Submitted to the Annals of Statistics

EFFICIENT ESTIMATION IN SEMIVARYING COEFFICIENT MODELS FOR LONGITUDINAL/CLUSTERED DATA

By Toshio Honda[¶,*], Ming-Yen Cheng [‖,†], and Jialiang Li[**,‡]

Hitotsubashi University[*], *National Taiwan University*[†], *and National University of Singapore*[‡]

In semivarying coefficient models for longitudinal/clustered data, usually of primary interest is usually the parametric component which involves unknown constant coefficients. First, we study semiparametric efficiency bound for estimation of the constant coefficients in a general setup. It can be achieved by spline regression provided that the within-cluster covariance matrices are all known, which is an unrealistic assumption in reality. Thus, we propose an adaptive estimator of the constant coefficients when the covariance matrices are unknown and depend only on the index random variable, such as time, and when the link function is the identity function. After preliminary estimation, based on working independence and both spline and local linear regression, we estimate the covariance matrices by applying local linear regression to the resulting residuals. Then we employ the covariance matrix estimates and spline regression to obtain our final estimators of the constant coefficients. The proposed estimator achieves the semiparametric efficiency bound under normality assumption, and it has the smallest covariance matrix among a class of estimators even when normality is violated. We also present results of numerical studies. The simulation results demonstrate that our estimator is superior to the one based on working independence. When applied to the CD4 count data, our method identifies an interesting structure that was not found by previous analyses.

1. Introduction. Longitudinal data consist of $(Y_{ij}, \boldsymbol{X}_{ij}, \boldsymbol{Z}_{ij}, T_{ij}), i = 1, \ldots, n, j = 1, \ldots, m_i$, where m_i is bounded uniformly in i, and $Y_{ij}, \boldsymbol{X}_{ij} =$

[§]This research was partially supported by the Hitotsubashi International Fellow Program.

[¶]Research was supported by the JSPS Grant-in-Aids for Scientific Research (A) 24243031 and (C) 25400197.

[‖]Corresponding author. Research was supported by the Ministry of Science and Technology grant MOST101-2118-M-002-001-MY3.

[**]Research was supported by grants AcRF R-155-000-130-112 and NMRC/CBRG/0014/2012.

MSC 2010 subject classifications: Primary 62G08

Keywords and phrases: Semiparametric efficiency bound, spline functions, local linear regression, covariance matrix estimation

$(X_{ij1}, \ldots, X_{ijp})^T$ and $\boldsymbol{Z}_{ij} = (Z_{ij1}, \ldots, Z_{ijq})^T$ are respectively the scalar response, the p-dimensional and q-dimensional covariate vectors of the ith subject at the jth observation time $T_{ij} \in [0, 1]$. Such kind of data are commonly acquired for various purposes, such as evidence based knowledge discovery and empirical study, in a wide range of subject areas. When the T_{ij}'s are observations on some index variables other than time, they are sometimes called clustered data.

A popular class of models for longitudinal data analysis are the semivarying coefficient models, which are specified by

$$\begin{aligned} &\mathrm{E}(Y_{ij}|\boldsymbol{X}_{ij}, \boldsymbol{Z}_{ij}, T_{ij}, \underline{\boldsymbol{X}}_i, \underline{\boldsymbol{Z}}_i, \underline{T}_i) \\ &\quad = \mathrm{E}(Y_{ij}|\boldsymbol{X}_{ij}, \boldsymbol{Z}_{ij}, T_{ij}) \equiv \mu(\boldsymbol{X}_{ij}^T\boldsymbol{\beta} + \boldsymbol{Z}_{ij}^T\boldsymbol{g}(T_{ij})) = \mu_{ij}, \end{aligned} \tag{1.1}$$

where A^T stands for the transpose of a matrix A. In model (1.1), $\mu(x)$ is a known strictly increasing smooth link function, $\boldsymbol{\beta}$ is an unknown regression coefficient vector, and $\boldsymbol{g}(t) = \big(g_1(t), \ldots, g_q(t)\big)^T$ is a vector of unknown twice continuously differentiable coefficient functions. We assume that all the covariates are uniformly bounded for technical reasons. Besides, we let $Z_{ij1} \equiv 1$ and suppose $\boldsymbol{X}_{ij}$ has no constant element for all i and j. We introduce some notation here. Denote

$$\underline{\boldsymbol{X}}_i = (\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{im_i})^T,\ \underline{\boldsymbol{Z}}_i = (\boldsymbol{Z}_{i1}, \ldots, \boldsymbol{Z}_{im_i})^T,\ \text{and } \underline{T}_i = (T_{i1}, \ldots, T_{im_i})^T,$$

and define

$$\underline{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{im_i})^T = \underline{Y}_i - \underline{\mu}_i\,,\ \text{and } \Sigma_i = \mathrm{Var}(\underline{\epsilon}_i|\underline{\boldsymbol{X}}_i, \underline{\boldsymbol{Z}}_i, \underline{T}_i), \tag{1.2}$$

where $\underline{Y}_i = (Y_{i1}, \ldots, Y_{im_i})^T$, $\underline{\mu}_i = (\mu_{i1}, \ldots, \mu_{im_i})^T$, and Σ_i is an $m_i \times m_i$ positive definite matrix depending on $\underline{\boldsymbol{X}}_i$, $\underline{\boldsymbol{Z}}_i$, and $\underline{T}_i$, $i = 1, \ldots, n$. This is a standard marginal model in the literature of longitudinal data analysis [17].

Model (1.1) consists of a parametric component, which provides information on the constant impacts $\boldsymbol{\beta}$ of some important covariates, and a nonparametric component which captures the dynamic impacts $\boldsymbol{g}(t)$ of the other covariates. In this way the model is able to reflect unknown nonlinear structures in the data while retaining similar interpretability as the classical linear models at the same time. As much as is needed in practical applications, there is an extensive literature on the variable selection, structure identification, estimation, and inference issues; see for example [3, 5, 9, 15, 18]. In particular, often of primary interest is to have access to the parametric component while the nonparametric component is viewed as the nuisance part. However, it is well known that assuming independence or some mis-specified

working covariance structure yields less efficient estimation of the constant coefficients. Therefore, a substantial portion of the existing literature are aimed at improving the efficiency via modeling and estimating the within-cluster covariance structure [3, 4, 6, 19, 20, 21], which is itself a challenging task.

The current work contributes to the efficient estimation problem for model (1.1) in two directions. First, we study explicit expression of the semiparametric efficiency bound and asymptotic normality of the Generalized Estimating Equations (GEE) spline estimators under general link functions, covariance structures and error distributions. These results, stated in Section 2, are parallel to that for partially linear additive models given by [1]. Partially linear additive models are a popular class of alternative semiparametric models for longitudinal data analysis, and they differ from semivarying coefficient models in that the nonparametric component admits an additive form. A special case is partially linear models [10, 16]. Spline estimation was used in [1] and [11].

Our second and the main contribution is about adaptive efficient estimation when the within-cluster covariances are unknown and estimated based on the data. Notice that this practical issue was ignored by [1] and they did not consider estimation of the covariances. We propose a simple and reliable procedure for estimating nonparametrically the error covariances, assuming they depend only on the observation times. This kind of assumptions are reasonable because we do not assume the observation times are regular across different clusters or they are dense. With irregular and/or sparse observation times, estimating the covariances in a completely nonparametric way is particularly problematic, and even almost intractable, when they depend on the covariates in addition to the observation times. Our main result states that our covariance estimates lead to efficient estimation of the parametric component under the identity link function and when the errors are normally distributed. Even when the normality assumption is violated, our results indicate that the efficiency cannot be improved under general error distributions. This result is partly motivated by [10], which dealt with efficient estimation in partially linear models and considered the same covariance structure. However, our approach is different from that of [10], and the proof of our main result is more elaborated since we deal with a linear regression model with a diverging dimension.

The approach of [10] involves local linear kernel estimation both when estimating the nonparametric component and when estimating the covariances. Thus, to achieve semiparametric efficiency, it requires some complicated iterative backfitting calculation even for the identity link function. We deal with

the identity link function in Section 3 and we are able to avoid such kind of complicated iterative procedures by using kernel estimation and spline function approximation at different steps. To be specific, we first employ spline function approximation when estimating the parametric component in (1.1) based on GEE and working independence. Next we apply local linear regression in our estimation of the nonparametric component in order to produce some residuals, which are then smoothed by local linear regression again to obtain estimates of the covariances. The final estimates of the constant coefficients are obtained by plugging-in the covariance estimates to the GEE spline estimation. This is a Feasible Generalized Least Squares (FGLS) procedure under the considered identity link function. Besides being simple and fast to compute, the FGLS approach allows us to obtain easily an estimate of the covariance matrix of the estimators of regression coefficients in (1.1). This is another significant advantage of our new procedure.

As in [10], we assume some specific structure of the covariance matrices Σ_i's in (1.2) and estimate them under the assumed structure. Then we carry out the FGLS procedure. As noted before, without the assumed covariance structure, it is almost impossible to consider and estimate the covariance matrices in a fully nonparametric way. In addition, even if the assumption on the Σ_i's fails to hold, our FGLS estimator still has the asymptotic normality under mild conditions and it still makes use of some information of the covariance matrices. For example, if the covariances depend on some time-dependent covariates, then such effects are still captured by our method to some extent. Compared with the methods of [4] and [14], which use respectively parametrically estimated and some ad-hoc covariance matrices, our approach is more adaptive to the unknown covariance matrices. In the independent case, i.e. there is one observation for each cluster, our assumption on the covariance matrices reduces to that of [13], which also suggested to improve the efficiency in a similar manner.

In summary, we provide a new method for adaptive efficient estimation of the parametric component in semivarying coefficient models (1.1). The ideas are clear and it is simple and fast to compute. We give rigorous theoretical arguments for its semiparametric efficiency under some general assumptions. Besides, our approach would work reasonably well even when the assumptions fail to hold. In addition, a simulation study shows that numerically the proposed method outperforms the working independence approach significantly and it behaves close to the oracle estimator which uses the true covariance matrices. We also applied our method to the CD4 count dataset and identified some interesting new effects not detected before.

The organization of this paper is as follows. In Section 2 we derive the

semiparametric efficiency bound in estimating the constant coefficient vector $\boldsymbol{\beta}$ and asymptotic normality of the GEE spline estimators. In Section 3 we propose an adaptive estimator of $\boldsymbol{\beta}$ under the identity link function $\mu(x) \equiv x$ and when the errors have some general covariance structure. In this section our main theoretical result regarding its asymptotic equivalence to the oracle estimator is stated in Theorem 1. Section 4 summarizes and discusses results of our simulation and empirical studies used to assess numerical performance of the proposed efficient estimator. Section 5 contains technical assumptions and proofs of semiparametric efficiency and asymptotic normality of the proposed estimator. Technical proofs of the other theoretical results and some lemmas are placed in the supplementary material [8].

2. Estimation of $\boldsymbol{\beta}$ and efficiency bound. In this section, V_i is a given inverse weight matrix depending only on $\underline{\boldsymbol{X}}_i$, $\underline{\boldsymbol{Z}}_i$, and $\underline{T}_i$, for $i = 1, \ldots, n$. We use spline functions to approximate the function $\boldsymbol{g}(t)$, specifically a K_n-dimensional equispaced B-spline basis $\boldsymbol{B}(t)$ on $[0, 1]$. See [12] for the definition and properties of B-spline bases. Since functions are assumed to be twice continuously differentiable in this paper, we recommend linear or quadratic spline approximation with the optimal order of K_n which is $K_n = c_K n^{1/5}$ for some constant c_K. We set

$$\boldsymbol{W}_{ij} = \boldsymbol{Z}_{ij} \otimes \boldsymbol{B}(T_{ij}) \text{ and } \underline{\boldsymbol{W}}_i = (\boldsymbol{W}_{i1}, \ldots, \boldsymbol{W}_{im_i})^T,$$

where $\otimes$ is the Kronecker product. Then we estimate the true values of $\boldsymbol{\beta}$ and $\boldsymbol{g}$ by minimizing with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ simultaneously the following objective function:

$$\sum_{i=1}^{n} (\underline{Y}_i - \underline{\mu}(\underline{\boldsymbol{X}}_i\boldsymbol{\beta} + \underline{\boldsymbol{W}}_i\boldsymbol{\gamma}))^T V_i^{-1} (\underline{Y}_i - \underline{\mu}(\underline{\boldsymbol{X}}_i\boldsymbol{\beta} + \underline{\boldsymbol{W}}_i\boldsymbol{\gamma})), \tag{2.1}$$

where $\boldsymbol{\gamma} \in R^{qK_n}$ and the j th element of $\underline{\mu}(\underline{\boldsymbol{X}}_i\boldsymbol{\beta} + \underline{\boldsymbol{W}}_i\boldsymbol{\gamma})$ is $\mu(\boldsymbol{X}_{ij}^T\boldsymbol{\beta} + \boldsymbol{W}_{ij}^T\boldsymbol{\gamma})$. Thus the generalized estimating equations are

$$\begin{aligned} &\sum_{i=1}^{n} \underline{\boldsymbol{X}}_i^T \Delta_i V_i^{-1} (\underline{Y}_i - \underline{\mu}(\underline{\boldsymbol{X}}_i\boldsymbol{\beta} + \underline{\boldsymbol{W}}_i\boldsymbol{\gamma})) = 0, \\ \text{and} \quad &\sum_{i=1}^{n} \underline{\boldsymbol{W}}_i^T \Delta_i V_i^{-1} (\underline{Y}_i - \underline{\mu}(\underline{\boldsymbol{X}}_i\boldsymbol{\beta} + \underline{\boldsymbol{W}}_i\boldsymbol{\gamma})) = 0, \end{aligned} \tag{2.2}$$

where Δ_i is an $m_i \times m_i$ diagonal matrix defined by

$$\Delta_i = \operatorname{diag}(\mu'(\boldsymbol{X}_{i1}^T\boldsymbol{\beta} + \boldsymbol{W}_{i1}^T\boldsymbol{\gamma}), \ldots, \mu'(\boldsymbol{X}_{im_i}^T\boldsymbol{\beta} + \boldsymbol{W}_{im_i}^T\boldsymbol{\gamma})).$$

Denote the solution to (2.2) by $\widehat{\boldsymbol{\beta}}_V$ and $\widehat{\boldsymbol{\gamma}}_V \equiv \left(\widehat{\gamma}_{1V}^T, \ldots, \widehat{\gamma}_{qV}^T\right)^T$. Then the GEE spline estimator with weight matrices V_i^{-1}, $i = 1, \ldots, n$, for $\boldsymbol{\beta}$ is $\widehat{\boldsymbol{\beta}}_V$, and that for $\boldsymbol{g}(t)$ is $\left(\widehat{\gamma}_{1V}^T \boldsymbol{B}(t), \ldots, \widehat{\gamma}_{qV}^T \boldsymbol{B}(t)\right)^T$.

We present the asymptotic normality of $\widehat{\boldsymbol{\beta}}_V$ in Proposition 1 and deal with the semiparametric efficiency bound in estimation of the true value of $\boldsymbol{\beta}$ in Proposition 2. In addition, Proposition 3 shows that if we know Σ_i, the estimator $\widehat{\boldsymbol{\beta}}_\Sigma$ with $V_i = \Sigma_i$ achieves the semiparametric efficiency bound. We introduce function spaces, inner products, and projections before we state Propositions 1-3. Hereafter we denote the true values of $\boldsymbol{\beta}$ and $\boldsymbol{g}(t)$ by $\boldsymbol{\beta}_0$ and $\boldsymbol{g}_0(t) = (g_{01}(t), \ldots, g_{0q}(t))^T$, respectively.

First we define two function spaces $\boldsymbol{G}$ and $\boldsymbol{G}_B$:

$$\boldsymbol{G} = \{(g_1, \ldots, g_p)^T \,|\, g_j \in L_2,\, j = 1, \ldots, p\},$$

where L_2 is the space of square integrable functions on $[0, 1]$ and

$$\boldsymbol{G}_B = \{(\boldsymbol{B}^T \boldsymbol{\gamma}_1, \ldots, \boldsymbol{B}^T \boldsymbol{\gamma}_q)^T \,|\, \boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^T, \ldots, \boldsymbol{\gamma}_q^T)^T \in R^{qK_n}\}\,.$$

Recall that $\boldsymbol{B}(t)$ is the equispaced B-spline basis on $[0, 1]$ and note that $\boldsymbol{G}_B \subset \boldsymbol{G}$.

Next we introduce inner products and the associated norms. Let v_1 and v_2 be two processes each taking a scalar stochastic value at T_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, m_i$. Then we define two inner products of v_1 and v_2 by

$$\langle v_1, v_2 \rangle_n^\Delta = \frac{1}{n} \sum_{i=1}^n \underline{v}_{1i}^T \Delta_{0i} V_i^{-1} \Delta_{0i} \underline{v}_{2i} \text{ and } \langle v_1, v_2 \rangle^\Delta = \mathrm{E}\{\langle v_1, v_2 \rangle_n^\Delta\}, \tag{2.3}$$

where $\underline{v}_{1i}$ and $\underline{v}_{2i}$ are defined in the same way as $\underline{\boldsymbol{X}}_i$ and

$$\Delta_{0i} = \mathrm{diag}\big(\mu'(\boldsymbol{X}_{i1}^T \boldsymbol{\beta}_0 + \boldsymbol{Z}_{i1}^T \boldsymbol{g}_0(T_{i1})), \ldots, \mu'(\boldsymbol{X}_{im_i}^T \boldsymbol{\beta}_0 + \boldsymbol{Z}_{im_i}^T \boldsymbol{g}_0(T_{im_i}))\big).$$

The associated norms are then defined by

$$\|v\|_n^\Delta = (\langle v, v \rangle_n^\Delta)^{1/2} \text{ and } \|v\|^\Delta = (\langle v, v \rangle^\Delta)^{1/2}. \tag{2.4}$$

Finally we define the projections, with respect to $\|\cdot\|^\Delta$, of the kth element of $\boldsymbol{X}_{ij}$ onto $\boldsymbol{Z}^T \boldsymbol{G}$ and $\boldsymbol{Z}^T \boldsymbol{G}_B$ by

$$\Pi_V X_k = \operatorname*{argmin}_{\boldsymbol{g} \in \boldsymbol{G}} \|X_k - \boldsymbol{Z}^T \boldsymbol{g}\|^\Delta \text{ and } \Pi_{Vn} X_k = \operatorname*{argmin}_{\boldsymbol{g} \in \boldsymbol{G}_B} \|X_k - \boldsymbol{Z}^T \boldsymbol{g}\|^\Delta, \tag{2.5}$$

where

$$\|X_k - \boldsymbol{Z}^T \boldsymbol{g}\|^\Delta = \frac{1}{n} \mathrm{E}\Big\{ \sum_{i=1}^n (\underline{X}_{ik} - \underline{(\boldsymbol{Z}^T \boldsymbol{g})}_i)^T \Delta_{0i} V_i^{-1} \Delta_{0i} (\underline{X}_{ik} - \underline{(\boldsymbol{Z}^T \boldsymbol{g})}_i) \Big\},$$

with $\underline{X}_{ik} = (X_{i1k}, \ldots, X_{im_ik})^T$ and $\underline{(\boldsymbol{Z}^T\boldsymbol{g})}_i = (\boldsymbol{Z}_{i1}^T\boldsymbol{g}(T_{i1}), \ldots, \boldsymbol{Z}_{im_i}^T\boldsymbol{g}(T_{im_i}))$. Note that

$$\langle X_k - \boldsymbol{Z}^T\Pi_V X_k, \boldsymbol{Z}^T\boldsymbol{g}\rangle^\Delta = 0 \quad \forall \boldsymbol{g} \in \boldsymbol{G} \tag{2.6}$$

and

$$\langle X_k - \boldsymbol{Z}^T\Pi_{Vn} X_k, \boldsymbol{Z}^T\boldsymbol{g}\rangle^\Delta = 0 \quad \forall \boldsymbol{g} \in \boldsymbol{G}_B. \tag{2.7}$$

Hereafter we write

$$\boldsymbol{\varphi}_{Vk}^* = \Pi_V X_k \in \boldsymbol{G} \quad \text{and} \quad \overline{\boldsymbol{\varphi}}_{Vk} = \Pi_{Vn} X_k \in \boldsymbol{G}_B.$$

When $V_i = \Sigma_i$, we denote $\boldsymbol{\varphi}_{\Sigma k}^*(t)$ by $\boldsymbol{\varphi}_{eff,k}^*(t)$ since it gives the seimparametric efficient score function. See Proposition 2. We assume that $\boldsymbol{\varphi}_{Vk}^*(t)$, $k = 1, \ldots, p$, are twice continuously differentiable on $[0, 1]$ and that they and their second order derivatives are uniformly bounded in n. We consider the existence and smoothness properties of $\boldsymbol{\varphi}_{Vk}^*(t)$ in Section 5.

Some matrices are necessary to present Proposition 1 and we define them here. Let

$$\begin{aligned} H &= \begin{pmatrix} \sum_{i=1}^n \underline{\boldsymbol{X}}_i^T \Delta_{0i} V_i^{-1} \Delta_{0i} \underline{\boldsymbol{X}}_i & \sum_{i=1}^n \underline{\boldsymbol{X}}_i^T \Delta_{0i} V_i^{-1} \Delta_{0i} \underline{\boldsymbol{W}}_i \\ \sum_{i=1}^n \underline{\boldsymbol{W}}_i^T \Delta_{0i} V_i^{-1} \Delta_{0i} \underline{\boldsymbol{X}}_i & \sum_{i=1}^n \underline{\boldsymbol{W}}_i^T \Delta_{0i} V_i^{-1} \Delta_{0i} \underline{\boldsymbol{W}}_i \end{pmatrix} \\ &= \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \quad \text{(say)}, \\ H_{11\cdot 2} &= H_{11} - H_{12} H_{22}^{-1} H_{21}\,, \quad \text{and} \quad H^{11} = (H_{11\cdot 2})^{-1}\,. \end{aligned} \tag{2.8}$$

Note that the (k, l) element of $n^{-1}H_{11\cdot 2}$ is an estimate of

$$\frac{1}{n}\sum_{i=1}^n \mathrm{E}\Big\{(\underline{X}_{ik} - \underline{(\boldsymbol{Z}^T\boldsymbol{\varphi}_{Vk}^*)}_i)^T \Delta_{0i} V_i^{-1} \Delta_{0i} (\underline{X}_{il} - \underline{(\boldsymbol{Z}^T\boldsymbol{\varphi}_{Vl}^*)}_i)\Big\}. \tag{2.9}$$

Let Ω_{Vn} be a $p \times p$ matrix whose (k, l)th element is (2.9). We assume that there exists a $p \times p$ positive definite matrix Ω_V such that

$$\lim_{n\to\infty} \Omega_{Vn} = \Omega_V. \tag{2.10}$$

The asymptotic normality of $\widehat{\boldsymbol{\beta}}_V$ is given in Proposition 1. We can establish the proposition in almost the same way as in the proof of Theorem 2 of [1], thus in Section 5 we describe only the necessary changes in the proof and omit the details. We denote the normal distribution with mean η and covariance Ω by $\mathrm{N}(\eta, \Omega)$, and by $\xrightarrow{d}$ we mean convergence in distribution. Let $\boldsymbol{I}_l$ be the l-dimensional identity matrix.

PROPOSITION 1. *(asymptotic normality of $\widehat{\boldsymbol{\beta}}_V$) Under the assumptions in this section and Assumptions A1-6 in Section 5, we have*

$$\widehat{\boldsymbol{\beta}}_V = \boldsymbol{\beta}_0 + H^{11}\sum_{i=1}^{n}(\underline{\boldsymbol{X}}_i - \underline{\boldsymbol{W}}_i H_{22}^{-1}H_{21})^T\Delta_{0i}V_i^{-1}\underline{\epsilon}_i + o_p\Big(\frac{1}{\sqrt{n}}\Big).$$

We also have

$$\sqrt{n}\widetilde{\Omega}_{\widehat{\boldsymbol{\beta}}_V}^{-1/2}(\widehat{\boldsymbol{\beta}}_V - \boldsymbol{\beta}_0) \xrightarrow{d} \mathrm{N}(0, \boldsymbol{I}_p),$$

where $\widetilde{\Omega}_{\widehat{\boldsymbol{\beta}}_V}$ is

(2.11)

$$nH^{11}\sum_{i=1}^{n}\Big\{(\underline{\boldsymbol{X}}_i - \underline{\boldsymbol{W}}_i H_{22}^{-1}H_{21})^T\Delta_{0i}V_i^{-1}\Sigma_i V_i^{-1}\Delta_{0i}(\underline{\boldsymbol{X}}_i - \underline{\boldsymbol{W}}_i H_{22}^{-1}H_{21})\Big\}H^{11}.$$

We can estimate $\widetilde{\Omega}_{\widehat{\boldsymbol{\beta}}_V}$ by replacing Δ_{0i} and Σ_i in (2.11) with some estimates based on $\widehat{\boldsymbol{\beta}}_V$ and $\widehat{\boldsymbol{\gamma}}_V$. For example, we can replace Σ_i with $\widetilde{\underline{\epsilon}}_i\widetilde{\underline{\epsilon}}_i^T$, where

$$\widetilde{\underline{\epsilon}}_i = \underline{Y}_i - \underline{\mu}(\underline{\boldsymbol{X}}_i^T\widehat{\boldsymbol{\beta}}_V + \underline{\boldsymbol{W}}_i^T\widehat{\boldsymbol{\gamma}}_V).$$

Alternatively, we can estimate $\boldsymbol{g}(t)$ by applying local linear regression to $\underline{Y}_i - \underline{\boldsymbol{X}}_i^T\widehat{\boldsymbol{\beta}}_V$, and then estimate Σ_i based on the resulting residuals and some assumption on the structure of Σ_i as described in Section 3.

We give in Proposition 2 the semiparametric efficiency bound for estimation of $\boldsymbol{\beta}$. It can be proved in the same way as Lemma 1 of [1] and the proof is omitted. We denote the semiparametric efficient score function of $\boldsymbol{\beta}$ by

$$\boldsymbol{l}_{\boldsymbol{\beta}}^* = (l_{\boldsymbol{\beta}1}^*, \ldots, l_{\boldsymbol{\beta}p}^*)^T.$$

Its expression is given in Proposition 2.

PROPOSITION 2. *(semiparametric efficiency bound) Under the assumptions in this section and Assumptions A1-6 in Section 5, we have*

$$l_{\boldsymbol{\beta}k}^* = \sum_{i=1}^{n}(\underline{X}_{ik} - \underline{(\boldsymbol{Z}^T\boldsymbol{\varphi}_{eff,k}^*)}_i)^T\Delta_{0i}\Sigma_i^{-1}\{\underline{Y}_i - \underline{\mu}(\underline{\boldsymbol{X}}_i\boldsymbol{\beta}_0 + \underline{(\boldsymbol{Z}^T\boldsymbol{g}_0)}_i)\},$$

and the semiparametric efficient information matrix for $\boldsymbol{\beta}$ is given by

$$\lim_{n\to\infty}\frac{1}{n}\mathrm{E}\{\boldsymbol{l}_{\boldsymbol{\beta}}^*(\boldsymbol{l}_{\boldsymbol{\beta}}^*)^T\} = \Omega_\Sigma \text{ with } V_i = \Sigma_i \text{ in (2.10).}$$

Proposition 3 is about the asymptotic normality of $\widehat{\boldsymbol{\beta}}_\Sigma$, the so called oracle estimator using the true covariance structure in the GEE spline regression, and it asserts that $\widehat{\boldsymbol{\beta}}_\Sigma$ achieves the semiparametric efficiency bound derived from Proposition 2. It can be proved in the same way as Corollary 1 of [1], and thus the proof is omitted.

PROPOSITION 3. *(oracle efficient estimator) Under the assumptions in this section and Assumptions A1-6 in Section 5, we have with $V_i = \Sigma_i$ in (2.2)*

$$\sqrt{n}\Omega_\Sigma^{-1/2}(\widehat{\boldsymbol{\beta}}_\Sigma - \boldsymbol{\beta}_0) \xrightarrow{d} \mathrm{N}(0, \boldsymbol{I}_p).$$

In practice, usually the Σ_i's are unknown and we have no direct access to the semiparametric efficient score function or the oracle estimator. In Section 3 we assume some structure on the Σ_i's. Specifically we assume that $\Sigma_i = \Sigma(\underline{T}_i)$ as in (3.1), and carry out nonparametric estimation of Σ_i. Then we use the covariance estimates and construct a FGLS procedure to improve the efficiency in estimation of $\boldsymbol{\beta}$. We prove the asymptotic equivalence between $\widehat{\boldsymbol{\beta}}_\Sigma$ and the FGLS estimator in Theorem 1. Some remarks on efficiency are in order before we introduce the FGLS estimator.

REMARK 1. *In Proposition 2, no assumption on the structure of Σ_i or the distribution of $\underline{\epsilon}_i$ is imposed. However, it is then almost impossible to estimate Σ_i in a fully nonparametric way. When $\Sigma_i = \Sigma(\underline{T}_i)$ as specified in (3.1) in Section 3, we should use this information in calculating the semiparametric efficient score function. Unfortunately, under general errors, this task seems still intractable and we have no results in this regard. Nevertheless, when assumption (3.1) and some regularity conditions hold, we come up with some remedies to improve the efficiency in estimation of $\boldsymbol{\beta}$, as compared to using some (arbitrary) working covariance structure. Indeed our estimator given in Section 3 has the smallest asymptotic variance among all $\widehat{\boldsymbol{\beta}}_V$ in this case. This property follows from Propositions 1-3, Theorem 1, and the semiparametric efficiency bound under normality assumption given in A.1 of [16]. Furthermore, our estimator is semiparametric efficient when $\underline{\epsilon}_i$ is normally distributed, conditional on $\underline{\boldsymbol{X}}_i$, $\underline{\boldsymbol{Z}}_i$, and $\underline{T}_i$.*

3. Efficient estimation of $\boldsymbol{\beta}$. In this section, we propose a procedure to improve the efficiency in estimation of $\boldsymbol{\beta}$ when the Σ_i's are unknown and assumed to have some specific structure. Hereafter, for simplicity, we consider only the identity link function $\mu(x) \equiv x$. In this case, we have only to carry out least squares estimation of $\boldsymbol{\beta}$ and the least squares estimate has an explicit expression once V_i in (2.10) is given.

We have considered the semiparametric efficiency bound of $\boldsymbol{\beta}$ in Proposition 2. It shows that knowledge of, or at least estimation, of Σ_i is necessary in order to construct a semiparametric efficient estimator. But, it is almost impossible to estimate Σ_i in a fully nonparametric way when no structure of Σ_i is assumed, noticing that each Σ_i has $m_i(m_i-1)/2$ parameters and the T_{ij}'s may differ from one cluster to another. Fortunately, for some data sets, it is reasonable to assume some structure on the covariance matrices such as

$$\Sigma_i = \Sigma(\underline{T}_i),\ i = 1, \ldots, n, \tag{3.1}$$

where the (j,j)th element of Σ_i is $\sigma^2(T_{ij})$ and the (j,j')th element is $\sigma(T_{ij}, T_{ij'})$, $j \neq j'$, for some smooth functions $\sigma^2(t)$ and $\sigma(s,t)$, respectively.

For each $i = 1, \ldots, n$, we estimate Σ_i by applying local linear regression under the assumption of (3.1) and denote the estimate by $\widehat{\Sigma}_i$. Our final estimator of $\boldsymbol{\beta}$, denoted by $\widehat{\boldsymbol{\beta}}_{\widehat{\Sigma}}$, is then obtained by taking $V_i = \widehat{\Sigma}_i$, $i = 1, \ldots, n$, in Section 2. We establish the asymptotic equivalence between $\widehat{\boldsymbol{\beta}}_{\widehat{\Sigma}}$ and $\widehat{\boldsymbol{\beta}}_{\Sigma}$ in Theorem 1 by exploiting some desirable properties of $\widehat{\Sigma}_i$ given in Proposition 4. Note that when m_i is fixed for all i and the T_{ij}'s are equispaced, we can estimate Σ_i without using the smoothing technique of this paper.

An estimator for $\boldsymbol{\beta}$ of a similar spirit was proposed by [10] for partially linear models for longitudinal data. However, it uses local linear regression in estimating the unknown function $\boldsymbol{g}_0(t)$, and thus some kind of iterative backfitting procedure is inevitable even for the identity link function. By contrast, since our procedure employes spline regression instead of local linear regression when estimating $\boldsymbol{g}_0(t)$, it requires no iterative calculation. In addition, it is easy to give an estimate of the covariance matrix of our estimator for $\boldsymbol{\beta}$ by just replacing Σ_i with $\widehat{\Sigma}_i$, for $i = 1, \ldots, n$.

Here we describe our assumptions on the smoothness of $\boldsymbol{g}_0(t)$, $\sigma^2(t)$ and $\sigma(s,t)$.

Assumption B.
(i) Assumption (3.1) holds.
(ii) $\boldsymbol{g}_0(t)$ is three times continuously differentiable on $[0,1]$.
(iii) $\sigma^2(t)$ is three times continuously differentiable on $[0,1]$.
(iv) $\sigma(s,t)$ is three times continuously differentiable on $[0,1]^2$.

Now we describe our estimation procedure. Let K be some continuously differentiable symmetric density function with compact support.

Step 1. Estimate $\boldsymbol{\beta}_0$ as in Section 2 with $V_i = \boldsymbol{I}_{m_i}$, $i = 1, \ldots, n$, and denote the resulting working independent estimate by $\widehat{\boldsymbol{\beta}}_I$.

Step 2. Estimate $\boldsymbol{g}_0(t)$ by applying local linear regression to $\{Y_{ij}-\boldsymbol{X}_{ij}^T\widehat{\boldsymbol{\beta}}_I,\ i=1,\ldots,n, j=1,\ldots,m_i\}$, using bandwidth $h_1=c_1n^{-a_h}$ where $1/6<a_h\leq 1/4$. We denote the resulting estimate by $\widehat{\boldsymbol{g}}(t)$, which is written as

(3.2)

$$\widehat{\boldsymbol{g}}(t)=D_q(A_{1n}(t))^{-1}\frac{1}{N_1h_1}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\boldsymbol{Z}_{ij}\otimes\begin{pmatrix}1\\ \frac{T_{ij}-t}{h_1}\end{pmatrix}K\Big(\frac{T_{ij}-t}{h_1}\Big)(Y_{ij}-\boldsymbol{X}_{ij}^T\widehat{\boldsymbol{\beta}}_I),$$

where $N_1=\sum_{i=1}^n m_i$, D_q is a $q\times(2q)$ matrix defined by $\boldsymbol{I}_q\otimes(1\ \ 0)$ and

$$A_{1n}(t)=\frac{1}{N_1h_1}\sum_{i=1}^{n}\sum_{j=1}^{m_i}(\boldsymbol{Z}_{ij}\boldsymbol{Z}_{ij}^T)\otimes\begin{pmatrix}1 & \frac{T_{ij}-t}{h_1}\\ \frac{T_{ij}-t}{h_1} & (\frac{T_{ij}-t}{h_1})^2\end{pmatrix}K\Big(\frac{T_{ij}-t}{h_1}\Big).$$

Step 3. Calculate the residuals, denoted as $\widehat{\epsilon}_{ij}$, $i=1,\ldots,n$ and $j=1,\ldots,m_i$, by

$$\widehat{\epsilon}_{ij}=Y_{ij}-\boldsymbol{X}_{ij}^T\widehat{\boldsymbol{\beta}}_I-\boldsymbol{Z}_{ij}^T\widehat{\boldsymbol{g}}(T_{ij}).$$

Step 4. Estimate $\sigma^2(t)$ by applying to the squared residuals local linear regression with bandwidth $h_2=c_2n^{-b_h}$, where $1/6<b_h\leq 1/4$. We denote the resulting estimate by $\widehat{\sigma^2}(t)$, which is expressed as

$$\widehat{\sigma^2}(t)=(1\,0)(A_{2n}(t))^{-1}\frac{1}{N_1h_2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\begin{pmatrix}1\\ \frac{T_{ij}-t}{h_2}\end{pmatrix}K\Big(\frac{T_{ij}-t}{h_2}\Big)(\widehat{\epsilon}_{ij})^2, \tag{3.3}$$

where

$$A_{2n}(t)=\frac{1}{N_1h_2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\begin{pmatrix}1 & \frac{T_{ij}-t}{h_2}\\ \frac{T_{ij}-t}{h_2} & (\frac{T_{ij}-t}{h_2})^2\end{pmatrix}K\Big(\frac{T_{ij}-t}{h_2}\Big).$$

Step 5. Estimate $\sigma(s,t)$ by applying to $\{\widehat{\epsilon}_{ij}\widehat{\epsilon}_{ij'}, j\neq j', i=1,\ldots,n\}$ local linear regression with bandwidth $h_3=c_3n^{-c_h}(1/6<c_h<1/4)$. We denote the resulting estimate by $\widehat{\sigma}(s,t)$, which has the following expression:

(3.4)

$$\begin{aligned}\widehat{\sigma}(s,t)&=(1\,0\,0)(A_{3n}(s,t))^{-1}\\ &\quad\times\frac{1}{N_2h_3^2}\sum_{i=1}^{n}\sum_{j\neq j'}\begin{pmatrix}1\\ \frac{T_{ij}-s}{h_3}\\ \frac{T_{ij'}-t}{h_3}\end{pmatrix}K\Big(\frac{T_{ij}-s}{h_3}\Big)K\Big(\frac{T_{ij'}-t}{h_3}\Big)\widehat{\epsilon}_{ij}\widehat{\epsilon}_{ij'},\end{aligned}$$

where $N_2 = \sum_{i=1}^n m_i(m_i-1)$ and

$$A_{3n}(s,t) = \frac{1}{N_2 h_3^2}\sum_i^n \sum_{j\neq j'} \begin{pmatrix} 1 \\ \frac{T_{ij}-s}{h_3} \\ \frac{T_{ij'}-t}{h_3} \end{pmatrix} \begin{pmatrix} 1 & \frac{T_{ij}-s}{h_3} & \frac{T_{ij'}-t}{h_3} \end{pmatrix} K\Big(\frac{T_{ij}-s}{h_3}\Big) K\Big(\frac{T_{ij'}-t}{h_3}\Big).$$

In general the covariance function $\widehat{\sigma}(s,t)$ is not positive semidefinite, and we can modify it by truncating the eigenfunctions in its spectral decomposition that have eigenvalues not exceeding some nonnegative constant λ_L.

Step 6. Calculate $\widehat{\Sigma}_i$ from steps **4** and **5** by letting

$$\widehat{\Sigma}_i(j,j') = \widehat{\sigma}(T_{ij},T_{ij'})I(j\neq j') + \widehat{\sigma^2}(T_{ij})I(j=j'),$$

and then estimate $\boldsymbol{\beta}_0$ with $V_i = \widehat{\Sigma}_i$ in the GEE (2.2). Denote the resulting estimate of $\boldsymbol{\beta}_0$ by $\widehat{\boldsymbol{\beta}}_{\widehat{\Sigma}}$.

We state some remarks on the above procedure; the first one is about the bandwidths; the second one is about estimation of $\sigma(s,t)$; the third one is about an alternative method of constructing residuals used in the estimation of Σ_i.

Remark 2. *When Assumption B holds, the conditions on the bandwidths h_1, h_2 and h_3 are not restrictive. For example, the optimal order of h_1 and h_2 is $n^{-1/5}$ which falls in the specified range. A larger bandwidth is recommended only for h_3 due to the two-dimensional smoothing. However, since the actual number of observations in local linear regression in steps* **4** *and* **5** *of the above procedure are N_1 and N_2, respectively, we anticipate that bandwidth choices will not seriously affect our final estimator.*

Remark 3. *When there is some kind of singularity of $\sigma(s,t)$ near the diagonal line, we may have to use only pairs of residuals $\widehat{\epsilon}_{ij}$ and $\widehat{\epsilon}_{ij'}$ with $T_{ij} < T_{ij'}$ when $s < t$. Then there may be some concern about the number of pairs for the two-dimensional smoothing, however.*

Remark 4. *When we calculate $\widehat{\boldsymbol{\beta}}_I$ by ordinary least squares as in (2.2), we get another set of residuals $Y_{ij} - \boldsymbol{X}_{ij}^T\widehat{\boldsymbol{\beta}}_I - \boldsymbol{W}_{ij}^T\widehat{\boldsymbol{\gamma}}_I$. We are then able to omit steps* **2** *and* **3** *of the above procedure by exploiting this set of residuals. Our simulation results summarized in Section 4 indicate that this approach is inferior to the proposed one, however.*

The asymptotic expression of $\widehat{\Sigma}_i$ is given in Proposition 4. We verify this proposition in the supplementary material [8]. Note that we need more elaborate representations than those in [10] since we deal with a linear regression model with a diverging dimension.

PROPOSITION 4. *(representations of the covariance estimators) Under the assumptions in Proposition 1 and Assumption B, we have the following representations of $\widehat{\sigma^2}(t)$ and $\widehat{\sigma}(s,t)$. Uniformly in t,*

$$\widehat{\sigma^2}(t) - \sigma^2(t) = B_1(t)h_2^2 + B_2(t)E_1(t) + O_p(h_1^3) + O_p\Big(\frac{\log n}{nh_1}\Big) + O_p(h_2^3) + O_p\Big(\frac{\log n}{nh_2}\Big),$$

where $B_1(t)$ and $B_2(t)$ are bounded functions, and

$$E_1(t) = \frac{1}{N_1 h_2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\begin{pmatrix}1\\ \frac{T_{ij}-t}{h_2}\end{pmatrix} K\Big(\frac{T_{ij}-t}{h_2}\Big)(\epsilon_{ij}^2 - \sigma^2(T_{ij})) = O_p\Big(\sqrt{\frac{\log n}{nh_2}}\Big) \quad uniformly\ in\ t.$$

Uniformly in s and t ($s \neq t$),

$$\widehat{\sigma}(s,t) - \sigma(s,t) = B_3(s,t)h_2^2 + B_4(s,t)E_2(s,t) + O_p(h_1^3) + O_p\Big(\frac{\log n}{nh_1}\Big) + O_p(h_3^3) + O_p\Big(\frac{\log n}{nh_3^2}\Big),$$

where $B_3(s,t)$ and $B_4(s,t)$ are bounded functions, and

$$E_2(s,t) = \frac{1}{N_2 h_3^2}\sum_{i=1}^{n}\sum_{j\neq j'}\begin{pmatrix}1\\ \frac{T_{ij}-s}{h_3}\\ \frac{T_{ij'}-t}{h_3}\end{pmatrix} K\Big(\frac{T_{ij}-s}{h_3}\Big)K\Big(\frac{T_{ij'}-t}{h_3}\Big)(\epsilon_{ij}\epsilon_{ij'} - \sigma(T_{ij},T_{ij'})) = O_p\Big(\sqrt{\frac{\log n}{nh_3^2}}\Big) \quad uniformly\ in\ s\ and\ t.$$

We can replace the three times continuously differentiability with the twice continuously differentiability and the Hölder continuity of the second derivatives of order α_1, α_2, and α_3 in assumptions B(ii), B(iii), and B(iv), respectively. In this case, the bandwidths in steps **2**, **4**, and **5** of our method have to satisfy the condition

$$\sqrt{n}(h_1^{2+\alpha_1} + h_2^{2+\alpha_2} + h_3^{2+\alpha_3}) \to 0.$$

Note that α_3 must be positive because step **5** of our procedure requires two-dimensional smoothing. Then we can prove similar results when $0 \le \alpha_1 < 1$, $0 \le \alpha_2 < 1$, and $0 < \alpha_3 < 1$. Specifically, the $O_p(h_j^3)$ terms in Proposition 4 will be replaced by $O_p(h_j^{2+\alpha_j})$, $j = 1, 2, 3$.

We state the desirable equivalence property of $\widehat{\boldsymbol{\beta}}_{\widehat{\Sigma}}$ in Theorem 1. In the proof, we evaluate the difference between $\widehat{\boldsymbol{\beta}}_{\widehat{\Sigma}}$ and $\widehat{\boldsymbol{\beta}}_{\Sigma}$ by applying Proposition 4 and explicit expressions of the estimators. The evaluation is technical and thus is postponed to Section 5.4. As for the semiparametric efficiency, see Remarks 1 after Proposition 3 and Remark 5.

THEOREM 1. *Under the assumptions in Proposition 1 and Assumption B, we have*

$$\widehat{\boldsymbol{\beta}}_{\widehat{\Sigma}} = \widehat{\boldsymbol{\beta}}_{\Sigma} + o_p(n^{-1/2}).$$

Even if (3.1) fails to hold, we still have the asymptotic equivalence in Theorem 1 and the asymptotic normality of Proposition 1 with $V_i = \mathrm{Var}(\underline{\epsilon}_i \,|\, \underline{T}_i)$ when $\mathrm{Var}(\underline{\epsilon}_i \,|\, \underline{T}_i)$ is represented by $\sigma^2(t)$ and $\sigma(s,t)$. We are still exploiting some information on Σ_i.

REMARK 5. *Note that $\widehat{\boldsymbol{\beta}}_{\widehat{\Sigma}}$ may not be semiparametric efficient under Assumption B and general error distributions, as mentioned in Remark 1. When the error vector $\underline{\epsilon}_i$ follows the* $\mathrm{N}(0, \Sigma_i)$ *distribution conditional on $\underline{\boldsymbol{X}}_i$, $\underline{\boldsymbol{Z}}_i$, and $\underline{T}_i$, $\widehat{\boldsymbol{\beta}}_{\widehat{\Sigma}}$ is semiparametric efficient.*

4. Numerical studies.

4.1. *Simulations.* In our simulation study summarized in this section, the data were generated from the following model:

$$Y_{ij} = \boldsymbol{X}_{ij}^T \boldsymbol{\beta}_0 + \boldsymbol{Z}_{ij}^T \boldsymbol{g}_0(T_{ij}) + \epsilon_i(T_{ij}),\ j = 1, \ldots, m_i,\ i = 1, \ldots, n,$$

with the first component of $\boldsymbol{Z}_{ij}$ being taken as 1. The number of observation time points in the ith subject was set as $m_i = m_0 + binomial(m_r, 0.65)$. The observation time points T_{ij} were uniformly distributed over the interval $[(j-1)/(m_0+m_r), j/(m_0+m_r)]$, $j = 1, \cdots, m_i$. We note that when $m_i = m_0 + m_r$, the subject is observed at all follow-up time points; when $m_i < m_0 + m_r$, the subject may be lost to follow up. We set $m_0 = 6$ and $m_r = 6$.

We generated the $(p+q-1)-$dimensional covariates from a multivariate Gaussian distribution, and we considered the following coefficients settings:

$p = 4$, $q = 4$, $\boldsymbol{\beta}_0 = (5, 5, -5, -5)^T$ and
$\boldsymbol{g}_0(t) = \big(3.5\sin(2\pi t), 5(1-t)^2, 3.5(\exp(-(3t-1)^2) + \exp(-(4t-3)^2)) - 1.5, 3.5t^{1/2}\big)^T$.

The random error process $\epsilon_i(t)$ was simulated from an ARMA$(1,1)$ Gaussian process with mean zero and covariance function $\text{cov}(\epsilon_i(s), \epsilon_i(t)) = \omega\rho^{|s-t|}$. We set $\omega = 4.95$ and considered $\rho = 0.4$ or 0.8.

The estimation results for various model components are summarized in Table 1. We considered two types of covariance structure: working independence covariance (under the column "Independent") and efficiently estimated covariance (under the column "Efficient"). For the sake of comparison, we also considered the oracle estimator (under the column "Oracle") where the true covariance is used. In addition, we considered using the covariance estimator with the crude raw residuals obtained from Step 1 and omitting Steps 2 and 3 (under the column "Crude"). Finally, we notice that we adjusted the covariance function $\widehat{\sigma}(s,t)$ by setting all negative eigenvalues to be zero. In contrast, we also considered a strictly positive threshold $\lambda_L = 0.05$ and set all eigenvalues lower than λ_L to be zero. This estimator is denoted by "Positive."

We report the estimation mean bias and mean standard error (SE) obtained from 100 repetitions in the table. In general, the efficient estimator could yield smaller estimation bias and variance, compared to the naive estimator assuming working independence. In particular, the standard error for the efficient estimator is only $20 \sim 50\%$ of that of the working independent estimator, indicating a remarkable reduction. In addition, we note that the efficient estimator has very similar performance to that of the oracle estimator.

For comparison, we also examined two estimators with different covariance estimation. The crude estimator is based on a simplified residual construction and therefore produces relatively less accurate covariance estimation. The estimation bias and standard error for the crude estimator are respectively larger than that for the efficient estimator. The positive estimator includes an adjustment in the estimation of the covariance function by setting eigenvalues lower than a positive cut-off to be zero while the efficient estimator only adjusts the negative eigenvalues. The resulting positive estimator is thus slightly more biased than the efficient estimator. In all cases, the crude and the positive estimators are still more efficient than the naive working independent estimators.

4.2. *Real Data Analysis.* We now present an application of our method to the CD4 count data from the AIDS Clinical Trial Group 193A Study [7]. The data came from a randomized, double-blind study of AIDS patients with CD4 counts of ≤ 50 cells/mm3. The patients were randomized to one of four treatments with roughly equal group sizes; each consisted of a daily

TABLE 1. *Estimation results of 100 simulations. "Independent" corresponds to $V_i = I$; "Efficient" refers to using estimated $\hat{\Sigma}_i$; "Oracle" refers to using the true Σ_i as V_i; "Crude" refers to using residuals directly from Step 1 to estimate the covariances; "Positive" means that we set a positive threshold for the covariance eigenvalues.*

			Independent		Efficient		Oracle		Crude		Positive	
n	ρ	Parameters	bias	SE	bias	SE	bias	SE	bias	SE	bias	SE
100	0.4	β_1	.0129	.0723	-.0014	.0265	-.0011	.0253	.0079	.0331	-.0072	.0324
		β_2	-.0200	.0708	-.0091	.0258	-.0074	.0248	-.0207	.0330	.0251	.0336
		β_3	-.0042	.0727	-.0058	.0265	-.0053	.0253	-.0131	.0330	.0210	.0321
		β_4	.0077	.0728	-.0136	.0268	-.0031	.0253	.0366	.0339	-.0047	.0325
200	0.4	β_1	-.0041	.0524	-.0046	.0169	-.0041	.0177	-.0139	.0253	-.0024	.0228
		β_2	-.0089	.0512	-.0020	.0168	-.0034	.0176	-.0106	.0254	-.0173	.0225
		β_3	-.0015	.0519	.0013	.0168	.0005	.0177	-.0166	.0250	-.0230	.0226
		β_4	-.0112	.0522	.0064	.0169	.0031	.0177	-.0297	.0250	-.0097	.0228
100	0.8	β_1	.0162	.0629	.0121	.0126	.0017	.0102	.0380	.0220	.0099	.0179
		β_2	-.0053	.0620	.0045	.0136	-.0006	.0101	.0447	.0232	-.0179	.0174
		β_3	.0092	.0636	-.0073	.0138	.0005	.0102	.0114	.0235	.0038	.0180
		β_4	-.0037	.0633	.0036	.0129	-.0021	.0102	-.0432	.0224	-.0041	.0173
200	0.8	β_1	.0071	.0493	-.0069	.0098	.0016	.0088	.0178	.0151	-.040	.0116
		β_2	-.0082	.0506	-.0055	.0099	-.0031	.0088	-.0156	.0134	.0024	.0116
		β_3	-.0052	.0495	.0055	.0102	.0003	.0088	.0176	.0126	.0047	.0115
		β_4	-.0119	.0497	-.0036	.0099	-.0004	.0088	-.0110	.0141	-.0032	.0115

regimen of 600 mg of zidovudine. Treatment 1 is zidovudine alternating monthly with 400 mg didanosine; Treatment 2 is zidovudine plus 225 mg of zalcitabine; Treatment 3 is zidovudine plus 400 mg of didanosine; Treatment 4 is a triple therapy consisting of zidovudine plus 400 mg of didanosine plus 400 mg of nevirapine. Measurements of CD4 counts were scheduled to be collected at baseline and at eight week intervals during the 40 weeks of follow-up. However, the real observation times were unbalanced due to mistimed measurements, skipped visits and dropouts. The number of measurements of CD4 counts during the 40 weeks of follow-up varied from 1 to 9, with a median of 4. The response variable was the log-transformed CD4 counts, Y =log(CD4 counts + 1). There was also gender and baseline age information about each patient. A total of 1309 patients were enrolled in the study. We eliminated the 122 patients who dropped out immediately after the baseline measurement.

We considered the following available covariates: treatment (coded by three indicator variables for treatment groups 2, 3 and 4, respectively), age (years), sex (coded as 1 for male and 0 for female), and interaction effects between these covariates. Using the group SCAD structural identification procedure of Cheng et al. (2014), we found that the coefficients for treatment 3, treatment 4 and the interaction between treatment 2 and sex are varying, and the coefficients given in Table 2 are constants. The group SCAD procedure also suggested that we remove all the other interaction effects. The estimated varying intercept (i.e. effect of treatment 1) and the varying coefficients are displayed in Figure 1 along with 95% confidence intervals. The constant coefficient estimates and their standard errors are provided in Table 2. To facilitate a comparison, we reported the results using the estimators assuming working independence and the efficient estimator proposed in this paper. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$ and $\underline{\boldsymbol{U}}_i = (\underline{\boldsymbol{X}}_i, \underline{\boldsymbol{W}}_i)$. In practice, the variances for the efficient parameter estimates were obtained from the first p diagonal elements of the following matrix:

$$\Big(\sum_{i=1}^{n} \underline{\boldsymbol{U}}_i^T \widehat{\Sigma}_i^{-1} \underline{\boldsymbol{U}}_i\Big)^{-1},$$

and for the working independent parameter estimates, the variances were obtained from the first p diagonal elements of the following matrix:

$$\Big(\sum_{i=1}^{n} \underline{\boldsymbol{U}}_i^T \underline{\boldsymbol{U}}_i\Big)^{-1} \sum_{i=1}^{n} \underline{\boldsymbol{U}}_i^T \widehat{\Sigma}_i \underline{\boldsymbol{U}}_i \Big(\sum_{i=1}^{n} \underline{\boldsymbol{U}}_i^T \underline{\boldsymbol{U}}_i\Big)^{-1}.$$

Eyeballing the estimation results in Table 2, we note that the estimated constant coefficients for treatment 2, age, and the interaction between treat-

TABLE 2
Estimation results of CD4 data.

	Independent		Efficient	
Covariates	Coefficients	SE	Coefficients	SE
treatment 2	.3614	.2257	.4038	.2027
age	.0946	.0274	.0818	.0245
sex	.1704	.1768	.2246	.1587
treatment 3:sex	-.2922	.2472	-.2908	.2209
treatment 4:sex	-.5321	.2416	-.5653	.2146

ment 4 and sex are all quite significant. The constant coefficient estimates for sex are not significant but are still kept in the model since we include the interactions between treatments and sex. The efficient estimates for all the constant and varying coefficients have smaller standard errors than the respective estimates assuming working independence. In fact, the Wald test statistic for the coefficient of treatment 2 is $.3614/.2257 = 1.60 < 1.96$ under the working independence, failing to declare a significant difference. On the other hand, the Wald test statistic for the same coefficient is $.4038/.2027 = 1.99 > 1.96$ from the efficient estimation, leading to a significant treatment difference. Other than these, because the sample size in this study was rather large, the two types of estimates for all the constant and varying coefficients appear to be very similar.

In general, the CD4 count tends to increase with age in the fitted model. Our estimation results suggest that there exist interaction effects between treatment and sex. Specifically, for the females (sex=0), subjects receiving treatments 2, 3 and 4 tend to have increasingly higher CD4 counts than those under treatment 1. The effect for treatment 2 (as compared with treatment 1) is estimated as a constant and is significant, while those for the other two treatment groups are varying (the upper right and the lower left panels in Figure 1) with even greater positive differences from treatment 1. For the males (sex=1), subjects receiving treatments 2, 3 and 4 also tend to have higher mean CD4 counts than those receiving treatment 1. The interaction between treatment 2 and sex is varying over time (the lower right panel in Figure 1) while those for treatments 3 and 4 are constant. The treatment effects from groups 3 and 4 are significantly different from treatment 1, judging from the values in Table 2. Also, we notice that the treatment differences seem to be greater among the females than among the males.

The estimated effects of four treatment groups are plotted in Figure 2 for the two estimators, respectively. Previous authors identified a similar pattern on the order of magnitude of the time-varying treatment effects [10]. However, they ignored the interaction between treatment and sex. Our

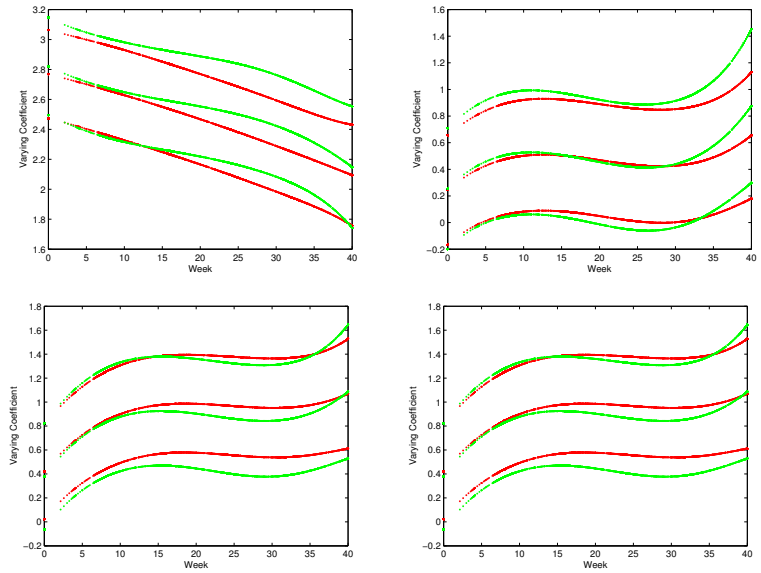

FIG 1. *Estimated varying-coefficients along with 95% confidence intervals for the intercept (upper left), treatment 3 (upper right), treatment 4 (lower left), and interaction between treatment 2 and sex (lower right). The red curves are efficient estimators while the green curves are estimators obtained under working independence.*

findings suggest the treatment effect curves might be quite different for the males and the females.

5. Proofs of main results.

5.1. *Assumptions and additional definitions.* We state some technical assumptions. In this paper, we denote the Euclidean norm of a vector a by $|a|$. Let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ stand for the minimum and maximum eigenvalues of a symmetric matrix A, respectively. Besides, C, C_1, C_2, ... are generic positive constants and the values may vary from line to line.

Assumption A1. We denote the density function of T_{ij}, $i = 1, \ldots, n$ and $j = 1, \cdots, m_i$, by $f_{ij}(t)$ and the joint density function of T_{ij} and $T_{ij'}$ $(j \neq j')$ by $f_{ijj'}(s,t)$. They are uniformly bounded and we have for some positive constants C_{A1} and C_{A2},

$$C_{A1} < \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} f_{ij}(t) < C_{A2} \text{ on } [0,1]$$

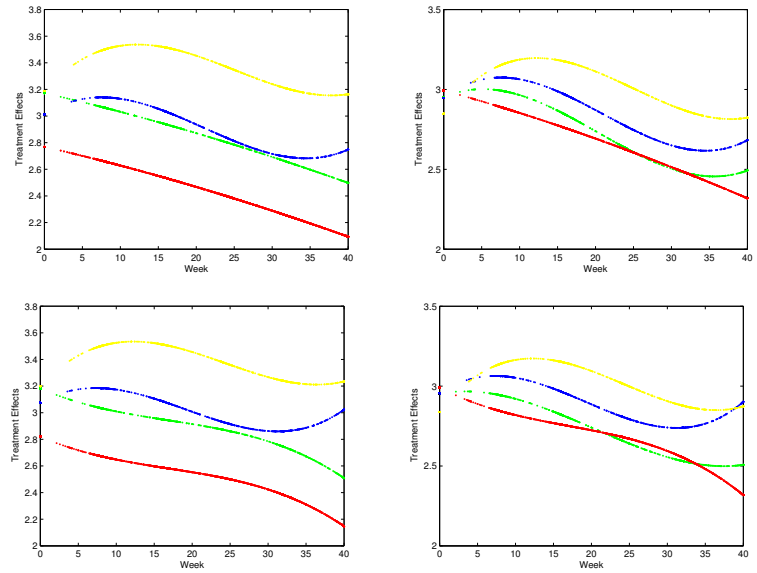

FIG 2. *Estimated treatment effects for the four treatment groups. The panels in the top row are the efficient estimates and those in the bottom row are the estimates assuming independence. The panels in the left column are for the females and those in the right column are for the males. Red, green, blue and yellow curves are for treatment groups 1, 2, 3 and 4, respectively.*

and

$$C_{A1} < \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq j'} f_{ijj'}(s,t) < C_{A2} \text{ on } [0,1]^2.$$

Assumption A2. For some positive constants C_{A3} and C_{A4}, we have uniformly in i and j,

$$C_{A3}\boldsymbol{I}_{p+q} \leq \mathrm{E}\left\{ \begin{pmatrix} \boldsymbol{X}_{ij}\boldsymbol{X}_{ij}^T & \boldsymbol{X}_{ij}\boldsymbol{Z}_{ij}^T \\ \boldsymbol{Z}_{ij}\boldsymbol{X}_{ij}^T & \boldsymbol{Z}_{ij}\boldsymbol{Z}_{ij}^T \end{pmatrix} \,\middle|\, T_{ij} \right\} \leq C_{A4}\boldsymbol{I}_{p+q}.$$

Assumption A3. For some positive constants C_{A5} and C_{A6}, we have in uniformly i,

$$C_{A5}\boldsymbol{I}_{m_i} \leq \lambda_{\min}(\Sigma_i) \leq \lambda_{\max}(\Sigma_i) \leq C_{A6}\boldsymbol{I}_{m_i}.$$

Assumption A4. For some positive constants C_{A7} and C_{A8}, we have in uniformly i,

$$C_{A7}\boldsymbol{I}_{m_i} \leq \lambda_{\min}(V_i) \leq \lambda_{\max}(V_i) \leq C_{A8}\boldsymbol{I}_{m_i}.$$

Assumption A5.
(i) $\mu(x)$ is twice continuously differentiable and $\inf_{x\in R}\mu'(x) > 0$.
(ii) For some positive constant C_{A9}, we have $\limsup_{|x|\to\infty}|\mu(x)|/|x|^{C_{A9}} < \infty$.
Assumption A6. For some positive constants C_{A10} and C_{A11}, we have uniformly in i,

$$\max_{1\le i\le n} C_{A10}\mathrm{E}\{\exp(|\epsilon_i|^2/C_{A10}) - 1|\underline{\boldsymbol{X}}_i, \underline{\boldsymbol{Z}}_i, \underline{T}_i\} \le C_{A11}.$$

We can prove Proposition 1 by just following closely the proof of Theorem 2 of [1], given in [2]. In this section, we describe only necessary definitions, assumptions, remarks on properties of projections, and changes necessary to prove Proposition 1.

We define the sup and L_2 norm on $\boldsymbol{G}$. For $\boldsymbol{g} = (g_1, \ldots, g_q)^T \in \boldsymbol{G}$, we define the norms by

$$\|\boldsymbol{g}\|_{G,\infty} = \sum_{j=1}^{q}\sup_{t\in[0,1]}|g_j(t)| \quad \text{and} \quad \|\boldsymbol{g}\|_{G,2}^2 = \sum_{j=1}^{q}\int_0^1 g_j^2(t)dt. \tag{5.1}$$

In Section 2, we defined $\langle\cdot,\cdot\rangle_n^{\Delta}$ and $\langle\cdot,\cdot\rangle^{\Delta}$. Here we define $\langle\cdot,\cdot\rangle_n$, $\langle\cdot,\cdot\rangle$, and the associated norms $\|\cdot\|_n$ and $\|\cdot\|$ by replacing Δ_{0i} with $\boldsymbol{I}_{m_i}$ in (2.3). Assumptions A1 and A2 imply there are positive constants C_1 and C_2 such that

$$C_1\|\boldsymbol{g}\|_{G,2} \le \min\{\|\boldsymbol{Z}^T\boldsymbol{g}\|^{\Delta}, \|\boldsymbol{Z}^T\boldsymbol{g}\|\} \le \max\{\|\boldsymbol{Z}^T\boldsymbol{g}\|^{\Delta}, \|\boldsymbol{Z}^T\boldsymbol{g}\|\} \le C_2\|\boldsymbol{g}\|_{G,2} \tag{5.2}$$

for any $\boldsymbol{g} \in \boldsymbol{G}$. The details are given in Lemma 1.

In (2.5), we defined two kinds of projections of X_k. We also define another one here by

$$\widehat{\boldsymbol{\varphi}}_{Vk} = \widehat{\Pi}_{Vn}X_k, \tag{5.3}$$

where

$$\widehat{\Pi}_{Vn}X_k = \operatorname*{argmin}_{\boldsymbol{g}\in\boldsymbol{G}_B}\|X_k - \boldsymbol{Z}^T\boldsymbol{g}\|_n^{\Delta}.$$

Now we have three kinds of projections :

$$\begin{aligned}
\boldsymbol{\varphi}_{Vk}^* &= \Pi_V X_k = \operatorname*{argmin}_{\boldsymbol{g}\in\boldsymbol{G}}\|X_k - \boldsymbol{Z}^T\boldsymbol{g}\|^{\Delta},\\
\overline{\boldsymbol{\varphi}}_{Vk} &= \Pi_{Vn}X_k = \operatorname*{argmin}_{\boldsymbol{g}\in\boldsymbol{G}_B}\|X_k - \boldsymbol{Z}^T\boldsymbol{g}\|^{\Delta},\\
\widehat{\boldsymbol{\varphi}}_{Vk} &= \widehat{\Pi}_{Vn}X_k = \operatorname*{argmin}_{\boldsymbol{g}\in\boldsymbol{G}_B}\|X_k - \boldsymbol{Z}^T\boldsymbol{g}\|_n^{\Delta}.
\end{aligned}$$

Note that $\overline{\boldsymbol{\varphi}}_{Vk}(t)$ is denoted by $\boldsymbol{\varphi}_{k,n}^*(t)$ in [1].

5.2. *Spline approximation and projections.* Now we consider the approximation error by spline functions. We have assumed all the relevant functions are twice continuously differentiable and that they and their second order derivatives are uniformly bounded. Hence the sup norm of approximation errors by spline functions is bounded from above by $C_{approx}K_n^{-2}$, where C_{approx} depends on the relevant functions. See Corollary 6.26 of [12].

In this section, we also establish existence and properties of the projections $\boldsymbol{\varphi}^*_{Vk}$, $k = 1, \ldots, p$. Note that $\langle \cdot, \cdot \rangle^\Delta$ and $\|\cdot\|^\Delta$ are defined on $\{v \mid \sum_{i,j} \mathrm{E}(v_{ij}^2) < \infty\}$ and that $\{\boldsymbol{Z}^T\boldsymbol{g}\}$ is a closed linear subspace due to (5.2). Therefore the projections $\boldsymbol{\varphi}^*_{Vk} = (\varphi^*_{Vk1}, \ldots, \varphi^*_{Vkq})^T$, $k = 1, \ldots, p$, uniquely exist. Next we drive integral equations for $\boldsymbol{\varphi}^*_{Vk}(t)$. We set

(5.4)
$$V_i^{-1} = (v_i^{j_1 j_2}),\ \mu'_{ij} = \mu'(\boldsymbol{X}_{ij}^T\boldsymbol{\beta}_0 + \boldsymbol{Z}_{ij}^T\boldsymbol{g}_0(T_{ij})),\ \text{and}\ \lambda_{ij_1j_2} = \mu'_{ij_1} v_i^{j_1j_2} \mu'_{ij_2}.$$

By representing (2.6) explicitly, we can derive the following integral equations for $\boldsymbol{\varphi}^*_{Vk}(t)$. For $d_1 = 1, \ldots, q$,

$$\sum_{d_2=1}^{q} a_{d_2}^{(d_1)}(t)\varphi^*_{Vkd_2}(t) = b^{(d_1)}(t) + \int_0^1 \sum_{d_2=1}^{q} c_{d_2}^{(d_1)}(s,t)\varphi^*_{Vkd_2}(s)ds, \tag{5.5}$$

where

$$a_{d_2}^{(d_1)}(t) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} \mathrm{E}\{Z_{ijd_2}\lambda_{ijj}Z_{ijd_1} \mid T_{ij} = t\} f_{ij}(t),$$

$$b^{(d_1)}(t) = \sum_{i=1}^{n} \sum_{1\le j_1, j_2 \le m_i} \mathrm{E}\{X_{ij_1k}\lambda_{ij_1j_2}Z_{ij_2d_1} \mid T_{ij_2} = t\} f_{ij_2}(t),$$

$$c_{d_2}^{(d_1)}(s,t) = -\sum_{i=1}^{n} \sum_{j_1 \ne j_2} \mathrm{E}\{Z_{ij_1d_2}\lambda_{ij_1j_2}Z_{ij_2d_1} \mid T_{ij_1} = s, T_{ij_2} = t\} f_{ij_1j_2}(s,t).$$

Let $A(t)$ be a $q \times q$ matrix whose (d_1, d_2)th element is $a_{d_2}^{(d_1)}(t)$. Assumptions A1 and A2 imply that $|A(t)| \ne 0$ on $[0, 1]$ and we set

$$\psi^*_{Vkd_1}(t) = \sum_{d_2=1}^{q} a_{d_2}^{(d_1)}(t)\varphi^*_{Vkd_2}(t).$$

Then (5.5) reduces to (S.2) of [2] and the same argument there applies. Therefore $\boldsymbol{\varphi}^*_{Vk}(t)$ will have the required smoothness properties with similar regularity conditions.

5.3. *Remarks on the proof of Proposition 1.* We have only to proceed as in [2] by replacing their $\boldsymbol{Z}_{ij}$, $\boldsymbol{Z}_i$, and $\boldsymbol{\varphi}_k^*(\boldsymbol{t})$ with $\boldsymbol{W}_{ij}$, $\underline{\boldsymbol{W}}_i$, and $\boldsymbol{Z}^T\boldsymbol{\varphi}_{Vk}^*(t)$, respectively. We have already defined necessary norms and projections. We just state the relevant changes and remarks:

(i) Lemmas S.2-S.4 of [2]: We reorganize these lemmas in Lemma 1 given later. Its (i)-(iii), (iv), and (vi) correspond to Lemma S.2, the latter half of Lemma S.3, and Lemma S.4 of [2], respectively. The former half of Lemma S.3 of [2] seems to be used in their Corollary 1. However, it can be relaxed to (v) of Lemma 1 here.

(ii) Lemma S.8 of [2]: Our regressors $\boldsymbol{X}_{ij}$ and $\boldsymbol{W}_{ij}$ still form a VC class and we can proceed completely in the same way as in [2].

We state Lemma 1 in the following. The proof is given in the supplementary material [8].

Lemma 1.

(i) There are positive constants C_1 and C_2 such that

$$C_1\|\boldsymbol{g}\|_{G,2} \le \min\{\|\boldsymbol{Z}^T\boldsymbol{g}\|^\Delta, \|\boldsymbol{Z}^T\boldsymbol{g}\|\} \le \max\{\|\boldsymbol{Z}^T\boldsymbol{g}\|^\Delta, \|\boldsymbol{Z}^T\boldsymbol{g}\|\} \le C_2\|\boldsymbol{g}\|_{G,2}$$

for any $\boldsymbol{g} \in \boldsymbol{G}$.

(ii) There are positive constants C_3 and C_4 such that

$$\|\boldsymbol{g}\|_{G,\infty}^2 \le C_3K_n\|\boldsymbol{g}\|_{G,2}^2 \le C_4K_n(\|\boldsymbol{Z}^T\boldsymbol{g}\|^\Delta)^2$$

for any $\boldsymbol{g} \in \boldsymbol{G}_B$.

(iii) There is a positive constant C_5 such that for any $\boldsymbol{\beta} \in R^p$ and $\boldsymbol{g} \in \boldsymbol{G}_B$,

$$\|\boldsymbol{X}^T\boldsymbol{\beta} + \boldsymbol{Z}^T\boldsymbol{g}\|_\infty \le C_5K_n^{1/2}\|\boldsymbol{X}^T\boldsymbol{\beta} + \boldsymbol{Z}^T\boldsymbol{g}\|^\Delta,$$

where $\|v\|_\infty = \max_{i,j}|v_{ij}|$.

(iv)

$$\sup_{\boldsymbol{g}_1,\boldsymbol{g}_2\in\boldsymbol{G}_B}\Big|\frac{\langle\boldsymbol{Z}^T\boldsymbol{g}_1, \boldsymbol{Z}^T\boldsymbol{g}_2\rangle_n^\Delta - \langle\boldsymbol{Z}^T\boldsymbol{g}_1, \boldsymbol{Z}^T\boldsymbol{g}_2\rangle^\Delta}{\|\boldsymbol{Z}^T\boldsymbol{g}_1\|^\Delta\|\boldsymbol{Z}^T\boldsymbol{g}_2\|^\Delta}\Big| = O_p(K_n\sqrt{\log n/n}).$$

(v) For any positive constant M, we have

$$\langle x_j - \boldsymbol{Z}^T\boldsymbol{g}_j, x_k - \boldsymbol{Z}^T\boldsymbol{g}_k\rangle_n^\Delta - \langle x_j - \boldsymbol{Z}^T\boldsymbol{g}_j, x_k - \boldsymbol{Z}^T\boldsymbol{g}_k\rangle^\Delta = o_p(1)$$

uniformly in $\boldsymbol{g}_j \in \boldsymbol{G}_B$ and $\boldsymbol{g}_k \in \boldsymbol{G}_B$ satisfying $\|\boldsymbol{g}_j\|_{G,2} \le M$ and $\|\boldsymbol{g}_k\|_{G,2} \le M$, respectively.

(vi) If $\|\delta_n\|_\infty$ is uniformly bounded in n and $\{\delta_{n,ij}\}_{j=1}^{m_i}$ are mutually independent in i, we have

$$\sup_{\boldsymbol{g}\in\boldsymbol{G}_B}\Big|\frac{\langle\delta_n, \boldsymbol{Z}^T\boldsymbol{g}\rangle_n^\Delta - \langle\delta_n, \boldsymbol{Z}^T\boldsymbol{g}\rangle^\Delta}{\|\boldsymbol{Z}^T\boldsymbol{g}\|^\Delta}\Big| = O_p(\sqrt{K_n/n})\|\delta_n\|_\infty.$$

5.4. *Proof of Theorem 1.* When we consider the identity link function, we have explicit expressions of $\widehat{\boldsymbol{\beta}}_\Sigma - \boldsymbol{\beta}_0$ and $\widehat{\boldsymbol{\beta}}_{\widehat{\Sigma}} - \boldsymbol{\beta}_0$:

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_\Sigma - \boldsymbol{\beta}_0 =& H^{11}\sum_{i=1}^n (\underline{\boldsymbol{X}}_i - \underline{\boldsymbol{W}}_i H_{22}^{-1}H_{21})^T \Sigma_i^{-1}\underline{\epsilon}_i \\ &- H^{11}\sum_{i=1}^n (\underline{\boldsymbol{X}}_i - \underline{\boldsymbol{W}}_i H_{22}^{-1}H_{21})^T \Sigma_i^{-1}(\underline{\boldsymbol{W}}_i\boldsymbol{\gamma}^* - \underline{(\boldsymbol{Z}^T\boldsymbol{g}_0)}_i) \\ =& I_1 - I_2 \quad \text{(say)}\end{aligned} \tag{5.6}$$

and

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{\widehat{\Sigma}} - \boldsymbol{\beta}_0 =& \widehat{H}^{11}\sum_{i=1}^n (\underline{\boldsymbol{X}}_i - \underline{\boldsymbol{W}}_i \widehat{H}_{22}^{-1}\widehat{H}_{21})^T \widehat{\Sigma}_i^{-1}\underline{\epsilon}_i \\ &- \widehat{H}^{11}\sum_{i=1}^n (\underline{\boldsymbol{X}}_i - \underline{\boldsymbol{W}}_i \widehat{H}_{22}^{-1}\widehat{H}_{21})^T \widehat{\Sigma}_i^{-1}(\underline{\boldsymbol{W}}_i\boldsymbol{\gamma}^* - \underline{(\boldsymbol{Z}^T\boldsymbol{g}_0)}_i) \\ =& \widehat{I}_1 - \widehat{I}_2 \quad \text{(say)},\end{aligned} \tag{5.7}$$

where $\widehat{H}^{11}$, $\widehat{H}_{22}$, $\widehat{H}_{21}$ are defined as in (2.8) with $V_i = \widehat{\Sigma}_i^{-1}$, $i = 1, \ldots, n$, and $\boldsymbol{\gamma}^* = (\boldsymbol{\gamma}_1^{*T}, \ldots, \boldsymbol{\gamma}_q^{*T})^T$ satisfies $|\boldsymbol{B}^T(t)\boldsymbol{\gamma}_j^* - g_{0j}(t)| \le C_g K_n^{-2}$, $j = 1, \ldots, q$, for some C_g. Note that this C_g depends on $\boldsymbol{g}_0(t)$. Proposition 4 and Assumption A3 imply that with probability tending to 1, we have

$$C_1\boldsymbol{I}_{m_i} \le \widehat{\Sigma}_i \le C_2\boldsymbol{I}_{m_i}$$

uniformly in i for some positive constants C_1 and C_2. As for $\widehat{\Sigma}_i^{-1}$, we have

$$\begin{aligned}\widehat{\Sigma}_i^{-1} - \Sigma_i^{-1} &= \widehat{\Sigma}_i^{-1}(\Sigma_i - \widehat{\Sigma}_i)\Sigma_i^{-1} \\ &= \Sigma_i^{-1}(\Sigma_i - \widehat{\Sigma}_i)\Sigma_i^{-1} + \widehat{\Sigma}_i^{-1}(\Sigma_i - \widehat{\Sigma}_i)\Sigma_i^{-1}(\Sigma_i - \widehat{\Sigma}_i)\Sigma_i^{-1}.\end{aligned}$$

It follows from Proposition 4 and the above identity that

$$\widehat{\Sigma}_i^{-1} - \Sigma_i^{-1} = \Sigma_i^{-1}(\Sigma_i - \widehat{\Sigma}_i)\Sigma_i^{-1} + O_p\Big(h_2^4 + h_3^4 + \frac{\log n}{nh_2} + \frac{\log n}{nh_3^2}\Big). \tag{5.8}$$

Since m_i is bounded, (5.8) holds both componentwise and in the sense of eigenvalue evaluation. Besides, Proposition 4 implies that each element of $\Sigma_i^{-1}(\Sigma_i - \widehat{\Sigma}_i)\Sigma_i^{-1}$ has the form of

$$D_i^{(1)}(\underline{T}_i)h_2^2 + D_i^{(2)}(\underline{T}_i)h_3^2 + \sum_{j=1}^{m_i} D_{ij}^{(3)}(\underline{T}_i)E_1(T_{ij}) + \sum_{j\neq j'} D_{ijj'}^{(4)}(\underline{T}_i)E_2(T_{ij}, T_{ij'}) + D_i^{(5)}, \tag{5.9}$$

where $D_i^{(1)}$, $D_i^{(2)}$, $D_{ij}^{(3)}$, and $D_{ijj'}^{(4)}$ are uniformly bounded functions in i, and

$$D_i^{(5)} = O_p\Big(h_1^3 + h_2^3 + h_3^3 + \frac{\log n}{nh_1} + \frac{\log n}{nh_2} + \frac{\log n}{nh_3^2}\Big)$$

uniformly in i.

Under the assumptions for Theorem 1 and from the local property of the B-spline basis, (5.8), and (5.9) we have the following lemmas. These lemmas are needed in order to evaluate $\widehat{I}_1 - I_1$ and their proofs are given in the supplementary material [8].

Recall that

$$\frac{1}{n}H_{12} = \frac{1}{n}\sum_{i=1}^{n} \underline{\boldsymbol{X}}_i^T \Sigma_i^{-1} \underline{\boldsymbol{W}}_i.$$

Lemma 2 is about this $p \times qK_n$ matrix. Let $h_{12,kl}$ and $\widehat{h}_{12,kl}$ be the (k,l) element of H_{12} and $\widehat{H}_{12}$, respectively.

LEMMA 2. *We have uniformly in k and l,*

$$\frac{1}{n}h_{12,kl} = O_p(K_n^{-1}),$$

$$\frac{1}{n}(h_{12,kl} - \widehat{h}_{12,kl}) = K_n^{-1} O_p\Big(h_2^2 + h_3^2 + \sqrt{\frac{\log n}{nh_2}} + \sqrt{\frac{\log n}{nh_3^2}}\Big),$$

$$\Big\{\sum_{l=1}^{qK_n} (n^{-1}h_{12,kl})^2\Big\}^{1/2} = O_p(K_n^{-1/2}),$$

$$\Big[\sum_{l=1}^{qK_n} \{n^{-1}(h_{12,kl} - \widehat{h}_{12,kl})\}^2\Big]^{1/2} = K_n^{-1/2} O_p\Big(h_2^2 + h_3^2 + \sqrt{\frac{\log n}{nh_2}} + \sqrt{\frac{\log n}{nh_3^2}}\Big).$$

Recall that

$$\frac{1}{n}H_{22} = \frac{1}{n}\sum_{i=1}^{n} \underline{\boldsymbol{W}}_i^T \Sigma_i^{-1} \underline{\boldsymbol{W}}_i.$$

Lemma 3 is about the eigenvalues of the $qK_n \times qK_n$ matrices $\frac{1}{n}H_{22}$ and $\frac{1}{n}(\widehat{H}_{22} - H_{22})$.

LEMMA 3. *With probability tending to 1, we have*

$$C_1 K_n^{-1} \le \lambda_{\min}(n^{-1}H_{22}) \le \lambda_{\max}(n^{-1}H_{22}) \le C_2 K_n^{-1}$$

for some positive constants C_1 and C_2. We also have

$$\begin{aligned}&\max\left\{|\lambda_{\min}(n^{-1}(\widehat{H}_{22}-H_{22}))|,|\lambda_{\max}(n^{-1}(\widehat{H}_{22}-H_{22}))|\right\}\\ &= K_n^{-1}O_p\Big(h_2^2+h_3^2+\sqrt{\frac{\log n}{nh_2}}+\sqrt{\frac{\log n}{nh_3^2}}\Big).\end{aligned}$$

Then we have

$$\max\left\{|\lambda_{\min}(n^{-1}\widehat{H}_{22})|,|\lambda_{\max}(n^{-1}\widehat{H}_{22})|\right\}=O_p(K_n^{-1})$$

and $\max\left\{|\lambda_{\min}((n^{-1}\widehat{H}_{22})^{-1}-(n^{-1}H_{22})^{-1})|,|\lambda_{\max}((n^{-1}\widehat{H}_{22})^{-1}-(n^{-1}H_{22})^{-1})|\right\}$ *is also bounded from above by* $K_nO_p\Big(h_2^2+h_3^2+\sqrt{\frac{\log n}{nh_2}}+\sqrt{\frac{\log n}{nh_3^2}}\Big)$.

Recall that

$$\frac{1}{n}H_{11}=\frac{1}{n}\sum_{i=1}^n\underline{\boldsymbol{X}}_i^T\Sigma_i^{-1}\underline{\boldsymbol{X}}_i\quad\text{and}\quad H^{11}=(H_{11}-H_{12}H_{22}^{-1}H_{21})^{-1}.$$

Lemma 4 is about the $p\times p$ matrices $\frac{1}{n}\widehat{H}_{11}$, $\frac{1}{n}\widehat{H}_{12}H_{22}^{-1}H_{21}$, and $n\widehat{H}^{11}$.

Lemma 4.

$$\begin{aligned}\frac{1}{n}\widehat{H}_{11}&=\frac{1}{n}H_{11}+o_p(1)\\ \frac{1}{n}\widehat{H}_{12}\Big(\frac{1}{n}\widehat{H}_{22}\Big)^{-1}\frac{1}{n}\widehat{H}_{21}&=\frac{1}{n}H_{12}\Big(\frac{1}{n}H_{22}\Big)^{-1}\frac{1}{n}H_{21}+o_p(1),\end{aligned}$$

where $o_p(1)$ means both componentwise and in the meaning of eigenvalue evaluation. Hence we have

$$n\widehat{H}^{11}=nH^{11}+o_p(1).$$

Lemma 5 is about the qK_n-dimensional vectors $\frac{1}{\sqrt{n}}\sum_{i=1}^n\underline{\boldsymbol{W}}_i^T\Sigma_i^{-1}\underline{\epsilon}_i$ and $\frac{1}{\sqrt{n}}\sum_{i=1}^n\underline{\boldsymbol{W}}_i^T(\widehat{\Sigma}_i^{-1}-\Sigma_i^{-1})\underline{\epsilon}_i$.

Lemma 5. *We have for some positive constants C_1 and C_2,*

$$\frac{C_1}{K_n}\boldsymbol{I}_{qK_n}\le\operatorname{cov}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n\underline{\boldsymbol{W}}_i^T\Sigma_i^{-1}\underline{\epsilon}_i\Big)\le\frac{C_2}{K_n}\boldsymbol{I}_{qK_n}.$$

In addition we have

$$\begin{aligned}
&\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^T(\widehat{\Sigma}_i^{-1}-\Sigma_i^{-1})\underline{\epsilon}_i\Big| \\
&= \sqrt{\frac{n}{K_n}}O_p\Big(\frac{\log n}{nh_1}+\frac{\log n}{nh_2}+\frac{\log n}{nh_3^2}\Big)+\sqrt{\frac{n}{K_n}}O_p(h_1^3+h_2^3+h_3^3) \\
&\quad +O_p(h_2^2+h_3^2)+O_p\Big(\frac{1}{\sqrt{nh_2}}+\frac{1}{\sqrt{nh_3^2}}+\frac{1}{\sqrt{nK_n}h_2}+\frac{1}{\sqrt{nK_n}h_3^2}\Big).
\end{aligned}$$

Lemma 6 is about the p-dimensional vectors $\frac{1}{\sqrt{n}}\sum_{i=1}^n \underline{\boldsymbol{X}}_i^T\Sigma_i^{-1}\underline{\epsilon}_i$ and $\frac{1}{\sqrt{n}}\sum_{i=1}^n \underline{\boldsymbol{X}}_i^T(\widehat{\Sigma}_i^{-1}-\Sigma_i^{-1})\underline{\epsilon}_i$.

LEMMA 6. *We have for some positive constants C_1 and C_2,*

$$C_1\boldsymbol{I}_p \le \operatorname{cov}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underline{\boldsymbol{X}}_i^T\Sigma_i^{-1}\underline{\epsilon}_i\Big) \le C_2\boldsymbol{I}_p.$$

In addition we have

$$\begin{aligned}
&\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underline{\boldsymbol{X}}_i^T(\widehat{\Sigma}_i^{-1}-\Sigma_i^{-1})\underline{\epsilon}_i\Big| \\
&= \sqrt{n}O_p\Big(\frac{\log n}{nh_1}+\frac{\log n}{nh_2}+\frac{\log n}{nh_3^2}\Big)+\sqrt{n}O_p(h_1^3+h_2^3+h_3^3) \\
&\quad +O_p(h_2^2+h_3^2)+O_p\Big(\frac{1}{\sqrt{n}h_2}+\frac{1}{\sqrt{n}h_3^2}\Big).
\end{aligned}$$

Now we prove that $\widehat{I}_1 - I_1 = o_p(n^{-1/2})$. Write

$$I_1 = H^{11}\sum_{i=1}^{n}\underline{\boldsymbol{X}}_i^T\Sigma_i^{-1}\underline{\epsilon}_i - H^{11}H_{12}H_{22}^{-1}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^T\Sigma_i^{-1}\underline{\epsilon}_i = H^{11}(I_{11}-I_{12}) \text{ (say)}.$$

We define $\widehat{I}_{11}$ and $\widehat{I}_{12}$ similarly. From Proposition 1 and Lemma 4, we have only to prove

$$\frac{1}{\sqrt{n}}(\widehat{I}_{11}-I_{11}) = o_p(1) \quad \text{and} \quad \frac{1}{\sqrt{n}}(\widehat{I}_{12}-I_{12}) = o_p(1). \tag{5.10}$$

The former result in (5.10) can be handled in the same way as the latter

and we consider only the latter. Write

$$\begin{aligned}\frac{1}{\sqrt{n}}(\widehat{I}_{12}-I_{12}) &= \frac{1}{n}\widehat{H}_{12}\Big(\frac{1}{n}\widehat{H}_{22}\Big)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^T(\widehat{\Sigma}_i^{-1}-\Sigma_i^{-1})\underline{\epsilon}_i \\ &\quad + \frac{1}{n}\widehat{H}_{12}\Big\{\Big(\frac{1}{n}\widehat{H}_{22}\Big)^{-1}-\Big(\frac{1}{n}H_{22}\Big)^{-1}\Big\}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^T\Sigma_i^{-1}\underline{\epsilon}_i \\ &\quad + \Big(\frac{1}{n}\widehat{H}_{12}-\frac{1}{n}H_{12}\Big)\Big(\frac{1}{n}H_{22}\Big)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^T\Sigma_i^{-1}\underline{\epsilon}_i \\ &= DI_{12}^{(1)}+DI_{12}^{(2)}+DI_{12}^{(3)} \quad \text{(say)}.\end{aligned}$$

Lemmas 2, 3, and 5 imply

$$\begin{aligned}DI_{12}^{(1)} =& \sqrt{n}O_p\Big(\frac{\log n}{nh_1}+\frac{\log n}{nh_2}+\frac{\log n}{nh_3^2}\Big)+\sqrt{n}O_p(h_1^3+h_2^3+h_3^3) \\ &+\sqrt{K_n}O_p\Big(\frac{1}{\sqrt{nh_2}}+\frac{1}{\sqrt{nh_3^2}}+\frac{1}{\sqrt{nK_nh_2}}+\frac{1}{\sqrt{nK_nh_3^2}}\Big) \\ &+\sqrt{K_n}O_p(h_2^2+h_3^2)=o_p(1), \\ DI_{12}^{(2)} =& \sqrt{K_n}O_p\Big(h_2^2+h_3^2+\sqrt{\frac{\log n}{nh_2}}+\sqrt{\frac{\log n}{nh_3^2}}\Big)=o_p(1), \\ DI_{12}^{(3)} =& \sqrt{K_n}O_p\Big(h_2^2+h_3^2+\sqrt{\frac{\log n}{nh_2}}+\sqrt{\frac{\log n}{nh_3^2}}\Big)=o_p(1).\end{aligned}$$

Hence we have established

$$\widehat{I}_1 - I_1 = o_p(n^{-1/2}). \tag{5.11}$$

Next we deal with $\widehat{I}_2 - I_2$ and two more lemmas are necessary. We evaluate the effects of the approximation bias in the lemmas.

LEMMA 7. *We have*

$$\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^T\Sigma_i^{-1}(\underline{\boldsymbol{W}}_i\boldsymbol{\gamma}^*-\underline{(\boldsymbol{Z}^T\boldsymbol{g}_0)}_i)\Big| = O_p(\sqrt{n}K_n^{-5/2})$$

and

$$\begin{aligned}&\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^T(\widehat{\Sigma}_i^{-1}-\Sigma_i^{-1})(\underline{\boldsymbol{W}}_i\boldsymbol{\gamma}^*-\underline{(\boldsymbol{Z}^T\boldsymbol{g}_0)}_i)\Big| \\ &= \sqrt{n}K_n^{-5/2}O_p\Big(h_2^2+h_3^2+\sqrt{\frac{\log n}{nh_2}}+\sqrt{\frac{\log n}{nh_3^2}}\Big).\end{aligned}$$

LEMMA 8. *We have*

$$\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underline{\boldsymbol{X}}_i^T\Sigma_i^{-1}(\underline{\boldsymbol{W}}_i\boldsymbol{\gamma}^*-\underline{(\boldsymbol{Z}^T\boldsymbol{g}_0)}_i)\Big| = O_p(\sqrt{n}K_n^{-2})$$

and

$$\begin{aligned}&\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underline{\boldsymbol{X}}_i^T(\widehat{\Sigma}_i^{-1}-\Sigma_i^{-1})(\underline{\boldsymbol{W}}_i\boldsymbol{\gamma}^*-\underline{(\boldsymbol{Z}^T\boldsymbol{g}_0)}_i)\Big|\\ &= \sqrt{n}K_n^{-2}O_p\Big(h_2^2+h_3^2+\sqrt{\frac{\log n}{nh_2}}+\sqrt{\frac{\log n}{nh_3^2}}\Big).\end{aligned}$$

Now we can show that $\widehat{I}_2 - I_2 = o_p(n^{-1/2})$. Write

$$\begin{aligned}I_2 =& H^{11}\sum_{i=1}^{n}\underline{\boldsymbol{X}}_i^T\Sigma_i^{-1}(\underline{\boldsymbol{W}}_i\boldsymbol{\gamma}^*-\underline{(\boldsymbol{Z}^T\boldsymbol{g}_0)}_i)\\ &- H^{11}H_{12}H_{22}^{-1}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^T\Sigma_i^{-1}(\underline{\boldsymbol{W}}_i\boldsymbol{\gamma}^*-\underline{(\boldsymbol{Z}^T\boldsymbol{g}_0)}_i)\\ =& H^{11}(I_{21}-I_{22}) \text{ (say)}.\end{aligned}$$

We define $\widehat{I}_{21}$ and $\widehat{I}_{22}$ similarly and write $\widehat{I}_2 = \widehat{H}^{11}(\widehat{I}_{21}-\widehat{I}_{22})$. From Proposition 1 and Lemma 4, we have only to prove

$$\frac{1}{\sqrt{n}}(\widehat{I}_{21}-I_{21}) = o_p(1) \quad \text{and} \quad \frac{1}{\sqrt{n}}(\widehat{I}_{22}-I_{22}) = o_p(1).$$

The former result in the above can be handled in the same way as the latter and we consider only the latter. Write

$$\begin{aligned}&\frac{1}{\sqrt{n}}(\widehat{I}_{22}-I_{22})\\ &= \frac{1}{n}\widehat{H}_{12}\Big(\frac{1}{n}\widehat{H}_{22}\Big)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^T(\widehat{\Sigma}_i^{-1}-\Sigma_i^{-1})(\underline{\boldsymbol{W}}_i\boldsymbol{\gamma}^*-\underline{(\boldsymbol{Z}^T\boldsymbol{g}_0)}_i)\\ &\quad+\frac{1}{n}\widehat{H}_{12}\Big\{\Big(\frac{1}{n}\widehat{H}_{22}\Big)^{-1}-\Big(\frac{1}{n}H_{22}\Big)^{-1}\Big\}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^T\Sigma_i^{-1}(\underline{\boldsymbol{W}}_i\boldsymbol{\gamma}^*-\underline{(\boldsymbol{Z}^T\boldsymbol{g}_0)}_i)\\ &\quad+\Big(\frac{1}{n}\widehat{H}_{12}-\frac{1}{n}H_{12}\Big)\Big(\frac{1}{n}H_{22}\Big)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^T\Sigma_i^{-1}(\underline{\boldsymbol{W}}_i\boldsymbol{\gamma}^*-\underline{(\boldsymbol{Z}^T\boldsymbol{g}_0)}_i).\\ &= DI_{22}^{(1)}+DI_{22}^{(2)}+DI_{22}^{(3)} \quad \text{(say)}\end{aligned}$$

Lemmas 2, 3, and 7 imply

$$DI_{22}^{(j)} = \sqrt{n}K_n^{-2}O_p\Big(h_2^2 + h_3^2 + \sqrt{\frac{\log n}{nh_2}} + \sqrt{\frac{\log n}{nh_3^2}}\Big) = o_p(1),\ j = 1, 2, 3.$$

Hence we have established

$$\widehat{I}_2 - I_2 = o_p(n^{-1/2}). \tag{5.12}$$

The desired result follows from (5.6), (5.7), (5.11), and (5.12). The proof of Theorem 1 is complete.

SUPPLEMENTARY MATERIAL

Supplement A: Some technical material
(doi: xx.xxxx/xx-AOSxxxxSUPP). Proofs of Proposition 4 and the lemmas.

References.

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

[2] Cheng, G., Zhou, L. and Huang, J. Z. (2014). Supplement to "Efficient semiparametric estimation in generalized partially linear additive models for longitudinal/clustered data". *Bernoulli* **20** 141-163.

[3] Cheng, M. Y., Honda, T., Li, J. and Peng, H. (2014). Nonparametric independence screening and structure identification for ultra-high dimensional longitudinal/clustered data. *Ann. Stat.* **42** 1819-1849.

[4] 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.

[5] 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.

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

[7] Henry, K., Erice, A., Tierney, C., Balfour, H. H. J., Fischl, M. A., Kmack, A., Liou, S. H., Kenton, A., Hirsch, M. S., Phair, J., Martinez, A., Kahn, J. O. and for the AIDS Clinical Trial Group 193A Study Team (1998). A randomized, controlled, double-blind study comparing the survival benefit of four different reverse transcriptase inhibitor therapies (three-drug, two-drug, and alternating drug) for the treatment of advanced AIDS. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* **19** 339-349.

[8] Honda, T., Cheng, M. Y. and Li, J. (2014). Supplement to "Efficient estimation in semivarying coefficient models for longitudinal data".

[9] Huang, J. Z., Wu, C. O. and Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statist. Sinica* **14** 763–788.

[10] Li, Y. (2011). Efficient semiparametric regression for longitudinal data with nonparametric covariance estimation. *Biometrika* **98** 355–370.

[11] Ma, S. (2012). Two-step spline estimating equations for generalized additive partially linear models with large cluster sizes. *Ann. Stat.* **40** 2943–2872.

[12] Schumaker, L. L. (2007). *Spline Functions: Basic Theory, 3rd ed.* Cambridge University Press, Cambridge.

[13] Shen, S. L., Cui, J. L., Mei, C. L. and Wang, C. W. (2014). Estimation and inference of semi-varying coefficient models with heteroscedastic errors. *J. Multivariate Anal.* **124** 70–93.

[14] Tian, R., Xue, L. and Liu, C. (2014). Penalized quadratic inference functions for semiparametric varying coefficient partially linear models with longitudinal data. *J. Multivariate Anal.* **132** 94–110.

[15] Wang, L. and Qu, A. (2009). Consistent model selection and data-driven smooth tests for longitudinal data in the estimating equations approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71** 177–190.

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

[17] Wu, H. and Zhang, J. T. (2006). *Nonparametric Regression Methods for Longitudinal Data: mixed-effects modeling approaches.* Wiley, New York.

[18] Xia, Y., Zhang, W. and Tong, H. (2004). Efficient estimation for semivarying-coefficient models. *Biometrika* **91** 661–681.

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

[20] Zhang, W., Fan, J. and Sun, Y. (2009). A semiparametric model for cluster data. *Ann. Stat.* **37** 2377-2408.

[21] Zhou, J. and Qu, A. (2012). Informative estimation and selection of correlation structure for longitudinal data. *J. Amer. Statist. Assoc.* **107** 701–710.

T. Honda
Graduate School of Economics
Hitotsubashi University
Kunitachi, Tokyo 186-8601, Japan
E-mail: t.honda@r.hit-u.ac.jp

M.-Y. Cheng
Department of Mathematics
National Taiwan University
Taipei 106, Taiwan
E-mail: cheng@math.ntu.edu.tw

J. Li
Department of Statistics & Applied Probability
National University of Singapore
Singapore 117546
E-mail: stalj@nus.edu.sg

Supplement to "Efficient estimation in semivarying coefficient models for longitudinal/clustered data"

by Toshio Honda, Ming-Yen Cheng, and Jialiang Li

S.1. Proof of Proposition 4. In the proof, we repeatedly use arguments based on exponential inequalities, truncation, and division of regions into small rectangles to prove uniform convergence results as in [3]. We don't give the details of these arguments since they are standard ones in nonparametric kernel methods like local linear regression and kernel density estimation. Recall that we have assumed three times continuous differentiability of the relevant functions.

The proof consists of four parts: (i) representation of $\widehat{\boldsymbol{g}}(t)$, (ii) representation of $\widehat{\epsilon}_{ij}$, (iii) representation of $\widehat{\sigma^2}(t)$, and (iv) representation of $\widehat{\sigma}(s,t)$.
(i) Representation of $\widehat{\boldsymbol{g}}(t)$. Applying the third order Taylor series expansion to $\boldsymbol{g}_0(t)$, we have

(S.1)
$$\boldsymbol{Z}_{ij}^T\boldsymbol{g}_0(T_{ij}) = \boldsymbol{Z}_{ij}^T\Big\{\boldsymbol{g}_0(t) + h_1\frac{T_{ij}-t}{h_1}\boldsymbol{g}_0'(t) + \frac{h_1^2}{2}\Big(\frac{T_{ij}-t}{h_1}\Big)^2\boldsymbol{g}_0''(t)\Big\} + O(h_1^3),$$

where $\boldsymbol{g}_0'(t) = (g_{01}'(t),\ldots,g_{0q}'(t))^T$ and $\boldsymbol{g}_0''(t) = (g_{01}''(t),\ldots,g_{0q}''(t))^T$. By plugging (S.1) into (3.2), we have uniformly in t,

$$\begin{aligned}\text{(S.2)}\quad \widehat{\boldsymbol{g}}(t) &= \boldsymbol{g}_0(t) + D_q(\widehat{L}_1(t))^{-1}\widehat{L}_2(t)(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}_I)\\ &\quad + \frac{h_1^2}{2}D_q(\widehat{L}_1(t))^{-1}\widehat{L}_3(t)\boldsymbol{g}_0''(t) + D_q(\widehat{L}_1(t))^{-1}E_0(t) + O_p(h_1^3),\end{aligned}$$

where $\widehat{L}_1(t) = A_{1n}(t)$ defined after (3.2),

$$\begin{aligned}\widehat{L}_2(t) &= \frac{1}{N_1h_1}\sum_{i=1}^n\sum_{j=1}^{m_i}\boldsymbol{Z}_{ij}\otimes\begin{pmatrix}1\\ \frac{T_{ij}-t}{h_1}\end{pmatrix}K\Big(\frac{T_{ij}-t}{h_1}\Big)\boldsymbol{X}_{ij}^T,\\ \widehat{L}_3(t) &= \frac{1}{N_1h_1}\sum_{i=1}^n\sum_{j=1}^{m_i}(\boldsymbol{Z}_{ij}\boldsymbol{Z}_{ij}^T)\otimes\begin{pmatrix}(\frac{T_{ij}-t}{h_1})^2\\ (\frac{T_{ij}-t}{h_1})^3\end{pmatrix}K\Big(\frac{T_{ij}-t}{h_1}\Big),\\ E_0(t) &= \frac{1}{N_1h_1}\sum_{i=1}^n\sum_{j=1}^{m_i}\boldsymbol{Z}_{ij}\otimes\begin{pmatrix}1\\ \frac{T_{ij}-t}{h_1}\end{pmatrix}K\Big(\frac{T_{ij}-t}{h_1}\Big)\epsilon_{ij}.\end{aligned}$$

By following standard arguments as in [3], we obtain for $j = 1,2,3$,

$$\text{(S.3)}\qquad \widehat{L}_j(t) = L_j(t) + O_p\Big(\sqrt{\frac{\log n}{nh_1}}\Big)\quad\text{uniformly in } t,$$

where $L_j = \mathrm{E}\{\widehat{L}_j(t)\}$ and

$$E_0(t) = O_p\Big(\sqrt{\frac{\log n}{nh_1}}\Big) \quad \text{uniformly in } t. \tag{S.4}$$

Assumption A1 implies that

$$C_1 \boldsymbol{I}_{2q} \le L_1(t) \le C_2 \boldsymbol{I}_{2q} \tag{S.5}$$

for some positive constants C_1 and C_2. From (S.2)-(S.5), we have uniformly in t,

(S.6)

$$\begin{aligned}
\widehat{\boldsymbol{g}}(t) &= \boldsymbol{g}_0(t) + D_q(L_1(t))^{-1}L_2(t)(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}_I) + \frac{h_1^2}{2}D_q(L_1(t))^{-1}L_3(t)\boldsymbol{g}_0''(t) \\
&\qquad + D_q(L_1(t))^{-1}E_0(t) + O_p(h_1^3) + O_p\Big(\frac{\log n}{nh_1}\Big) + O_p\Big(h_1^2\sqrt{\frac{\log n}{nh_1}}\Big) \\
&= \boldsymbol{g}_0(t) + L_4(t)(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}_I) + h_1^2 L_5(t)\boldsymbol{g}_0''(t) + L_6(t)E_0(t) \\
&\qquad + O_p(h_1^3) + O_p\Big(\frac{\log n}{nh_1}\Big) + O_p\Big(h_1^2\sqrt{\frac{\log n}{nh_1}}\Big) \quad \text{(say)}.
\end{aligned}$$

Note that all the elements of $L_j(t)$, $j = 4, 5, 6$, are bounded functions of t.
(ii) Representation of $\widehat{\epsilon}_{ij}$. We have

$$\widehat{\epsilon}_{ij} = \epsilon_{ij} + \boldsymbol{X}_{ij}^T(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}_I) + \boldsymbol{Z}_{ij}^T(\boldsymbol{g}_0(T_{ij}) - \widehat{\boldsymbol{g}}(T_{ij})). \tag{S.7}$$

By plugging (S.6) into (S.7), we obtain uniformly in i and j,

(S.8)

$$\begin{aligned}
\widehat{\epsilon}_{ij} &= \epsilon_{ij} + (\boldsymbol{X}_{ij}^T - \boldsymbol{Z}_{ij}^T L_4(T_{ij}))(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}_I) - h_1^2 \boldsymbol{Z}_{ij}^T L_5(T_{ij})\boldsymbol{g}''(T_{ij}) \\
&\qquad - \boldsymbol{Z}_{ij}^T L_6(T_{ij})E_0(T_{ij}) + O_p(h_1^3) + O_p\Big(\frac{\log n}{nh_1}\Big) + O_p\Big(h_1^2\sqrt{\frac{\log n}{nh_1}}\Big) \\
&= \epsilon_{ij} + M_{ij}^{(1)}(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}_I) + h_1^2 M_{ij}^{(2)}\boldsymbol{g}''(T_{ij}) + M_{ij}^{(3)}E_0(T_{ij}) \\
&\qquad + O_p(h_1^3) + O_p\Big(\frac{\log n}{nh_1}\Big) + O_p\Big(h_1^2\sqrt{\frac{\log n}{nh_1}}\Big) \quad \text{(say)}.
\end{aligned}$$

Note that all the elements of $M_{ij}^{(1)}$, $M_{ij}^{(2)}$, and $M_{ij}^{(3)}$ are uniformly bounded functions of $\boldsymbol{X}_{ij}$, $\boldsymbol{Z}_{ij}$, and T_{ij}.

(iii) Representation of $\widehat{\sigma^2}(t)$. We have uniformly in i and j,

$$(\widehat{\epsilon}_{ij})^2 = \epsilon_{ij}^2 - \sigma^2(T_{ij}) + \sigma^2(T_{ij}) + 2\epsilon_{ij}M_{ij}^{(3)}E_0(T_{ij}) + 2\epsilon_{ij}M_{ij}^{(1)}(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}_I) + 2\epsilon_{ij}h_1^2M_{ij}^{(2)}\boldsymbol{g}_0''(T_{ij}) + O_p(h_1^3) + O_p\Big(\frac{\log n}{nh_1}\Big) + O_p\Big(h_1^2\sqrt{\frac{\log n}{nh_1}}\Big). \tag{S.9}$$

Recall that $M_{ij}^{(l)}$, $l = 1, 2, 3$, are defined in (S.8). It is easy to see that the contributions of $2\epsilon_{ij}M_{ij}^{(1)}(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}_I)$ and $2\epsilon_{ij}h_1^2M_{ij}^{(2)}\boldsymbol{g}''(T_{ij})$ to $\widehat{\sigma^2}(t)$ are

$$O_p\Big(\frac{1}{\sqrt{n}}\sqrt{\frac{\log n}{nh_2}}\Big) \quad \text{and} \quad O_p\Big(h_1^2\sqrt{\frac{\log n}{nh_2}}\Big)$$

uniformly in t, respectively. Thus we have only to consider $\epsilon_{ij}^2 - \sigma^2(T_{ij})$, $\sigma^2(T_{ij})$, and $2\epsilon_{ij}M_{ij}^{(3)}E_0(T_{ij})$ in (S.9).

Setting $\widehat{L}_7(t) = A_{2n}(t)$ defined after (3.3), we have for some positive constants C_1 and C_2,

$$\widehat{L}_7(t) = L_7(t) + O_p\Big(\sqrt{\frac{\log n}{nh_2}}\Big) \quad \text{and} \quad C_1\boldsymbol{I}_2 \le L_7(t) \le C_2\boldsymbol{I}_2 \tag{S.10}$$

uniformly in t, where $L_7(t) = \mathrm{E}\{\widehat{L}_7(t)\}$. Now we have uniformly in t,

$$\widehat{\sigma^2}(t) = (1\,0)(\widehat{L}_7(t))^{-1}(E_1(t) + Bias_1(t) + R_1(t)) + O_p(h_1^3) + O_p\Big(\frac{\log n}{nh_1}\Big) + O_p\Big(h_1^2\sqrt{\frac{\log n}{nh_1}}\Big), \tag{S.11}$$

where $E_1(t)$ is defined in Proposition 4, $Bias_1(t)$ is the term of $\sigma^2(T_{ij})$, and $R_1(t)$ is the term of $2\epsilon_{ij}M_{ij}^{(3)}E_0(T_{ij})$. It is easy to see that uniformly in t,

$$E_1(t) = O_p\Big(\sqrt{\frac{\log n}{nh_2}}\Big). \tag{S.12}$$

By applying the Taylor series expansion, we have

$$\sigma^2(T_{ij}) = \sigma^2(t) + h_2(\sigma^2)'(t)\frac{T_{ij} - t}{h_2} + \frac{h_2^2}{2}(\sigma^2)''(t)\Big(\frac{T_{ij} - t}{h_2}\Big)^2 + O(h_2^3).$$

Therefore $Bias_1(t)$ can be represented as

$$Bias_1(t) = \widehat{L}_7(t)\begin{pmatrix}\sigma^2(t)\\ h_2(\sigma^2)'(t)\end{pmatrix} + \frac{h_2^2(\sigma^2)''(t)}{2N_1h_2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\begin{pmatrix}(\frac{T_{ij}-t}{h_2})^2\\ (\frac{T_{ij}-t}{h_2})^3\end{pmatrix}K\Big(\frac{T_{ij} - t}{h_2}\Big) + O_p(h_2^3).$$

uniformly in t. Setting

$$\widehat{L}_8(t) = \frac{1}{N_1 h_2} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \begin{pmatrix} (\frac{T_{ij}-t}{h_2})^2 \\ (\frac{T_{ij}-t}{h_2})^3 \end{pmatrix} K\Big(\frac{T_{ij}-t}{h_2}\Big),$$

we have uniformly in t,

$$\widehat{L}_8(t) = L_8(t) + O_p\Big(\sqrt{\frac{\log n}{n h_2}}\Big),$$

where $L_8(t) = \mathrm{E}\{\widehat{L}_8(t)\}$ and $L_8(t)$ is a bounded vector function of t. Hence we have uniformly in t,

(S.13)

$$Bias_1(t) = \widehat{L}_7(t) \begin{pmatrix} \sigma^2(t) \\ h_2(\sigma^2)'(t) \end{pmatrix} + \frac{h_2^2(\sigma^2)''(t)}{2} L_8(t) + O_p(h_2^3) + O_p\Big(h_2^2 \sqrt{\frac{\log n}{n h_2}}\Big).$$

Next we deal with $R_1(t)$, which is written as

(S.14)

$$\frac{1}{N_1^2 h_1 h_2} \sum_{a,b} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{i'=1}^{n} \sum_{j'=1}^{m_{i'}} \epsilon_{ij}\epsilon_{i'j'} A_{ab,ij} B_{ab,i'j'} K_a\Big(\frac{T_{ij}-t}{h_2}\Big) K_b\Big(\frac{T_{i'j'}-T_{ij}}{h_1}\Big),$$

where $K_l(t) = t^l K(t)$, $a = 0, 1$, and $b = 0, 1$. Note that $A_{ab,ij}$ and $B_{ab,ij}$ are uniformly bounded functions of $\boldsymbol{X}_{ij}$, $\boldsymbol{Z}_{ij}$, and T_{ij}.

By using an exponential inequality (3.5) in [1] and standard arguments in nonparametric regression as in [3], we evaluate

(S.15)

$$\begin{aligned}
&\frac{1}{N_1^2 h_1 h_2} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{i'=1}^{n} \sum_{j'=1}^{m_{i'}} \epsilon_{ij}\epsilon_{i'j'} A_{ab,ij} B_{ab,i'j'} K_a\Big(\frac{T_{ij}-t}{h_2}\Big) K_b\Big(\frac{T_{i'j'}-T_{ij}}{h_1}\Big) \\
&\quad = \frac{1}{N_1^2 h_1 h_2} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \epsilon_{ij}^2 A_{ab,ij} B_{ab,ij} K_a\Big(\frac{T_{ij}-t}{h_2}\Big) K_b(0) \\
&\qquad + \frac{1}{N_1^2 h_1 h_2} \sum_{i=1}^{n} \sum_{j \neq j'} \epsilon_{ij}\epsilon_{ij'} A_{ab,ij} B_{ab,ij'} K_a\Big(\frac{T_{ij}-t}{h_2}\Big) K_b\Big(\frac{T_{ij'}-T_{ij}}{h_1}\Big) \\
&\qquad + \frac{1}{N_1^2 h_1 h_2} \sum_{i \neq i'} \sum_{j,j'} \epsilon_{ij}\epsilon_{i'j'} A_{ab,ij} B_{ab,i'j'} K_a\Big(\frac{T_{ij}-t}{h_2}\Big) K_b\Big(\frac{T_{i'j'}-T_{ij}}{h_1}\Big) \\
&\quad = R_{1ab}^{(1)}(t) + R_{1ab}^{(2)}(t) + R_{1ab}^{(3)}(t) \quad \text{(say)}.
\end{aligned}$$

Note that we cannot apply classical exponential inequalities for U-statistics since kernel functions depend on i and i' and observations are not identical.

It is easy to see that uniformly in t,

$$R_{1ab}^{(1)}(t) = O_p((nh_1)^{-1}) \quad \text{and} \quad R_{1ab}^{(2)}(t) = O_p(n^{-1}). \tag{S.16}$$

We evaluate $R_{1ab}^{(3)}(t)$ by using an exponential inequality as (3.5) in [1] with $A = C_1(\log n)^k/(N_1^2 h_1 h_2)$, $B = C_2(\log n)^k/(N_1^{3/2} h_1 h_2)$, $C = C_3/(N_1 h_1^{1/2} h_2^{1/2})$, and $x = M \log n/(N_1 h_1^{1/2} h_2^{1/2})$ in the inequality and standard arguments in nonparametric regression as in [3]. Note that we used a kind of truncation technique to handle ϵ_{ij} and that we have to take sufficiently large k and M here. Hence we have

$$R_{1ab}^{(3)}(t) = O_p\Big(\frac{\log n}{n(h_1 h_2)^{1/2}}\Big).$$

The above equation and (S.14)-(S.16) imply that

$$R_1(t) = O_p\Big(\frac{\log n}{n(h_1 h_2)^{1/2}}\Big) + O_p\Big(\frac{1}{nh_1}\Big) \tag{S.17}$$

uniformly in t.

It follows from (S.10)-(S.13) and (S.17) that

$$\begin{aligned}\widehat{\sigma^2}(t) = \sigma^2(t) &+ (1\,0)(L_7(t))^{-1} E_1(t) + \frac{h_2^2}{2}(1\,0)(L_7(t))^{-1} L_8(t)(\sigma^2)''(t) \\ &+ O_p(h_1^3) + O_p(h_2^3) + O_p\Big(\frac{\log n}{nh_1}\Big) + O_p\Big(\frac{\log n}{nh_2}\Big).\end{aligned}$$

The expression of $\widehat{\sigma^2}(t)$ in Proposition 4 also follows from the above expression.

(iv) Representation of $\widehat{\sigma}(s,t)$. We can proceed almost in the same way as when we deal with $\widehat{\sigma^2}(t)$. First we examine $\widehat{\epsilon}_{ij}\widehat{\epsilon}_{ij'}$ closely. Then we have uniformly in i, j, and j',

$$\begin{aligned}\widehat{\epsilon}_{ij}\widehat{\epsilon}_{ij'} = \epsilon_{ij}\epsilon_{ij'} &- \sigma(T_{ij}, T_{ij'}) + \sigma(T_{ij}, T_{ij'}) \\ &+ \epsilon_{ij} M_{ij'}^{(3)} E_0(T_{ij'}) + \epsilon_{ij'} M_{ij}^{(3)} E_0(T_{ij}) \\ &+ \epsilon_{ij} M_{ij'}^{(1)}(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}_I) + \epsilon_{ij'} M_{ij}^{(1)}(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}_I) \qquad \text{(S.18)} \\ &+ \epsilon_{ij} h_1^2 M_{ij'}^{(2)} \boldsymbol{g}_0''(T_{ij'}) + \epsilon_{ij'} h_1^2 M_{ij}^{(2)} \boldsymbol{g}_0''(T_{ij}) \qquad \text{(S.19)} \\ &+ O_p(h_1^3) + O_p\Big(\frac{\log n}{nh_1}\Big) + O_p\Big(h_1^2\sqrt{\frac{\log n}{nh_1}}\Big).\end{aligned}$$

It is easy to see that the contributions of (S.18) and (S.19) to $\widehat{\sigma}(s,t)$ are

$$O_p\Big(\frac{1}{\sqrt{n}}\sqrt{\frac{\log n}{nh_3^2}}\Big) \quad \text{and} \quad O_p\Big(h_1^2\sqrt{\frac{\log n}{nh_3^2}}\Big)$$

uniformly in s and t, respectively. Therefore we have only to consider $\epsilon_{ij}\epsilon_{ij'} - \sigma(T_{ij}, T_{ij'})$, $\sigma(T_{ij}, T_{ij'})$, and $\epsilon_{ij}M_{ij'}^{(3)}E_0(T_{ij'}) + \epsilon_{ij'}M_{ij}^{(3)}E_0(T_{ij})$ in $\widehat{\epsilon}_{ij}\widehat{\epsilon}_{ij'}$.

Setting $\widehat{L}_9(s,t) = A_{3n}(s,t)$, we have for some positive constants C_1 and C_2,

$$\widehat{L}_9(s,t) = L_9(s,t) + O_p\Big(\sqrt{\frac{\log n}{nh_3^2}}\Big) \quad \text{and} \quad C_1\boldsymbol{I}_3 \le L_9(s,t) \le C_2\boldsymbol{I}_3 \tag{S.20}$$

uniformly in s and t, where $L_9(s,t) = \mathrm{E}\{\widehat{L}_9(s,t)\}$. Now we have uniformly in s and t,

$$\begin{aligned}\widehat{\sigma}(s,t) = (1\,0\,0)(\widehat{L}_9(s,t))^{-1}(E_2(s,t) + Bias_2(s,t) + R_2(s,t)) \\ + O_p(h_1^3) + O_p\Big(\frac{\log n}{nh_1}\Big) + O_p\Big(h_1^2\sqrt{\frac{\log n}{nh_1}}\Big),\end{aligned} \tag{S.21}$$

where $E_2(s,t)$ is defined in Proposition 4, $Bias_2(s,t)$ is the term of $\sigma(T_{ij}, T_{ij'})$, and $R_2(s,t)$ is the term of $\epsilon_{ij}M_{ij'}^{(3)}E_0(T_{ij'}) + \epsilon_{ij'}M_{ij}^{(3)}E_0(T_{ij})$. It is easy to see that uniformly in s and t,

$$E_2(s,t) = O_p\Big(\sqrt{\frac{\log n}{nh_3^2}}\Big). \tag{S.22}$$

Setting

$$\begin{aligned}&\widehat{L}_{10}(s,t) \\ &= \frac{1}{N_2h_3^2}\sum_{i=1}^{n}\sum_{j\neq j'}\begin{pmatrix}1\\ \frac{T_{ij}-s}{h_3}\\ \frac{T_{ij'}-t}{h_3}\end{pmatrix}\Big(\Big(\frac{T_{ij}-s}{h_3}\Big)^2\ \frac{2(T_{ij}-s)(T_{ij'}-t)}{h_3^2}\ \Big(\frac{T_{ij'}-t}{h_3}\Big)^2\Big) \\ &\qquad\qquad \times K\Big(\frac{T_{ij}-s}{h_3}\Big)K\Big(\frac{T_{ij'}-t}{h_3}\Big),\end{aligned}$$

we have uniformly in s and t,

$$\widehat{L}_{10}(s,t) = L_{10}(s,t) + O_p\Big(\sqrt{\frac{\log n}{nh_3^2}}\Big),$$

where $L_{10}(s,t) = \mathrm{E}\{\widehat{L}_{10}(s,t)\}$ and $L_{10}(s,t)$ is a bounded matrix function of (s,t). Then we have, as in the proof of the representation of $\widehat{\sigma^2}(t)$, uniformly in s and t

$$\text{(S.23)} \qquad Bias_2(s,t) = \widehat{L}_9(s,t) \begin{pmatrix} \sigma(s,t) \\ h_3 \frac{\partial \sigma}{\partial s}(s,t) \\ h_3 \frac{\partial \sigma}{\partial t}(s,t) \end{pmatrix} + \frac{h_3^2}{2} L_{10}(s,t) \begin{pmatrix} \frac{\partial^2 \sigma}{\partial s^2}(s,t) \\ \frac{\partial^2 \sigma}{\partial s \partial t}(s,t) \\ \frac{\partial^2 \sigma}{\partial t^2}(s,t) \end{pmatrix} + O_p(h_3^3) + O_p\Big(h_3^2 \sqrt{\frac{\log n}{n h_3^2}}\Big).$$

Finally we deal with $R_2(s,t)$ in the same way as in the proof of the representation of $\widehat{\sigma^2}(t)$. We use the same exponential inequality for U-statistics. We should consider

$$\text{(S.24)} \quad \frac{1}{N_1 N_2 h_1 h_3^2} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \sum_{j_1 \neq j_2} \sum_{j_3} \epsilon_{i_1 j_1} \epsilon_{i_2 j_3} A_{abc,i_1 j_2} B_{abc,i_2 j_3} \times K_a\Big(\frac{T_{i_2 j_3} - T_{i_1 j_2}}{h_1}\Big) K_b\Big(\frac{T_{i_1 j_1} - t}{h_3}\Big) K_c\Big(\frac{T_{i_1 j_2} - s}{h_3}\Big),$$

where $K_l(t) = t^l K(t)$, $a = 0,1$, $b = 0,1$, and $c = 0,1$. Note that $A_{abc,ij}$ and $B_{abc,ij}$ are uniformly bounded functions of $\boldsymbol{X}_{ij}$, $\boldsymbol{Z}_{ij}$, and T_{ij}. This is a generalized U-statistics when we remove the summands of $i_1 = i_2$ and we recall (1.1) when we evaluate (S.24). It is easy to see that uniformly in s and t,

(S.25)

$$\frac{1}{N_1 N_2 h_1 h_3^2} \sum_{i_1=1}^{n} \sum_{j_1 \neq j_2} \sum_{j_3} \epsilon_{i_1 j_1} \epsilon_{i_1 j_3} A_{abc,i_1 j_2} B_{abc,i_1 j_3} \times K_a\Big(\frac{T_{i_1 j_3} - T_{i_1 j_2}}{h_1}\Big) K_b\Big(\frac{T_{i_1 j_1} - t}{h_3}\Big) K_c\Big(\frac{T_{i_1 j_2} - s}{h_3}\Big) = O_p\Big(\frac{1}{n h_1}\Big).$$

In the same way as when dealing with $R_{1ab}^{(3)}(t)$, we obtain

(S.26)

$$\frac{1}{N_1 N_2 h_1 h_3^2} \sum_{i_1 \neq i_2} \sum_{j_1 \neq j_2} \sum_{j_3} \epsilon_{i_1 j_1} \epsilon_{i_2 j_3} A_{abc,i_1 j_2} B_{abc,i_2 j_3} \times K_a\Big(\frac{T_{i_2 j_3} - T_{i_1 j_2}}{h_1}\Big) K_b\Big(\frac{T_{i_1 j_1} - t}{h_3}\Big) K_c\Big(\frac{T_{i_1 j_2} - s}{h_3}\Big) = O_p\Big(\frac{\log n}{n h_1^{1/2} h_3}\Big)$$

with $A = C_1(\log n)^k/(n^2 h_1 h_3^2)$, $B = C_2(\log n)^k/(n^{3/2} h_1^{1/2} h_3^2)$, $C = C_3/(n h_1^{1/2} h_3)$, and $x = M \log n/(n h_1^{1/2} h_3)$ in the exponential inequality. Note that we should choose sufficiently large k and M. It follows from (S.25) and (S.26) that uniformly in s and t,

$$R_2(s,t) = O_p\Big(\frac{\log n}{n h_1^{1/2} h_3}\Big). \tag{S.27}$$

Hence (S.20)- (S.23) and (S.27) imply that uniformly in s and t,

$$\begin{aligned}
&\widehat{\sigma}(s,t) - \sigma(s,t) \\
&= (1\,0\,0)(L_9(s,t))^{-1} E_2(s,t) + \frac{h_3^2}{2}(1\,0\,0)(L_9(s,t))^{-1} L_{10}(s,t) \begin{pmatrix} \frac{\partial^2 \sigma}{\partial s^2}(s,t) \\ \frac{\partial^2 \sigma}{\partial s \partial t}(s,t) \\ \frac{\partial^2 \sigma}{\partial t^2}(s,t) \end{pmatrix} \\
&\qquad + O_p(h_1^3) + O_p(h_3^3) + O_p\Big(\frac{\log n}{n h_1}\Big) + O_p\Big(\frac{\log n}{n h_3^2}\Big).
\end{aligned}$$

The expression in Proposition 4 follows from the above expression. The proof of Proposition 4 is complete.

S.2. Proofs of Lemmas 1-8. *Proof of Lemma 1.* The proof consists of six parts.
(i) Recall that

$$(\|\boldsymbol{Z}^T \boldsymbol{g}\|^{\Delta})^2 = \frac{1}{n} \mathrm{E}\Big\{ \sum_{i=1}^n \underline{(\boldsymbol{Z}^T \boldsymbol{g})}_i^T \Delta_{0i} V_i^{-1} \Delta_{0i} \underline{(\boldsymbol{Z}^T \boldsymbol{g})}_i \Big\}.$$

We have from Assumptions A4 and A5 that

$$\begin{aligned}
&\frac{C_1}{n} \mathrm{E}\Big\{ \sum_{i=1}^n \sum_{j=1}^{m_i} \boldsymbol{g}^T(T_{ij}) \boldsymbol{Z}_{ij} \boldsymbol{Z}_{ij}^T \boldsymbol{g}(T_{ij}) \Big\} \\
&\le (\|\boldsymbol{Z}^T \boldsymbol{g}\|^{\Delta})^2 \le \frac{C_2}{n} \mathrm{E}\Big\{ \sum_{i=1}^n \sum_{j=1}^{m_i} \boldsymbol{g}^T(T_{ij}) \boldsymbol{Z}_{ij} \boldsymbol{Z}_{ij}^T \boldsymbol{g}(T_{ij}) \Big\}
\end{aligned} \tag{S.28}$$

for some positive constants C_1 and C_2. Assumptions A1 and A2 imply that for some positive constants C_3 and C_4,
(S.29)

$$C_3 \sum_{l=1}^q \int g_l^2(t) dt \le \frac{1}{n} \mathrm{E}\Big\{ \sum_{i=1}^n \sum_{j=1}^{m_i} \boldsymbol{g}^T(T_{ij}) \boldsymbol{Z}_{ij} \boldsymbol{Z}_{ij}^T \boldsymbol{g}(T_{ij}) \Big\} \le C_4 \sum_{l=1}^q \int g_l^2(t) dt.$$

The results on $\|\cdot\|^{\Delta}$ follows from (S.28) and (S.29). We can verify those on $\|\cdot\|$ in the same way.
(ii) This is a well-known result in the literature of spline regression. See for example A.2 of [2].
(iii) Here, $a_n \sim b_n$ means that $C_1 a_n < b_n < C_2 a_n$ for some positive constants C_1 and C_2. The reslut in (ii) implies

$$\|\boldsymbol{X}^T\boldsymbol{\beta} + \boldsymbol{Z}^T\boldsymbol{g}\|_{\infty}^2 \le CK_n\Big(|\boldsymbol{\beta}|^2 + \|\boldsymbol{g}\|_{G,2}^2\Big)$$

for some positive constant C. Recall that p and q are fixed in this paper. On the other hand, we have

$$\begin{aligned}
&(\|\boldsymbol{X}^T\boldsymbol{\beta} + \boldsymbol{Z}^T\boldsymbol{g}\|^{\Delta})^2 \\
&\quad\sim \frac{1}{n}\mathrm{E}\Big\{\sum_{i=1}^{n}\sum_{j=1}^{m_i}(\boldsymbol{\beta}^T\,\boldsymbol{g}^T(T_{ij}))\begin{pmatrix}\boldsymbol{X}_{ij}\boldsymbol{X}_{ij}^T & \boldsymbol{X}_{ij}\boldsymbol{Z}_{ij}^T\\ \boldsymbol{Z}_{ij}\boldsymbol{X}_{ij}^T & \boldsymbol{Z}_{ij}\boldsymbol{Z}_{ij}^T\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{g}(T_{ij})\end{pmatrix}\Big\} \\
&\quad\sim \frac{1}{n}\mathrm{E}\Big\{\sum_{i=1}^{n}\sum_{j=1}^{m_i}(\boldsymbol{\beta}^T\,\boldsymbol{g}^T(T_{ij}))\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{g}(T_{ij})\end{pmatrix}\Big\} \sim |\boldsymbol{\beta}|^2 + \|\boldsymbol{g}\|_{G,2}^2.
\end{aligned}$$

Hence the desired result is established.
(iv) For $\boldsymbol{g}_1 \in \boldsymbol{G}_B$ and $\boldsymbol{g}_2 \in \boldsymbol{G}_B$, we have
(S.30)

$$\langle \boldsymbol{Z}^T\boldsymbol{g}_1, \boldsymbol{Z}^T\boldsymbol{g}_2\rangle_n^{\Delta} = \boldsymbol{\gamma}_1^T\Big\{\frac{1}{n}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^T\Delta_{0i}V_i^{-1}\Delta_{0i}\underline{\boldsymbol{W}}_i\Big\}\boldsymbol{\gamma}_2 = \boldsymbol{\gamma}_1^T\overline{\Delta}_{Vn}\boldsymbol{\gamma}_2 \quad \text{(say)},$$

where $\overline{\Delta}_{Vn}$ is a $qK_n \times qK_n$ matrix and $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ correspond to $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$, respectively. Elements of $\frac{1}{n}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^T\Delta_{0i}V_i^{-1}\Delta_{0i}\underline{\boldsymbol{W}}_i$ are written as

$$\text{(S.31)}\quad \frac{1}{n}\sum_{i=1}^{n}\sum_{j_1,j_2}\lambda_{ij_1j_2}B_{k_1}(T_{ij_1})B_{k_2}(T_{ij_2})Z_{ij_1l_1}Z_{ij_2l_2} = \overline{\Delta}_{Vn}^{(k_1,l_1,k_2,l_2)} \quad \text{(say)},$$

where $\lambda_{ij_1j_2}$ is defined in (5.4), $1 \le k_1, k_2 \le K_n$, and $1 \le l_1, l_2 \le q$.

By evaluating the variance of (S.31) and using the Bernstein inequality for independent bounded random variables, we have uniformly in k_1, k_2, l_1, and l_2,

$$\text{(S.32)}\quad \overline{\Delta}_{Vn}^{(k_1,l_1,k_2,l_2)} - \mathrm{E}(\overline{\Delta}_{Vn}^{(k_1,l_1,k_2,l_2)}) = O_p\Big(\sqrt{\frac{\log n}{nK_n^2}}\Big) \quad \text{if } B_{k_1}(t)B_{k_2}(t) \equiv 0$$

and

$$\text{(S.33)}\quad \overline{\Delta}_{Vn}^{(k_1,l_1,k_2,l_2)} - \mathrm{E}(\overline{\Delta}_{Vn}^{(k_1,l_1,k_2,l_2)}) = O_p\Big(\sqrt{\frac{\log n}{nK_n}}\Big) \quad \text{if } B_{k_1}(t)B_{k_2}(t) \not\equiv 0.$$

By exploiting (S.32), (S.33), and the local property of the B-spline basis, we obtain
(S.34)

$$\max\{|\lambda_{\min}(\overline{\Delta}_{Vn} - \mathrm{E}(\overline{\Delta}_{Vn}))|, |\lambda_{\max}(\overline{\Delta}_{Vn} - \mathrm{E}(\overline{\Delta}_{Vn}))|\} = O_p\Big(\sqrt{\frac{\log n}{n}}\Big).$$

We also have

$$\frac{C_1}{K_n} \le \lambda_{\min}(\mathrm{E}(\overline{\Delta}_{Vn})) \le \lambda_{\max}(\mathrm{E}(\overline{\Delta}_{Vn})) \le \frac{C_2}{K_n} \tag{S.35}$$

since

$$\begin{aligned}\frac{C_3}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}(\boldsymbol{Z}_{ij}\otimes\boldsymbol{B}(T_{ij}))^T(\boldsymbol{Z}_{ij}\otimes\boldsymbol{B}(T_{ij}))\\ \le \overline{\Delta}_{Vn} \le \frac{C_4}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}(\boldsymbol{Z}_{ij}\otimes\boldsymbol{B}(T_{ij}))^T(\boldsymbol{Z}_{ij}\otimes\boldsymbol{B}(T_{ij}))\end{aligned}$$

for some positive constants C_3 and C_4. See the proof of Lemma A.3 of [2]. Hence the desired result follows from (S.34) and (S.35).
(v) This follows from (iv) and (vi).
(vi) Note that

$$\langle \delta_n, Z_l B_k\rangle_n^{\Delta} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j_1,j_2}\delta_{n,ij_1}\lambda_{ij_1j_2}Z_{ij_2l}B_k(T_{ij_2}).$$

Since we have

$$\sum_{l=1}^{q}\sum_{k=1}^{K_n}\mathrm{var}(\langle \delta_n, Z_l B_k\rangle_n^{\Delta}) \le \frac{C}{n}\|\delta\|_\infty$$

for some positive constant C, the desired result follows from (S.35).

Proof of Lemma 2. We can verify the result on $n^{-1}h_{12,kl}$ by using the local property of the B-spline basis and the Bernstein inequality for independent bounded random variables. Since

$$\begin{aligned}\frac{1}{n}(\widehat{H}_{12} - H_{12}) = \frac{1}{n}\sum_{i=1}^{n}\underline{\boldsymbol{X}}_i^T\{\Sigma_i^{-1}(\Sigma_i - \widehat{\Sigma}_i)\Sigma_i^{-1}\}\underline{\boldsymbol{W}}_i\\ + \frac{1}{n}\sum_{i=1}^{n}\underline{\boldsymbol{X}}_i^T\{\widehat{\Sigma}_i^{-1}(\Sigma_i - \widehat{\Sigma}_i)\Sigma_i^{-1}(\Sigma_i - \widehat{\Sigma}_i)\Sigma_i^{-1}\}\underline{\boldsymbol{W}}_i,\end{aligned}$$

the desired result on $n^{-1}(\widehat{h}_{12,kl} - h_{12,kl})$ follows from Proposition 4 and the local property of the B-spline basis. The results on the Euclidean norm follow from those on the elements. Hence the proof is complete.

Proof of Lemma 3. We have from Assumption A3 that

$$\frac{C_1}{n}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^T\underline{\boldsymbol{W}}_i \le \frac{1}{n}H_{22} \le \frac{C_2}{n}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^T\underline{\boldsymbol{W}}_i \tag{S.36}$$

for some positive constants C_1 and C_2 and

$$\frac{1}{n}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^T\underline{\boldsymbol{W}}_i = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}(\boldsymbol{Z}_{ij}\boldsymbol{Z}_{ij}^T)\otimes(\boldsymbol{B}(T_{ij})\boldsymbol{B}^T(T_{ij})).$$

Thus the first result follows from Assumptions A1 and A2 and the standard arguments on B-spline bases as in the proofs of Lemmas A.1 and A.2 of [2].

Since we have

$$\begin{aligned}\frac{1}{n}(\widehat{H}_{22} - H_{22}) = \frac{1}{n}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^T\{\Sigma_i^{-1}(\Sigma_i - \widehat{\Sigma}_i)\Sigma_i^{-1}\}\underline{\boldsymbol{W}}_i \\ + \frac{1}{n}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^T\{\widehat{\Sigma}_i^{-1}(\Sigma_i - \widehat{\Sigma}_i)\Sigma_i^{-1}(\Sigma_i - \widehat{\Sigma}_i)\Sigma_i^{-1}\}\underline{\boldsymbol{W}}_i,\end{aligned}$$

the second result follows from Proposition 4 and the inequalities similar to (S.36).

The third result follows from the first and second results here. Finally we deal with the fourth result. Note that

$$\begin{aligned}&(n^{-1}\widehat{H}_{22})^{-1} - (n^{-1}H_{22})^{-1} \\ &= (n^{-1}H_{22})^{-1}(n^{-1}H_{22} - n^{-1}\widehat{H}_{22})(n^{-1}H_{22})^{-1} \\ &\quad +(n^{-1}\widehat{H}_{22})^{-1}(n^{-1}H_{22} - n^{-1}\widehat{H}_{22}) \\ &\qquad \times(n^{-1}H_{22})^{-1}(n^{-1}H_{22} - n^{-1}\widehat{H}_{22})(n^{-1}H_{22})^{-1}.\end{aligned} \tag{S.37}$$

By using the first, second, and third results here and (S.37), we obtain the fourth one. Hence the proof is complete.

Proof of Lemma 4. The first result follows from Proposition 4. The second one follows from Lemmas 2 and 3. The last one follows from the first and second ones here.

Proof of Lemma 5. The first result follows from the fact

$$\frac{C_1}{n}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^T\underline{\boldsymbol{W}}_i \le \frac{1}{n}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^T\Sigma_i^{-1}\underline{\boldsymbol{W}}_i \le \frac{C_2}{n}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^T\underline{\boldsymbol{W}}_i$$

for some positive constants C_1 and C_2. Next note that

$$\begin{aligned}
\text{(S.38)}\qquad & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^T\{\Sigma_i^{-1}(\Sigma_i-\widehat{\Sigma}_i)\Sigma_i^{-1}\}\underline{\epsilon}_i \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^T\{\Sigma_i^{-1}(\Sigma_i-\widehat{\Sigma}_i)\Sigma_i^{-1}\}\underline{\epsilon}_i \\
&\quad +\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^T\{\widehat{\Sigma}_i^{-1}(\Sigma_i-\widehat{\Sigma}_i)\Sigma_i^{-1}(\Sigma_i-\widehat{\Sigma}_i)\Sigma_i^{-1}\}\underline{\epsilon}_i.
\end{aligned}$$

Since the elements of $\widehat{\Sigma}_i^{-1}(\Sigma_i-\widehat{\Sigma}_i)\Sigma_i^{-1}(\Sigma_i-\widehat{\Sigma}_i)\Sigma_i^{-1}$ have the stochastic order of $O_p(h_2^4+h_3^4+\log n/(nh_2)+\log n/(nh_3^2))$ uniformly in i, the stochastic order of the elements of the second term of the right-hand side is uniformly $O_p(\sqrt{n}K_n^{-1}(h_2^4+h_3^4+\log n/(nh_2)+\log n/(nh_3^2)))$. Here we used the local property of the B-spline basis and applied the Bernstein inequality and a truncation argument to $n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}|W_{ijl}||\epsilon_{ij}|$, $l=1,\ldots,qK_n$. Thus the norm of this qK_n-dimensional vector has the stochastic order of

$$\text{(S.39)}\qquad \sqrt{\frac{n}{K_n}}O_p\Big(h_2^4+h_3^4+\frac{\log n}{nh_2}+\frac{\log n}{nh_3^2}\Big).$$

We deal with the first term of the right-hand side of (S.38) by using Proposition 4. According to the results of Proposition 4, it can be decomposed into

$$\text{(S.40)}\qquad \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^TQ_{1i}\underline{\epsilon}_i+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^TQ_{2i}\underline{\epsilon}_i+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underline{\boldsymbol{W}}_i^TQ_{3i}\underline{\epsilon}_i,$$

where the (k,l) element of Q_{1i} has the form of $\widetilde{B}_{kl}^{(1)}(\underline{T}_i)h_2^2+\widetilde{B}_{kl}^{(3)}(\underline{T}_i)h_3^2$, the (k,l) element of Q_{2i} has the form of

$$\sum_{j=1}^{m_i}\widetilde{B}_{j,kl}^{(2)}(\underline{T}_i)E_1(T_{ij})+\sum_{j\neq j'}\widetilde{B}_{jj',kl}^{(4)}(\underline{T}_i)E_2(T_{ij},T_{ij'}),$$

and the elements of Q_{3i} have the stochastic order of

$$O_p\Big(h_1^3+h_2^3+h_3^3+\frac{\log n}{nh_1}+\frac{\log n}{nh_2}+\frac{\log n}{nh_3^2}\Big)$$

uniformly in i. Note that $\widetilde{B}_{kl}^{(1)}$, $\widetilde{B}_{kl}^{(3)}$, $\widetilde{B}_{j,kl}^{(2)}$, and $\widetilde{B}_{jj',kl}^{(4)}$ are uniformly bounded functions.

It is easy to see that

$$C_1 \frac{h_2^4 + h_3^4}{K_n} \boldsymbol{I}_{qK_n} \leq \operatorname{cov}\Big(n^{-1/2} \sum_{i=1}^{n} \underline{\boldsymbol{W}}_i^T Q_{1i} \underline{\epsilon}_i\Big) \leq C_2 \frac{h_2^4 + h_3^4}{K_n} \boldsymbol{I}_{qK_n}$$

for some positive constants C_1 and C_2. Hence we have

$$\Big| n^{-1/2} \sum_{i=1}^{n} \underline{\boldsymbol{W}}_i^T Q_{1i} \underline{\epsilon}_i \Big| = O_p(h_2^2 + h_3^2). \tag{S.41}$$

Similarly to the second term of the right-hand side of (S.38), we can demonstrate that

$$\Big| n^{-1/2} \sum_{i=1}^{n} \underline{\boldsymbol{W}}_i^T Q_{3i} \underline{\epsilon}_i \Big| = \sqrt{\frac{n}{K_n}} O_p\Big(h_1^3 + h_2^3 + h_3^3 + \frac{\log n}{nh_1} + \frac{\log n}{nh_2} + \frac{\log n}{nh_3^2}\Big). \tag{S.42}$$

Finally we evaluate the second term of (S.40). Since it has a structure of V-statistics, we can easily evaluate the expectations and the variances of the elements. Thus we have

$$\Big| n^{-1/2} \sum_{i=1}^{n} \underline{\boldsymbol{W}}_i^T Q_{2i} \underline{\epsilon}_i \Big| = O_p\Big(\frac{1}{\sqrt{nh_2}} + \frac{1}{\sqrt{nh_3^2}} + \frac{1}{\sqrt{nK_n}h_2} + \frac{1}{\sqrt{nK_n}h_3^2}\Big). \tag{S.43}$$

Hence the second result follows from (S.39), (S.41), (S.42), and (S.43).

Proof of Lemma 6. This lemma can be proved in the same way as Lemma 5 and the details are omitted.

Proof of Lemma 7. From the definition of $\boldsymbol{\gamma}^*$, we have

$$|\underline{\boldsymbol{W}}_i \boldsymbol{\gamma}^* - \underline{(\boldsymbol{Z}^T \boldsymbol{g}_0)}_i| = O_p(K_n^{-2}) \tag{S.44}$$

uniformly in i. We obtain the first result since the elements of

$$\frac{1}{n} \sum_{i=1}^{n} \underline{\boldsymbol{W}}_i \Sigma_i^{-1} (\underline{\boldsymbol{W}}_i \boldsymbol{\gamma}^* - \underline{(\boldsymbol{Z}^T \boldsymbol{g}_0)}_i)$$

is uniformly $O_p(K_n^{-3})$ from (S.44) and the local property of the B-spine basis. The second one follows from the fact that

$$|(\widehat{\Sigma}_i^{-1} - \Sigma_i^{-1})(\underline{\boldsymbol{W}}_i \boldsymbol{\gamma}^* - \underline{(\boldsymbol{Z}^T \boldsymbol{g}_0)}_i)| = K_n^{-2} O_p\Big(h_2^2 + h_3^2 + \sqrt{\frac{\log n}{nh_2}} + \sqrt{\frac{\log n}{nh_3^2}}\Big)$$

uniformly in i. Hence the proof is complete.

Proof of Lemma 8. This lemma can be proved in the same way as Lemma 7 and the details are omitted.

References.

[1] Giné, E., Latała, R. and Zinn, J. (2000). Exponential and moment inequalities for U-statistics. In *High Dimensional Probability* II (pp. 13-38). Boston: Birkhäuser.

[2] Huang, J. Z., Wu, C.O. and Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Stat. Sinica* **14** 763–788.

[3] Masry, E.(1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *J. Time Series Anal.* **17** 571-599.