and Wang, J. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**, 577–590.

[44] Yu, K. and Jones, M. C. (2004). Likelihood-based local linear estimation of the conditional variance function. *Journal of the American Statistical Association* **99**, 139–144.

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

[46] Zhang, J. and Chen, J. (2007). Statistical inferences for functional data. *Annals of Statistics* **35**, 1052–1079.

[47] Zhang, W., Fan, J. and Sun, Y. (2009). A semiparametric model for cluster data. *Annals of Statistics* **37**, 2377–2408.

Table 5.1. Performance of SGEE and PULS estimation methods under correct specification of an underlying AR(1) correlation structure

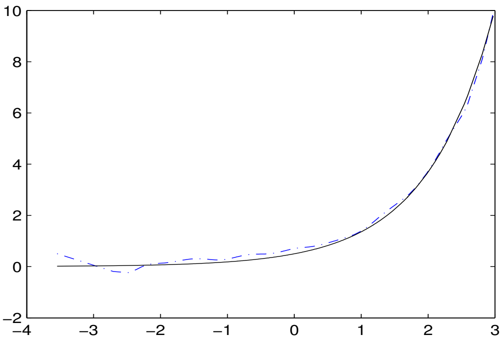
	n		30			50		
$\overline{m}$	Parameters	Methods	Bias	SD	MAD	Bias	SD	MAD
10	β_1	PULS	0.0048	0.0402	0.0288	-0.0030	0.0308	0.0195
		SGEE	-0.0026	0.0508	0.0081	-0.0016	0.0259	0.0074
	β_2	PULS	-0.0024	0.0409	0.0243	0.0049	0.0267	0.0180
		SGEE	-0.0018	0.0298	0.0110	0.0033	0.0310	0.0077
	θ_1	PULS	-0.0049	0.0299	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

Table 5.2. Performance of SGEE and PULS estimation methods under correct specification of an underlying ARMA(1,1) correlation structure

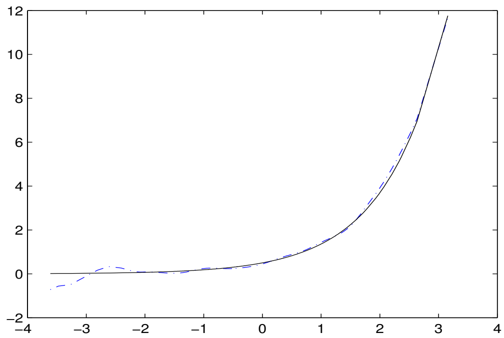
	n		30			50		
$\overline{m}$	Parameters	Methods	Bias	SD	MAD	Bias	SD	MAD
10	β_1	PULS	-0.0029	0.0400	0.0280	0.0006	0.0322	0.0221
		SGEE	-0.0025	0.0244	0.0155	0.0418×10^{-3}	0.0193	0.0124
	β_2	PULS	0.0032	0.0386	0.0282	-0.0045	0.0299	0.0205
		SGEE	0.0009	0.0249	0.0171	0.1378×10^{-3}	0.0212	0.0126
	θ_1	PULS	-0.0004	0.0267	0.0181	-0.2799×10^{-3}	0.0188	0.0126
		SGEE	-0.0002	0.0161	0.0104	0.5767×10^{-3}	0.0146	0.0073
	θ_2	PULS	-0.0047	0.0343	0.0209	0.5278×10^{-3}	0.0223	0.0156
		SGEE	-0.0031	0.0192	0.0113	-0.2302×10^{-3}	0.0145	0.0087
	θ_3	PULS	0.0008	0.0253	0.0158	-0.9189×10^{-3}	0.0201	0.0121
		SGEE	0.0011	0.0148	0.0102	-0.9375×10^{-3}	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

Table 5.3. Performance of parameter estimation methods under misspecification of an underlying ARMA(1,1) correlation structure

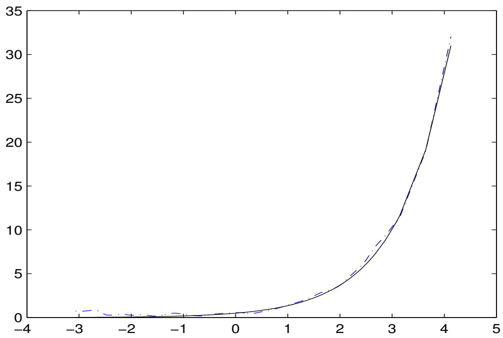
	n		30			50		
$\overline{m}$	Parameters	Methods	Bias	SD	MAD	Bias	SD	MAD
10	β_1	PULS	0.0072	0.0410	0.0357	-0.0038	0.0299	0.0201
		SGEE	-0.0054	0.0261	0.0210	-0.0055	0.0211	0.0147
	β_2	PULS	0.0068	0.0336	0.0256	0.0037	0.0290	0.0163
		SGEE	0.0025	0.0267	0.0157	0.0023	0.0190	0.0136
	θ_1	PULS	0.0037	0.0166	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



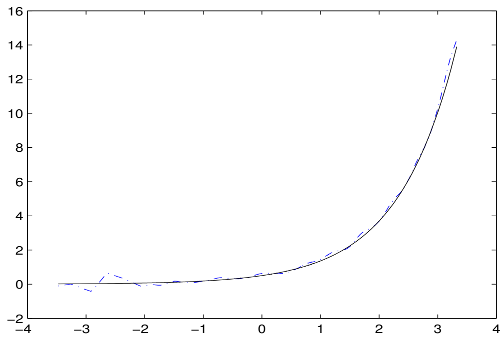
(a) $n = 30$, $\overline{m} = 10$



(b) $n = 50$, $\overline{m} = 10$

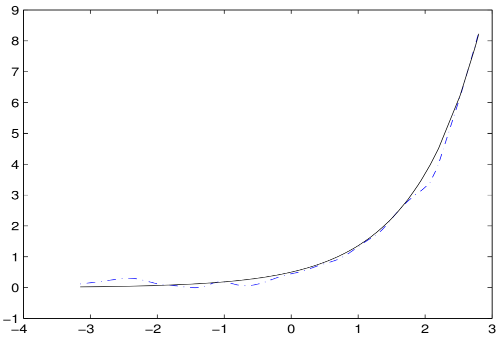


(c) $n = 30$, $\overline{m} = 30$

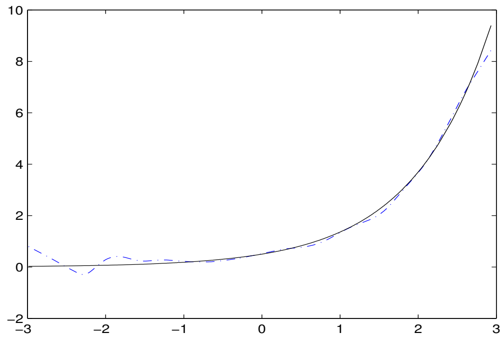


(d) $n = 50$, $\overline{m} = 30$

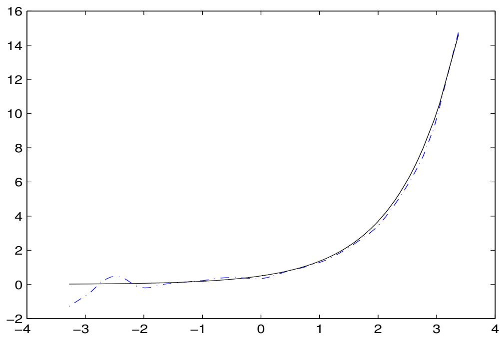
Figure 5.1. Estimated link function (dot-dashed line), together with the true link function (solid line), from a typical realization of model (1.2) with AR(1) correlation structure for each combination of n and $\overline{m}$: (a) $n = 30$, $\overline{m} = 10$; (b) $n = 50$, $\overline{m} = 10$; (c) $n = 30$, $\overline{m} = 30$; (d) $n = 50$, $\overline{m} = 30$.



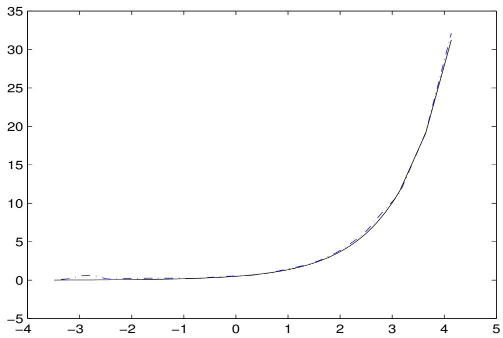
(a) $n = 30$, $\overline{m} = 10$



(b) $n = 50$, $\overline{m} = 10$



(c) $n = 30$, $\overline{m} = 30$



(d) $n = 50$, $\overline{m} = 30$

Figure 5.2. 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) $n = 50$, $\overline{m} = 30$.

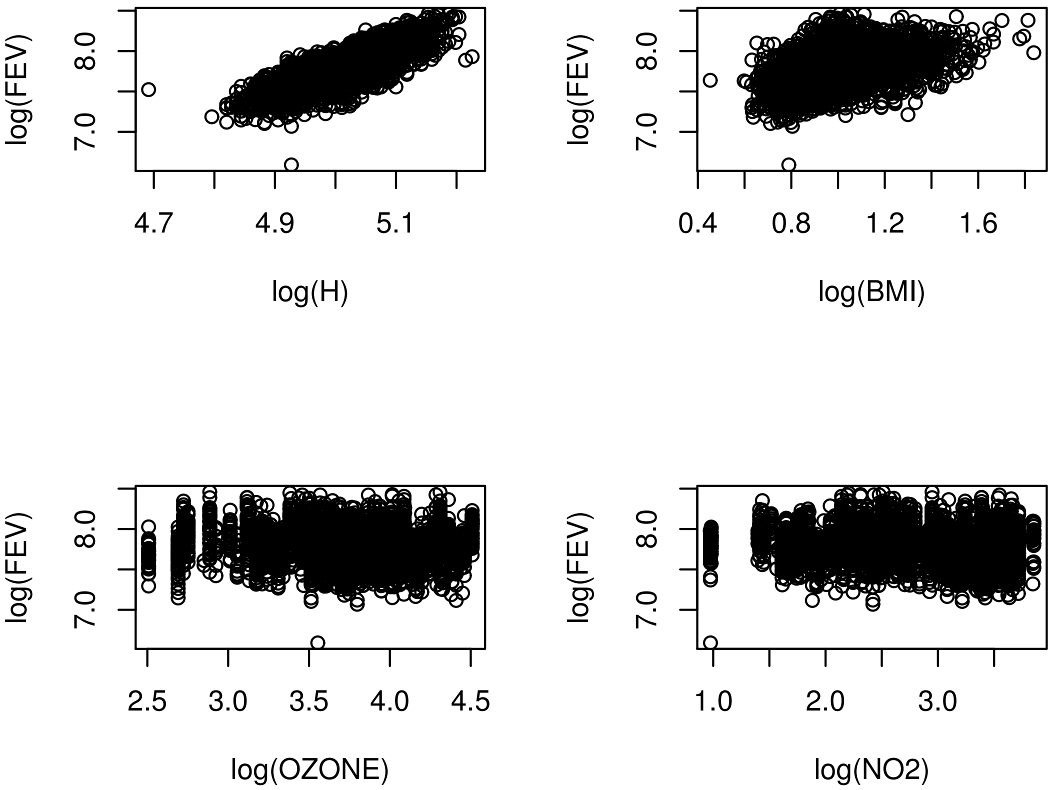


Figure 5.3. The scatter plots of the response variable log(FEV) against the continuous regressors, i.e., (clockwise from top left) log(H), log(BMI), log(NO_2), log(OZONE).

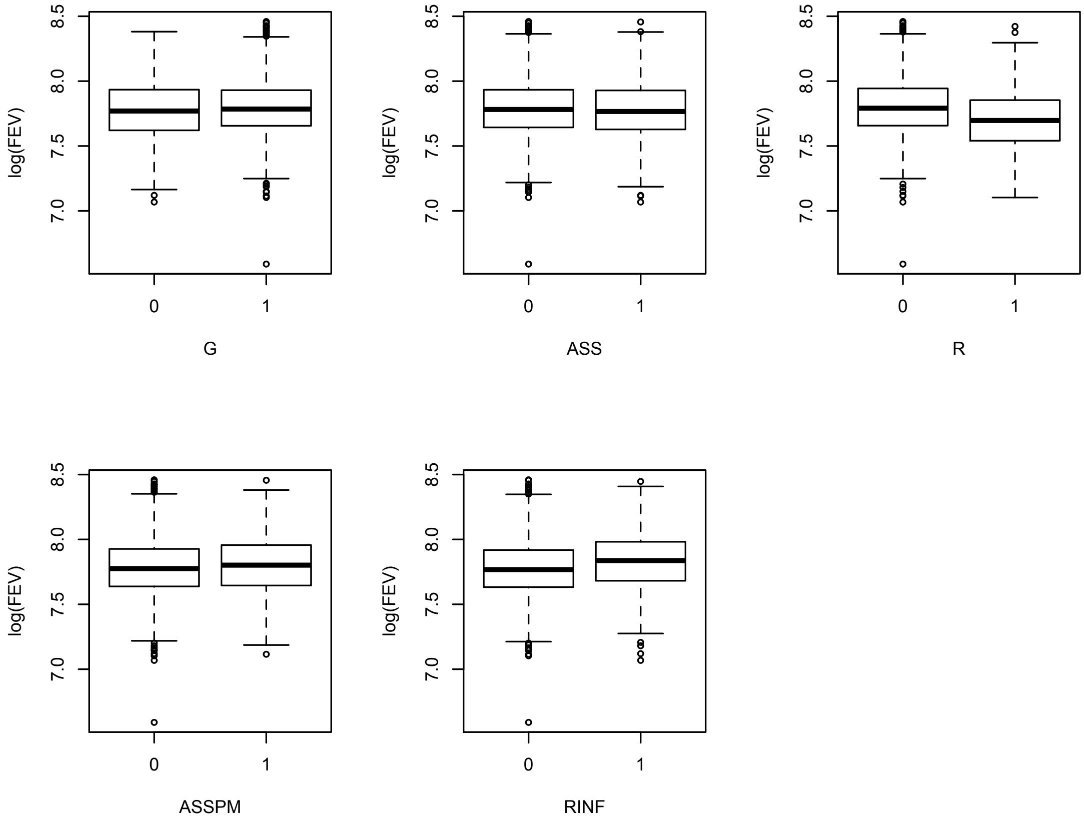


Figure 5.4. The box plots of the response variable log(FEV) against the binary regressors, i.e., (clockwise from top left) G, ASS, R, RINF, ASSPM.

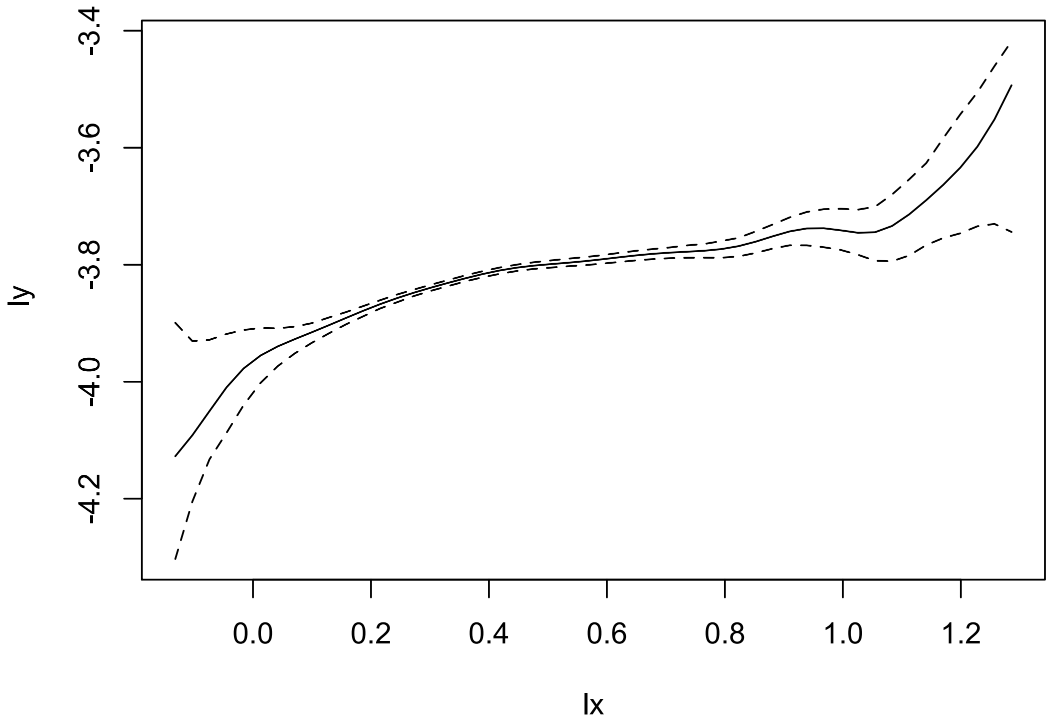


Figure 5.5. The estimated link function and its 95% confidence band.