**Electronic Journal of Statistics**
Vol. 14 (2020) 2395–2438
ISSN: 1935-7524
https://doi.org/10.1214/20-EJS1723

# Oracally efficient estimation and simultaneous inference in partially linear single-index models for longitudinal data

**Li Cai**[1] **and Lei Jin**[2] **and Suojin Wang**[3*]

[1] *School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou, China*
*e-mail:* caili16@126.com

[2] *Department of Mathematics and Statistics, Texas A&M University, Corpus Christi, Texas, USA e-mail:* Lei.Jin@tamucc.edu

[3] *Department of Statistics, Texas A&M University, College Station, TX 77843, USA*
*e-mail:* sjwang@stat.tamu.edu

**Abstract:** Oracally efficient estimation and an asymptotically accurate simultaneous confidence band are established for the nonparametric link function in the partially linear single-index models for longitudinal data. The proposed procedure works for possibly unbalanced longitudinal data under general conditions. The link function estimator is shown to be oracally efficient in the sense that it is asymptotically equivalent in the order of $n^{-1/2}$ to that with all true values of the parameters being known oracally. Furthermore, the asymptotic distribution of the maximal deviation between the estimator and the true link function is provided, and hence a simultaneous confidence band for the link function is constructed. Finite sample simulation studies are carried out which support our asymptotic theory. The proposed SCB is applied to analyze a CD4 data set.

**Keywords and phrases:** Local linear smoothing, longitudinal data, oracle efficiency, partially linear single-index model, simultaneous confidence band.

Received January 2020.

## 1. Introduction

There has been substantial interest in semiparametric partially linear models in the last three decades as they enjoy the advantages of both the flexibility of nonparametric modeling and easy interpretation of parametric modeling. In this paper we consider the following partially linear single-index model:

$$Y(t) = \mathbf{X}(t)^T \beta + \phi\left(\mathbf{Z}(t)^T \theta\right) + \varepsilon(t), t \in \mathscr{T}, \tag{1.1}$$

where $\beta$ and $\theta$ are unknown $p$ and $q$ dimensional parameter vectors, $\phi(\cdot)$ is an unknown link function, $\mathscr{T}$ is a compact set of time $t$, $Y(t)$ is a scalar stochastic

*Corresponding Author.

process, $\mathbf{X}(t)$ and $\mathbf{Z}(t)$ are $p$ and $q$ dimensional covariates, and $\varepsilon(t)$ is a random error process.

For this model in the context of longitudinal data, many different approaches have been studied to estimate the unknown coefficient vector $\mathbf{\Theta} = (\beta^T, \theta^T)^T$ and link function $\phi(\cdot)$. For instance, Chen et al. (2013) proposed semiparametric minimum average variance estimation for the partially linear single-index panel data models with fixed effects for dense longitudinal data. Ma et al. (2014) approximated the link function by the polynomial spline, and then applied the quadratic inference function with the profile principle to estimate the linear and sing-index coefficients for sparse longitudinal data. But their methods were limited to the balanced longitudinal data case. Later, Chen et al. (2015) proposed a unified semiparametric estimation method called SGEE which combines the local linear smoothing and generalized estimating equations (GEE) for possibly unbalanced both sparse and dense longitudinal data. Moreover, Cai and Wang (2019) introduced a refined three-stage approach called SMGEE to estimate the nonparametric link function using marginal kernel regression and the parametric components with generalized estimating equations. Their approach produces more efficient point and curve estimators by incorporating the within-subject correlation for the link function estimation, but with more computational cost. See also Lin and Carroll (2000, 2001, 2006), Fan and Li (2004), and He et al. (2005) for a series of efforts on semiparametric models for longitudinal or cluster data, and Carroll et al. (1997), Yu and Ruppert (2002), Xia and Härdle (2006), Liang et al. (2010) and Gu and Yang (2015) for studies on the single-index models in the context of independent and identically distributed data.

All aforementioned works mainly focus on the pointwise properties of estimation and/or the problem of dimension reduction. To assess the shape of the unknown link curve, however, it is desired to construct a uniform or simultaneous confidence band (SCB), which can be viewed as a collection of sliding confidence intervals over the whole domain of the function with a predetermined error probability. It is a powerful tool for making inferences on the global shape of a true curve and for making decisions if some feature of the estimated curve should be considered as a structure of the unknown function.

The construction of SCBs has been attempted in various contexts. A pioneering work can be traced back to Bickel and Rosenblatt (1973) for probability density functions, followed by a series of studies such as Johnston (1982), Härdle (1989), Eubank and Speckman (1993), Claeskens and Van Keilegom (2003), Wang and Yang (2009), Cai and Yang (2015) and Cai et al. (2019) for nonparametric regression curves, and Wu and Zhao (2007), Liu and Wu (2010) and Zhou and Wu (2010) for time series data. Recently, Degras (2011) constructed an SCB for the mean function of functional data based on local linear regression, while Cao et al. (2012) and Cao et al. (2016) constructed SCBs for the mean and covariance functions of functional data using B-spline regression. One limitation of their methods is the setting of fixed design and balanced measurement points. In addition, Ma et al. (2012) proposed an SCB for the mean function of the longitudinal data based on constant piecewise spline smoothing. Cao et al. (2018) proposed an SCB for varying coefficient models for sparse longitudinal data via local linear regression.

In this paper we aim to construct an asymptotically accurate SCB for the link function $\phi(\cdot)$ in the partially linear single-index model for possibly unbalanced longitudinal data under general conditions. Technical challenges on formulating such bands arise due to the highly intrinsic dimensionality of longitudinal data, the irregular and possibly subject specific time points at which the data are collected and the difficulty to obtain the uniform oracle properties of the estimation error (also known as maximal deviation) in our current setting. That is also why the SCBs have been rather under explored in the nonparametric curve estimation literature, especially in the context of multidimensional longitudinal data.

The proposed SCB for the link function is innovative and useful since it is the first of its kind for possibly unbalanced longitudinal data in the partially linear single-index modeling structure with sound theoretical justifications. It also provides satisfactory numerical performance in finite sample sizes. Moreover, the approach is flexible to cover a variety of situations, e.g., the single-index model (Xia and Li (1999)) and the semiparametric partially linear model (Härdle et al. (2012)). It can also be applied to test many kinds of predetermined models for data, e.g., multiple linear regression model. As a motivating example, the SCB is applied to study and make global inference on how the CD4 cell number depends on several predictors over time via analyzing a CD4 data set.

The rest of the paper is organized as follows. In Section 2 we present an asymptotic theory for constructing our proposed SCB. Section 3 provides further analysis of the estimation error structure. In Section 4, concrete steps are given to implement the proposed SCB. Simulation studies and a real data analysis are reported in Sections 5 and 6, respectively. All the technical proofs are relegated to the Appendix.

## 2. Asymptotic theory for constructing a new SCB

By convention, for any $m$-dimensional vector $\mathbf{v}$ with $i$th entry being $v_i$, we write $\mathbf{v} = (v_1, \dots, v_m)^T$ as a column vector. Suppose that a random sample consists of $n$ subjects. For $i$-th subject, the response variable $Y_i(t)$ and the covariates $\mathbf{X}_i(t) = (X_{i1}(t), \dots, X_{ip}(t))^T$ and $\mathbf{Z}_i(t) = (Z_{i1}(t), \dots, Z_{iq}(t))^T$ are collected at points $t_{ij}, 1 \le j \le N_i$. To simplify the representation, denote $Y_{ij} = Y_i(t_{ij})$, $X_{ij,l} = X_{il}(t_{ij}), l = 1, 2, \dots, p, Z_{ij,d} = Z_{id}(t_{ij}), d = 1, 2, \dots, q$, $\mathbf{X}_{ij} = \mathbf{X}_i(t_{ij}) = (X_{ij,1}, \dots, X_{ij,p})^T$, $\mathbf{Z}_{ij} = \mathbf{Z}_i(t_{ij}) = (Z_{ij,1}, \dots, Z_{ij,q})^T$ and $\varepsilon_{ij} = \varepsilon_i(t_{ij})$. Meanwhile, let $\beta = (\beta_1, \dots, \beta_p)^T$, $\theta = (\theta_1, \dots, \theta_q)^T$, and define $N_T = \sum_{i=1}^n N_i$, $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iN_i})^T, \mathbf{X}_i = (\mathbf{X}_{i1}, \dots, \mathbf{X}_{iN_i})^T, \mathbf{E}_i = (\varepsilon_{i1}, \dots, \varepsilon_{iN_i})^T$, and $\mathbf{Z}_i = (\mathbf{Z}_{i1}, \dots, \mathbf{Z}_{iN_i})^T$.

For the longitudinal data introduced above, model (1.1) can be written as

$$Y_{ij} = \mathbf{X}_{ij}^T \beta + \phi\left(\mathbf{Z}_{ij}^T \theta\right) + \varepsilon_{ij}, 1 \le j \le N_i, 1 \le i \le n. \tag{2.1}$$

Similar to related works, such as Wang et al. (2005), Fan et al. (2007) and Chen et al. (2015), we assume that the subjects are mutually independent but there is a within-subject correlation structure for each subject. We also assume that $\max_{1\le i\le n} N_i$ is bounded as $n\to\infty$. To ensure identifiability of the link function $\phi(\cdot)$, the parameter vector $\theta$ is assumed to be a unit vector with the first nonzero positive element, i.e., $\|\theta\|=1, \theta_1>0$, and no intercept is given in model (2.1) (it is effectively included in $\phi(\cdot)$). We also assume that the range of the single-index $u=\mathbf{Z}^T(t)\theta$ (and thus the domain of $\phi(u)$) is a finite interval $[a,b]$.

For model (2.1), if the coefficient vector $\mathbf{\Theta}$ were known oracally, a local linear estimator $\hat{\phi}(u,\mathbf{\Theta})$ of $\phi(u)$ is obtained by minimizing the following quadratic form with respect to $(c,d)$,

$$N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\left\{Y_{ij}-\mathbf{X}_{ij}^T\beta-c-d\left(\mathbf{Z}_{ij}^T\theta-u\right)\right\}^2 K_h\left(\mathbf{Z}_{ij}^T\theta-u\right), \tag{2.2}$$

where $K_h(u)=h^{-1}K(u/h)$ is a rescaled kernel function of kernel $K(u)$ with bandwidth $h$. By some standard calculations, one has that

$$\hat{\phi}(u,\mathbf{\Theta})=e_0^T\left(N_T^{-1}\mathbf{Z}^T\mathbf{W}\mathbf{Z}\right)^{-1}N_T^{-1}\mathbf{Z}^T\mathbf{W}\mathbf{Y}^*, \tag{2.3}$$

where $e_0=(1,0)^T, \mathbf{Y}^*=\left(Y_{11}^*,\ldots,Y_{1N_1^*},\ldots,Y_{nN_n}^*\right)^T, Y_{ij}^*=Y_{ij}-\mathbf{X}_{ij}^T\beta$,

$$\mathbf{Z}=\begin{pmatrix}1 & \cdots & 1 & \cdots & 1\\ \mathbf{Z}_{11}^T\theta-u & \cdots & \mathbf{Z}_{1N_1}^T\theta-u & \cdots & \mathbf{Z}_{nN_n}^T\theta-u\end{pmatrix}^T,$$

and the weight matrix $\mathbf{W}=\operatorname{diag}\left(K_h\left(\mathbf{Z}_{11}^T\theta-u\right),\ldots,K_h\left(\mathbf{Z}_{nN_n}^T\theta-u\right)\right)$.

However, $\mathbf{\Theta}$ is generally unknown. In this case, we replace the unknown $\mathbf{\Theta}$ in (2.2) by a $\sqrt{n}$ consistent estimator $\hat{\mathbf{\Theta}}=(\hat{\beta}^T,\hat{\theta}^T)^T$ to obtain a feasible estimator

$$\hat{\phi}\left(u,\hat{\mathbf{\Theta}}\right)=e_0^T\left(N_T^{-1}\hat{\mathbf{Z}}^T\hat{\mathbf{W}}\hat{\mathbf{Z}}\right)^{-1}N_T^{-1}\hat{\mathbf{Z}}^T\hat{\mathbf{W}}\hat{\mathbf{Y}}^*,$$

where the symbols with a hat on the right side of the equation are the same as those in equation (2.3) but with $\mathbf{\Theta}$ replaced by $\hat{\mathbf{\Theta}}$. Here the infeasible estimator $\hat{\phi}(u,\mathbf{\Theta})$ is introduced as a useful benchmark for feasible ones to compare with, and as a pivotal medium by which an asymptotic SCB can be constructed. Note that the estimation of $\mathbf{\Theta}$ is an auxiliary step to estimate $\phi(u)$, so any $\sqrt{n}$-consistent estimator $\hat{\mathbf{\Theta}}$ will be effective in our procedure.

For integer $p\ge 0$, denote by $C^{(p)}[a,b]$ the space of functions that have continuous $p$-th derivative on $[a,b]$ and let $C[a,b]=C^{(0)}[a,b]$. We make the following technical assumptions:

(A1) *The link function* $\phi(u)\in C^{(2)}[a,b]$. *The functions* $\mathrm{E}\left[\mathbf{Z}_{ij}|\mathbf{Z}_{ij}^T\theta=t\right]$ *and* $\mathrm{E}\left[\mathbf{X}_{ij}|\mathbf{Z}_{ij}^T\theta=t\right]$ *of* $t$ *have continuous derivatives up to the second order.*

(A2) *The errors* $\varepsilon_{ij}, 1\le j\le N_i, 1\le i\le n$, *are independent of covariates* $\mathbf{Z}_{ij}$ *and* $\mathbf{X}_{ij}$ *and are independent across* $i$, *satisfying* $\mathrm{E}\,\varepsilon_{ij}=0$ and $\mathrm{E}\,\varepsilon_{ij}\varepsilon_{ik}=\sigma_i^{jk}$ with $c_\sigma\le \mathrm{E}\,\varepsilon_{ij}^2=\sigma_i^{jj}\le C_\sigma$ *for some* $c_\sigma>0$ and $C_\sigma>0$. *Moreover,* *there exist some* $m > 2/3, m_0 > 2/3$, and $m_1 > 2/3$ *such that* $\mathrm{E}\left|\varepsilon_{ij}\right|^{2+m} < M_\varepsilon$, $\mathrm{E}\left|Z_{ij,d}\right|^{2+m_0} < M_Z$, and $\mathrm{E}\left|X_{ij,l}\right|^{2+m_1} < M_X$ *for some constants* $M_\varepsilon > 0, M_X > 0$, and $M_Z > 0$ *with* $d = 1, \ldots, p$, $l = 1, \ldots, q$. *Meanwhile, there exist constants* $G_1 > 0$ and $G_2 > 0$ *such that* $\mathrm{E}\left[\left|X_{ij,l}\right|^2 |\mathbf{Z}_{ij}^T\theta = t\right] \leq G_1$ *and* $\mathrm{E}\left[\left|Z_{ij,d}\right|^2 |\mathbf{Z}_{ij}^T\theta = t\right] \leq G_2$ *for all* $t \in [a, b]$.

(A3) *The kernel function* $K(u)$ *is a symmetric probability density function supported on* $[-1, 1]$ *and has the second Lipschitz continuous derivative.*

(A4) *The covariate vector* $\mathbf{Z}(t)$ *takes values in a* $q$*-dimensional bounded closed region. The density function* $f(u)$ *of* $\mathbf{Z}_{ij}^T\theta$ *satisfies* $f(u) \in C^{(2)}[a, b]$ *and there exist* $c_f > 0$ and $C_f > 0$ *such that* $c_f \leq f(u) \leq C_f$. *The joint density functions of* $\left(\mathbf{Z}_{ij}^T\theta, \mathbf{Z}_{ij'}^T\theta\right)$, $\left(\mathbf{Z}_{ij}^T\theta, Z_{ij,l}\right)$, $\left(\mathbf{Z}_{ij}^T\theta, X_{ij,l}\right)$, *and* $\left(\mathbf{Z}_{ij}^T\theta, Z_{ij,l}, Z_{ij,l'}\right)$ *have continuous partial derivatives of order one.*

(A5) *The bandwidth* $h$ *satisfies* $nh^4 \to \infty$ and $nh^5 \log n \to 0$ as $n \to \infty$.

Assumption (A1) is a mild smoothness condition on the link function $\phi(u)$ to apply local linear regression. Assumption (A2) is about the moment conditions of the error terms and the covariates that are similar to those used in Chen et al. (2015). It allows the error terms to have within-subject correlations. Assumption (A3) is a typical condition for the kernel function. Clearly, if we denote $m_l(K) = \int u^l K(u)\, du$, by (A3) one then has that $m_0(K) = 1$ and for all odd $l$, $m_l(K) = 0$.

Assumption (A4) gives some common conditions in partially linear single-index models similar to Xia and Härdle (2006), Chen et al. (2015), and Gu and Yang (2015). Assumption (A5) is a general restriction on the choice of bandwidth $h$ which keeps the bias of $\hat{\phi}(u, \mathbf{\Theta})$ at a lower rate than its standard error and keeps the order of oracle efficiency at $n^{-1/2}$; see Johnston (1982), Härdle (1989) and Gu and Yang (2015) for similar undersmoothing bandwidth conditions. In Section 4, we will describe a detailed procedure to select an appropriate bandwidth.

The existence of a $\sqrt{n}$-consistent estimator $\hat{\mathbf{\Theta}}$ is needed in our main results. The following condition that ensures any of the estimators $\hat{\mathbf{\Theta}}$ in Chen et al. (2015), Lin and Carroll (2000) and Cai and Wang (2019) is $\sqrt{n}$-consistent:

(C1) *The link function* $\phi(u) \in C^{(2)}[a, b]$*; the density function* $f(u)$ *of* $Z_{ij}^T\theta$ *is positive and has a continuous second-order derivative in* $U = \{\mathbf{Z}^T\theta : \mathbf{Z} \in \mathcal{Z}, \theta \in \Omega\}$ *where* $\Omega$ *is a compact parameter space for* $\theta$ *and* $\mathcal{Z}$ *is a compact support of* $\mathbf{Z}_{ij}$*;* $\mathrm{E}\left(\mathbf{X}_{ij} | \mathbf{Z}_{ij}^T\theta = u\right)$ *and* $\mathrm{E}\left(\mathbf{Z}_{ij} | \mathbf{Z}_{ij}^T\theta = u\right)$ *are a bounded smooth function of* $u$ *with continuous second-order derivative and the constant* $m$ *given in (A2) is larger than* 0.

Condition (C1) is satisfied by Assumptions (A1), (A2) and (A4). In the numerical studies described in Sections 5 and 6, we used the SGEE method in Chen et al. (2015) to obtain a $\sqrt{n}$-consistent estimator $\hat{\mathbf{\Theta}}$.

For any functions $f_n(x)$ and $g_n(x)$ defined in domain $\mathscr{X}$, we use $f_n(x) = O(g_n(x))$ and $f_n(x) = o(g_n(x))$ to denote that $f_n(x)/g_n(x)$ is bounded and $f_n(x)/g_n(x) \to 0$ as $n \to \infty$ for any fixed $x \in \mathscr{X}$, respectively, while we use $O_p(\cdot), o_p(\cdot), O_{a.s.}(\cdot), o_{a.s.}(\cdot)$ to represent the corresponding terms in probability and almost surely. Moreover, we use $f_n(x) = U_p(g_n(x))$ and $f_n(x) = u_p(g_n(x))$ to denote that $\sup_{x\in\mathscr{X}} |f_n(x)/g_n(x)|$ is bounded and $\sup_{x\in\mathscr{X}} |f_n(x)/g_n(x)| \to 0$ as $n \to \infty$ in probability, respectively. Meanwhile, define $\tau_n = n^{-1/2}h^{-1/2}\log^{1/2} n$ and let $[a_0, b_0]$ be an arbitrary closed subinterval of $(a, b)$ so that it excludes the end points of $a$ and $b$.

**Theorem 2.1.** *Under Assumptions (A1)–(A5), as $n \to \infty$, for any $u \in [a_0, b_0]$, one has that*

$$\hat{\phi}(u, \mathbf{\Theta}) - \phi(u) = R_n(u) + 2^{-1}\phi^{(2)}(u)\, m_2(K)\, h^2 + u_p\left(h^2\right),$$

*where* $R_n(u) = f^{-1}(u)\, N_T^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T\theta - u\right)\varepsilon_{ij}$.

Theorem 2.1 shows the global convergence property of $\hat{\phi}(u, \mathbf{\Theta})$ whose proof given in Appendix A.2 relies on Propositions 3.1 and 3.2 in Section 3. By Theorem 2.1, one has

$$\sup_{u\in[a_0,b_0]} |\hat{\phi}(u, \mathbf{\Theta}) - \phi(u)| = \sup_{u\in[a_0,b_0]} |R_n(u)| + O_p(h^2).$$

By Lemma A.2 in the Appendix A.1 that $\sup_{u\in[a_0,b_0]} |R_n(u)| = O_p(\tau_n)$ and $nh^5 \log n \to 0$ in Assumption (A5), one only needs to consider the dominating term $\sup_{u\in[a_0,b_0]} |R_n(u)|$ to investigate the asymptotic distributional properties of $\sup_{u\in[a_0,b_0]} |\hat{\phi}(u, \mathbf{\Theta}) - \phi(u)|$. For $u \in [a_0, b_0]$, define

$$C_n(u) = f^{-2}(u)\, N_T^{-1} h \sum_{i=1}^{n} \sum_{j=1}^{N_i} \sum_{j'=1}^{N_i} \mathrm{E}\left\{K_h\left(\mathbf{Z}_{ij}^T\theta - u\right) K_h\left(\mathbf{Z}_{ij'}^T\theta - u\right)\right\} \mathrm{E}(\varepsilon_{ij}\varepsilon_{ij'}).$$

**Theorem 2.2.** *Under Assumptions (A1)–(A5), one has that $\forall z \in \mathbb{R}$,*

$$\lim_{n\to\infty} P\left[a_h\left\{\sup_{u\in[a_0,b_0]}\left|(N_T h)^{1/2}\left\{\hat{\phi}(u, \mathbf{\Theta}) - \phi(u)\right\}C_n^{-1/2}(u)\right| - b_h\right\} \le z\right] = e^{-2e^{-z}},$$

*where*

$$a_h = \left\{-2\log\left(\frac{h}{b_0 - a_0}\right)\right\}^{1/2}, b_h = a_h + 2^{-1}a_h^{-1}\log\frac{C_K}{2\pi^2},$$
$$C_K = \int \left\{K^{(1)}(v)\right\}^2 dv \Big/ \left(2\int K^2(v)\, dv\right).$$

Theorem 2.2 describes the limiting distribution of the maximal deviation of $\hat{\phi}(u, \mathbf{\Theta})$. Its proof is readily obtained by following that of Theorem 1 in Cao et al. (2018) for $R_n(u)$. Note that while Theorem 1 in Cao et al. (2018) assumes that the regression variables (times $t_{ij}$ in their paper) to be independent over time even within-subjects, it is seen that the independence condition of the regression variables can be weakened to be dependent within-subjects when the number of observations per subject is bounded. This is because the independence assumption needed in using the Bernstein inequality can be relaxed by using the independence among the subjects and treating each subject's finite sum as a subjectwise random variable as in (A.14) in the proof of Proposition 3.1.

**Theorem 2.3.** *Under Assumptions (A1)–(A5), for any $\sqrt{n}$-consistent estimator $\hat{\mathbf{\Theta}}$, as $n \to \infty$, one has that*

$$\sup_{u\in[a_0,b_0]} \left|\hat{\phi}(u,\hat{\mathbf{\Theta}}) - \hat{\phi}\left(u,\mathbf{\Theta}\right)\right| = O_p\left(n^{-1/2}\right). \tag{2.4}$$

The proof of Theorem 2.3 is given in Appendix A.2 which relies on Propositions 3.3 and 3.4 in Section 3. Theorem 2.3 shows that the estimator $\hat{\phi}(u,\hat{\mathbf{\Theta}})$ with any $\sqrt{n}$-consistent estimator $\hat{\mathbf{\Theta}}$ is as efficient as the estimator $\hat{\phi}\left(u,\mathbf{\Theta}\right)$ with the true coefficient vector $\mathbf{\Theta}$ known by an "oracle". Therefore, asymptotically there is no need to distinguish the difference between $\hat{\phi}(u,\hat{\mathbf{\Theta}})$ and $\hat{\phi}\left(u,\mathbf{\Theta}\right)$ when constructing SCBs. We have hence established the oracle efficiency of $\hat{\phi}(u,\hat{\mathbf{\Theta}})$. This along with Theorem 2.2 leads to the following theorem.

**Theorem 2.4.** *Under Assumptions (A1)–(A5), for any $\sqrt{n}$-consistent estimator $\hat{\mathbf{\Theta}}$, one has that $\forall z \in \mathbb{R}$*

$$\lim_{n\to\infty} P\left[a_h\left\{\sup_{u\in[a_0,b_0]}\left|(N_T h)^{1/2}\left\{\hat{\phi}\left(u,\hat{\mathbf{\Theta}}\right)-\phi\left(u\right)\right\}C_n^{-1/2}\left(u\right)\right|-b_h\right\}\le z\right]=e^{-2e^{-z}}.$$

By Theorem 2.4, one can easily obtain a theoretical SCB for $\phi(u)$ which depends on the unknown quantity $C_n\left(u\right)$. In order to construct a feasible SCB based on this result, one further needs to estimate $C_n\left(u\right)$ with a proper uniform convergence property.

We first estimate the unknown density function $f\left(u\right)$ of $\mathbf{Z}_{ij}^T\theta$. The following kernel density pilot estimator is applied to estimate $f\left(u\right)$:

$$\hat{f}\left(u\right) = N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} K_{h_f}\left(\mathbf{Z}_{ij}^T\hat{\theta}-u\right), \tag{2.5}$$

where $h_f$ is the Silverman's rule-of-thumb bandwidth (Silverman (1986), p. 48) with the order of $n^{-1/5}$. By the proof of Proposition 3.1 and Assumption (A5), one has

$$\sup_{u\in[a_0,b_0]}\left|\hat{f}\left(u\right) - f\left(u\right)\right| = o_p\left(h_f\right) + O_p\left(n^{-1/2}h_f^{-1/2}\log^{1/2} n\right) = o_p\left(h_f\right). \tag{2.6}$$

Let $\hat{\varepsilon}_{ij} = Y_{ij} - \mathbf{X}_{ij}^T\hat{\beta} - \hat{\phi}\left(\mathbf{Z}_{ij}^T\hat{\theta}\right)$, $1 \le j \le N_i, 1 \le i \le n$, and define

$$\hat{C}_n\left(u\right) = \hat{f}^{-2}\left(u\right)N_T^{-1}h\sum_{i=1}^{n}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T\hat{\theta}-u\right)K_h\left(\mathbf{Z}_{ij'}^T\hat{\theta}-u\right)\hat{\varepsilon}_{ij}\hat{\varepsilon}_{ij'}.$$

as an estimator of $C_n(u)$. Then we have the following theorem whose proof is given in Appendix A.2.

**Theorem 2.5.** *Under Assumptions (A1)–(A5), as $n \to \infty$, one has that*

$$\sup_{u \in [a_0, b_0]} \left| \hat{C}_n(u) - C_n(u) \right| = o_p(h_f).$$

Since $h_f = O\left(n^{-1/5}\right) \ll \log^{-1} n$, by combining Theorems 2.4 and 2.5 one obtains the main result in the following corollary.

**Corollary 2.1.** *Under Assumptions (A1)–(A5), for any $\alpha \in (0,1)$ and any $\sqrt{n}$-consistent estimator $\hat{\boldsymbol{\Theta}}$, as $n \to \infty$, an asymptotically correct $100(1-\alpha)\%$ SCB for $\phi(u), u \in [a_0, b_0]$, is*

$$\hat{\phi}\left(u, \hat{\boldsymbol{\Theta}}\right) \pm (N_T h)^{-1/2} \hat{C}_n^{1/2}(u) \left(a_h^{-1} Q_{1-\alpha} + b_h\right),$$

*where $Q_{1-\alpha} = -\log\left\{-2^{-1}\log(1-\alpha)\right\}$ and $a_h, b_h$ are given in Theorem 2.2.*

## 3. Error decomposition

In order to prove the global property of $\hat{\phi}(u, \boldsymbol{\Theta})$ in Theorem 2.1 and the oracle efficiency of $\hat{\phi}(u, \hat{\boldsymbol{\Theta}})$ in Theorem 2.3, we make the following decomposition of the estimation errors $\hat{\phi}(u, \boldsymbol{\Theta}) - \phi(u)$ and $\hat{\phi}(u, \hat{\boldsymbol{\Theta}}) - \phi(u)$.

For simplicity, we first denote

$$T_{n,l}(u) = N_T^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T \theta - u\right) \left(\mathbf{Z}_{ij}^T \theta - u\right)^l, l \geq 0, \tag{3.1}$$

and

$$W_{n,l}(u) = N_T^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T \theta - u\right) \left(\mathbf{Z}_{ij}^T \theta - u\right)^l \left\{Y_{ij}^* - \phi(u) - \phi^{(1)}(u)\left(\mathbf{Z}_{ij}^T \theta - u\right)\right\}. \tag{3.2}$$

Thus,

$$N_T^{-1} \mathbf{Z}^T \mathbf{W} \mathbf{Z} = \begin{pmatrix} T_{n,0}(u) & T_{n,1}(u) \\ T_{n,1}(u) & T_{n,2}(u) \end{pmatrix}$$

and $N_T^{-1} \mathbf{Z}^T \mathbf{W}\left(\mathbf{Y}^* - \phi(u) \mathbf{Z} e_0^T - \phi^{(1)}(u) \mathbf{Z} e_1\right) = \left(W_{n,0}(u), W_{n,1}(u)\right)^T$, $e_1 = (0,1)^T$, which with (2.3) imply that

$$\begin{aligned} \hat{\phi}(u, \boldsymbol{\Theta}) - \phi(u) &= e_0^T \left(N_T^{-1} \mathbf{Z}^T \mathbf{W} \mathbf{Z}\right)^{-1} N_T^{-1} \mathbf{Z}^T \mathbf{W}\left(\mathbf{Y}^* - \phi(u) \mathbf{Z} e_0^T\right) \\ &= e_0^T \begin{pmatrix} T_{n,0}(u) & T_{n,1}(u) \\ T_{n,1}(u) & T_{n,2}(u) \end{pmatrix}^{-1} \begin{pmatrix} W_{n,0}(u) \\ W_{n,1}(u) \end{pmatrix}. \end{aligned} \tag{3.3}$$

**Proposition 3.1.** *Under Assumptions (A3)–(A5), as $n \to \infty$, for $u \in [a_0, b_0]$ and integer $l = 0, 1, 2$, one has that*

$$T_{n,l}(u) = f(u) m_l(K) h^l + h^{l+1} m_{l+1}(K) f^{(1)}(u) + u_p\left(h^{l+1}\right) + U_p\left(h^l \tau_n\right).$$

**Proposition 3.2.** *Under Assumptions (A1)–(A5), as $n \to \infty$, for $u \in [a_0, b_0]$, one has that*

$$W_{n,0}(u) = N_T^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T \theta - u\right) \varepsilon_{ij} + 2^{-1} \phi^{(2)}(u) f(u) m_2(K) h^2 + u_p\left(h^2\right)$$

*and*

$$W_{n,1}(u) = U_p\left(h^3 + h\tau_n\right).$$

For $l = 0, 1, 2$, denote

$$\hat{T}_{n,l}(u) = N_T^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T \hat{\theta} - u\right)\left(\mathbf{Z}_{ij}^T \hat{\theta} - u\right)^l, \tag{3.4}$$

and

$$\begin{aligned} \hat{W}_{n,l}(u) = & N_T^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T \hat{\theta} - u\right)\left(\mathbf{Z}_{ij}^T \hat{\theta} - u\right)^l \\ & \times \left\{Y_{ij} - \mathbf{X}_{ij}^T \hat{\beta} - \phi(u) - \phi^{(1)}(u)\left(\mathbf{Z}_{ij}^T \hat{\theta} - u\right)\right\}. \end{aligned}$$

Similar to (3.3), one has

$$\hat{\phi}\left(u, \hat{\mathbf{\Theta}}\right) - \phi(u) = e_0^T \begin{pmatrix} \hat{T}_{n,0}(u) & \hat{T}_{n,1}(u) \\ \hat{T}_{n,1}(u) & \hat{T}_{n,2}(u) \end{pmatrix}^{-1} \begin{pmatrix} \hat{W}_{n,0}(u) \\ \hat{W}_{n,1}(u) \end{pmatrix}.$$

**Proposition 3.3.** *Under Assumptions (A1)–(A5), as $n \to \infty$, one has*

$$\sup_{u \in [a_0, b_0]} \left|\hat{T}_{n,l}(u) - T_{n,l}(u)\right| = O_p\left(n^{-1/2}\right), \ l = 0, 1, 2.$$

**Proposition 3.4.** *Under Assumptions (A1)–(A5), as $n \to \infty$, one has*

$$\sup_{u \in [a_0, b_0]} \left|\hat{W}_{n,l}(u) - W_{n,l}(u)\right| = O_p\left(n^{-1/2}\right), l = 0, 1.$$

The proofs of Propositions 3.1–3.4 are given in Appendix A.2.

## 4. Implementation

In this section, we describe a concrete procedure to implement the SCB for $\phi(u)$ given in Corollary 2.1. The procedure will be used in Sections 5 and 6 for simulation studies and a real data analysis.

We use $(\hat{a}, \hat{b}) = (\min_{i=1,j=1}^{n,N_i} \boldsymbol{Z}_{ij}^T \hat{\boldsymbol{\theta}}, \max_{i=1,j=1}^{n,N_i} \boldsymbol{Z}_{ij}^T \hat{\boldsymbol{\theta}})$ as the index domain and let $\hat{a}_0$ and $\hat{b}_0$ be the 99% and 1% quantiles of $\mathbf{Z}_{ij}^T \hat{\boldsymbol{\theta}}, j = 1, 2, \ldots, N_i, i = 1, 2, \ldots, n$, respectively. The SCB for the link function is constructed over the compact interval $[\hat{a}_0, \hat{b}_0]$.

The quartic kernel $K(u) = 15\left(1-u^2\right)^2 I(|u| \leq 1)/16$ is used for the local linear regression in (2.2) which satisfies Assumption (A3). It is also employed to estimate the density function of $\mathbf{Z}_{ij}^T \boldsymbol{\theta}$ in (2.5). The bandwidth of the local linear regression is selected by minimizing the estimated integrated mean squared error (MISE) with undersmoothing by $\log^{-1/2} n$, i.e., $h = h_{\text{opt}} \log^{-1/2} n$, where

$$h_{\text{opt}} = \left\{ \frac{35\sigma^2(b-a)}{n \int \{\phi^{(2)}(u)\}^2 f(u) du} \right\}^{1/5}, \tag{4.1}$$

which follows the direct plug-in approach in Ruppert et al. (1995), with the two unknown functionals in $h_{\text{opt}}$ being replaced by the kernel estimates described in Ruppert et al. (1995). From (4.1), one can see that the order of $h_{\text{opt}}$ is $n^{-1/5}$. Thus $h = h_{\text{opt}} \log^{-1/2} n$ has the order of $n^{-1/5} \log^{-1/2} n$, fulfilling condition (A5).

To obtain a $\sqrt{n}$-consistent estimator $\hat{\boldsymbol{\Theta}}$ of $\boldsymbol{\Theta}$, one can employ the two-step SGEE method in Chen et al. (2015) or the method of WI kernel GEE and WI least square in Lin and Carroll (2000). Under Assumptions (A1)–(A5), according to Chen et al. (2015) and Lin and Carroll (2000), one has as $n \to \infty$,

$$\left\| \hat{\beta} - \beta \right\| = O_p\left(n^{-1/2}\right), \left\| \hat{\theta} - \theta \right\| = O_p\left(n^{-1/2}\right). \tag{4.2}$$

In general, one can also apply the SMGEE method in Cai and Wang (2019) to compute $\hat{\boldsymbol{\Theta}}$ which also satisfies (4.2), but its computational procedure is time-consuming and can be prohibitive for simulation studies. Hence the methods in Chen et al. (2015) and Lin and Carroll (2000) are recommended here for computational simplicity.

One reviewer commented on how in practice one would reasonably divide the covariates into two parts before a partially linear single-index model is fitted to data. While this is a very good general question when considering this type of models, we provide a brief discussion here. Firstly, it would be useful to examine the pairwise scatterplots of the data to see the trend between the response variable and each covariate. In general, if a plot shows a clear linear trend, then one would treat the covariate as a linear component. Meanwhile, by convention, the dummy variables are often put in the linear part. This heuristic model selection method for partially linear single-index models was applied in a lot of the existing literature, see, for example, Carroll et al. (1997), Ma et al. (2014), Chen et al. (2015) and Cai and Wang (2019). See also Section 6 for a CD4 data analysis. Secondly, one could apply prior knowledge on the relationship between the variables provided by expert opinions, historical information or previous relevant studies to divide the covariates into two parts. This method was applied, for example, in Xue and Zhang (2020) when using a partially linear single-index model to analyze a data set from an AIDS clinical trial group (ACTG) study. Thirdly, one could employ a penalized testing procedure in Liang et al. (2010) or Ma et al. (2014) to simultaneously select significant variables and test the nonparametric component. However, they are not able to detect misspecification after the linear part and the single-index part are specified. It would be useful to develop a formal model checking diagnostic test in the current context. This is generally an interesting topic for further study.

## 5. Simulation studies

In this section we examine finite sample performance of the proposed SCB by Monte Carlo simulations.

We generated data using model (2.1) with $\mathbf{X}(t)$ and $\mathbf{Z}(t)$ being two-dimensional and three-dimensional covariates respectively, and $\beta = (2,1)^T$, $\theta = 3^{-1} \times(2,1,2)^T$. The covariates $(\mathbf{X}_i^T(t_{ij}), \mathbf{Z}_i^T(t_{ij}))^T$ were generated independently from a five-dimensional normal distribution with marginal mean 0, variance 1 and pairwise correlation 0.1. In addition, following a reviewer's suggestion, we also considered some discrete variables for the $\mathbf{X}(t)$ part. Specifically, we let $\mathbf{X}(t)$ be a three-dimensional variable with the true parameter vector $\beta = (1,1,1)^T$. Covariate $X_{i1}(t_{ij})$ was generated from Bernoulli distribution with probability 0.7 of being 1, while covariate $X_{i2}(t_{ij})$ was drawn from $\{0,1,2\}$ with equal probability and covariate $X_{i3}(t_{ij})$ was generated independently from normal distribution $N(0,1)$. The covariate $\mathbf{Z}(t)$ was still a three-dimensional covariate generated in the same way as above.

The following two link functions were considered: (1) $\phi(u) = 0.5\exp(u)$ and (2) $\phi(u) = 2\sin(u)$. Following Chen et al. (2015) the number of subjects was first set to be $n = 30, 50$. In addition, $n = 100$ was also considered. The confidence levels for the SCBs were set to be $1-\alpha = 0.95, 0.99$. The observation time $t_{ij}, i = 1,2,\ldots,n, j = 1,2,\ldots,N_i$ for each subject was generated in the following way as in Fan et al. (2007): each subject has a set of "scheduled" time points $\{0,1,\ldots,12\}$, and each of them has a 20% probability of being skipped with the average of $N_i$ being 10.4; the actual time $t_{ij}$ is a perturbation of a non-skipped "scheduled" time by adding a uniform $[0,1]$ random variable. Furthermore, $\varepsilon(t)$ was generated from a Gaussian process with mean $\mathbf{0}$, variance function $\sigma^2(t) = 0.5\exp(t/12)$, and ARMA(1,1) correlation structure

$$\operatorname{corr}(\varepsilon(t), \varepsilon(s)) = \begin{cases} \rho\gamma^{|t-s|}, & \text{if } t \neq s, \\ 1, & \text{if } t = s. \end{cases}$$

Notice that when $\rho = 1$, the correlation structure above is reduced to an $AR(1)$ structure. We took four pairs of $(\rho, \gamma)$: $(1, 0.3)$, $(1, 0.75)$, $(0.85, 0.3)$, $(0.85, 0.75)$, which include both strongly and weakly correlated AR(1) and ARMA(1,1) errors.

The construction of the SCB according to Corollary 2.1 is over an interior interval $[a_0, b_0] \subset (a,b)$. When the number of subjects $n = 30$, over 1000 replications, the average interval of $[\hat{a}_0, \hat{b}_0]$ was $[-2.650, 2.653]$ and the average interval $[\hat{a}, \hat{b}] = [-3.032, 3.082]$. When $n$ increases to 100, the average interval of $[\hat{a}_0, \hat{b}_0]$ was $[-2.900, 2.903]$ and the average interval $[\hat{a}, \hat{b}] = [-3.112, 3.135]$. The results here are based on the case of normally distributed covariates and $\phi(u) = 0.5\exp(u)$ under the AR(1) correlation structure with $\rho = 1, \gamma =$

Table 1
*Empirical coverage frequencies and average widths (in parentheses) of the proposed SCB with $\hat{\mathbf{\Theta}}$ obtained by the SGEE method with 1000 replications under the correct specification of the underlying correlation structure.*

| $\phi(u) = 0.5\exp(u)$ | | | | | |
|---|---|---|---|---|---|
| $n$ | $1-\alpha$ | $\varepsilon(t) \sim$ AR(1) | | $\varepsilon(t) \sim$ ARMA(1,1) | |
| | | $\rho=1, \gamma=0.3$ | $\rho=1, \gamma=0.75$ | $\rho=0.85, \gamma=0.3$ | $\rho=0.85, \gamma=0.75$ |
| 30 | 0.95 | 0.863 (1.810) | 0.865 (1.906) | 0.860 (2.140) | 0.864 (1.879) |
| | 0.99 | 0.929 (2.226) | 0.943 (2.341) | 0.928 (2.629) | 0.929 (2.309) |
| 50 | 0.95 | 0.918 (1.275) | 0.911 (1.390) | 0.910 (1.396) | 0.896 (1.330) |
| | 0.99 | 0.959 (1.549) | 0.954 (1.687) | 0.968 (1.695) | 0.957 (1.616) |
| 100 | 0.95 | 0.970 (0.917) | 0.964 (0.979) | 0.976 (0.911) | 0.964 (0.971) |
| | 0.99 | 0.988 (1.100) | 0.986 (1.174) | 0.989 (1.093) | 0.987 (1.164) |
| $\phi(u) = 2\sin(u)$ | | | | | |
| $n$ | $1-\alpha$ | $\varepsilon(t) \sim$ AR(1) | | $\varepsilon(t) \sim$ ARMA(1,1) | |
| | | $\rho=1, \gamma=0.3$ | $\rho=1, \gamma=0.75$ | $\rho=0.85, \gamma=0.3$ | $\rho=0.85, \gamma=0.75$ |
| 30 | 0.95 | 0.892 (1.599) | 0.866 (1.884) | 0.888 (1.763) | 0.872 (1.932) |
| | 0.99 | 0.954 (1.958) | 0.939 (2.308) | 0.950 (2.159) | 0.928 (2.367) |
| 50 | 0.95 | 0.937 (1.437) | 0.918 (1.360) | 0.946 (1.274) | 0.917 (1.429) |
| | 0.99 | 0.965 (1.740) | 0.964 (1.647) | 0.970 (1.541) | 0.967 (1.729) |
| 100 | 0.95 | 0.977 (0.980) | 0.974 (1.028) | 0.982 (0.948) | 0.966 (1.006) |
| | 0.99 | 0.988 (1.172) | 0.987 (1.229) | 0.990 (1.134) | 0.990 (1.202) |

0.3. Other settings yield similar results and hence are omitted. In principle, $[a_0, b_0]$ can cover essentially the entire interior $(a, b)$ as the number of subjects $n \to \infty$.

We first look at the performance of the proposed SCB in the cases that the correlation structure in the estimation of covariance function is correctly specified. Tables 1 and 5 report the coverage frequencies over 1000 replications that the true curve was covered by the SCBs (with $\hat{\mathbf{\Theta}}$ being obtained by the method of SGEE in Chen et al. (2015)), at the equally spaced 401 points $\{\hat{a}_0 + k(\hat{b}_0 - \hat{a}_0)/400, k = 0, 1, \ldots, 400\}$ for continuous and partly discrete covariates, respectively. It can be seen that in all the cases (i) as the number of subjects $n$ increases, the coverage frequencies improve and approach the nominal level $1 - \alpha$ while the average widths decrease, supporting the asymptotic theoretical results in Corollary 2.1; (ii) the coverage frequencies for the simulated data with some discrete covariates are similar to those for the simulated data with normally distributed covariates. In addition, we also compared the performance of the SCB when $\mathbf{\Theta}$ is known and unknown. Table 3 reports the coverage frequencies and the average widths of the bands when the true $\mathbf{\Theta}$ was used over 1000 replications for the case of the simulated data with normally distributed covariates. By comparing Table 1 with Table 3, one can see that the coverage frequencies and the widths of the bands are quite close to each other which agrees with the theoretical results in Theorems 2.2 and 2.3. The comparisons for the case of having some discrete covariates are similar, and are thus omitted.

We next look at the impact of misspecification of the correlation structure on the the performance of the SCB. Similar to Chen et al. (2015) and Fan et al. (2007), we used the AR(1) working correlation structure in the covariance matrix

Table 2

*Empirical coverage frequencies and average widths (in parentheses) of the proposed SCB with $\hat{\boldsymbol{\Theta}}$ obtained by the SGEE method with 1000 replications in the misspecified case of using the AR(1) correlation structure to model the true ARMA(1,1) correlation structure.*

| $n$ | $1-\alpha$ | $\phi(u)=0.5\exp(u)$ | | $\phi(u)=2\sin(u)$ | |
|---|---|---|---|---|---|
| | | $\rho=0.85,\gamma=0.3$ | $\rho=0.85,\gamma=0.75$ | $\rho=0.85,\gamma=0.3$ | $\rho=0.85,\gamma=0.75$ |
| 30 | 0.95 | 0.848 (1.636) | 0.864 (2.215) | 0.887 (1.758) | 0.870 (2.011) |
| | 0.99 | 0.932 (2.011) | 0.935 (2.727) | 0.953 (2.154) | 0.926 (2.465) |
| 50 | 0.95 | 0.907 (1.242) | 0.898 (1.437) | 0.946 (1.376) | 0.915 (1.393) |
| | 0.99 | 0.960 (1.508) | 0.954 (1.746) | 0.972 (1.665) | 0.964 (1.685) |
| 100 | 0.95 | 0.977 (0.911) | 0.963 (0.966) | 0.981 (0.948) | 0.963 (1.001) |
| | 0.99 | 0.988 (1.093) | 0.983 (1.158) | 0.988 (1.133) | 0.988 (1.196) |

Table 3

*Empirical coverage frequencies and average widths (in parentheses) of the proposed SCB with using the true $\boldsymbol{\Theta}$ with 1000 replications under the correct specification of the underlying correlation structure.*

| | | $\phi(u)=0.5\exp(u)$ | | | |
|---|---|---|---|---|---|
| $n$ | $1-\alpha$ | $\varepsilon(t)\sim$ AR(1) | | $\varepsilon(t)\sim$ ARMA(1,1) | |
| | | $\rho=1,\gamma=0.3$ | $\rho=1,\gamma=0.75$ | $\rho=0.85,\gamma=0.3$ | $\rho=0.85,\gamma=0.75$ |
| 30 | 0.95 | 0.844 (1.488) | 0.856 (1.625) | 0.870 (1.473) | 0.860 (1.431) |
| | 0.99 | 0.924 (1.830) | 0.933 (1.997) | 0.932 (1.811) | 0.939 (1.760) |
| 50 | 0.95 | 0.919 (1.183) | 0.903 (1.281) | 0.904 (1.185) | 0.907 (1.211) |
| | 0.99 | 0.964 (1.438) | 0.955 (1.555) | 0.964 (1.440) | 0.953 (1.471) |
| 100 | 0.95 | 0.969 (0.923) | 0.966 (0.985) | 0.975 (0.916) | 0.964 (0.970) |
| | 0.99 | 0.990 (1.107) | 0.987 (1.182) | 0.992 (1.099) | 0.991 (1.163) |
| | | $\phi(u)=2\sin(u)$ | | | |
| $n$ | $1-\alpha$ | $\varepsilon(t)\sim$ AR(1) | | $\varepsilon(t)\sim$ ARMA(1,1) | |
| | | $\rho=1,\gamma=0.3$ | $\rho=1,\gamma=0.75$ | $\rho=0.85,\gamma=0.3$ | $\rho=0.85,\gamma=0.75$ |
| 30 | 0.95 | 0.874 (1.529) | 0.863 (1.686) | 0.872 (1.528) | 0.856 (1.649) |
| | 0.99 | 0.938 (1.873) | 0.936 (2.065) | 0.944 (1.871) | 0.923 (2.020) |
| 50 | 0.95 | 0.931 (1.231) | 0.921 (1.319) | 0.941 (1.233) | 0.928 (1.309) |
| | 0.99 | 0.968 (1.490) | 0.964 (1.597) | 0.970 (1.492) | 0.964 (1.584) |
| 100 | 0.95 | 0.975 (0.963) | 0.977 (1.017) | 0.980 (0.954) | 0.969 (1.008) |
| | 0.99 | 0.985 (1.151) | 0.984 (1.216) | 0.989 (1.141) | 0.989 (1.204) |

Table 4

*Empirical coverage frequencies and average widths (in parentheses) of the proposed SCB with using the true $\boldsymbol{\Theta}$ with 1000 replications in the misspecified case of using the AR(1) correlation structure to model the true ARMA(1,1) correlation structure.*

| $n$ | $1-\alpha$ | $\phi(u)=0.5\exp(u)$ | | $\phi(u)=2\sin(u)$ | |
|---|---|---|---|---|---|
| | | $\rho=0.85,\gamma=0.3$ | $\rho=0.85,\gamma=0.75$ | $\rho=0.85,\gamma=0.3$ | $\rho=0.85,\gamma=0.75$ |
| 30 | 0.95 | 0.872 (1.475) | 0.863 (1.593) | 0.881 (1.533) | 0.856 (1.650) |
| | 0.99 | 0.936 (1.814) | 0.945 (1.959) | 0.948 (1.878) | 0.929 (2.022) |
| 50 | 0.95 | 0.902 (1.184) | 0.910 (1.259) | 0.943 (1.225) | 0.924 (1.307) |
| | 0.99 | 0.957 (1.438) | 0.953 (1.529) | 0.972 (1.482) | 0.962 (1.582) |
| 100 | 0.95 | 0.975 (0.915) | 0.964 (0.981) | 0.981 (0.961) | 0.969 (1.008) |
| | 0.99 | 0.990 (1.098) | 0.990 (1.176) | 0.988 (1.149) | 0.988 (1.205) |

Table 5

*Empirical coverage frequencies and average widths (in parentheses) of the proposed SCB with $\hat{\mathbf{\Theta}}$ obtained by the SGEE method for $\phi(u) = 0.5\exp(u)$ with 1000 replications under the correct specification of the underlying correlation structure. The simulated data have some discrete variables in the linear covariates $\mathbf{X}$ part.*

| $\phi(u) = 0.5\exp(u)$ | | | | | |
|---|---|---|---|---|---|
| $n$ | $1-\alpha$ | $\varepsilon(t) \sim$ AR(1) | | $\varepsilon(t) \sim$ ARMA(1,1) | |
| | | $\rho=1, \gamma=0.3$ | $\rho=1, \gamma=0.75$ | $\rho=0.85, \gamma=0.3$ | $\rho=0.85, \gamma=0.75$ |
| 30 | 0.95 | 0.826 (1.588) | 0.844 (1.671) | 0.805 (2.148) | 0.845 (1.947) |
| | 0.99 | 0.912 (1.962) | 0.931 (2.064) | 0.901 (2.645) | 0.936 (2.405) |
| 50 | 0.95 | 0.889 (1.204) | 0.917 (1.341) | 0.874 (1.560) | 0.903 (1.314) |
| | 0.99 | 0.957 (1.470) | 0.961 (1.636) | 0.949 (1.908) | 0.956 (1.602) |
| 100 | 0.95 | 0.951 (0.902) | 0.954 (0.946) | 0.950 (0.879) | 0.962 (0.928) |
| | 0.99 | 0.991 (1.086) | 0.979 (1.139) | 0.984 (1.058) | 0.984 (1.117) |
| $\phi(u) = 2\sin(u)$ | | | | | |
| $n$ | $1-\alpha$ | $\varepsilon(t) \sim$ AR(1) | | $\varepsilon(t) \sim$ ARMA(1,1) | |
| | | $\rho=1, \gamma=0.3$ | $\rho=1, \gamma=0.75$ | $\rho=0.85, \gamma=0.3$ | $\rho=0.85, \gamma=0.75$ |
| 30 | 0.95 | 0.819 (2.020) | 0.869 (1.944) | 0.788 (1.714) | 0.850 (1.696) |
| | 0.99 | 0.920 (2.484) | 0.947 (2.393) | 0.902 (2.105) | 0.939 (2.086) |
| 50 | 0.95 | 0.904 (1.377) | 0.919 (1.308) | 0.901 (1.233) | 0.913 (1.378) |
| | 0.99 | 0.965 (1.672) | 0.966 (1.591) | 0.959 (1.498) | 0.962 (1.674) |
| 100 | 0.95 | 0.962 (1.022) | 0.955 (0.986) | 0.958 (0.932) | 0.960 (1.080) |
| | 0.99 | 0.991 (1.226) | 0.984 (1.184) | 0.990 (1.119) | 0.990 (1.295) |

Table 6

*Empirical coverage frequencies and average widths (in parentheses) of the proposed SCB with $\hat{\mathbf{\Theta}}$ obtained by the SGEE method with 1000 replications in the misspecified case of using the AR(1) correlation structure to model the true ARMA(1,1) correlation structure. The simulated data have some discrete variables in the linear covariates $\mathbf{X}$ part.*

| $n$ | $1-\alpha$ | $\phi(u) = 0.5\exp(u)$ | | $\phi(u) = 2\sin(u)$ | |
|---|---|---|---|---|---|
| | | $\rho=0.85, \gamma=0.3$ | $\rho=0.85, \gamma=0.75$ | $\rho=0.85, \gamma=0.3$ | $\rho=0.85, \gamma=0.75$ |
| 30 | 0.95 | 0.806 (1.642) | 0.845 (1.792) | 0.805 (1.577) | 0.846 (1.829) |
| | 0.99 | 0.904 (2.028) | 0.921 (2.211) | 0.900 (1.939) | 0.933 (2.249) |
| 50 | 0.95 | 0.874 (1.223) | 0.892 (3.069) | 0.906 (1.250) | 0.917 (1.336) |
| | 0.99 | 0.958 (1.492) | 0.956 (3.751) | 0.967 (1.519) | 0.960 (1.624) |
| 100 | 0.95 | 0.955 (0.876) | 0.956 (0.933) | 0.960 (0.937) | 0.956 (1.111) |
| | 0.99 | 0.984 (1.055) | 0.983 (1.124) | 0.989 (1.124) | 0.988 (1.332) |

estimation when the true underlying correlation structure is in fact ARMA(1,1). Tables 2 and 6 summarize the simulation results under this misspecification, while Table 4 reports those when the true $\mathbf{\Theta}$ was used. One can see that in all the scenarios (i) as $n$ increases, the coverage frequencies go to the nominal level and are quite close to those under the correct specification of the underlying correlation structure, supporting the theoretical results in Corollary 2.1 and Theorem 2.2; (ii) the coverage frequencies and the widths of the bands obtained by employing the estimated $\hat{\mathbf{\Theta}}$ are close to those based on the true $\mathbf{\Theta}$, confirming the oracle efficiency in Theorem 2.4. Moreover, the comparison results are again similar for the case of having some discrete covariates in $\mathbf{X}(t)$. Hence, even when the correlation structure is misspecified, the performance of the proposed SCB is still desirable in this case. The simulation results suggest that the SCBs are robust to the error correlation structure. In addition, we also considered

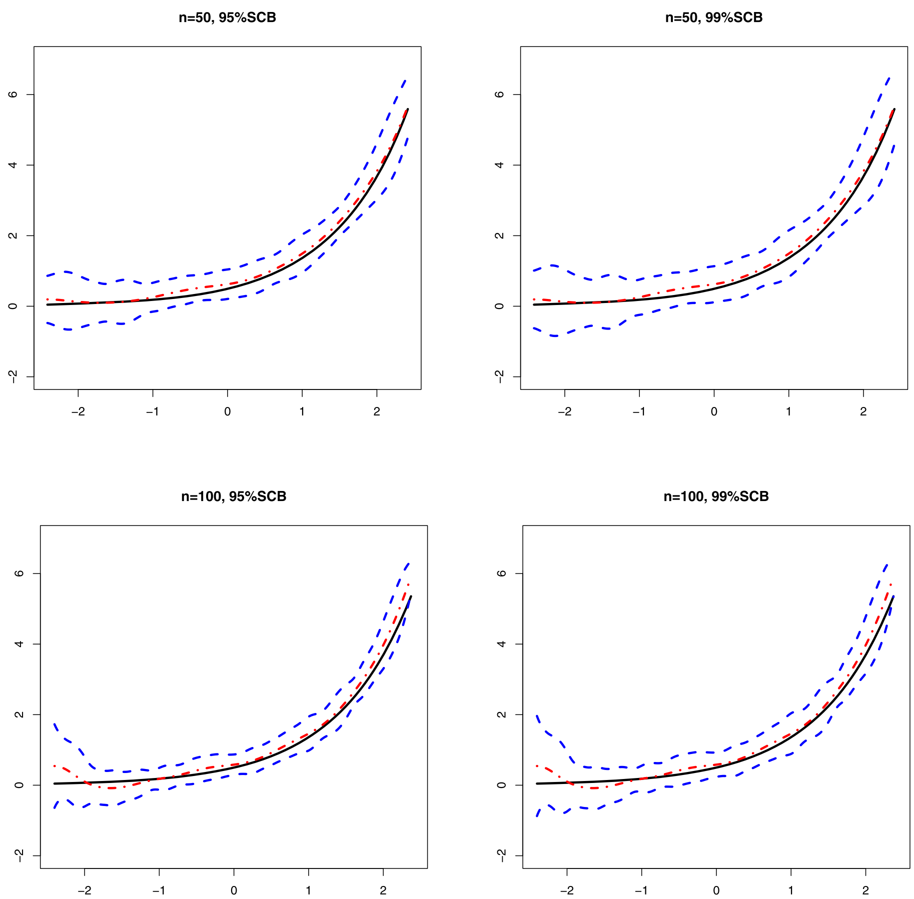


FIG 1. *Plots of the estimated link function (dash-dotted line), the true link function* $\phi(u) = 0.5\exp(u)$ *(solid line) and the SCBs (dashed line) with AR(1) error* ($\rho = 1, \gamma = 0.75$).

the SCBs with $\hat{\mathbf{\Theta}}$ obtained by the WI kernel GEE method in Lin and Carroll (2000). The results are similar but somewhat wider widths than those in Tables 1–6 employing the SGEE method in Chen et al. (2015).

To visualize the actual function estimates and the SCBs, Figures 1 and 2 were created for $\phi(u) = 0.5\exp(u)$ and $\phi(u) = 2\sin(u)$ under the correct specification of the underlying AR(1) correlation structure with $\rho = 1, \gamma = 0.75$. It can be seen that the SCBs for $n = 100$ are thinner and fit better than those for $n = 50$. Other settings yield similar results.

Responding to a reviewer's comment, we also examined the global discrepancy of $\hat{\phi}(u, \hat{\mathbf{\Theta}})$ and $\hat{\phi}\,(u, \mathbf{\Theta})$ measured by the mean squared integrated error (MISE):

$$\mathrm{MISE}(\hat{\varphi}(u)) = \int \mathrm{E}\left(\hat{\varphi}(u) - \phi\,(u)\right)^2 du$$

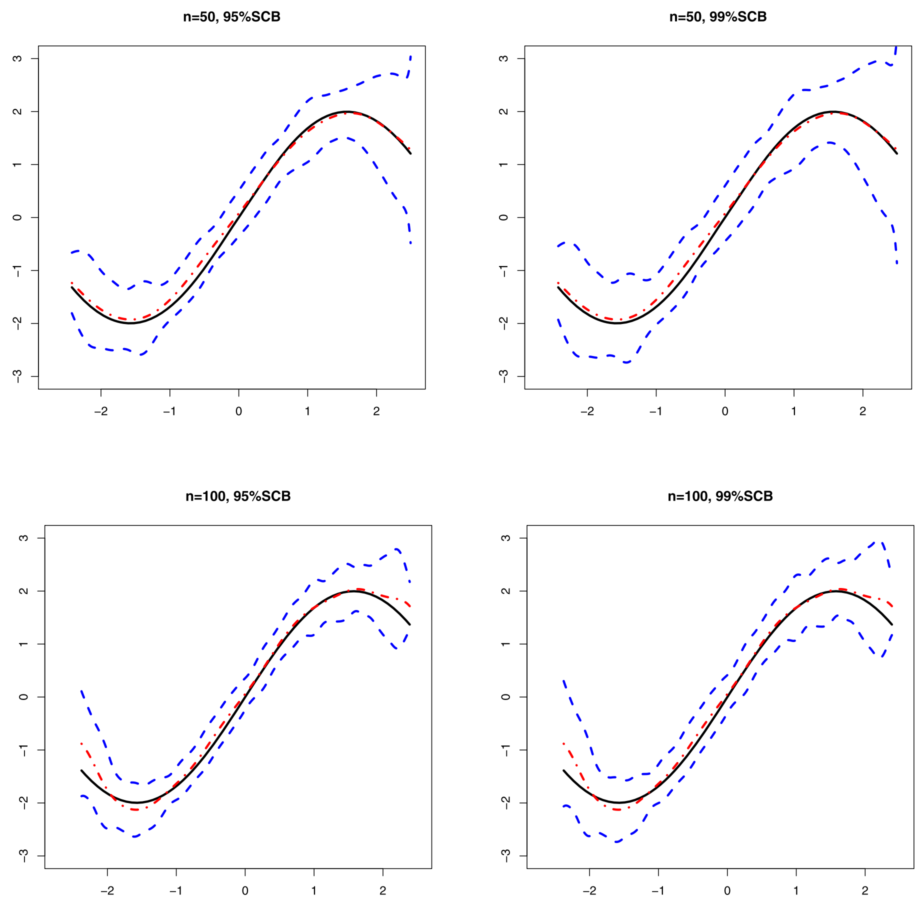


FIG 2. *Plots of the estimated link function (dash-dotted line), the true link function* $\phi(u) = 2\sin(u)$ *(solid line) and the SCBs (dashed line) with AR(1) error* ($\rho = 1, \gamma = 0.75$).

for $\hat{\varphi}(u) = \hat{\phi}(u, \hat{\mathbf{\Theta}})$ or $\hat{\phi}\,(u, \mathbf{\Theta})$, where the expectation $\mathrm{E}$ is approximated by the average of $(\hat{\varphi}(u) - \varphi\,(u))^2$ by 1000 replications and the integration $\int$ is approximated by the numerical integration over 401 points $\{\hat{a}_0 + k(\hat{b}_0 - \hat{a}_0)/400, k = 0, 1, \ldots, 400\}$. Table 7 shows $\mathrm{MISE}(\hat{\phi}(u, \hat{\mathbf{\Theta}}))$ and the ratio $\mathrm{MISE}(\hat{\phi}(u, \hat{\mathbf{\Theta}}))/\mathrm{MISE}(\hat{\phi}\,(u, \mathbf{\Theta}))$, while Table 8 shows those under the misspecification of the underlying correlation structure. The results for the data with some discrete covariates are similar and hence they are omitted. It can be seen that $\mathrm{MISE}(\hat{\phi}(u, \hat{\mathbf{\Theta}}))$ goes to 0 and the ratio $\mathrm{MISE}(\hat{\phi}(u, \hat{\mathbf{\Theta}}))/\mathrm{MISE}(\hat{\phi}\,(u, \mathbf{\Theta}))$ tends to 1 when the sample size increases regardless whether the correlation structure is correctly specified or not. In addition, Figure 3 shows the boxplots of $\sqrt{n}\sup_{k=0}^{400}|\hat{\phi}(u_k, \hat{\mathbf{\Theta}}) - \hat{\phi}\,(u_k, \mathbf{\Theta})\,|$ over 1000 replications with $u_k = \hat{a}_0 + k(\hat{b}_0 - \hat{a}_0)/400$ for the cases of $\phi(u) = 0.5\exp(u)$ (on the left panel) and $\phi(u) = 2\sin(u)$ (on the right panel)

TABLE 7
*Comparing MISE($\hat{\phi}(u,\hat{\mathbf{\Theta}})$) and MISE($\hat{\phi}(u,\mathbf{\Theta})$) for the data generated by normally distributed covariates based on 1000 replications under the correct specification of the underlying correlation structure.*

| $\phi(u)=0.5\exp(u)$ | | | | | | |
|---|---|---|---|---|---|---|
| $(\rho,\gamma)$ | MISE($\hat{\phi}(u,\hat{\mathbf{\Theta}})$) | | | MISE($\hat{\phi}(u,\hat{\mathbf{\Theta}})$)/MISE($\hat{\phi}(u,\mathbf{\Theta})$) | | |
| | $n=30$ | $n=50$ | $n=100$ | $n=30$ | $n=50$ | $n=100$ |
| (1, 0.3) | 0.0416 | 0.0275 | 0.0142 | 1.0927 | 1.0553 | 1.0159 |
| (1, 0.75) | 0.0465 | 0.0321 | 0.0170 | 1.0764 | 1.0514 | 1.0274 |
| (0.85, 0.3) | 0.0416 | 0.0271 | 0.0140 | 1.0960 | 1.0523 | 1.0092 |
| (0.85, 0.75) | 0.0467 | 0.0308 | 0.0165 | 1.0750 | 1.0458 | 1.0170 |
| $\phi(u)=2\sin(u)$ | | | | | | |
| $(\rho,\gamma)$ | MISE($\hat{\phi}(u,\hat{\mathbf{\Theta}})$) | | | MISE($\hat{\phi}(u,\hat{\mathbf{\Theta}})$)/MISE($\hat{\phi}(u,\mathbf{\Theta})$) | | |
| | $n=30$ | $n=50$ | $n=100$ | $n=30$ | $n=50$ | $n=100$ |
| (1, 0.3) | 0.0351 | 0.0229 | 0.0127 | 1.0211 | 1.0232 | 1.0112 |
| (1, 0.75) | 0.0428 | 0.0285 | 0.0155 | 1.0162 | 1.0139 | 1.0072 |
| (0.85, 0.3) | 0.0344 | 0.0229 | 0.0127 | 1.0280 | 1.0172 | 1.0166 |
| (0.85, 0.75) | 0.0422 | 0.0268 | 0.0151 | 1.0164 | 1.0076 | 1.0118 |

TABLE 8
*Comparing MISE($\hat{\phi}(u,\hat{\mathbf{\Theta}})$) and MISE($\hat{\phi}(u,\mathbf{\Theta})$) for the data generated by normally distributed covariates based on 1000 replications under misspecification of the underlying correlation structure.*

| $\phi(u)=0.5\exp(u)$ | | | | | | |
|---|---|---|---|---|---|---|
| $(\rho,\gamma)$ | MISE($\hat{\phi}(u,\hat{\mathbf{\Theta}})$) | | | MISE($\hat{\phi}(u,\hat{\mathbf{\Theta}})$)/MISE($\hat{\phi}(u,\mathbf{\Theta})$) | | |
| | $n=30$ | $n=50$ | $n=100$ | $n=30$ | $n=50$ | $n=100$ |
| (0.85, 0.3) | 0.0415 | 0.0272 | 0.0141 | 1.1074 | 1.0560 | 1.0142 |
| (0.85, 0.75) | 0.0470 | 0.0309 | 0.0165 | 1.0852 | 1.0479 | 1.0186 |
| $\phi(u)=2\sin(u)$ | | | | | | |
| $(\rho,\gamma)$ | MISE($\hat{\phi}(u,\hat{\mathbf{\Theta}})$) | | | MISE($\hat{\phi}(u,\hat{\mathbf{\Theta}})$)/MISE($\hat{\phi}(u,\mathbf{\Theta})$) | | |
| | $n=30$ | $n=50$ | $n=100$ | $n=30$ | $n=50$ | $n=100$ |
| (0.85, 0.3) | 0.0340 | 0.0227 | 0.0128 | 1.0178 | 1.0180 | 1.0168 |
| (0.85, 0.75) | 0.0424 | 0.0269 | 0.0152 | 1.0174 | 1.0110 | 1.0113 |

under the correct specification of AR(1) error ($\rho=1,\gamma=0.3$). It indicates that $\sqrt{n}\sup_{u\in[a_0,b_0]}|\hat{\phi}(u,\hat{\mathbf{\Theta}})-\hat{\phi}\,(u,\mathbf{\Theta})\,|$ is bounded in probability. All this supports the asymptotic property (oracle efficiency) of the proposed estimator $\hat{\phi}(u,\hat{\mathbf{\Theta}})$ in Theorem 2.3. We also created the boxplots in other cases and the results are similar.

Moreover, to see the performance of the parameter estimate $\hat{\mathbf{\Theta}}$, we list the averaged mean squared error (MSE) over 1000 replications. Tables 9 and 10 exhibit the MSE results for the data with normally distributed covariates. One can see that for all the cases the MSE becomes smaller and goes to 0 as $n$ increases, which is consistent with the theoretical results in Chen et al. (2015). The case with some discrete covariates has similar results.

## 6. Real data analysis

In this section we apply the proposed SCB to analyze a CD4 data set. The data set resulted from a survey of 369 men who were infected with HIV. Each man's

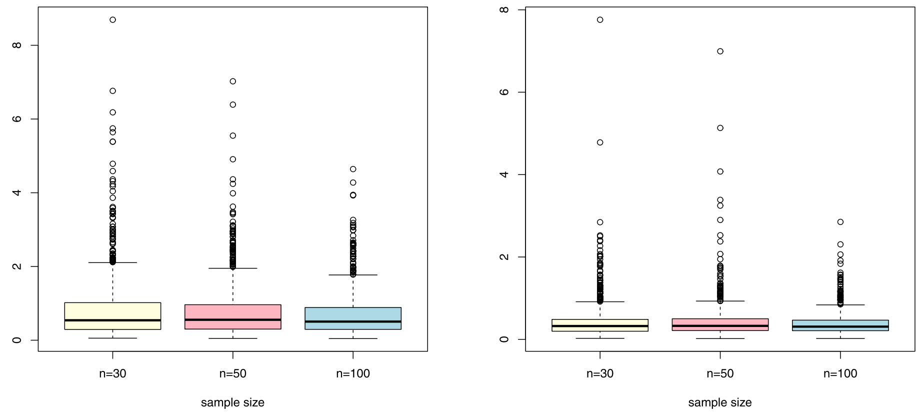


Fig 3. *Boxplots of* $\sqrt{n}\sup_{k=0}^{400}|\hat{\phi}(u_k,\hat{\boldsymbol{\Theta}})-\hat{\phi}(u_k,\boldsymbol{\Theta})|$ *over* 1000 *replications with* $u_k=\hat{a}_0+k(\hat{b}_0-\hat{a}_0)/400$ *for* $\phi(u)=0.5\exp(u)$ *(left panel) and* $\phi(u)=2\sin(u)$ *(right panel) under the correct specification of AR(1) error* $(\rho=1,\gamma=0.3)$.

Table 9

*Performance of the estimates of the parameters measured by the averaged mean squared error (MSE) for the data generated by normally distributed covariates based on* 1000 *replications under the correct specification of the underlying correlation structure (all the values are in percentage).*

| | | $\phi(u)=0.5\exp(u)$ | | | | |
|---|---|---|---|---|---|---|
| $n$ | $(\rho,\gamma)$ | MSE($\beta_1$) | MSE($\beta_2$) | MSE($\theta_1$) | MSE($\theta_2$) | MSE($\theta_3$) |
| 30 | (1, 0.3) | 0.377 | 0.380 | 0.226 | 0.363 | 0.233 |
| | (1, 0.75) | 0.281 | 0.302 | 0.219 | 0.345 | 0.215 |
| | (0.85, 0.3) | 0.386 | 0.423 | 0.244 | 0.378 | 0.236 |
| | (0.85, 0.75) | 0.273 | 0.285 | 0.200 | 0.341 | 0.215 |
| 50 | (1, 0.3) | 0.237 | 0.207 | 0.141 | 0.209 | 0.135 |
| | (1, 0.75) | 0.195 | 0.321 | 0.134 | 0.239 | 0.133 |
| | (0.85, 0.3) | 0.232 | 0.228 | 0.144 | 0.237 | 0.133 |
| | (0.85, 0.75) | 0.195 | 0.203 | 0.132 | 0.186 | 0.131 |
| 100 | (1, 0.3) | 0.110 | 0.108 | 0.065 | 0.106 | 0.067 |
| | (1, 0.75) | 0.100 | 0.097 | 0.063 | 0.098 | 0.065 |
| | (0.85, 0.3) | 0.119 | 0.112 | 0.067 | 0.116 | 0.065 |
| | (0.85, 0.75) | 0.098 | 0.099 | 0.063 | 0.094 | 0.064 |
| | | $\phi(u)=2\sin(u)$ | | | | |
| $n$ | $(\rho,\gamma)$ | MSE($\beta_1$) | MSE($\beta_2$) | MSE($\theta_1$) | MSE($\theta_2$) | MSE($\theta_3$) |
| 30 | (1, 0.3) | 0.262 | 0.274 | 0.096 | 0.152 | 0.090 |
| | (1, 0.75) | 0.196 | 0.212 | 0.072 | 0.104 | 0.065 |
| | (0.85, 0.3) | 0.295 | 0.352 | 0.107 | 0.185 | 0.163 |
| | (0.85, 0.75) | 0.218 | 0.222 | 0.074 | 0.122 | 0.076 |
| 50 | (1, 0.3) | 0.182 | 0.178 | 0.064 | 0.099 | 0.064 |
| | (1, 0.75) | 0.133 | 0.140 | 0.053 | 0.079 | 0.047 |
| | (0.85, 0.3) | 0.184 | 0.175 | 0.064 | 0.102 | 0.064 |
| | (0.85, 0.75) | 0.154 | 0.151 | 0.051 | 0.090 | 0.047 |
| 100 | (1, 0.3) | 0.083 | 0.093 | 0.029 | 0.044 | 0.029 |
| | (1, 0.75) | 0.069 | 0.073 | 0.027 | 0.038 | 0.026 |
| | (0.85, 0.3) | 0.092 | 0.088 | 0.030 | 0.048 | 0.031 |
| | (0.85, 0.75) | 0.078 | 0.079 | 0.028 | 0.051 | 0.028 |

TABLE 10

*Performance of the estimates of the parameters measured by the averaged mean squared error (MSE) for the data generated by normally distributed covariates based on* 1000 *replications under misspecification of the underlying correlation structure (all the values are in percentage).*

| | | $\phi(u) = 0.5\exp(u)$ | | | | |
|---|---|---|---|---|---|---|
| $n$ | $(\rho, \gamma)$ | MSE($\beta_1$) | MSE($\beta_2$) | MSE($\theta_1$) | MSE($\theta_2$) | MSE($\theta_3$) |
| 30 | (0.85, 0.3) | 0.370 | 0.404 | 0.245 | 0.387 | 0.231 |
| | (0.85, 0.75) | 0.299 | 0.315 | 0.221 | 0.384 | 0.228 |
| 50 | (0.85, 0.3) | 0.236 | 0.227 | 0.146 | 0.243 | 0.137 |
| | (0.85, 0.75) | 0.201 | 0.206 | 0.137 | 0.196 | 0.135 |
| 100 | (0.85, 0.3) | 0.119 | 0.108 | 0.066 | 0.106 | 0.066 |
| | (0.85, 0.75) | 0.102 | 0.104 | 0.064 | 0.097 | 0.065 |
| | | $\phi(u) = 2\sin(u)$ | | | | |
| $n$ | $(\rho, \gamma)$ | MSE($\beta_1$) | MSE($\beta_2$) | MSE($\theta_1$) | MSE($\theta_2$) | MSE($\theta_3$) |
| 30 | (0.85, 0.3) | 0.284 | 0.289 | 0.101 | 0.153 | 0.100 |
| | (0.85, 0.75) | 0.236 | 0.240 | 0.081 | 0.127 | 0.079 |
| 50 | (0.85, 0.3) | 0.175 | 0.176 | 0.061 | 0.099 | 0.062 |
| | (0.85, 0.75) | 0.155 | 0.157 | 0.052 | 0.093 | 0.049 |
| 100 | (0.85, 0.3) | 0.091 | 0.087 | 0.030 | 0.048 | 0.031 |
| | (0.85, 0.75) | 0.081 | 0.085 | 0.029 | 0.051 | 0.030 |

CD4 cell numbers were measured repeatedly from 3 years before to 6 years after seroconversion. Each person has a different number of repeated observations varying from 1 to 12, resulting in a total of 2376 CD4 observations on the following variables: CD4 cell numbers, years after seroconversion, recreational drug use (yes=1/no=0), number of sexual partners, packs of cigarettes a day, depression symptoms as measured by the Center for Epidemiologic Studies Depression (CESD) scale (larger values indicate increased depressive symptoms) and age centered around 30. The data were also studied in Zeger and Diggle (1994) and Fitzmaurice et al. (2008) by using a semiparametric model for longitudinal data.

Figure 4 shows the scatter plots of the CD4 cell numbers against the covariates: years after seroconversion, patient's age centered around 30, patient's depression symptoms, and the boxplots of CD4 cell numbers by the smoking factor. The thick solid line in the plot is the local linear fit for each pair of the data. It can be seen that the relationship between the CD4 cell numbers and the patient's depression symptoms has a horizontal linear trend except near the right end where the data are sparse, while the trend of other fitted curves is less clear. Thus in the data analysis we put the covariate of patient's depression symptoms in the linear part and the other two continuous covariates, years after seroconversion and patient's age, in the single-index part. Next, the boxplots for the discrete covariates were created. All of them are similar to those in Figure 4, which has a roughly linear trend. Therefore, the discrete variables are treated as the linear components.

Denote $Y_{ij}$ as the $j$-th observed CD4 cell number of the $i$-th person, $X_{ij,1}$ as recreational drug use, $X_{ij,2}$ as number of sexual partners, $X_{ij,3}$ as the corresponding packs of cigarettes a day, and $X_{ij,4}$ as the CESD scale. Let $Z_{ij,1}$ be the years since seroconversion and $Z_{ij,2}$ be the age centered around 30. Then

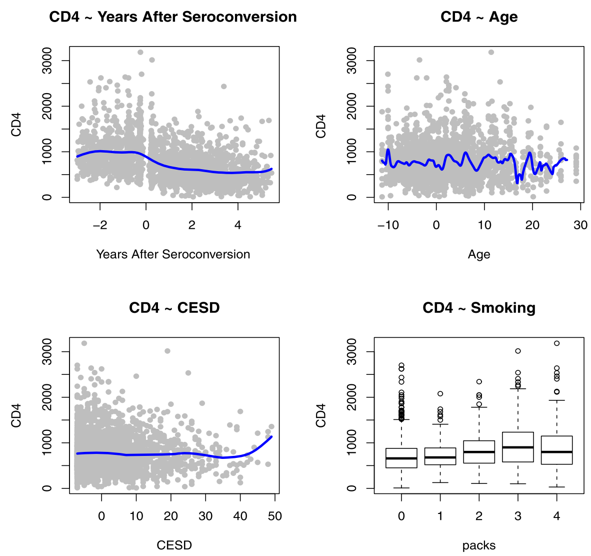


Fig 4. *Scatter plots and local linear fits (thick solid line) of the response variable CD4 cell numbers against covariates: years since seroconversion, patient's age, depression symptoms as measured by the CESD scale, and the boxplots of CD4 cell numbers by packs of cigarettes a day.*

Table 11
*Parameter estimates and the corresponding standard errors for the CD4 data.*

| Parameter | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\theta_1$ | $\theta_2$ |
|---|---|---|---|---|---|---|
| Estimate | 332.63 | −6.21 | 22.89 | −1.63 | 0.9937 | 0.111 |
| Standard error | 32.16 | 2.43 | 9.91 | 0.85 | 0.0064 | 0.024 |

the assumed partially linear single-index model for this data set is

$$Y_{ij} = \sum_{k=1}^{4} \beta_k X_{ij,k} + \phi\left(\sum_{l=1}^{2} \theta_l Z_{ij,l}\right) + \varepsilon_{ij}.$$

The resulting parameter estimates and standard errors are listed in Table 11 by using the proposed method with $\hat{\mathbf{\Theta}}$ obtained by the SGEE method assuming the AR(1) working correlation structure. The analysis on linear coefficients indicates that (i) recreational drug use and smoking are significantly positively associated with the CD4 cell number, while the number of sexual partners has a significantly negative association with the CD4 number; (ii) increased depression symptoms are marginally significantly associated with decreased CD4 cell numbers with $p$-value 0.055.

To further study the relationship between the CD4 cell number and time after seroconversion and patient's age, we consider the single-index parameter estimates and the link function estimate together. Figure 5 shows the plots of the link function estimate $\hat{\phi}(u)$ (solid), 95% and 67.7% SCBs (dashed line) for the CD4 data. The straight thick line is the fitted null hypothesis curve

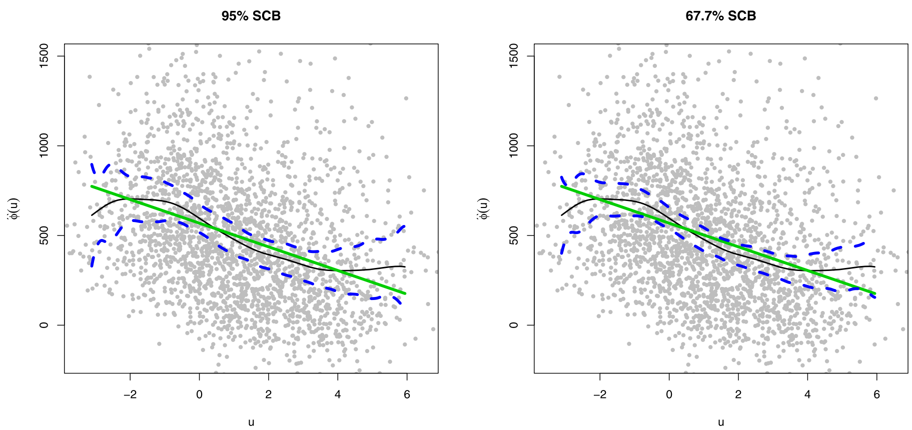


FIG 5. *Plots of the estimated link function* $\hat{\phi}(u)$ *(solid line), the* 95% *and* 67.7% *SCBs (dashed lines), and the fitted null hypothesis curve of* $\phi(u) = \gamma_0 + \gamma_1 u$ *(thick solid line).*

obtained by the least squares regression on the data points $(\sum_{l=1}^{2} \hat{\theta}_l Z_{ij,l}, Y_{ij} - \sum_{k=1}^{4} \hat{\beta}_l X_{ij,k})_{i,j=1}^{n,N_i}$. Since the lowest confidence level of SCB containing the null curve is 67.7%, one cannot reject the null hypothesis with the asymptotic $p$-value $= 0.323$, i.e., the link function might be treated as a decreasing linear function. The asymptotic $p$-value 0.323, implying that the minimum confidence level of the SCB totally containing the null curve equals to 67.7%, is approximated by inverting the limiting Gumbel distribution given in Theorem 2.4 as follows:

$$p=1-\exp\left[-2\exp\left(-a_h\left\{\max_{0\le k\le 400}\left|\sqrt{\frac{N_T h}{\hat{C}_n\left(u_k\right)}}\left[\hat{\phi}\left(u_k,\hat{\mathbf{\Theta}}\right)-(\gamma_0+\gamma_1 u_k)\right]\right|-b_h\right\}\right)\right],$$

where $u_k = \hat{a}_0 + k(\hat{b}_0 - \hat{a}_0)/400, k = 0, 1, ..., 400$, are equally spaced grid points of $[\hat{a}_0, \hat{b}_0]$ and $\gamma_0 + \gamma_1 u_k$ represents the linear least squares fit for the data $(\sum_{l=1}^{2} \hat{\theta}_l Z_{ij,l}, Y_{ij} - \sum_{k=1}^{4} \hat{\beta}_k X_{ij,k})_{i,j=1}^{n,N_i}$.

With the sign and magnitude of the parameter estimates, one concludes that the time after seroconversion and patient's age play negative roles on the CD4 cell number, i.e., patients' CD4 cell numbers tend to decrease as time goes by. Most of these conclusions are consistent with the previous studies in Zeger and Diggle (1994) and Fitzmaurice et al. (2008). However, our conclusions about the negative association of the age and the number of sexual partners with the CD4 cell number are different from theirs and might have better scientific interpretations.

## 7. Concluding remarks

In this paper, an asymptotically accurate SCB was proposed under general conditions for the link function in partially linear single-index models for possibly unbalanced longitudinal data. It was shown rigorously that the estimator for the link function with any $\sqrt{n}$-consistent estimate $\hat{\mathbf{\Theta}}$ is oracally efficient in the sense that it is asymptotically indistinguishable from the estimator with the true $\mathbf{\Theta}$ known as a prior. Using the oracle efficiency, the limiting distribution of the maximal deviation was obtained and hence an asymptotically correct SCB was constructed. Our Monte Carlo experiments with commonly encountered sample sizes support our theoretical findings.

The methodology is also suitable to the single-index model (Xia and Li (1999)) and the semiparametric partially linear model (Härdle et al. (2012)). It can be employed to make hypothesis testing in these models. On the other hand, this work has focused on the case of sparse longitudinal data. Further research problems include investigating whether similar strategies can be extended to the more complex setting of dense longitudinal data and to generalized partially linear single-index models for dense/sparse longitudinal data. It would also be interesting to study possibly superior within-subject covariance estimation and direct precision matrix estimation to reduce the potential risk of severe correlation structure misspecification and improve the efficiency of the parameter and link function estimators. We expect that future studies following these lines to yield further useful methods that not only are on solid theoretical footing but also have interesting applications.

## Acknowledgments

We would like to thank the Editor and two reviewers for their constructive comments that substantially improved an earlier version of this paper. This research was supported in part by the National Natural Science Foundation of China Award #11901521, First Class Discipline of Zhejiang–A (Zhejiang Gongshang University–Statistics) and the Simons Foundation Mathematics and Physical Sciences–Collaboration Grants for Mathematicians Program Award #499650. A part of the research was carried out while Li Cai was a Visiting Scholar at Texas A&M University with financial support from the China Scholarship Council (CSC).

## Appendix

In the following, denote $u_n \sim v_n$ as $\lim_{n\to\infty} u_n/v_n = c$, where $c$ is some nonzero constant and $[x]$ as the integer part of $x$. For any Lebesgue measurable function $\varphi(u)$ on $[a,b]$, define $\|\varphi\|_\infty = \sup_{u\in[a,b]} |\varphi(u)|$.

### *A.1. Preliminaries*

In this subsection, we present some lemmas that are needed in our theoretical development.

**Lemma A.1** (Theorem 1.2 of Bosq (1998)–the Bernstein inequality)**.** *Let $X_1, \ldots, X_n$ be independent zero-mean real-valued random variables. If there exists $c > 0$ such that*

$$\mathrm{E}\left|X_i\right|^k \leq c^{k-2} k! \,\mathrm{E}\, X_i^2 < +\infty; i = 1, 2, \ldots, n; k = 3, 4, \ldots,$$

*(Cramér's conditions), then*

$$P\left\{\left|\sum_{i=1}^{n} X_i\right| > t\right\} \leq 2\exp\left\{-\frac{t^2}{4\sum_{i=1}^{n} \mathrm{E}\, X_i^2 + 2ct}\right\}, t > 0.$$

**Lemma A.2.** *Under Assumptions (A2)–(A5), for $l = 0, 1$, as $n \to \infty$,*

$$\sup_{u\in[a_0,b_0]}\left|N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T\theta - u\right)\left(\mathbf{Z}_{ij}^T\theta - u\right)^l \varepsilon_{ij}\right| = O_p\left(h^l \tau_n\right).$$

**Proof of Lemma A.2.** Let $T_n = n^{\delta}$ with $1/\left(2+m\right) < \delta < 3/8$, where $m > 2/3$ is in Assumption (A2). By $nh^5 \log n \to 0$ in Assumption (A5) and $1/\left(2+m\right) < \delta$, one has that $T_n^{-(1+m)} \ll \tau_n$ and $\sum_{n=1}^{\infty} T_n^{-(2+m)} < +\infty$, while by $\delta < 3/8$ and $nh^4 \to \infty$ in Assumption (A5), one has that $T_n\tau_n \to 0$.

First, we truncate the noise $\varepsilon_{ij}$ by $T_n$:

$$\varepsilon_{ij} = \varepsilon_{ij,1} + \varepsilon_{ij,2} + \mu_{ij},$$

where

$$\varepsilon_{ij,1} = \varepsilon_{ij} I\left\{\left|\varepsilon_{ij}\right| > T_n\right\}, \ \varepsilon_{ij,2} = \varepsilon_{ij} I\left\{\left|\varepsilon_{ij}\right| \leq T_n\right\} - \mu_{ij}, \ \mu_{ij} = \mathrm{E}\left[\varepsilon_{ij} I\left\{\left|\varepsilon_{ij}\right| \leq T_n\right\}\right].$$

Therefore,

$$\begin{aligned}
&N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T\theta - u\right)\left(\mathbf{Z}_{ij}^T\theta - u\right)^l \varepsilon_{ij} \\
&= N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T\theta - u\right)\left(\mathbf{Z}_{ij}^T\theta - u\right)^l \varepsilon_{ij,1} \\
&\quad + N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T\theta - u\right)\left(\mathbf{Z}_{ij}^T\theta - u\right)^l \varepsilon_{ij,2} \\
&\quad + N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T\theta - u\right)\left(\mathbf{Z}_{ij}^T\theta - u\right)^l \mu_{ij}. \qquad \text{(A.1)}
\end{aligned}$$

Note that

$$\mu_{ij} = \mathrm{E}\left[\varepsilon_{ij} I\left\{\left|\varepsilon_{ij}\right| \leq T_n\right\}\right] = \mathrm{E}\,\varepsilon_{ij} - \mathrm{E}\left[\varepsilon_{ij} I\left\{\left|\varepsilon_{ij}\right| > T_n\right\}\right] = -\,\mathrm{E}\left[\varepsilon_{ij} I\left\{\left|\varepsilon_{ij}\right| > T_n\right\}\right],$$

which implies that

$$\left|\mu_{ij}\right| \leq M_{\varepsilon} T_n^{-(1+m)},$$

where $M_\varepsilon$ is given in Assumption (A2). According to Proposition 3.1 which will be shown later, one has that

$$\begin{aligned}&\left|N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}K_h\left(\mathbf{Z}_{ij}^T\theta-u\right)\left(\mathbf{Z}_{ij}^T\theta-u\right)^l\mu_{ij}\right|\\&\leq M_\varepsilon T_n^{-(1+m)}h^l N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}K_h\left(\mathbf{Z}_{ij}^T\theta-u\right)\\&=U_p\left(T_n^{-(1+m)}h^l\right)=u_p\left(h^l\tau_n\right).\end{aligned}\tag{A.2}$$

Next, since

$$\sum_{n=1}^{\infty}P\left\{\max_{1\leq j\leq N_n}|\varepsilon_{nj}|>T_n\right\}\leq\sum_{n=1}^{\infty}\sum_{j=1}^{N_n}P\left\{|\varepsilon_{nj}|>T_n\right\}\leq M_\varepsilon\sum_{n=1}^{\infty}N_nT_n^{-(2+m)}<+\infty,$$

the Borel-Cantelli Lemma implies that

$$P\left\{\omega|\text{ there exists }n_1(\omega),\ \max_{1\leq j\leq N_n}|\varepsilon_{nj}|\leq T_n,\text{ for all }n>n_1(\omega)\right\}=1.$$

Hence, it is easily seen that

$$P\left\{\omega|\text{ there exists }n_2(\omega),\ \max_{1\leq i\leq n}\max_{1\leq j\leq N_i}|\varepsilon_{ij}|\leq T_n,\text{ for all }n>n_2(\omega)\right\}=1,$$

which implies that $P\{\omega|\text{ there exists }n_2(\omega)\text{ such that }\varepsilon_{ij,1}=0, 1\leq j\leq N_i,$ $1\leq i\leq n,\text{ for }n>n_2(\omega)\}=1$. Thus,

$$N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}K_h\left(\mathbf{Z}_{ij}^T\theta-u\right)\left(\mathbf{Z}_{ij}^T\theta-u\right)^l\varepsilon_{ij,1}=0\ a.s.\tag{A.3}$$

We next consider $N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}K_h\left(\mathbf{Z}_{ij}^T\theta-u\right)\left(\mathbf{Z}_{ij}^T\theta-u\right)^l\varepsilon_{ij,2}$. It is clear that $\mathrm{E}\,\varepsilon_{ij,2}=0$,

$$\begin{aligned}\mathrm{E}\left(\varepsilon_{ij,2}\right)^2&=\mathrm{E}\left\{\varepsilon_{ij}I\left(|\varepsilon_{ij}|\leq T_n\right)\right\}^2-\left(\mu_{ij}\right)^2\\&=\mathrm{E}\,\varepsilon_{ij}^2-\mathrm{E}\left\{\varepsilon_{ij}^2I\left(|\varepsilon_{ij}|>T_n\right)\right\}-\left(\mu_{ij}\right)^2\\&=\sigma_i^{jj}+O\left(T_n^{-m}+T_n^{-2(1+m)}\right),\end{aligned}$$

and

$$\begin{aligned}&\mathrm{E}\left(\varepsilon_{ij,2}\varepsilon_{ij',2}\right)\leq\left\{\mathrm{E}\left(\varepsilon_{ij,2}\right)^2\mathrm{E}\left(\varepsilon_{ij',2}\right)^2\right\}^{1/2}\\&=\left\{\sigma_i^{jj}+O\left(T_n^{-m}+T_n^{-2(1+m)}\right)\right\}^{1/2}\left\{\sigma_i^{j'j'}+O\left(T_n^{-m}+T_n^{-2(1+m)}\right)\right\}^{1/2}\\&=\left(\sigma_i^{jj}\sigma_i^{j'j'}\right)^{1/2}+O\left(T_n^{-m}+T_n^{-2(1+m)}\right).\end{aligned}$$

By Assumption (A2), one has that $\mathrm{E}\left(\varepsilon_{ij,2}\varepsilon_{ij',2}\right) < C_\sigma$ for $n$ large enough. Denote $\eta_{i,l}\left(u\right) = N_T^{-1}\sum_{j=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T\theta - u\right)\left(\mathbf{Z}_{ij}^T\theta - u\right)^l \varepsilon_{ij,2}$. One then has that

$$N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T\theta - u\right)\left(\mathbf{Z}_{ij}^T\theta - u\right)^l \varepsilon_{ij,2} = \sum_{i=1}^{n}\eta_{i,l}\left(u\right).$$

It is clear that $\mathrm{E}\,\eta_{i,l}\left(u\right) = 0$ and

$$\begin{aligned}
\mathrm{E}\left\{\eta_{i,l}\left(u\right)\right\}^2 &= N_T^{-2}\sum_{j=1}^{N_i}\mathrm{E}\left\{K_h^2\left(\mathbf{Z}_{ij}^T\theta - u\right)\left(\mathbf{Z}_{ij}^T\theta - u\right)^{2l}\right\}\mathrm{E}\,\varepsilon_{ij,2}^2 \\
&+N_T^{-2}\sum_{j=1}^{N_i}\sum_{j'\neq j}^{N_i}\mathrm{E}\left\{K_h\left(\mathbf{Z}_{ij}^T\theta - u\right)K_h\left(\mathbf{Z}_{ij'}^T\theta - u\right)\left(\mathbf{Z}_{ij}^T\theta - u\right)^l\left(\mathbf{Z}_{ij'}^T\theta - u\right)^l\right\} \\
&\quad\times\mathrm{E}\left(\varepsilon_{ij,2}\varepsilon_{ij',2}\right) \\
&= N_T^{-2}h^{2l-1}f\left(u\right)\int K^2\left(v\right)v^{2l}dv\sum_{j=1}^{N_i}\sigma_i^{jj} + u_p\left(n^{-2}h^{2l-1}\right).
\end{aligned}$$

Thus $\mathrm{E}\sum_{i=1}^{n}\eta_{i,l}\left(u\right) = 0$ and

$$\mathrm{E}\left\{\sum_{i=1}^{n}\eta_{i,l}\left(u\right)\right\}^2 = N_T^{-2}h^{2l-1}f\left(u\right)\int K^2\left(v\right)dv\sum_{i=1}^{n}\sum_{j=1}^{N_i}\sigma_i^{jj} + u_p\left(n^{-1}h^{2l-1}\right),$$

which implies that $\mathrm{E}\left\{\sum_{i=1}^{n}\eta_{i,l}\left(u\right)\right\}^2 \leq Cn^{-1}h^{2l-1}$ for some $C > 0$. Furthermore, for $k \geq 3$, one has that

$$\begin{aligned}
\left|\eta_{i,l}\left(u\right)\right|^{k-2} &= \left|N_T^{-1}\sum_{j=1}^{N_i}K_h\left(\mathbf{Z}_{ij}^T\theta - u\right)\left(\mathbf{Z}_{ij}^T\theta - u\right)^l\varepsilon_{ij,2}\right|^{k-2} \\
&\leq \left\{N_T^{-1}N_ih^{l-1}\left\|K\right\|_\infty 2T_n\right\}^{k-2}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathrm{E}\left|\eta_{i,l}\left(u\right)\right|^k &= \mathrm{E}\left\{\left|\eta_{i,l}\left(u\right)\right|^{k-2}\left|\eta_{i,l}\left(u\right)\right|^2\right\} \\
&\leq \left\{N_T^{-1}N_ih^{l-1}\left\|K\right\|_\infty 2T_n\right\}^{(k-2)}\mathrm{E}\left|\eta_{i,l}\left(u\right)\right|^2,
\end{aligned}$$

which means that $\eta_{i,l}\left(u\right), 1 \leq i \leq n$, satisfy the Cramér conditions in Lemma A.1 with $c = 2N_T^{-1}C_Nh^{l-1}\left\|K\right\|_\infty T_n$, $C_N$ being an upper bound of all $N_i$. Then for any given large enough $\kappa > 0$ and sufficiently large $n$,

$$\begin{aligned}
&P\left\{\left|\sum_{i=1}^{n}\eta_{i,l}\left(u\right)\right| > \kappa h^l\tau_n\right\} \\
&\quad\leq 2\exp\left\{-\frac{\kappa^2h^{2l}\tau_n^2}{4\sum_{i=1}^{n}\mathrm{E}\left\{\eta_{i,l}\left(u\right)\right\}^2 + 4\kappa N_T^{-1}C_Nh^{l-1}\left\|K\right\|_\infty T_nh^l\tau_n}\right\}
\end{aligned}$$

$$= 2\exp\left\{-\frac{\kappa^2\log n}{4nh^{1-2l}\sum_{i=1}^{n}\mathrm{E}\left\{\eta_{i,l}(u)\right\}^2 + 4nN_T^{-1}\kappa C_N\|K\|_\infty T_n\tau_n}\right\}$$
$$\leq 2n^{-4}, \tag{A.4}$$

which holds since $nh^{1-2l}\sum_{i=1}^{n}\mathrm{E}\left\{\eta_{i,l}(u)\right\}^2$ is bounded and $T_n\tau_n \to 0$.

To bound $\sum_{i=1}^{n}\eta_{i,l}(u)$ uniformly for all $u \in [a_0, b_0]$, we discretize $[a_0, b_0]$ by equally spaced points $a_0 = u_0 \leq u_1 \leq \cdots \leq u_{M_n} = b_0$ with $M_n = n^2$. By applying (A.4), one obtains that

$$\sum_{n=1}^{\infty} P\left\{\max_{0\leq k\leq M_n}\left|\sum_{i=1}^{n}\eta_{i,l}(u_k)\right| > 2h^l\tau_n\right\} \leq \sum_{n=1}^{\infty}\sum_{k=0}^{M_n} 2n^{-4} < +\infty.$$

The Borel-Cantelli Lemma implies that

$$\max_{0\leq k\leq M_n}\left|\sum_{i=1}^{n}\eta_{i,l}(u_k)\right| = O_{a.s.}\left(h^l\tau_n\right).$$

Therefore,

$$\sup_{u\in[a_0,b_0]}\left|N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}K_h\left(\mathbf{Z}_{ij}^T\theta - u\right)\left(\mathbf{Z}_{ij}^T\theta - u\right)^l\varepsilon_{ij,2}\right| = \sup_{u\in[a_0,b_0]}\left|\sum_{i=1}^{n}\eta_{i,l}(u)\right|$$
$$\leq \max_{1\leq k\leq M_n}\left|\sum_{i=1}^{n}\eta_{i,l}(u_k)\right| + \max_{1\leq k\leq M_n-1}\sup_{u\in[u_k,u_{k+1}]}\left|\sum_{i=1}^{n}\eta_{i,l}(u) - \sum_{i=1}^{n}\eta_{i,l}(u_k)\right|$$
$$= O_{a.s.}\left(h^l\tau_n\right) + \max_{k=1}^{M_n-1}\sup_{u\in[u_k,u_{k+1}]}$$
$$\left|N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\left\{K_h\left(\mathbf{Z}_{ij}^T\theta - u\right)\left(\mathbf{Z}_{ij}^T\theta - u\right)^l - K_h\left(\mathbf{Z}_{ij}^T\theta - u_k\right)\left(\mathbf{Z}_{ij}^T\theta - u_k\right)^l\right\}\varepsilon_{ij,2}\right|$$
$$\leq O_{a.s.}\left(h^l\tau_n\right) + 2nN_T^{-1}C_N h^{l-2}\left\|K^{(1)}\right\|_\infty T_n(b_0 - a_0)M_n^{-1}$$
$$= O_{a.s.}\left(h^l\tau_n\right), \tag{A.5}$$

which holds since $h^{l-2}T_nM_n^{-1} \ll h^l\tau_n$. Putting (A.1), (A.2), (A.3) and (A.5) together, one concludes the result.

**Lemma A.3.** *Under Assumptions (A1)–(A5), for $l = 0, 1$, as $n \to \infty$, one has*

$$\sup_{u\in[a_0,b_0]}\left|N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}K_h\left(\mathbf{Z}_{ij}^T\hat{\theta} - u\right)\left(\mathbf{Z}_{ij}^T\hat{\theta} - u\right)^l\hat{\varepsilon}_{ij}\right.$$
$$\left. - N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}K_h\left(\mathbf{Z}_{ij}^T\theta - u\right)\left(\mathbf{Z}_{ij}^T\theta - u\right)^l\varepsilon_{ij}\right| = O_p\left(n^{-1/2}\right),$$

*where* $\hat{\varepsilon}_{ij} = Y_{ij} - \mathbf{X}_{ij}^T\hat{\beta} - \phi\left(\mathbf{Z}_{ij}^T\hat{\theta}\right)$.

**Proof of Lemma A.3.** Notice that

$$\begin{aligned}
& N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T\hat{\theta}-u\right)\left(\mathbf{Z}_{ij}^T\hat{\theta}-u\right)^l\hat{\varepsilon}_{ij} \\
&\qquad - N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T\hat{\theta}-u\right)\left(\mathbf{Z}_{ij}^T\hat{\theta}-u\right)^l\varepsilon_{ij} \\
&= N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T\hat{\theta}-u\right)\left(\mathbf{Z}_{ij}^T\hat{\theta}-u\right)^l\left\{\mathbf{X}_{ij}^T\beta-\mathbf{X}_{ij}^T\hat{\beta}+\phi\left(\mathbf{Z}_{ij}^T\theta\right)-\phi\left(\mathbf{Z}_{ij}^T\hat{\theta}\right)\right\} \\
&= N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T\hat{\theta}-u\right)\left(\mathbf{Z}_{ij}^T\hat{\theta}-u\right)^l\mathbf{X}_{ij}^T\left(\beta-\hat{\beta}\right) \\
&\qquad + N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T\hat{\theta}-u\right)\left(\mathbf{Z}_{ij}^T\hat{\theta}-u\right)^l\left\{\phi\left(\mathbf{Z}_{ij}^T\theta\right)-\phi\left(\mathbf{Z}_{ij}^T\hat{\theta}\right)\right\}.
\end{aligned} \tag{A.6}$$

Applying the first order Tayor expansion, one has that

$$\begin{aligned}
& N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T\hat{\theta}-u\right)\left(\mathbf{Z}_{ij}^T\hat{\theta}-u\right)^l\mathbf{X}_{ij}^T\left(\beta-\hat{\beta}\right) \\
&= N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} h^{-1}K\left(\frac{\mathbf{Z}_{ij}^T\theta-u}{h}\right)\left(\mathbf{Z}_{ij}^T\theta-u\right)^l\mathbf{X}_{ij}^T\left(\beta-\hat{\beta}\right) \\
&+ N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} h^{-1}\left\{h^{-1}K^{(1)}\left(\frac{\mathbf{Z}_{ij}^T\theta^*-u}{h}\right)\left(\mathbf{Z}_{ij}^T\theta^*-u\right)^l\right. \\
&\left.+lK\left(\frac{\mathbf{Z}_{ij}^T\theta^*-u}{h}\right)\left(\mathbf{Z}_{ij}^T\theta^*-u\right)^{l-1}\right\}\mathbf{X}_{ij}^T\left(\beta-\hat{\beta}\right)\mathbf{Z}_{ij}^T\left(\hat{\theta}-\theta\right),
\end{aligned} \tag{A.7}$$

where $\left(\mathbf{Z}_{ij}^T\theta^*-u\right)/h$ is some value between $\left(\mathbf{Z}_{ij}^T\hat{\theta}-u\right)/h$ and $\left(\mathbf{Z}_{ij}^T\theta-u\right)/h$. Clearly,

$$\begin{aligned}
\sup_{u\in[a_0,b_0]} & \left| N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} h^{-1}\left\{h^{-1}K^{(1)}\left(\frac{\mathbf{Z}_{ij}^T\theta^*-u}{h}\right)\left(\mathbf{Z}_{ij}^T\theta^*-u\right)^l\right.\right. \\
&\left.\left. +lK\left(\frac{\mathbf{Z}_{ij}^T\theta^*-u}{h}\right)\left(\mathbf{Z}_{ij}^T\theta^*-u\right)^{l-1}\right\}\mathbf{X}_{ij}^T\left(\beta-\hat{\beta}\right)\mathbf{Z}_{ij}^T\left(\hat{\theta}-\theta\right)\right| \\
&\le N_T^{-1}h^{l-2}\left\{\left\|K^{(1)}\right\|_\infty + l\left\|K\right\|_\infty\right\}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\left|\mathbf{X}_{ij}^T\left(\beta-\hat{\beta}\right)\mathbf{Z}_{ij}^T\left(\hat{\theta}-\theta\right)\right| \\
&= O_p\left(n^{-1}h^{l-2}\right) = O_p\left(n^{-1/2}\right),
\end{aligned} \tag{A.8}$$

since $\mathrm{E}\left|Z_{ij,d}\right|^{2+m_0} < M_Z$ and $\mathrm{E}\left|X_{ij,l}\right|^{2+m_1} < M_X$, $m_0, m_1 > 2/3$ in Assumption (A2).

We next consider $N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} h^{-1}K\left(\frac{\mathbf{Z}_{ij}^T\theta-u}{h}\right)\left(\mathbf{Z}_{ij}^T\theta-u\right)^l \mathbf{X}_{ij}^T\left(\beta-\hat{\beta}\right)$. Let $D_n = n^{\alpha}$ with $1/\left(2+m_1\right) < \alpha < 3/8$. The $n^{-1}h^{-4} \ll 1$ in Assumption (A5) and $\alpha < 3/8$ implies that $D_n\tau_n \to 0$, while $1/\left(2+m_1\right) < \alpha$ implies that $\sum\limits_{n=1}^{\infty} D_n^{-(2+m_1)} < +\infty$. We first truncate $X_{ij,k}, k = 1, 2, ..., p$, by $D_n$ with

$$X_{ij,k} = X_{ij,k}^{(1)} + X_{ij,k}^{(2)},$$

where $X_{ij,k}^{(1)} = X_{ij,k}I\left\{\left|X_{ij,k}\right| > D_n\right\}, X_{ij,k}^{(2)} = X_{ij,k}I\left\{\left|X_{ij,k}\right| \le D_n\right\}$. Notice that

$$\begin{aligned}\sum_{n=1}^{\infty} P\left\{\max_{1\le j\le N_n}\left|X_{nj,k}\right| > D_n\right\} &\le \sum_{n=1}^{\infty}\sum_{j=1}^{N_n} P\left\{\left|X_{nj,k}\right| > D_n\right\} \\ &\le \sum_{n=1}^{\infty}\sum_{j=1}^{N_n} D_n^{-(2+m_1)}\,\mathrm{E}\left|X_{nj,k}\right|^{(2+m_1)} \\ &\le M_X\sum_{n=1}^{\infty} N_n D_n^{-(2+m_1)} < \infty,\end{aligned}$$

which implies that

$$P\left\{\omega|\text{there exist a } n_0(\omega) \text{ such that } \max_{1\le j\le N_n}\left|X_{nj,k}\right| \le D_n \text{ for } n > n_0(\omega)\right\} = 1.$$

Then one has that

$$\begin{aligned}P\Bigg\{\omega|\text{there exist a } n_1(\omega) \text{ such that } \max_{1\le i\le n}\max_{1\le j\le N_i}\left|X_{ij,k}\right| \le D_n \text{ for } n > n_1(\omega)\Bigg\} \\ = 1,\end{aligned}$$

and hence

$$\begin{aligned}P\Big\{\omega|\text{there exist a } n_1(\omega) \text{ such that } X_{ij,k}^{(1)} = 0, 1\le j\le N_i, 1\le i\le n, \\ \text{for } n > n_1(\omega)\Big\} = 1.\end{aligned}$$

Therefore,

$$\sum_{i=1}^{n} N_T^{-1}\sum_{j=1}^{N_i} h^{-1}K\left(\frac{\mathbf{Z}_{ij}^T\theta-u}{h}\right)\left(\mathbf{Z}_{ij}^T\theta-u\right)^l X_{ij,k}^{(1)} = 0 \ a.s.$$

Next, define $\eta_{i,l,k}(u) = N_T^{-1}\sum_{j=1}^{N_i} h^{-1}K\left(\frac{\mathbf{Z}_{ij}^T\theta-u}{h}\right)\left(\mathbf{Z}_{ij}^T\theta-u\right)^l X_{ij,k}^{(2)}$. Thus,

$$\left|\mathrm{E}\,\eta_{i,l,k}(u)\right| \le N_T^{-1}\sum_{j=1}^{N_i} h^{-1}\,\mathrm{E}\left|K\left(\frac{\mathbf{Z}_{ij}^T\theta-u}{h}\right)\left(\mathbf{Z}_{ij}^T\theta-u\right)^l X_{ij,k}\right|$$

$$\begin{aligned}
&= N_T^{-1}\sum_{j=1}^{N_i} h^{-1}\int K\left(\frac{s-u}{h}\right)|s-u|^l\,|t|\, f_{\mathbf{Z}_{ij}^T\theta|X_{ij,k}}(s|t)\, f_{X_{ij,k}}(t)\,dsdt \\
&= N_T^{-1}\sum_{j=1}^{N_i}\int K(v)\, h^l\,|v|^l\,|t|\, f_{\mathbf{Z}_{ij}^T\theta|X_{ij,k}}(u+hv|t)\, f_{X_{ij,k}}(t)\,dvdt \\
&= N_T^{-1}\sum_{j=1}^{N_i} h^l\int K(v)\,|v|^l\,|t|\times \\
&\qquad \left\{ f_{\mathbf{Z}_{ij}^T\theta|X_{ij,k}}(u|t) + \frac{\partial f_{\mathbf{Z}_{ij}^T\theta|X_{ij,k}}(u|t)}{\partial u} hv + u(h)\right\} f_{X_{ij,k}}(t) dvdt,
\end{aligned}$$

where $f_{\mathbf{Z}_{ij}^T\theta|X_{ij,k}}(u|t)$ and $f_{X_{ij,k}}(t)$ represent the density function of $\mathbf{Z}_{ij}^T\theta$ given $X_{ij,k}$ and the density function of $X_{ij,k}$, respectively. Since $\mathrm{E}\left[|X_{ij,k}|^2\,|\mathbf{Z}_{ij}^T\theta = t\right] \le G_1$ in Assumption (A2), one then has that $|\mathrm{E}\,\eta_{i,l,k}(u)| \le C_0 n^{-1} h^l$ for some constant $C_0$. Next, $\mathrm{E}\,\eta_{i,l,k}^2(u)$ equals

$$\begin{aligned}
& N_T^{-2}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_i} h^{-2} \\
&\times \mathrm{E}\left\{ K\left(\frac{\mathbf{Z}_{ij}^T\theta - u}{h}\right)\left(\mathbf{Z}_{ij}^T\theta - u\right)^l X_{ij,k}^{(2)} K\left(\frac{\mathbf{Z}_{ij'}^T\theta - u}{h}\right)\left(\mathbf{Z}_{ij'}^T\theta - u\right)^l X_{ij',k}^{(2)}\right\} \\
&\le N_T^{-2}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_i} h^{-2} \\
&\times \mathrm{E}\left| K\left(\frac{\mathbf{Z}_{ij}^T\theta - u}{h}\right)\left(\mathbf{Z}_{ij}^T\theta - u\right)^l X_{ij,k} K\left(\frac{\mathbf{Z}_{ij'}^T\theta - u}{h}\right)\left(\mathbf{Z}_{ij'}^T\theta - u\right)^l X_{ij',k}\right| \\
&= N_T^{-2}\sum_{j=1}^{N_i} h^{-2}\,\mathrm{E}\left\{ K\left(\frac{\mathbf{Z}_{ij}^T\theta - u}{h}\right)\left(\mathbf{Z}_{ij}^T\theta - u\right)^l X_{ij,k}\right\}^2 \\
&+ N_T^{-2}\sum_{j=1}^{N_i}\sum_{j'\neq j}^{N_i} h^{-2} \\
&\times \mathrm{E}\left| K\left(\frac{\mathbf{Z}_{ij}^T\theta - u}{h}\right)\left(\mathbf{Z}_{ij}^T\theta - u\right)^l X_{ij,k} K\left(\frac{\mathbf{Z}_{ij'}^T\theta - u}{h}\right)\left(\mathbf{Z}_{ij'}^T\theta - u\right)^l X_{ij',k}\right|.
\end{aligned}$$

By $\mathrm{E}\left[|X_{ij,k}|^2\,|\mathbf{Z}_{ij}^T\theta = t\right] \le G_1$ in Assumption (A2), one has that $\mathrm{E}\,\eta_{i,l,k}^2(u) \le C_1 n^{-2} h^{l-1}$ with $C_1$ being some positive constant. Moreover, notice that

$$|\eta_{i,l,k}(u)| \le N_T^{-1} h^{l-1}\,\|K\|_\infty\, 2D_n N_i \le C_2 n^{-1} h^{l-1} D_n,$$

for some $C_2 > 0$. Let $\tilde{\eta}_{i,l,k}(u) = \eta_{i,l,k}(u) - \mathrm{E}\,\eta_{i,l,k}(u)$. Thus $\mathrm{E}\,\tilde{\eta}_{i,l,k}(u) = 0$ and $\mathrm{E}\,\tilde{\eta}_{i,l,k}^2(u) = \mathrm{E}\,\eta_{i,l,k}^2(u) - (\mathrm{E}\,\eta_{i,l,k}(u))^2 \le C_1 n^{-2} h^{l-1}$. Since $|\tilde{\eta}_{i,l,k}(u)| \le$

$|\eta_{i,l,k}(u)| + |\mathrm{E}\,\eta_{i,l,k}(u)| \leq 2C_2 n^{-1} h^{l-1} D_n,$

$$\mathrm{E}\,|\tilde{\eta}_{i,l,k}(u)|^m \leq \left(2C_2 n^{-1} h^{l-1} D_n\right)^{m-2} \mathrm{E}\,|\tilde{\eta}_{i,l,k}(u)|^2$$

for $m > 2$. Thus $\tilde{\eta}_{i,l,k}(u)$ satisfies the Cramér conditions in Lemma A.1 with $c = 2C_2 n^{-1} h^{l-1} D_n$. Employing the Bernstein inequality in Lemma A.1, the Borel-Cantelli Lemma and discretization method again, one immediately obtains that

$$\sup_{u\in[a_0,b_0]} \left|\sum_{i=1}^{n} \tilde{\eta}_{i,l,k}(u)\right| = O_p\left(h^{l/2}\tau_n\right),$$

and hence

$$\begin{aligned}\sup_{u\in[a_0,b_0]} \left|\sum_{i=1}^{n} \eta_{i,l,k}(u)\right| &\leq \sup_{u\in[a_0,b_0]} \left|\sum_{i=1}^{n} \tilde{\eta}_{i,l,k}(u)\right| + \sup_{u\in[a_0,b_0]} \left|\sum_{i=1}^{n} \mathrm{E}\,\eta_{i,l,k}(u)\right| \\ &= O_p\left(h^l \tau_n\right) + O_p\left(h^l\right).\end{aligned}$$

Thus,

$$\begin{aligned}&\sup_{u\in[a_0,b_0]} \left|N_T^{-1} \sum_{i=1}^{n}\sum_{j=1}^{N_i} h^{-1} K\left(\frac{\mathbf{Z}_{ij}^T\theta - u}{h}\right) (\mathbf{Z}_{ij}^T\theta - u)^l \mathbf{X}_{ij}^T \left(\beta - \hat{\beta}\right)\right| \\ &= \sup_{u\in[a_0,b_0]} \left|\left(\sum_{i=1}^{n} \eta_{i,l,1}(u), \ldots, \sum_{i=1}^{n} \eta_{i,l,p}(u)\right)\left(\beta - \hat{\beta}\right)\right| \\ &\leq \sup_{u\in[a_0,b_0],1\leq k\leq p} \left|\sum_{i=1}^{n} \eta_{i,l,k}(u)\right| \sqrt{p}\left\|\beta - \hat{\beta}\right\| \\ &= \left[O_p\left(h^{l/2}\tau_n\right) + O_p\left(h^l\right)\right] O_p\left(n^{-1/2}\right) = O_p\left(n^{-1/2}\right).\end{aligned} \tag{A.9}$$

Meanwhile, it is easy to verify that

$$\begin{aligned}\sup_{u\in[a_0,b_0]} \left|N_T^{-1} \sum_{i=1}^{n}\sum_{j=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T\hat{\theta} - u\right)\left(\mathbf{Z}_{ij}^T\hat{\theta} - u\right)^l \left\{\phi\left(\mathbf{Z}_{ij}^T\theta\right) - \phi\left(\mathbf{Z}_{ij}^T\hat{\theta}\right)\right\}\right| \\ = O_p\left(n^{-1/2}\right),\end{aligned}$$

which together with (A.6), (A.7), (A.8) and (A.9) concludes that

$$\begin{aligned}&\sup_{u\in[a_0,b_0]} \left|N_T^{-1} \sum_{i=1}^{n}\sum_{j=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T\hat{\theta} - u\right)\left(\mathbf{Z}_{ij}^T\hat{\theta} - u\right)^l \hat{\varepsilon}_{ij} - \right. \\ &\left. N_T^{-1} \sum_{i=1}^{n}\sum_{j=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T\hat{\theta} - u\right)\left(\mathbf{Z}_{ij}^T\hat{\theta} - u\right)^l \varepsilon_{ij}\right| = O_p\left(n^{-1/2}\right).\end{aligned} \tag{A.10}$$

Next, for simplicity, we denote $\varphi_l(t) = K(t)\,t^l$. By the second order Taylor expansion of $\varphi_l(t)$ at $t_0 = \left(\mathbf{Z}_{ij}^T\theta - u\right)/h$, one obtains that

$$
\begin{aligned}
& N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T\hat{\theta} - u\right)\left(\mathbf{Z}_{ij}^T\hat{\theta} - u\right)^l \varepsilon_{ij} \\
&\quad - N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T\theta - u\right)\left(\mathbf{Z}_{ij}^T\theta - u\right)^l \varepsilon_{ij} \\
&= N_T^{-1}h^{l-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\varphi_l\left(\frac{\mathbf{Z}_{ij}^T\hat{\theta}-u}{h}\right)\varepsilon_{ij} - N_T^{-1}h^{l-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\varphi_l\left(\frac{\mathbf{Z}_{ij}^T\theta-u}{h}\right)\varepsilon_{ij} \\
&= N_T^{-1}h^{l-2}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\varphi_l^{(1)}\left(\frac{\mathbf{Z}_{ij}^T\theta-u}{h}\right)\varepsilon_{ij}\mathbf{Z}_{ij}^T\left(\hat{\theta}-\theta\right) \\
&\quad + 2^{-1}N_T^{-1}h^{l-3}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\varphi_l^{(2)}\left(\frac{\mathbf{Z}_{ij}^T\theta-u}{h}\right)\left(\hat{\theta}-\theta\right)^T\varepsilon_{ij}\mathbf{Z}_{ij}\mathbf{Z}_{ij}^T\left(\hat{\theta}-\theta\right) + N_T^{-1}h^{l-1} \\
&\quad \times\sum_{i=1}^{n}\sum_{j=1}^{N_i}\varepsilon_{ij}\int_{\left(\mathbf{Z}_{ij}^T\theta-u\right)/h}^{\left(\mathbf{Z}_{ij}^T\hat{\theta}-u\right)/h}\left\{\varphi_l^{(2)}(t) - \varphi_l^{(2)}\left(\frac{\mathbf{Z}_{ij}^T\theta-u}{h}\right)\right\}\left(\frac{\mathbf{Z}_{ij}^T\hat{\theta}-u}{h} - t\right)dt. \qquad \text{(A.11)}
\end{aligned}
$$

Notice that

$$
\begin{aligned}
& \sup_{u\in[a_0,b_0]}\Bigg| N_T^{-1}h^{l-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\varepsilon_{ij}\times \\
&\qquad \int_{\left(\mathbf{Z}_{ij}^T\theta-u\right)/h}^{\left(\mathbf{Z}_{ij}^T\hat{\theta}-u\right)/h}\left\{\varphi_l^{(2)}(t) - \varphi_l^{(2)}\left(\frac{\mathbf{Z}_{ij}^T\theta-u}{h}\right)\right\}\left(\frac{\mathbf{Z}_{ij}^T\hat{\theta}-u}{h} - t\right)dt\Bigg| \\
&\le CN_T^{-1}h^{l-1}\sup_{u\in[a_0,b_0]}\sum_{i=1}^{n}\sum_{j=1}^{N_i}|\varepsilon_{ij}|\int_{\left(\mathbf{Z}_{ij}^T\theta-u\right)/h}^{\left(\mathbf{Z}_{ij}^T\hat{\theta}-u\right)/h}\left(t - \frac{\mathbf{Z}_{ij}^T\theta-u}{h}\right)\left(\frac{\mathbf{Z}_{ij}^T\hat{\theta}-u}{h} - t\right)dt \\
&= C'N_T^{-1}h^{l-4}\sup_{u\in[a_0,b_0]}\sum_{i=1}^{n}\sum_{j=1}^{N_i}|\varepsilon_{ij}|\left|\mathbf{Z}_{ij}^T\left(\hat{\theta}-\theta\right)\right|^3 \\
&= O_p\left(n^{-3/2}h^{l-4}\right) = O_p\left(n^{-1/2}\right) \qquad \text{(A.12)}
\end{aligned}
$$

for some constants $C, C' > 0$. It holds since $\varphi_l^{(2)}(t)$ is Lipschitz continuous by Assumption (A3). One next truncates $\varepsilon_{ij}$ into three parts as in (A.1). Then again applying the Bernstein inequality in Lemma A.1, the Borel-Cantilli Lemma and discretization method, one has that

$$
\sup_{u\in[a_0,b_0]}\left| N_T^{-1}h^{l-2}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\varphi_l^{(1)}\left(\frac{\mathbf{Z}_{ij}^T\theta-u}{h}\right)\varepsilon_{ij}\mathbf{Z}_{ij}^T\left(\hat{\theta}-\theta\right)\right| = O_p\left(n^{-1/2}\right),
$$

and

$$\sup_{u\in[a_0,b_0]}\left|N_T^{-1}h^{l-3}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\varphi_l^{(2)}\left(\frac{\mathbf{Z}_{ij}^T\theta-u}{h}\right)\varepsilon_{ij}\left(\hat{\theta}-\theta\right)^T\mathbf{Z}_{ij}\mathbf{Z}_{ij}^T\left(\hat{\theta}-\theta\right)\right| = \mathcal{O}_p\left(n^{-1/2}\right).$$

Therefore,

$$\sup_{u\in[a_0,b_0]}\left|N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}K_h\Big(\mathbf{Z}_{ij}^T\hat{\theta}-u\Big)\Big(\mathbf{Z}_{ij}^T\hat{\theta}-u\Big)^l\varepsilon_{ij} - N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}K_h(\mathbf{Z}_{ij}^T\theta-u)(\mathbf{Z}_{ij}^T\theta-u)^l\varepsilon_{ij}\right| = O_p\left(n^{-1/2}\right),$$

which together with (A.10) implies that

$$\sup_{u\in[a_0,b_0]}\left|N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}K_h\Big(\mathbf{Z}_{ij}^T\hat{\theta}-u\Big)\Big(\mathbf{Z}_{ij}^T\hat{\theta}-u\Big)^l\hat{\varepsilon}_{ij} - N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}K_h(\mathbf{Z}_{ij}^T\theta-u)(\mathbf{Z}_{ij}^T\theta-u)^l\varepsilon_{ij}\right| = O_p\left(n^{-1/2}\right).$$

The proof is completed.

**Lemma A.4.** *Under Assumptions (A1)–(A5), as $n\to\infty$, one has*

$$\sup_{u,u'\in[a_0,b_0]}\left|N_T^{-1}h^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_i}\left\{K^*\Big(\mathbf{Z}_{ij}^T\hat{\theta},\mathbf{Z}_{ij'}^T\hat{\theta};u\Big)-K^*\left(\mathbf{Z}_{ij}^T\theta,\mathbf{Z}_{ij'}^T\theta;u\right)\right\}\sigma_i^{jj'}\right| = O_p\left(n^{-1}h^{-3}\right),$$

*where* $K^*(s,t;u)=K\left(\frac{s-u}{h}\right)K\left(\frac{t-u}{h}\right)$.

**Proof of Lemma A.4.** By the Mean Value Theorem, one has that

$$\begin{aligned} &N_T^{-1}h^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_i}\left\{K^*\left(\mathbf{Z}_{ij}^T\hat{\theta},\mathbf{Z}_{ij'}^T\hat{\theta};u\right)-K^*\left(\mathbf{Z}_{ij}^T\theta,\mathbf{Z}_{ij'}^T\theta;u\right)\right\}\sigma_i^{jj'}\\ &= N_T^{-1}h^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_i}\frac{\partial K^*\left(\mathbf{Z}_{ij}^T\theta,\mathbf{Z}_{ij'}^T\theta;u\right)}{\partial s}\mathbf{Z}_{ij}^T\left(\hat{\theta}-\theta\right)\sigma_i^{jj'}\\ &\quad + N_T^{-1}h^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_i}\frac{\partial K^*\left(\mathbf{Z}_{ij}^T\theta,\mathbf{Z}_{ij'}^T\theta;u\right)}{\partial t}\mathbf{Z}_{ij'}^T\left(\hat{\theta}-\theta\right)\sigma_i^{jj'}\end{aligned}$$

$$+ N_T^{-1}h^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_i} R_{ijj',K^*}(u)\,\sigma_i^{jj'}$$
$$\equiv S_I(u) + S_{II}(u) + S_{III}(u),$$

where

$$\begin{aligned} R_{ijj',K^*}(u) &= 2^{-1}\frac{\partial^2 K^*\left(\mathbf{Z}_{ij}^T\theta^*, \mathbf{Z}_{ij'}^T\theta^*; u\right)}{\partial s^2}\left(\mathbf{Z}_{ij}^T\hat{\theta} - \mathbf{Z}_{ij}^T\theta\right)^2 \\ &\quad + 2^{-1}\frac{\partial^2 K^*\left(\mathbf{Z}_{ij}^T\theta^*, \mathbf{Z}_{ij'}^T\theta^*; u\right)}{\partial t^2}\left(\mathbf{Z}_{ij'}^T\hat{\theta} - \mathbf{Z}_{ij'}^T\theta\right)^2 \\ &\quad + 2^{-1}\frac{\partial^2 K^*\left(\mathbf{Z}_{ij}^T\theta^*, \mathbf{Z}_{ij'}^T\theta^*; u\right)}{\partial s\partial t}\left(\mathbf{Z}_{ij}^T\hat{\theta} - \mathbf{Z}_{ij}^T\theta\right)\left(\mathbf{Z}_{ij'}^T\hat{\theta} - \mathbf{Z}_{ij'}^T\theta\right) \end{aligned}$$

and $\mathbf{Z}_{ij}^T\theta^*$ is a value between $\mathbf{Z}_{ij}^T\hat{\theta}$ and $\mathbf{Z}_{ij}^T\theta$. By the $\sqrt{n}$-consistency of $\hat{\theta}$ in (4.2), one can easily obtain that

$$\max_{u\in[a_0,b_0]}\max_{1\le i\le n, 1\le j,j'\le N_i}\left|R_{ijj'^*}(u)\right| \le Ch^{-2}\left\|\hat{\theta} - \theta\right\|^2 = O_p\left(n^{-1}h^{-2}\right)$$

for some constant $C$. Thus,

$$\sup_{u\in[a_0,b_0]}\left|S_{III}(u)\right| = O_p\left(n^{-1}h^{-3}\right).$$

Next,

$$\begin{aligned} &\sup_{u\in[a_0,b_0]}\left|S_I(u)\right| \\ &= \sup_{u\in[a_0,b_0]}\left|N_T^{-1}h^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_i}\frac{\partial K^*\left(\mathbf{Z}_{ij}^T\theta, \mathbf{Z}_{ij'}^T\theta; u\right)}{\partial s}\mathbf{Z}_{ij}^T\left(\hat{\theta} - \theta\right)\sigma_i^{jj'}\right| \\ &\le \max_{i,j,j'}\left|\mathbf{Z}_{ij}^T\left(\hat{\theta} - \theta\right)\sigma_i^{jj'}\right|\sup_{u\in[a_0,b_0]}\left|N_T^{-1}h^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_i}\frac{\partial K^*\left(\mathbf{Z}_{ij}^T\theta, \mathbf{Z}_{ij'}^T\theta; u\right)}{\partial s}\right| \\ &\le O_p\left(n^{-1/2}\right)\sup_{u\in[a_0,b_0]}\left|N_T^{-1}h^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_i}\mathrm{E}\left\{\frac{\partial K^*\left(\mathbf{Z}_{ij}^T\theta, \mathbf{Z}_{ij'}^T\theta; u\right)}{\partial s}\right\}\right| \\ &\quad + O_p\left(n^{-1/2}\right)\sup_{u\in[a_0,b_0]}\left|N_T^{-1}h^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_i}\left\{\frac{\partial K^*\left(\mathbf{Z}_{ij}^T\theta, \mathbf{Z}_{ij'}^T\theta; u\right)}{\partial s}\right.\right. \\ &\quad \left.\left. - \mathrm{E}\,\frac{\partial K^*\left(\mathbf{Z}_{ij}^T\theta, \mathbf{Z}_{ij'}^T\theta; u\right)}{\partial s}\right\}\right| \\ &= O_p\left(n^{-1/2}\right)O(1) + O_p\left(n^{-1/2}\right)O_p(\tau_n) = O_p\left(n^{-1/2}\right), \end{aligned}$$

where the average in the fourth line above is obtained by applying the Bernstein inequality in Lemma A.1 with the variance bounded by $O\left(n^{-1}h^{-1}\right)$. Similarly, one obtains $\sup_{u\in[a_0,b_0]}\left|S_{II}\left(u\right)\right| = O_p\left(n^{-1/2}\right)$. By Assumption (A5), $n^{-1/2} \ll n^{-1}h^{-3}$, completing the proof.

### *A.2. Proofs of propositions and theorems*

**Proof of Proposition 3.1.** Recall that $m_l\left(K\right) = \int u^l K\left(u\right)du$. For integer $l = 0, 1, 2$, one has that

$$\begin{aligned}
\mathrm{E}\left\{T_{n,l}\left(u\right)\right\} &= \mathrm{E}\left\{N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} K_h\left(\mathbf{Z}_{ij}^T\theta - u\right)\left(\mathbf{Z}_{ij}^T\theta - u\right)^l\right\} \\
&= N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} h^{-1}\int K\left(\frac{t-u}{h}\right)\left(t-u\right)^l f\left(t\right)dt \\
&= \int K\left(v\right)h^l v^l\left\{f\left(u\right) + f^{(1)}\left(u\right)hv + u\left(h\right)\right\}dv \\
&= f\left(u\right)m_l\left(K\right)h^l + h^{l+1}m_{l+1}\left(K\right)f^{(1)}\left(u\right) + u\left(h^{l+1}\right). \qquad \text{(A.13)}
\end{aligned}$$

Let

$$\xi_{i,l}\left(u\right) = N_T^{-1}\sum_{j=1}^{N_i}\left[K_h\left(\mathbf{Z}_{ij}^T\theta - u\right)\left(\mathbf{Z}_{ij}^T\theta - u\right)^l - \mathrm{E}\left\{K_h\left(\mathbf{Z}_{ij}^T\theta - u\right)\left(\mathbf{Z}_{ij}^T\theta - u\right)^l\right\}\right].$$

Thus,

$$T_{n,l}\left(u\right) - \mathrm{E}\,T_{n,l}\left(u\right) = \sum_{i=1}^{n}\xi_{i,l}\left(u\right). \qquad \text{(A.14)}$$

It is clear that $\xi_{i,l}\left(u\right), 1 \le i \le n$, are independent with $\mathrm{E}\,\xi_{i,l}\left(u\right) = 0$, which together with the fact that $N_T \sim n$ concludes that

$$\begin{aligned}
\mathrm{E}\,\xi_{i,l}^2\left(u\right) &= N_T^{-2}\sum_{j=1}^{N_i}\mathrm{E}\,K_h^2\left(\mathbf{Z}_{ij}^T\theta - u\right)\left(\mathbf{Z}_{ij}^T\theta - u\right)^{2l} \\
&+ N_T^{-2}\sum_{j=1}^{N_i}\sum_{j'\neq j}^{N_i}\mathrm{E}\left\{K_h\left(\mathbf{Z}_{ij}^T\theta - u\right)K_h\left(\mathbf{Z}_{ij'}^T\theta - u\right)\left(\mathbf{Z}_{ij}^T\theta - u\right)^l\left(\mathbf{Z}_{ij'}^T\theta - u\right)^l\right\} \\
&- N_T^{-2}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_i}\mathrm{E}\left\{K_h\left(\mathbf{Z}_{ij}^T\theta - u\right)\left(\mathbf{Z}_{ij}^T\theta - u\right)^l\right\}\mathrm{E}\left\{K_h\left(\mathbf{Z}_{ij'}^T\theta - u\right)\left(\mathbf{Z}_{ij'}^T\theta - u\right)^l\right\} \\
&= N_T^{-2}h^{2l-1}N_i f\left(u\right)\int K^2\left(v\right)v^{2l}dv + u\left(n^{-2}h^{2l-1}\right).
\end{aligned}$$

Thus,

$$\mathrm{var}\left\{\sum_{i=1}^{n}\xi_{i,l}\left(u\right)\right\} = \sum_{i=1}^{n}\mathrm{E}\,\xi_{i,l}^2(u) = N_T^{-1}h^{2l-1}f\left(u\right)\int K^2\left(v\right)v^{2l}dv + u\left(n^{-1}h^{2l-1}\right).$$

By Assumptions (A3) and (A4), one has $\operatorname{var}\{\sum_{i=1}^{n}\xi_{i,l}(u)\} \le Cn^{-1}h^{2l-1}$ for some $C>0$. Moreover, by Minkowski's inequality, for $k>2$,

$$\begin{aligned}
&|\xi_{i,l}(u)|^{k-2}\\
&= \left| N_T^{-1}\sum_{j=1}^{N_i}\Big[K_h(\mathbf{Z}_{ij}^T\theta-u)(\mathbf{Z}_{ij}^T\theta-u)^l - \mathrm{E}\Big\{K_h(\mathbf{Z}_{ij}^T\theta-u)(\mathbf{Z}_{ij}^T\theta-u)^l\Big\}\Big]\right|^{k-2}\\
&\le N_T^{-(k-2)}2^{k-3}\left[\left|\sum_{j=1}^{N_i}K_h\left(\mathbf{Z}_{ij}^T\theta-u\right)\left(\mathbf{Z}_{ij}^T\theta-u\right)^l\right|^{k-2}\right.\\
&\qquad\qquad \left.+\left|\sum_{j=1}^{N_i}\mathrm{E}\left\{K_h\left(\mathbf{Z}_{ij}^T\theta-u\right)\left(\mathbf{Z}_{ij}^T\theta-u\right)^l\right\}\right|^{k-2}\right]\\
&\le N_T^{-(k-2)}2^{k-2}\left\{N_i h^{l-1}\|K\|_\infty\right\}^{k-2}.
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathrm{E}\,|\xi_{i,l}(u)|^k &= \mathrm{E}\left\{|\xi_{i,l}(u)|^{k-2}|\xi_{i,l}(u)|^2\right\}\\
&\le N_T^{-(k-2)}2^{k-2}\left\{N_i h^{l-1}\|K\|_\infty\right\}^{k-2}\mathrm{E}\left\{\xi_{i,l}(u)\right\}^2\\
&= \left\{2N_T^{-1}h^{l-1}N_i\|K\|_\infty\right\}^{k-2}\mathrm{E}\left\{\xi_{i,l}(u)\right\}^2 \le c_0^{k-2}k!\,\mathrm{E}\left\{\xi_{i,l}(u)\right\}^2
\end{aligned}$$

for $c_0 = 2N_T^{-1}h^{l-1}C_N\|K\|_\infty$ with $C_N$ being an upper bound of all $N_i$. That means for each $u\in[a_0,b_0]$, $\xi_{i,l}(u), 1\le i\le n$, satisfy Cramér's conditions in Lemma A.1 with $c=c_0$. Then applying the Bernstein inequality in Lemma A.1, the Borel-Cantelli Lemma and discretization technique, one obtains that

$$\sup_{u\in[a_0,b_0]}\left|\sum_{i=1}^{n}\xi_{i,l}(u)\right| = O_{a.s.}\left(h^l\tau_n\right). \tag{A.15}$$

Putting (A.13), (A.14) and (A.15) together, one has that

$$\begin{aligned}
T_{n,l}(u) &= \mathrm{E}\,T_{n,l}(u) + T_{n,l}(u) - \mathrm{E}\,T_{n,l}(u)\\
&= f(u)\,m_l(K)\,h^l + h^{l+1}m_{l+1}(K)\,f^{(1)}(u) + u_p\left(h^{l+1}\right) + U_p\left(h^l\tau_n\right).
\end{aligned}$$

The proof is completed.

**Proof of Proposition 3.2.** Recall that

$$\begin{aligned}
&W_{n,l}(u)\\
&= N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}K_h\left(\mathbf{Z}_{ij}^T\theta-u\right)\left(\mathbf{Z}_{ij}^T\theta-u\right)^l\left\{Y_{ij}^*-\phi(u)-\phi^{(1)}(u)\left(\mathbf{Z}_{ij}^T\theta-u\right)\right\}
\end{aligned}$$

$$= N_T^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} K_h \left( \mathbf{Z}_{ij}^T \theta - u \right) \left( \mathbf{Z}_{ij}^T \theta - u \right)^l \varepsilon_{ij} + N_T^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} K_h \left( \mathbf{Z}_{ij}^T \theta - u \right) \times \left( \mathbf{Z}_{ij}^T \theta - u \right)^l \left\{ \phi \left( \mathbf{Z}_{ij}^T \theta \right) - \phi (u) - \phi^{(1)} (u) \left( \mathbf{Z}_{ij}^T \theta - u \right) \right\}. \quad \text{(A.16)}$$

Notice that when $l = 0$,

$$\begin{aligned} &N_T^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} K_h \left( \mathbf{Z}_{ij}^T \theta - u \right) \left\{ \phi \left( \mathbf{Z}_{ij}^T \theta \right) - \phi (u) - \phi^{(1)} (u) \left( \mathbf{Z}_{ij}^T \theta - u \right) \right\} \\ &= 2^{-1} \phi^{(2)} (u) T_{n,2} (u) + u_p \left( h^2 \right) \\ &= 2^{-1} \phi^{(2)} (u) f (u) m_2 (K) h^2 + u_p \left( h^2 \right), \end{aligned} \quad \text{(A.17)}$$

and when $l = 1$,

$$\begin{aligned} &N_T^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} K_h \left( \mathbf{Z}_{ij}^T \theta - u \right) \left( \mathbf{Z}_{ij}^T \theta - u \right) \left\{ \phi \left( \mathbf{Z}_{ij}^T \theta \right) - \phi (u) - \phi^{(1)} (u) \left( \mathbf{Z}_{ij}^T \theta - u \right) \right\} \\ &= N_T^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} K_h \left( \mathbf{Z}_{ij}^T \theta - u \right) \left( \mathbf{Z}_{ij}^T \theta - u \right) \left\{ 2^{-1} \phi^{(2)} (u) \left( \mathbf{Z}_{ij}^T \theta - u \right)^2 + u_p \left( h^2 \right) \right\} \\ &= U_p \left( h^3 \right). \end{aligned} \quad \text{(A.18)}$$

Therefore, equations (A.16), (A.17), (A.18), and Lemma A.2 immediately conclude the result.

**Proof of Proposition 3.3.** By $\varphi_l (t) = K (t) t^l, l = 0, 1, 2$ and (A.11) in the proof of Lemma A.3, one has that

$$\begin{aligned} &\hat{T}_{n,l} (u) - T_{n,l} (u) \\ &= N_T^{-1} h^{l-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \varphi_l \left( \frac{\mathbf{Z}_{ij}^T \hat{\theta} - u}{h} \right) - N_T^{-1} h^{l-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \varphi_l \left( \frac{\mathbf{Z}_{ij}^T \theta - u}{h} \right) \\ &= N_T^{-1} h^{l-2} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \varphi_l^{(1)} \left( \frac{\mathbf{Z}_{ij}^T \theta - u}{h} \right) \mathbf{Z}_{ij}^T \left( \hat{\theta} - \theta \right) \\ &\quad + 2^{-1} N_T^{-1} h^{l-3} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \varphi_l^{(2)} \left( \frac{\mathbf{Z}_{ij}^T \theta - u}{h} \right) \left( \hat{\theta} - \theta \right)^T \mathbf{Z}_{ij} \mathbf{Z}_{ij}^T \left( \hat{\theta} - \theta \right) \\ &\quad + N_T^{-1} h^{l-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \int_{\left( \mathbf{Z}_{ij}^T \theta - u \right)/h}^{\left( \mathbf{Z}_{ij}^T \hat{\theta} - u \right)/h} \left\{ \varphi_l^{(2)} (t) - \varphi_l^{(2)} \left( \frac{\mathbf{Z}_{ij}^T \theta - u}{h} \right) \right\} \left( \frac{\mathbf{Z}_{ij}^T \hat{\theta} - u}{h} - t \right) dt \\ &\equiv A_{T,l} (u) + B_{T,l} (u) + C_{T,l} (u). \end{aligned}$$

According to (A.12), one has that $\sup_{u \in [a_0, b_0]} \left| C_{T,l} (u) \right|$ equals

$$\sup_{u \in [a_0, b_0]} \left| N_T^{-1} h^{l-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \int_{\left( \mathbf{Z}_{ij}^T \theta - u \right)/h}^{\left( \mathbf{Z}_{ij}^T \hat{\theta} - u \right)/h} \left\{ \varphi_l^{(2)} (t) - \varphi_l^{(2)} \left( \frac{\mathbf{Z}_{ij}^T \theta - u}{h} \right) \right\} \left( \frac{\mathbf{Z}_{ij}^T \hat{\theta} - u}{h} - t \right) dt \right|$$

$$= O_p\left(n^{-3/2}h^{l-4}\right) = O_p\left(n^{-1/2}\right).$$

Next, define

$$\zeta_{i,l,k}\left(u\right) = N_T^{-1}h^{l-2}\sum_{j=1}^{N_i}\varphi_l^{(1)}\left(\frac{\mathbf{Z}_{ij}^T\theta - u}{h}\right)Z_{ij,k}, k = 1, 2, \ldots, q, l = 0, 1, 2.$$

By Assumptions (A3) and (A4), one easily obtains that $\left|\mathrm{E}\,\zeta_{i,0,k}\left(u\right)\right| = O\left(n^{-1}\right)$ and $\left|\mathrm{E}\,\zeta_{i,l,k}\left(u\right)\right| = O\left(n^{-1}h^{l-1}\right)$ for $l = 1, 2$. Thus, there exists $C_1 > 0$ such that $\mathrm{E}\,\zeta_{i,l,k}\left(u\right) \le C_1 n^{-1}h^{l-1}$ for all $l = 0, 1, 2$. Meanwhile,

$$\begin{aligned}
&\mathrm{E}\,\zeta_{i,l,k}^2\left(u\right)\\
&= N_T^{-2}h^{2l-4}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_i}\mathrm{E}\left\{\varphi_l^{(1)}\left(\frac{\mathbf{Z}_{ij}^T\theta - u}{h}\right)Z_{ij,k}\varphi_l^{(1)}\left(\frac{\mathbf{Z}_{ij'}^T\theta - u}{h}\right)Z_{ij',k}\right\}\\
&= N_T^{-2}h^{2l-4}\sum_{j=1}^{N_i}\mathrm{E}\left\{\varphi_l^{(1)}\left(\frac{\mathbf{Z}_{ij}^T\theta - u}{h}\right)Z_{ij,k}\right\}^2\\
&+ N_T^{-2}h^{2l-4}\sum_{j=1}^{N_i}\sum_{j'\neq j}^{N_i}\mathrm{E}\left\{\varphi_l^{(1)}\left(\frac{\mathbf{Z}_{ij}^T\theta - u}{h}\right)Z_{ij,k}\varphi_l^{(1)}\left(\frac{\mathbf{Z}_{ij'}^T\theta - u}{h}\right)Z_{ij',k}\right\}.
\end{aligned}$$

Hence there exists $C_2 > 0$ such that $\mathrm{E}\,\zeta_{i,l,k}^2\left(u\right) \le C_2 n^{-2}h^{2l-3}$. Let $\zeta_{i,l,k}^*\left(u\right) = \zeta_{i,l,k}\left(u\right) - \mathrm{E}\,\zeta_{i,l,k}\left(u\right)$. So $\mathrm{E}\,\zeta_{i,l,k}^*\left(u\right) = 0$ and $\mathrm{E}\left\{\zeta_{i,l,k}^*\left(u\right)\right\}^2 = \mathrm{E}\,\zeta_{i,l,k}^2\left(u\right) - \left\{\mathrm{E}\,\zeta_{i,l,k}\left(u\right)\right\}^2 \le C_2 n^{-2}h^{2l-3}$. Notice that

$$\begin{aligned}
\left|\zeta_{i,l,k}^*\left(u\right)\right| \le \left|\zeta_{i,l,k}\left(u\right)\right| + \left|\mathrm{E}\,\zeta_{i,l,k}\left(u\right)\right| &\le C N_T^{-1}h^{l-2}\left\|\varphi_l^{(1)}\right\|_\infty + C_1 n^{-1}h^{l-1}\\
&\le C_0 n^{-1}h^{l-2}
\end{aligned}$$

for some $C > 0, C_0 > 0$. Thus for $k > 2$,

$$\mathrm{E}\left|\zeta_{i,l,k}^*\left(u\right)\right|^k = \mathrm{E}\left|\zeta_{i,l,k}^*\left(u\right)\right|^{k-2}\left|\zeta_{i,l,k}^*\left(u\right)\right|^2 \le \left(C_0 n^{-1}h^{l-2}\right)^{(k-2)}k!\,\mathrm{E}\left|\zeta_{i,l,k}^*\left(u\right)\right|^2,$$

which implies that $\zeta_{i,l,k}^*\left(u\right)$ satisfies the Cramér conditions in Lemma A.1 with $c = C_0 n^{-1}h^{l-2}$. Hence, applying the Bernstein inequality in Lemma A.1, the Borel-Cantelli Lemma and Discretization method again, one has that

$$\sup_{u\in[a_0,b_0]}\left|\sum_{i=1}^n \zeta_{i,l,k}^*\left(u\right)\right| = O_p\left(h^{l-1}\tau_n\right).$$

Therefore,

$$\sup_{u\in[a_0,b_0]}\left|A_{T,l}\left(u\right)\right| \le \sup_{u\in[a_0,b_0]}\left|\left(\sum_{i=1}^n \zeta_{i,l,1}\left(u\right), \ldots, \sum_{i=1}^n \zeta_{i,l,q}\left(u\right)\right)\left(\hat{\theta} - \theta\right)\right|$$

$$\leq \sup_{u\in[a_0,b_0],1\leq k\leq q}\left|\sum_{i=1}^{n}\zeta_{i,l,k}\left(u\right)\right|\sqrt{q}\left\|\hat{\theta}-\theta\right\|$$

$$= \sup_{u\in[a_0,b_0],1\leq k\leq q}\left|\sum_{i=1}^{n}\zeta^{*}_{i,l,k}\left(u\right)+\sum_{i=1}^{n}\mathrm{E}\,\zeta_{i,l,k}\left(u\right)\right|\sqrt{q}\left\|\hat{\theta}-\theta\right\|$$

$$= \left[O_p\left(h^{l-1}\tau_n\right)+O\left(1\right)\right]O_p\left(n^{-1/2}\right)=O_p\left(n^{-1/2}\right).$$

We next consider $B_{T,l}\left(u\right)$. Denote

$$\zeta_{i,l,k,k'}\left(u\right)=N_T^{-1}\sum_{j=1}^{N_i}h^{l-3}\varphi_l^{(2)}\left(\frac{\mathbf{Z}_{ij}^T\theta-u}{h}\right)Z_{ij,k}Z_{ij,k'},$$

$k,k'=1,2,\ldots,q,l=0,1,2$. Applying the Bernstein inequality in Lemma A.1, the Borel-Cantelli Lemma and discretization method as before, one has that

$$\sup_{u\in[a_0,b_0]}\left|\sum_{i=1}^{n}\zeta_{i,l,k,k'}\left(u\right)\right|=O_p\left(h^{l-2}\tau_n\right)+O_p\left(h^{l-2}\right),$$

and hence

$$\sup_{u\in[a_0,b_0]}\left|B_{T,l}\left(u\right)\right|=\left[O_p\left(h^{l-2}\tau_n\right)+\mathcal{O}_p\left(h^{l-2}\right)\right]O_p\left(n^{-1}\right)=o_p\left(n^{-1/2}\right).$$

The proof is completed.

**Proof of Proposition 3.4.** By the Mean Value Theorem, for $l=0,1$, one has that

$$N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}K_h\left(\mathbf{Z}_{ij}^T\hat{\theta}-u\right)\left(\mathbf{Z}_{ij}^T\hat{\theta}-u\right)^l\left\{\phi\left(\mathbf{Z}_{ij}^T\hat{\theta}\right)-\phi\left(u\right)-\phi^{(1)}(u)\left(\mathbf{Z}_{ij}^T\hat{\theta}-u\right)\right\}$$

$$-N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}K_h\left(\mathbf{Z}_{ij}^T\theta-u\right)\left(\mathbf{Z}_{ij}^T\theta-u\right)^l\left\{\phi\left(\mathbf{Z}_{ij}^T\theta\right)-\phi\left(u\right)-\phi^{(1)}(u)\left(\mathbf{Z}_{ij}^T\theta-u\right)\right\}$$

$$=N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}K_h\left(\mathbf{Z}_{ij}^T\theta-u\right)\left(\mathbf{Z}_{ij}^T\theta-u\right)^l\left\{\phi\left(\mathbf{Z}_{ij}^T\hat{\theta}\right)-\phi\left(\mathbf{Z}_{ij}^T\theta\right)\right\}$$

$$+N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}K_h\left(\mathbf{Z}_{ij}^T\theta-u\right)\left(\mathbf{Z}_{ij}^T\theta-u\right)^l\phi^{(1)}\left(u\right)\mathbf{Z}_{ij}^T\left(\theta-\hat{\theta}\right)+N_T^{-1}h^{-1}$$

$$\times\sum_{i=1}^{n}\sum_{j=1}^{N_i}\left\{h^{-1}K^{(1)}\left(\frac{\mathbf{Z}_{ij}^T\theta^{*}-u}{h}\right)\left(\mathbf{Z}_{ij}^T\theta^{*}-u\right)^l+lK\left(\frac{\mathbf{Z}_{ij}^T\theta^{*}-u}{h}\right)\left(\mathbf{Z}_{ij}^T\theta^{*}-u\right)^{l-1}\right\}$$

$$\times\mathbf{Z}_{ij}^T\left(\hat{\theta}-\theta\right)\left\{\phi\left(\mathbf{Z}_{ij}^T\hat{\theta}\right)-\phi\left(u\right)-\phi^{(1)}\left(u\right)\left(\mathbf{Z}_{ij}^T\hat{\theta}-u\right)\right\},$$

where $\left(\mathbf{Z}_{ij}^T\theta^{*}-u\right)/h$ is some value between $\left(\mathbf{Z}_{ij}^T\hat{\theta}-u\right)/h$ and $\left(\mathbf{Z}_{ij}^T\theta-u\right)/h$.

Notice that

$$\sup_{u\in[a_0,b_0]}\left|N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}K_h\left(\mathbf{Z}_{ij}^T\theta-u\right)\left(\mathbf{Z}_{ij}^T\theta-u\right)^l\left\{\phi\left(\mathbf{Z}_{ij}^T\hat{\theta}\right)-\phi\left(\mathbf{Z}_{ij}^T\theta\right)\right\}\right|$$
$$\leq\left\|\phi^{(1)}\left(u\right)\right\|_\infty\sup_{u\in[a_0,b_0]}N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}K_h\left(\mathbf{Z}_{ij}^T\theta-u\right)\left|\mathbf{Z}_{ij}^T\theta-u\right|^l\left|\mathbf{Z}_{ij}^T\left(\hat{\theta}-\theta\right)\right|$$
$$=O_p\left(n^{-1/2}h^l\right)=O_p\left(n^{-1/2}\right)\tag{A.19}$$

and

$$\sup_{u\in[a_0,b_0]}\left|N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}K_h\left(\mathbf{Z}_{ij}^T\theta-u\right)\left(\mathbf{Z}_{ij}^T\theta-u\right)^l\phi^{(1)}\left(u\right)\mathbf{Z}_{ij}^T\left(\theta-\hat{\theta}\right)\right|$$
$$\leq\left\|\phi^{(1)}\left(u\right)\right\|_\infty\sup_{u\in[a_0,b_0]}N_T^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}K_h\left(\mathbf{Z}_{ij}^T\theta-u\right)\left|\mathbf{Z}_{ij}^T\theta-u\right|^l\left|\mathbf{Z}_{ij}^T\left(\hat{\theta}-\theta\right)\right|$$
$$=O_p\left(n^{-1/2}h^l\right)=O_p\left(n^{-1/2}\right).\tag{A.20}$$

Meanwhile, by Assumptions (A3) and (A4), it is easy to show that

$$\sup_{u\in[a_0,b_0]}\Bigg|N_T^{-1}h^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\Bigg\{h^{-1}K^{(1)}\left(\frac{\mathbf{Z}_{ij}^T\theta^*-u}{h}\right)\left(\mathbf{Z}_{ij}^T\theta^*-u\right)^l$$
$$+lK\left(\frac{\mathbf{Z}_{ij}^T\theta^*-u}{h}\right)\left(\mathbf{Z}_{ij}^T\theta^*-u\right)^{l-1}\Bigg\}\mathbf{Z}_{ij}^T\left(\hat{\theta}-\theta\right)\times$$
$$\left\{\phi\left(\mathbf{Z}_{ij}^T\hat{\theta}\right)-\phi\left(u\right)-\phi^{(1)}\left(u\right)\left(\mathbf{Z}_{ij}^T\hat{\theta}-u\right)\right\}\Bigg|=O_p\left(n^{-1/2}\right).\tag{A.21}$$

Thus, equations (A.19), (A.20), and (A.21) and Lemma A.3 conclude that

$$\sup_{u\in[a_0,b_0]}\left|\hat{W}_{n,l}\left(u\right)-W_{n,l}\left(u\right)\right|=O_p\left(n^{-1/2}\right),$$

completing the proof.

**Proof of Theorem 2.1.** By Proposition 3.1 and Assumption (A5), one has that

$$\begin{pmatrix}T_{n,0}\left(u\right)&T_{n,1}\left(u\right)\\T_{n,1}\left(u\right)&T_{n,2}\left(u\right)\end{pmatrix}^{-1}=f^{-1}\left(u\right)\begin{pmatrix}1+u_p\left(h\right)&U_p\left(h^2\right)\\U_p\left(h^2\right)&m_2\left(K\right)h^2+u_p\left(h^3\right)\end{pmatrix}^{-1}.$$

This along with Lemma A.2 and Proposition 3.2 implies that

$$\hat{\phi}\left(u,\mathbf{\Theta}\right)-\phi\left(u\right)=e_0^T\begin{pmatrix}T_{n,0}\left(u\right)&T_{n,1}\left(u\right)\\T_{n,1}\left(u\right)&T_{n,2}\left(u\right)\end{pmatrix}^{-1}\begin{pmatrix}W_{n,0}\left(u\right)\\W_{n,1}\left(u\right)\end{pmatrix}$$

$$= e_0^T \left\{ f^{-1}(u) \begin{pmatrix} 1 + u_p(h) & U_p(1) \\ U_p(1) & m_2^{-1}(K) h^{-2} + u_p(h^{-1}) \end{pmatrix} \right\} \times$$
$$\begin{pmatrix} N_T^{-1} \sum_{i=1}^n \sum_{j=1}^{N_i} K_h \left( \mathbf{Z}_{ij}^T \theta - u \right) \varepsilon_{ij} + 2^{-1} \phi^{(2)}(u) f(u) m_2(K) h^2 + u_p(h^2) \\ U_p(h^3 + h\tau_n) \end{pmatrix}$$
$$= R_n(u) + 2^{-1} \phi^{(2)}(u) m_2(K) h^2 + u_p\left(h^2 + h\tau_n\right) + U_p\left(h^3 + h\tau_n\right)$$
$$= R_n(u) + 2^{-1} \phi^{(2)}(u) m_2(K) h^2 + u_p\left(h^2\right),$$

where $R_n(u) = f^{-1}(u) N_T^{-1} \sum_{i=1}^n \sum_{j=1}^{N_i} K_h \left( \mathbf{Z}_{ij}^T \theta - u \right) \varepsilon_{ij}$. Thus the proof is completed.

**Proof of Theorem 2.3.** Note that

$$\hat{\phi}\left(u, \hat{\mathbf{\Theta}}\right) - \hat{\phi}(u, \mathbf{\Theta}) = \hat{\phi}\left(u, \hat{\mathbf{\Theta}}\right) - \phi(u) - \left\{ \hat{\phi}(u, \mathbf{\Theta}) - \phi(u) \right\}$$
$$= e_0^T \left( N_T^{-1} \hat{\mathbf{Z}}^T \hat{\mathbf{W}} \hat{\mathbf{Z}} \right)^{-1} \begin{pmatrix} \hat{W}_{n,0}(u) \\ \hat{W}_{n,1}(u) \end{pmatrix} - e_0^T \left( N_T^{-1} \mathbf{Z}^T \mathbf{W} \mathbf{Z} \right)^{-1} \begin{pmatrix} W_{n,0}(u) \\ W_{n,1}(u) \end{pmatrix}$$
$$= e_0^T \begin{pmatrix} \hat{T}_{n,0}(u) & \hat{T}_{n,1}(u) \\ \hat{T}_{n,1}(u) & \hat{T}_{n,2}(u) \end{pmatrix}^{-1} \begin{pmatrix} \hat{W}_{n,0}(u) \\ \hat{W}_{n,1}(u) \end{pmatrix}$$
$$- e_0^T \begin{pmatrix} T_{n,0}(u) & T_{n,1}(u) \\ T_{n,1}(u) & T_{n,2}(u) \end{pmatrix}^{-1} \begin{pmatrix} W_{n,0}(u) \\ W_{n,1}(u) \end{pmatrix}.$$

According to Proposition 3.3, one has that

$$\begin{pmatrix} \hat{T}_{n,0}(u) & \hat{T}_{n,1}(u) \\ \hat{T}_{n,1}(u) & \hat{T}_{n,2}(u) \end{pmatrix}^{-1} = \begin{pmatrix} T_{n,0}(u) & T_{n,1}(u) \\ T_{n,1}(u) & T_{n,2}(u) \end{pmatrix}^{-1} + U_p\left(n^{-1/2}\right),$$

which with Proposition 3.4 implies that

$$\hat{\phi}\left(u, \hat{\mathbf{\Theta}}\right) - \hat{\phi}(u, \mathbf{\Theta}) = e_0^T \left\{ \begin{pmatrix} T_{n,0}(u) & T_{n,1}(u) \\ T_{n,1}(u) & T_{n,2}(u) \end{pmatrix}^{-1} + U_p\left(n^{-1/2}\right) \right\}$$
$$\times \begin{pmatrix} W_{n,0}(u) + U_p\left(n^{-1/2}\right) \\ W_{n,1}(u) + U_p\left(n^{-1/2}\right) \end{pmatrix} - e_0^T \begin{pmatrix} T_{n,0}(u) & T_{n,1}(u) \\ T_{n,1}(u) & T_{n,2}(u) \end{pmatrix}^{-1} \begin{pmatrix} W_{n,0}(u) \\ W_{n,1}(u) \end{pmatrix}.$$

Thus,

$$\sup_{u \in [a_0, b_0]} \left| \hat{\phi}\left(u, \hat{\mathbf{\Theta}}\right) - \hat{\phi}(u, \mathbf{\Theta}) \right| = O_p\left(n^{-1/2}\right).$$

The proof is completed.

**Proof of Theorem 2.5.** It is clear that

$$N_T^{-1} h^{-1} \sum_{i=1}^n \sum_{j=1}^{N_i} \sum_{j'=1}^{N_i} K^* \left( \mathbf{Z}_{ij}^T \hat{\theta}, \mathbf{Z}_{ij'}^T \hat{\theta}; u \right) \hat{\varepsilon}_{ij} \hat{\varepsilon}_{ij'}$$
$$- N_T^{-1} h^{-1} \sum_{i=1}^n \sum_{j=1}^{N_i} \sum_{j'=1}^{N_i} \mathrm{E}\, K^* \left( \mathbf{Z}_{ij}^T \theta, \mathbf{Z}_{ij'}^T \theta; u \right) \sigma_i^{jj'}$$

$$= N_T^{-1}h^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_i}\left\{K^*\left(\mathbf{Z}_{ij}^T\hat{\theta},\mathbf{Z}_{ij'}^T\hat{\theta};u\right)-K^*\left(\mathbf{Z}_{ij}^T\theta,\mathbf{Z}_{ij'}^T\theta;u\right)\right\} \times\left(\hat{\varepsilon}_{ij}\hat{\varepsilon}_{ij'}-\sigma_i^{jj'}\right)$$
$$+N_T^{-1}h^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_i}K^*\left(\mathbf{Z}_{ij}^T\theta,\mathbf{Z}_{ij'}^T\theta;u\right)\left(\hat{\varepsilon}_{ij}\hat{\varepsilon}_{ij'}-\sigma_i^{jj'}\right)$$
$$+N_T^{-1}h^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_i}\left\{K^*\left(\mathbf{Z}_{ij}^T\hat{\theta},\mathbf{Z}_{ij'}^T\hat{\theta};u\right)-K^*\left(\mathbf{Z}_{ij}^T\theta,\mathbf{Z}_{ij'}^T\theta;u\right)\right\}\sigma_i^{jj'}$$
$$+N_T^{-1}h^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_i}\left\{K^*\left(\mathbf{Z}_{ij}^T\theta,\mathbf{Z}_{ij'}^T\theta;u\right)-\mathrm{E}\,K^*\left(\mathbf{Z}_{ij}^T\theta,\mathbf{Z}_{ij'}^T\theta;u\right)\right\}\sigma_i^{jj'}$$

where $K^*(s,t;u)=K\left(\frac{s-u}{h}\right)K\left(\frac{t-u}{h}\right)$ is in Lemma A.4. Notice that

$$\begin{aligned}\hat{\varepsilon}_{ij}\hat{\varepsilon}_{ij'} = & \left[\varepsilon_{ij}-\mathbf{X}_{ij}^T\left(\hat{\beta}-\beta\right)-\left\{\hat{\phi}\left(\mathbf{Z}_{ij}^T\hat{\theta}\right)-\hat{\phi}\left(\mathbf{Z}_{ij}^T\theta\right)\right\}\right]\times \\ & \left[\varepsilon_{ij'}-\mathbf{X}_{ij'}^T\left(\hat{\beta}-\beta\right)-\left\{\hat{\phi}\left(\mathbf{Z}_{ij'}^T\hat{\theta}\right)-\hat{\phi}\left(\mathbf{Z}_{ij'}^T\theta\right)\right\}\right],\end{aligned}$$

and $\sigma_i^{jj'}=\mathrm{E}\,\varepsilon_{ij}\varepsilon_{ij'}$. As in the proof of Lemma A.4, it is easy to show that

$$\sup_{u\in[a_0,b_0]}\left|N_T^{-1}h^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_i}\left\{K^*\left(\mathbf{Z}_{ij}^T\hat{\theta},\mathbf{Z}_{ij'}^T\hat{\theta};u\right)-K^*\left(\mathbf{Z}_{ij}^T\theta,\mathbf{Z}_{ij'}^T\theta;u\right)\right\} \times\left(\hat{\varepsilon}_{ij}\hat{\varepsilon}_{ij'}-\sigma_i^{jj'}\right)\right|=O_p\left(n^{-1}h^{-3}\right).$$

Meanwhile, using the inequality in Lemma A.1 and the discretization method again as in Lemmas A.2 and A.3, one can obtain that

$$\sup_{u\in[a_0,b_0]}\left|N_T^{-1}h^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_i}K^*\left(\mathbf{Z}_{ij}^T\theta,\mathbf{Z}_{ij'}^T\theta;u\right)\left(\hat{\varepsilon}_{ij}\hat{\varepsilon}_{ij'}-\sigma_i^{jj'}\right)\right| = O_p\left(n^{-1/2}\log^{1/2}n\right)$$

and

$$\sup_{u,u'\in[a_0,b_0]}\left|N_T^{-1}h^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_i}\left\{K^*\left(\mathbf{Z}_{ij}^T\theta,\mathbf{Z}_{ij'}^T\theta;u\right)-\mathrm{E}K^*\left(\mathbf{Z}_{ij}^T\theta,\mathbf{Z}_{ij'}^T\theta;u\right)\right\}\sigma_i^{jj'}\right| = O_p\left(n^{-1/2}\log^{1/2}n\right).$$

In addition, according to Lemma A.4, one has that

$$\sup_{u\in[a_0,b_0]}\left|N_T^{-1}h^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_i}\left\{K^*\left(\mathbf{Z}_{ij}^T\hat{\theta},\mathbf{Z}_{ij'}^T\hat{\theta};u\right)-K^*\left(\mathbf{Z}_{ij}^T\theta,\mathbf{Z}_{ij'}^T\theta;u\right)\right\}\sigma_i^{jj'}\right|$$

$$= O_p\left(n^{-1}h^{-3}\right).$$

Therefore, the four equations above imply that

$$\sup_{u,u'\in[a_0,b_0]}\left|N_T^{-1}h^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_i}K^*\left(\mathbf{Z}_{ij}^T\hat{\theta},\mathbf{Z}_{ij'}^T\hat{\theta};u\right)\hat{\varepsilon}_{ij}\hat{\varepsilon}_{ij'}\right.$$
$$\left.-N_T^{-1}h^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\sum_{j'=1}^{N_i}\mathrm{E}\,K^*\left(\mathbf{Z}_{ij}^T\theta,\mathbf{Z}_{ij'}^T\theta;u\right)\sigma_i^{jj'}\right|=O_p\left(n^{-1}h^{-3}\right).$$

This along with $\sup_{u\in[a_0,b_0]}\left|\hat{f}(u)-f(u)\right| = o_p(h_f)$ in (2.6) and Assumption (A5) concludes that

$$\sup_{u\in[a_0,b_0]}\left|\hat{C}_n(u)-C_n(u)\right| = o_p(h_f)+O_p\left(n^{-1}h^{-3}\right) = o_p(h_f+h) = o_p(h_f).$$

The proof is completed.

## References


Bickel, P. and Rosenblatt, M. (1973) On some global measures of deviations of density function estimates. *Ann. Statist.* **31** 1852–1884. MR0383633

Bosq, D. (1998) *Nonparametric Statistics for Stochastic Processes*. Springer-Verlag, New York. MR1640691

Cai, L., Liu, R., Wang, S. and Yang, L. (2019) Simultaneous confidence bands for mean and variance functions based on deterministic design. *Statist. Sinica* **29** 505–525. MR3889378

Cai, Q. and Wang, S. (2019) Efficient estimation in partially linear single-index models for longitudinal data. *Scand. J. Stat.* **46** 116–141. MR3915269

Cai, L. and Yang, L. (2015) A smooth simultaneous confidence band for conditional variance function. *TEST* **24** 632–655. MR3377897

Cao, G., Yang, L. and Todem, D. (2012) Simultaneous inference for the mean function based on dense functional data. *J. Nonparametr. Stat.* **24** 359–377. MR2921141

Cao, G., Wang, L., Li, Y. and Yang, L. (2016) Oracle-efficient confidence envelopes for covariance functions in dense functional data. *Statist. Sinica* **26** 359–383. MR3468356

Cao, H., Liu, W. and Zhou, Z. (2018) Simultaneous nonparametric regression analysis of sparse longitudinal data. *Bernoulli* **24** 3013–3038. MR3779709

Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. (1997) Generalized partially linear single-index models. *J. Amer. Statist. Assoc.* **92** 477–489. MR1467842

Chen, J., Gao, J. and Li, D. (2013) Estimation in partially linear single-index panel data models with fxed effects. *J. Bus. Econ. Stat.* **31** 315–330. MR3173684

Chen, J., Li, D., Liang, H. and Wang, S. (2015) Semiparametric GEE analysis in partially linear singleindex models for longitudinal data. *Ann. Statist.* **43** 1682–1715. MR3357875

Claeskens, G. and Van Keilegom, I. (2003) Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.* **31** 1852–1884. MR2036392

Degras, D. (2011) Simultaneous confidence bands for nonparametric regression with functional data. *Statist. Sinica* **21** 1735–1765. MR2895997

Eubank, R. and Speckman, P. (1993) Confidence bands in nonparametric regression. *J. Amer. Statist. Assoc.* **88** 1287–1301. MR1245362

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

Fan, J., Huang, T. and Li, R. (2007) Analysis of longitudinal data with semiparametric estimation of covariance function. *J. Amer. Statist. Assoc.* **102** 632–641. MR2370857

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

Fitzmaurice, G., Davidian, M., Verbeke, G. and Molenberghs, G. (2008) *Longitudinal Data Analysis*. CRC Press, Florida. MR1500110

Gu, L. and Yang, L. (2015) Oracally efficient estimation for single-index link function with simultaneous confidence band. *Electron. J. Stat.* **9** 1540–1561. MR3376116

Härdle, W. (1989) Asmptotic maximal deviation of M-smoothers. *J. Multivariate Anal.* **29** 163–179. MR1004333

Härdle, W., Liang, H. and Gao, J. (2012) *Partially Linear Models*. Springer-Verlag, New York. MR2376405

He, X., Fung, W. and Zhu, Z. (2005) Robust estimation in generalized partial linear models for clustered data. *J. Amer. Statist. Assoc.* **100** 1176–1184. MR2236433

Johnston, G. (1982) Probabilities of maximal deviations for nonparametric regression function estimates. *J. Multivariate Anal.* **12** 402–414. MR0666014

Liang, H., Liu, X., Li, R. and Tsai, C. (2010) Estimation and testing for partially linear single-index models for longitudinal data. *Ann. Statist.* **38** 3811–3836. MR2766869

Lin, X. and Carroll, R. J. (2000) Nonparametric function estimation for clustered data when the predictor is measured without/with error. *J. Amer. Statist. Assoc.* **95** 520–534. MR1803170

Lin, X. and Carroll, R. J. (2001) Semiparametric regression for clustered data using generalized estimating equations. *J. Amer. Statist. Assoc.* **96** 455–467. MR1947252

Lin, X. and Carroll, R. J. (2006) Semiparametric estimation in general repeated measures problems. *J. Roy. Statist. Soc. Ser. B* **68** 69–88. MR2212575

Liu, W. and Wu, W. (2010) Simultaneous nonparametric inference of time series. *Ann. Statist.* **38** 2388–2421. MR2676893

Ma, S., Yang, L. and Carroll, R. J. (2012) A simultaneous confidence band for sparse longitudinal regression. *Statist. Sinica* **22** 95–122. MR2933169

Ma, S., Liang, H. and Tsai, C. (2014) Partially linear single index models for repeated measurements. *J. Multivariate Anal.* **130** 354–375. MR3229543

Ruppert, D., Sheather, S. and Wand, M. (1995) An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.* **90** 1257–1270. MR1379468

Silverman, B. (1986) *Density Estimation*. Chapman and Hall, London.

Wang, J. and Yang, L. (2009) Polynomial spline confidence bands for regression curves. *Statist. Sinica* **19** 325–342. MR2487893

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

Wu, W. and Zhao, Z. (2007) Inference of trends in time series. *J. Roy. Statist. Soc. B* **69** 391–410. MR2323759

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

Xia, Y. and Li, W. (1999) On single-index coefficient regression models. *J. Amer. Statist. Assoc.* **94** 1275–1285. MR1731489

Xue, L. and Zhang, J. (2020) Empirical likelihood for partially linear single-index models with missing observations. *Comput. Statist. Data Anal.* `https://doi.org/10.1016/j.csda.2019.106877`. MR4029015

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

Zeger, S. and Diggle, P. (1994) Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics* **50** 689–699.

Zhou, Z. and Wu, W. (2010) Simultaneous inference of linear models with time-varying coefficients. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 513–531. MR2758526