The Annals of Statistics
2012, Vol. 40, No. 6, 2943–2972
DOI: 10.1214/12-AOS1056

TWO-STEP SPLINE ESTIMATING EQUATIONS FOR GENERALIZED ADDITIVE PARTIALLY LINEAR MODELS WITH LARGE CLUSTER SIZES

BY SHUJIE MA

University of California, Riverside

We propose a two-step estimating procedure for generalized additive partially linear models with clustered data using estimating equations. Our proposed method applies to the case that the number of observations per cluster is allowed to increase with the number of independent subjects. We establish oracle properties for the two-step estimator of each function component such that it performs as well as the univariate function estimator by assuming that the parametric vector and all other function components are known. Asymptotic distributions and consistency properties of the estimators are obtained. Finite-sample experiments with both simulated continuous and binary response variables confirm the asymptotic results. We illustrate the methods with an application to a U.S. unemployment data set.

1. Introduction. The generalized estimating equations (GEE) approach has been widely applied to the analysis of clustered data. Reference [15] introduced the GEEs to estimate the regression parameters of generalized linear models with possible unknown correlations between responses. The GEE approach only requires the first two marginal moments and a working correlation matrix that accounts for the form of within-subject correlations of responses, and it can yield consistent parameter estimators even when the covariance structure is misspecified, as long as the mean function is correctly specified.

Parametric GEEs enjoy simplicity by assuming a fully predetermined parametric form for the mean function, but they have suffered from inflexibility in modeling complicated relationships between the response and covariates in clustered data studies. To allow for flexibility, [9, 32] and [16] proposed to model covariate effects nonparametrically via GEE. The proposed nonparametric GEE method enables us to capture the underlying structure that otherwise can be missed. Reference [17] extended the kernel estimating equations in [16] to generalized partially linear models (GPLMs), which assume that the mean of the outcome variable depends on a vector of covariates parametrically and a scalar predictor nonparametrically to overcome the "curse of dimensionality" of nonparametric models. As an extension, [6] and [14] approximated the nonparametric function in GPLMs

Received January 2012; revised June 2012.
MSC2010 subject classifications. Primary 62G08; secondary 62G20.
Key words and phrases. Estimating equations, generalized additive partially linear models, clustered data, longitudinal data, infinite cluster sizes, spline.

by regression splines. It is pointed out in [31] and [18] that splines effectively account for the correlations of clustered data and are more efficient in nonparametric models with longitudinal data than conventional local-polynomials. Splines also provide optimal convergence rates in partially linear models [7, 8]. To allow the nonparametric part in partially linear models to include multivariate covariates, [21] extended the estimating equations method to generalized additive partially linear models (GAPLMs) with an identity link for continuous response cases, and obtained estimators for the parametric vector and the nonparametric additive functions via a one-step spline estimation.

To introduce GAPLMs for clustered data, denote $\{(Y_{ij}, \mathbf{X}_{ij}, \mathbf{Z}_{ij}), 1 \le i \le n, 1 \le j \le m_i\}$ as the jth repeated observation for the ith subject or experimental unit, where Y_{ij} is the response variable, $\mathbf{X}_{ij} = (1, X_{ij1}, \ldots, X_{ij(d_1-1)})^{\mathrm{T}}$ and $\mathbf{Z}_{ij} = (Z_{ij1}, \ldots, Z_{ijd_2})^{\mathrm{T}}$ are d_1-dimensional and d_2-dimensional vectors of covariates, respectively. The marginal model assumes that $Y_{ij} = \mu_{ij} + \varepsilon_{ij}$, and the marginal mean μ_{ij} depends on $\mathbf{X}_{ij}$ and $\mathbf{Z}_{ij}$ through a known monotonic and differentiable link function ϑ, so that the GAPLM is given as

$$(1) \qquad \eta_{ij} = \vartheta(\mu_{ij}) = \mathbf{X}_{ij}^{\mathrm{T}} \boldsymbol{\beta} + \sum_{l=1}^{d_2} \theta_l(Z_{ijl}), \qquad j = 1, \ldots, m_i, i = 1, \ldots, n,$$

where $\boldsymbol{\beta}$ is a d_1-dimensional regression parameter, and θ_l, $l = 1, \ldots, d_2$, are unknown but smooth functions. We assume $\underline{\boldsymbol{\varepsilon}}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{im_i})^{\mathrm{T}} \sim (\mathbf{0}, \boldsymbol{\Sigma}_i)$. For identifiability, both the additive and linear components must be centered, that is, $E\theta_l(Z_{ijl}) \equiv 0$, $l = 1, \ldots, d_2$, $EX_{ijk} = 0$, $k = 1, \ldots, d_1$. Model (1) can either become a generalized additive model [5] if the parameter vector $\boldsymbol{\beta} = \mathbf{0}$ or be a generalized linear model if $\theta_l(\cdot) = 0, 1 \le l \le d_2$. Model (1) is more parsimonious and easier to interpret than purely generalized additive models by allowing a subset of predictors to be discrete and unbounded, modeled as some of the variables $(X_{ijk})_{k=0}^{d_1-1}$ and more flexible than generalized linear models by allowing nonlinear relationships.

The GEE methods have been widely applied to analyze clustered data with small cluster sizes and a large number of subjects n. However, data with large cluster sizes have occurred frequently in various fields such as machine learning, pattern recognition, image analysis, information retrieval and bioinformatics. Reference [33] first studied the asymptotics for parametric GEE estimators with large cluster sizes. As an extension, we develop asymptotic properties of the spline GEE estimators in the GAPLMs (1) when the cluster sizes are allowed to increase with n, that is, the maximum cluster size $m_{(n)} = \max_{1 \le i \le n} m_i$ is a function of n, such that $m_{(n)} \to \infty$ as $n \to \infty$.

The one-step spline estimation in [21] for GAPLMs with identity link is fast to compute but lacks limiting distribution. The traditional backfitting approach has been widely used to estimate additive models for independent and identically distributed (i.i.d.) and weekly-dependent data [5, 23, 25]. It, however, has computational burden issues, due to its iterative nature. Moreover, it is pointed out in [12] that derivation of the asymptotic properties of a backfitting estimator for a model with a link function is very complicated. As an alternative, [10, 12, 19] and [11] proposed two-stage kernel based estimators for i.i.d. data including one step backfitting of the integration estimators in [19] and one step backfitting of the projection estimators in [10], one Newton step from the nonlinear least squares estimators in [12], and the extension of the method in [12] to additive quantile regression models. The two-stage estimator enjoys the oracle property which backfitting estimators do not have, that is, it performs as well as the univariate function estimator by assuming that other components are known.

In this paper, we propose a two-step spline GEE approach to approximate $\theta_l(\cdot)$ for $1 \le l \le d_2$ in model (1) with $m_{(n)}$ going to infinity or bounded, and establish oracle efficiency such that the two-step spline GEE estimator of $\theta_l(\cdot)$ achieves the same asymptotic distribution of the oracle estimator obtained by assuming that $\boldsymbol{\beta}$ and other functions $\theta_{l'}(\cdot)$ for $1 \le l' \le d_2$ and $l' \ne l$ are known. In the first step, the additive components $\theta_{l'}(\cdot)$ for $1 \le l' \le d_2$ and $l' \ne l$ are pre-estimated by their pilot estimators through an undersmoothed spline procedure. In the second step, a more smoothed spline estimating procedure is applied to the univariate data to estimate $\theta_l(\cdot)$ with asymptotic distribution established. The proposed two-step estimators achieve uniform oracle efficiency by "reducing bias via undersmoothing" in the first step and "averaging out the variance" in the second step. We establish asymptotic consistency and normality of the one-step estimator for the parameter vector and the two-step estimators of the nonparametric components. The two-step spline GEE approach is inspired by the idea of "spline-backfitted kernel/spline smoothing" of [20, 26, 29] and [22] for additive models, additive coefficient models and additive partially linear models with i.i.d or weekly-dependent data by using least squares. The complex correlations within the clusters as well as the non-Gaussian nature of discrete data make the estimation and development of asymptotic properties in the framework studied in this paper much more challenging.

2. Two-step spline estimating equations. For simplicity, we denote vectors $\underline{\mathbf{Y}}_i = \{(Y_{i1}, \ldots, Y_{im_i})^{\mathrm{T}}\}_{m_i \times 1}$ and $\underline{\boldsymbol{\eta}}_i = \{(\eta_{i1}, \ldots, \eta_{im_i})^{\mathrm{T}}\}_{m_i \times 1}$, $1 \le m_i \le m_{(n)}$, $1 \le i \le n$. Let $\varepsilon_{ij} = Y_{ij} - \mu_{ij}$, and $\underline{\boldsymbol{\varepsilon}}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{im_i})^{\mathrm{T}}$. Similarly, let $\underline{\mathbf{X}}_i = \{(\mathbf{X}_{i1}, \ldots, \mathbf{X}_{im_i})^{\mathrm{T}}\}_{m_i \times d_1}$ and $\underline{\mathbf{Z}}_i = \{(\mathbf{Z}_{i1}, \ldots, \mathbf{Z}_{im_i})^{\mathrm{T}}\}_{m_i \times d_2}$. Assume that Z_{ijl} is distributed on a compact interval $[a_l, b_l]$, $1 \le l \le d_2$, and, without loss of generality, we take all intervals $[a_l, b_l] = [0, 1]$, $1 \le l \le d_2$. We further let $\theta_l(\mathbf{Z}_{il}) = \{\{\theta_l(Z_{i1l}), \ldots, \theta_l(Z_{im_il})\}^{\mathrm{T}}\}_{m_i \times 1}$, for $l = 1, \ldots, d_2$. The mean function in model (1) can be written in matrix notation as $\underline{\boldsymbol{\eta}}_i = \underline{\mathbf{X}}_i \boldsymbol{\beta} + \sum_{l=1}^{d_2} \theta_l(\mathbf{Z}_{il})$, which is the marginal model [15]. Let $\mu(\cdot) = \vartheta^{-1}(\cdot)$ be the inverse of the link function and $\mu(\underline{\boldsymbol{\eta}}_i) = [\{\mu(\eta_{i1}), \ldots, \mu(\eta_{im_i})\}^{\mathrm{T}}]_{m_i \times 1}$.

As in [30], we allow $\underline{\mathbf{X}}_i$ and $\underline{\mathbf{Z}}_i$ to be dependent. Let $\mathbf{V}_i = \mathbf{V}_i(\underline{\mathbf{X}}_i, \underline{\mathbf{Z}}_i)$ be the assumed "working" covariance of $\underline{\mathbf{Y}}_i$, where $\mathbf{V}_i = \mathbf{A}_i^{1/2}\mathbf{R}_i(\boldsymbol{\alpha})\mathbf{A}_i^{1/2}$, $\mathbf{A}_i = \mathbf{A}_i(\underline{\mathbf{X}}_i, \underline{\mathbf{Z}}_i)$ denotes an $m_i \times m_i$ diagonal matrix that contains the marginal variances of Y_{ij}, and $\mathbf{R}_i$ is an invertible working correlation matrix, which depends on a nuisance parameter vector $\boldsymbol{\alpha}$. Let $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_i(\underline{\mathbf{X}}_i, \underline{\mathbf{Z}}_i)$ be the true covariance of $\underline{\mathbf{Y}}_i$. If $\mathbf{R}_i$ is equal to the true correlation matrix $\overline{\mathbf{R}}_i$, then $\mathbf{V}_i = \boldsymbol{\Sigma}_i$.

Following [29], we approximate the nonparametric functions θ_l's by centered polynomial splines. Let G_n be the space of polynomial splines of degree $q \geq 1$. We introduce a knot sequence with N_n interior knots

$$t_{-q} = \cdots = t_{-1} = t_0 = 0 < t_1 < \cdots < t_N < 1 = t_{N+1} = \cdots = t_{N+q+1},$$

where $N \equiv N_n$ increases when the number of subjects n increases, with order assumption given in condition (A4). Then G_n consists of functions ϖ satisfying the following: (i) ϖ is a polynomial of degree q on each of the subintervals $I_s = [t_s, t_{s+1})$, $s = 0, \ldots, N_n - 1$, $I_{Nn} = [t_{N_n}, 1]$; (ii) for $q \geq 1$, ϖ is $q-1$ time continuously differentiable on [0, 1]. Let $J_n = N_n + q + 1$. Let $\{b_{s,l} : 1 \leq l \leq d_2, 1 \leq s \leq J_n + 1\}^{\mathrm{T}}$ be a basis system of the space G_n. We adopt the centered B-spline space G_n^0 introduced in [34], where $\mathbf{B}(\mathbf{z}) = \{B_{s,l}(z_l) : 1 \leq l \leq d_2, 1 \leq s \leq J_n\}^{\mathrm{T}}$ is a basis system of the space G_n^0 with $B_{s,l}(z_l) = \sqrt{N_n}[b_{s+1,l}(z_l) - \{E(b_{s+1,l})/E(b_{1,l})\}b_{1,l}(z_l)]$ and $\mathbf{z} = (z_l)_{l=1}^{d_2}$.

Equally-spaced knots are used in this article for simplicity of proof. Other regular knot sequences can also be used, with similar asymptotic results.

Step I. Pilot estimators of $\boldsymbol{\beta}$ and $\theta_l(\cdot)$. Suppose that θ_l can be approximated well by a spline function in G_n^0, so that

$$\theta_l(z_l) \approx \widetilde{\theta}_l(z_l) = \sum_{s=1}^{J_n} \gamma_{sl} B_{s,l}(z_l). \tag{2}$$

Let $\boldsymbol{\gamma} = (\gamma_{sl} : s = 1, \ldots, J_n, l = 1, \ldots, d_2)^{\mathrm{T}}$ be the collection of the coefficients in (2), and denote $\mathbf{B}_{ijl} = [\{B_{s,l}(Z_{ijl}) : s = 1, \ldots, J_n\}^{\mathrm{T}}]_{J_n \times 1}$ and $\mathbf{B}_{ij} = \{(\mathbf{B}_{ij1}^{\mathrm{T}}, \ldots, \mathbf{B}_{ijd_2}^{\mathrm{T}})^{\mathrm{T}}\}_{d_2 J_n \times 1}$, then we have an approximation $\eta_{ij} \approx \widetilde{\eta}_{ij} = \mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{B}_{ij}^{\mathrm{T}}\boldsymbol{\gamma}$. We can also write the approximation in matrix notation as $\underline{\boldsymbol{\eta}}_i \approx \underline{\widetilde{\boldsymbol{\eta}}}_i = \underline{\mathbf{X}}_i\boldsymbol{\beta} + \underline{\mathbf{B}}_i\boldsymbol{\gamma}$, where $\underline{\mathbf{B}}_i = \{(\mathbf{B}_{i1}, \ldots, \mathbf{B}_{im_i})^{\mathrm{T}}\}_{m_i \times d_2 J_n}$. Let $\mu(\underline{\widetilde{\boldsymbol{\eta}}}_i) = [\{\mu(\widetilde{\eta}_{i1}), \ldots, \mu(\widetilde{\eta}_{im_i})\}^{\mathrm{T}}]_{m_i \times 1}$. Let $\widehat{\boldsymbol{\beta}}_n = (\widehat{\beta}_{n,1}, \ldots, \widehat{\beta}_{n,d_1})^{\mathrm{T}}$ and $\widehat{\boldsymbol{\gamma}}_n = \{\widehat{\gamma}_{n,sl} : s = 1, \ldots, J_n, l = 1, \ldots, d_2\}^{\mathrm{T}}$ be the minimizer of

$$Q_n(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \frac{1}{2}\sum_{i=1}^{n}\{\underline{\mathbf{Y}}_i - \mu(\underline{\mathbf{X}}_i\boldsymbol{\beta} + \underline{\mathbf{B}}_i\boldsymbol{\gamma})\}^{\mathrm{T}}\mathbf{V}_i^{-1}(\boldsymbol{\beta}, \boldsymbol{\gamma})\{\underline{\mathbf{Y}}_i - \mu(\underline{\mathbf{X}}_i\boldsymbol{\beta} + \underline{\mathbf{B}}_i\boldsymbol{\gamma})\}, \tag{3}$$

which is corresponding to the class of working covariance matrices $\{\mathbf{V}_i, 1 \leq i \leq n\}$. Then $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{\boldsymbol{\gamma}}_n$ solve the estimating equations

$$\mathbf{g}_n(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{i=1}^{n} \underline{\mathbf{D}}_i^{\mathrm{T}} \Delta_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) \mathbf{V}_i^{-1}(\boldsymbol{\beta}, \boldsymbol{\gamma})\{\underline{\mathbf{Y}}_i - \mu(\underline{\mathbf{X}}_i\boldsymbol{\beta} + \underline{\mathbf{B}}_i\boldsymbol{\gamma})\} = \mathbf{0}, \tag{4}$$

where $\underline{\mathbf{D}}_i = (\underline{\mathbf{X}}_i, \underline{\mathbf{B}}_i)_{m_i \times (d_1+d_2 J_n)}$, and

$$\Delta_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \operatorname{diag}\bigl(\Delta_{i1}(\boldsymbol{\beta}, \boldsymbol{\gamma}), \ldots, \Delta_{im_i}(\boldsymbol{\beta}, \boldsymbol{\gamma})\bigr)$$

is a diagonal matrix with the diagonal elements being the first derivative of $\mu(\cdot)$ evaluated at $\widetilde{\eta}_{ij}$, $j = 1, \ldots, m_i$. Then we let $\widehat{\boldsymbol{\beta}}_n$ be the estimator of the parameter vector $\boldsymbol{\beta}$. For each $1 \le l \le d_2$, the pilot estimator of the lth nonparametric function $\theta_l(z_l)$ is $\widehat{\theta}_{n,l}(z_l) = \sum_{s=1}^{J_n} \widehat{\gamma}_{n,sl} B_{s,l}(z_l)$. The one-step spline estimator of each function component has consistency properties, but lacks limiting distribution [21, 22, 29].

Step II. Two-step spline GEE estimator of $\theta_l(\cdot)$. Next, we propose a two-step spline estimator of $\theta_l(\cdot)$ for given $1 \le l \le d_2$. The basic idea is that for every $1 \le l \le d_2$, we estimate the lth function $\theta_l(\cdot)$ in model (1) nonparametrically with the GEE method by assuming that the parameter vector $\boldsymbol{\beta}$ and other nonparametric components $\boldsymbol{\theta}_{-l} = \{\theta_{l'}(\cdot) : 1 \le l' \le d_2, l' \ne l\}$ are known. The problem turns into a univariate function estimation problem. Because the true parameter vector $\boldsymbol{\beta}$ and functions $\boldsymbol{\theta}_{-l}$ are not known in reality, we replace them by their pilot estimators from step I to obtain the two-step estimator of $\theta_l(\cdot)$. Both kernel and spline based methods can be employed in the second step to estimate $\theta_l(\cdot)$. Here we choose the spline method described in the beginning of this section. We use the splines of the same degree q as in step I. Denote $\mathbf{B}_{ijl}^{\mathcal{S}} = [\{B_{s,l}^{\mathcal{S}}(Z_{ijl}) : s = 1, \ldots, J_n^{\mathcal{S}}\}^{\mathrm{T}}]_{J_n^{\mathcal{S}} \times 1}$, where $B_{s,l}^{\mathcal{S}}(z_l)$ is the spline function defined in the same way as $B_{s,l}(z_l)$ in step I, but with $N^{\mathcal{S}} \equiv N_n^{\mathcal{S}}$ the number of interior knots and let $J_n^{\mathcal{S}} = N^{\mathcal{S}} + q + 1$. Denote $\mathbf{B}_l^{\mathcal{S}}(z_l) = \{B_{s,l}^{\mathcal{S}}(z_l), s = 1, \ldots, J_n^{\mathcal{S}}\}^{\mathrm{T}}$, $\mathbf{B}_{i\cdot l}^{\mathcal{S}} = \{(\mathbf{B}_{i1l}, \ldots, \mathbf{B}_{im_il})^{\mathrm{T}}\}_{m_i \times J_n^{\mathcal{S}}}$, and $\boldsymbol{\gamma}_l^{\mathcal{S}} = (\gamma_{sl} : s = 1, \ldots, J_n^{\mathcal{S}})^{\mathrm{T}}$. By assuming that $\boldsymbol{\beta}$ and $\boldsymbol{\theta}_{-l} = \{\theta_{l'}(\cdot) : l' \ne l, 1 \le l' \le d_2\}$ are known, $\theta_l(z_l)$ is estimated by the oracle estimator

$$\widehat{\theta}_{n,l}^{\mathcal{S}}(z_l, \boldsymbol{\beta}, \boldsymbol{\theta}_{-l}) = \sum_{s=1}^{J_n} \widehat{\gamma}_{n,sl}^{\mathcal{S}}(\boldsymbol{\beta}, \boldsymbol{\theta}_{-l}) B_{s,l}^{\mathcal{S}}(z_l) = \mathbf{B}_l^{\mathcal{S}}(z_l)^{\mathrm{T}} \widehat{\boldsymbol{\gamma}}_{n,l}^{\mathcal{S}}(\boldsymbol{\beta}, \boldsymbol{\theta}_{-l}) \tag{5}$$

with $\widehat{\boldsymbol{\gamma}}_{n,l}^{\mathcal{S}}(\boldsymbol{\beta}, \boldsymbol{\theta}_{-l}) = \{\widehat{\gamma}_{n,sl}^{\mathcal{S}}(\boldsymbol{\beta}, \boldsymbol{\theta}_{-l})\}_{s=1}^{J_n^{\mathcal{S}}}$ solving the estimating equation

$$\begin{aligned} &\mathbf{g}_{n,l}^{\mathcal{S}}(\boldsymbol{\gamma}_l^{\mathcal{S}}, \boldsymbol{\beta}, \boldsymbol{\theta}_{-l}) \\ &\quad = \sum_{i=1}^{n} (\mathbf{B}_{i\cdot l}^{\mathcal{S}})^{\mathrm{T}} \Delta_i(\boldsymbol{\beta}, \boldsymbol{\theta}_{-l}, \boldsymbol{\gamma}_l^{\mathcal{S}}) \mathbf{V}_i^{-1}(\boldsymbol{\beta}, \boldsymbol{\theta}_{-l}, \boldsymbol{\gamma}_l^{\mathcal{S}}) \\ &\qquad \times \left\{ \underline{\mathbf{Y}}_i - \mu\left(\underline{\mathbf{X}}_i \boldsymbol{\beta} + \sum_{l'=1, l' \ne l}^{d_2} \theta_{l'}(\mathbf{Z}_{il'}) + (\mathbf{B}_{i\cdot l}^{\mathcal{S}})^{\mathrm{T}} \boldsymbol{\gamma}_l^{\mathcal{S}} \right) \right\} \\ &\quad = \mathbf{0}, \end{aligned} \tag{6}$$

where $\Delta_i(\boldsymbol{\beta}, \boldsymbol{\theta}_{-l}, \boldsymbol{\gamma}_l^{\mathcal{S}}) = \operatorname{diag}(\Delta_{i1}(\eta_{i1}^{\mathcal{S}}), \ldots, \Delta_{im_i}(\eta_{im_i}^{\mathcal{S}}))$, and $\Delta_{ij}(\eta_{ij}^{\mathcal{S}})$ is the first derivative of $\mu(\cdot)$ evaluated at $\eta_{ij}^{\mathcal{S}} = \mathbf{X}_{ij}^{\mathrm{T}} \boldsymbol{\beta} + \sum_{l'=1, l' \ne l}^{d_2} \theta_{l'}(Z_{ijl'}) + (\mathbf{B}_{ijl}^{\mathcal{S}})^{\mathrm{T}} \boldsymbol{\gamma}_l^{\mathcal{S}}$, $j =$

$1, \dots, m_i$. We replace the true parameter vector $\boldsymbol{\beta}$ and the true functions $\boldsymbol{\theta}_{-l} = \{\theta_{l'}(\cdot), 1 \le l' \le d_2, l' \ne l\}$ with the pilot estimators $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{\boldsymbol{\theta}}_{n,-l} = \{\widehat{\theta}_{n,l'}(\cdot), 1 \le l' \le d_2, l' \ne l\}$, where $\widehat{\theta}_{n,l'}(z_{l'}) = \sum_{s=1}^{J_n} \widehat{\gamma}_{n,sl'} B_{s,l'}(z_{l'})$, so that $\theta_l(z_l)$ is estimated by the two-step spline estimator

$$\widehat{\theta}_{n,l}^{\mathcal{S}}(z_l, \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\theta}}_{n,-l}) = \mathbf{B}_l^{\mathcal{S}}(z_l)^{\mathrm{T}} \widehat{\boldsymbol{\gamma}}_{n,l}^{\mathcal{S}}(\widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\theta}}_{n,-l}). \tag{7}$$

The Newton–Raphson algorithm of GEE is applied to obtain $\widehat{\boldsymbol{\gamma}}_{n,l}^{\mathcal{S}}$. Define

$$\mathcal{D}_n(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \{-\partial \mathbf{g}_n(\boldsymbol{\beta}, \boldsymbol{\gamma}) / \partial(\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\gamma}^{\mathrm{T}})\}_{(d_1 + d_2 J_n) \times (d_1 + d_2 J_n)},$$

$$\Psi_n(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \left\{ \sum_{i=1}^n \underline{\mathbf{D}}_i^{\mathrm{T}} \Delta_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) \mathbf{V}_i^{-1}(\boldsymbol{\beta}, \boldsymbol{\gamma}) \Delta_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) \underline{\mathbf{D}}_i \right\}_{(d_1 + d_2 J_n) \times (d_1 + d_2 J_n)}.$$

3. Asymptotic properties of the estimators. For any $s \times s$ symmetric matrix $\mathbf{A}$, denote by $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ its smallest and largest eigenvalues. For any vector $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_s)^{\mathrm{T}}$, let its Euclidean norm be $\|\boldsymbol{\alpha}\| = \sqrt{\alpha_1^2 + \cdots + \alpha_s^2}$. Let $C^{0,1}(\mathcal{X}_w)$ be the space of Lipschitz continuous functions on $\mathcal{X}_w$, that is,

$$C^{0,1}(\mathcal{X}_w) = \left\{ \varphi : \|\varphi\|_{0,1} = \sup_{w \ne w', w, w' \in \mathcal{X}_w} \frac{|\varphi(w) - \varphi(w')|}{|w - w'|} < +\infty \right\},$$

in which $\|\varphi\|_{0,1}$ is the $C^{0,1}$-norm of φ. Throughout the paper, we assume the following regularity conditions:

(C1) The random variables Z_{ijl} are bounded, uniformly in $1 \le j \le m_i$, $1 \le i \le n$, $1 \le l \le d_2$. The marginal density $f_{ijl}(\cdot)$ of Z_{ijl} is bounded away from 0 and ∞ on $[0, 1]$, uniformly in $1 \le j \le m_i$, $1 \le i \le n$. The joint density $f_{ijlj'l'}(\cdot, \cdot)$ of $(Z_{ijl}, Z_{ij'l'})$ is bounded away from 0 and ∞ on $[0, 1]$, uniformly in $1 \le i \le n$, $1 \le j, j' \le m_i$, and $1 \le l \ne l' \le d_2$.

(C2) The eigenvalues of the true correlation matrices $\overline{\mathbf{R}}_i$ are bounded away from 0, uniformly in $1 \le i \le n$.

(C3) The eigenvalues of the inverse of the working correlation matrices $\mathbf{R}_i(\alpha)^{-1}$ are bounded away from 0, uniformly in $1 \le i \le n$.

(C4) Let $n_{\mathrm{T}} = \sum_{i=1}^n m_i$. There are constants $0 < c < C < \infty$, such that $c n_{\mathrm{T}} \le \lambda_{\min}(\sum_{i=1}^n \underline{\mathbf{X}}_i^{\mathrm{T}} \underline{\mathbf{X}}_i) \le \lambda_{\max}(\sum_{i=1}^n \underline{\mathbf{X}}_i^{\mathrm{T}} \underline{\mathbf{X}}_i) \le C n_{\mathrm{T}}$.

(C5) For $1 \le l \le d_2$, $\theta_l^{(p-1)}(z_l) \in C^{0,1}[0, 1]$, for given integer $p \ge 1$. The spline degree satisfies $q + 1 \ge p$, and $\mu'(\eta) \in C^{0,1}(\mathcal{X}_\eta)$. The number of interior knots $N_n \to \infty$, as $n_{\mathrm{T}} \to \infty$.

Conditions (C1)–(C4) are similar to conditions (A1)–(A4) in [21], and condition (C5) is weaker than the first part of condition (A5) in [21]. Let $\boldsymbol{\beta}_0$ be the true parameter vector and $\theta_{l0}(\cdot)$ be the true lth additive function in model (1). According to the result on page 149 of [3], for $\theta_{l0}(\cdot)$ satisfying condition (C5), there is a function

$$\widetilde{\theta}_{l0}(z_l) = \sum_{s=1}^{J_n} \gamma_{sl,0} B_{s,l}(z_l) \in G_n^0, \tag{8}$$

such that $\sup_{z_l \in [0,1]} |\widetilde{\theta}_{l0}(z_l) - \theta_{l0}(z_l)| = O(J_n^{-p})$. Thus, by letting $\boldsymbol{\gamma}_0 = (\gamma_{sl,0} : s = 1, \ldots, J_n, l = 1, \ldots, d_2)^{\mathrm{T}}$,

$$\sup_{\mathbf{z} \in [0,1]^{d_2}} \left| \mathbf{B}(\mathbf{z})^{\mathrm{T}} \boldsymbol{\gamma}_0 - \sum_{l=1}^{d_2} \theta_{l0}(z_l) \right| \le \sum_{l=1}^{d_2} \sup_{z_l \in [0,1]} \left| \widetilde{\theta}_{l0}(z_l) - \theta_{l0}(z_l) \right| = O\left(d_2 J_n^{-p}\right).$$

In addition to the regularity conditions above, we need extra conditions to ensure the existence and weak consistency of the estimators in (4). Let $\lambda_n^{\min} = \min_{1 \le i \le n} \lambda_{\min}\{\mathbf{R}_i^{-1}(\alpha)\}$, $\lambda_n^{\max} = \max_{1 \le i \le n} \lambda_{\max}\{\mathbf{R}_i^{-1}(\alpha)\}$, $\tau_n^{\max} = \max_{1 \le i \le n}\{\lambda_{\max}(\mathbf{R}_i^{-1}(\alpha)\overline{\mathbf{R}}_i)\}$ and $\tau_n^{\min} = \min_{1 \le i \le n}\{\lambda_{\min}(\mathbf{R}_i^{-1}(\alpha)\overline{\mathbf{R}}_i)\}$. The additional conditions are as follows:

(A1) $(\lambda_n^{\min}/\tau_n^{\max}) n_{\mathrm{T}} / J_n^{1/2} \to \infty$.

(A2) There is a constant $c_0 > 0$, for any $r > 0$, such that $P\{\mathcal{D}_n(\boldsymbol{\beta}, \boldsymbol{\gamma}) \ge c_0 \Psi_n(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$ and $\mathcal{D}_n(\boldsymbol{\beta}, \boldsymbol{\gamma})$ is nonsingular, for all $(\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\gamma}^{\mathrm{T}})^{\mathrm{T}} \in \xi_n(r)\} \to 1$, where $\xi_n(r) = \{(\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\gamma}^{\mathrm{T}})^{\mathrm{T}} : \|\{\Psi_n(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)\}^{1/2}((\boldsymbol{\beta} - \boldsymbol{\beta}_0)^{\mathrm{T}}, (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^{\mathrm{T}})^{\mathrm{T}}\| \le (\tau_n^{\max})^{1/2} r\}$.

Conditions (A1) and (A2) are used to ensure the existence and weak consistency of the solutions in (4). Condition (A2) corresponds to condition ($\mathrm{L}_{\mathrm{w}}^*$) in [33] for generalized linear models. Conditions (A1) and (C4) imply condition ($\mathrm{I}_{\mathrm{w}}^*$) in [33], which will be proved in the Appendix. Condition (A2) relates to the true and the working correlation structures $\overline{\mathbf{R}}_i$ and $\mathbf{R}_i(\alpha)$. Since the true correlations $\overline{\mathbf{R}}_i$ are often not completely specified and $\max_{1 \le i \le n} \lambda_{\max}(\overline{\mathbf{R}}_i) \le m_{(n)}$, then condition (A1) is implied by

(A1*) $(\lambda_n^{\min}/\lambda_n^{\max}) m_{(n)}^{-1} n_{\mathrm{T}} / J_n^{1/2} \to \infty$.

Condition (A1*) does not contain $\overline{\mathbf{R}}_i$. Thus, the order requirements of n, $m_{(n)}$ and J_n depend on the choice of the working correlations $\mathbf{R}_i(\alpha)$. For instance, if the working correlation structures are independent or AR(1) within each subject, then there exist constants $0 < c_R \le C_R < \infty$, such that $c_R \le (\lambda_n^{\max})^{-1} \lambda_n^{\min} \le C_R$. Thus, condition (A1*) is equivalent to $m_{(n)}^{-1} n_{\mathrm{T}} / J_n^{1/2} \to \infty$. For exchangeable working correlation structures, there exist constants $0 < C_R' < \infty$, such that $\max_{1 \le i \le n} \lambda_{\max}\{\mathbf{R}_i(\alpha)\} \le C_R' m_{(n)}$, then $(\lambda_n^{\max})^{-1} \lambda_n^{\min} \ge c_R' m_{(n)}^{-1}$, for some constant $0 < c_R' < \infty$. Condition (A1*) is implied by $m_{(n)}^{-2} n_{\mathrm{T}} / J_n^{1/2} \to \infty$.

THEOREM 1. *Under conditions* (A1) *and* (A2) *or* (A1*) *and* (A2), *as* $n_{\mathrm{T}} \to \infty$, *there exist sequences of random variables* $\widehat{\boldsymbol{\beta}}_n$ *and* $\widehat{\boldsymbol{\gamma}}_n$, *such that* $P\{\mathbf{g}_n(\widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\gamma}}_n) = 0\} \to 1$, *and* $\|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\| \to 0$ *and* $\|\widehat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_0\| \to 0$ *in probability.*

Next we derive the asymptotic properties of $\widehat{\boldsymbol{\beta}}_n$. Let $\mathcal{X}$ and $\mathcal{Z}$ be the collections of all X_{ijk}'s and Z_{ijl}'s, respectively, that is, $\mathcal{X}_{n_{\mathrm{T}}\times d_1} = (\underline{\mathbf{X}}_1^{\mathrm{T}}, \ldots, \underline{\mathbf{X}}_n^{\mathrm{T}})^{\mathrm{T}}$ and $\mathcal{Z}_{n_{\mathrm{T}}\times d_2} = (\underline{\mathbf{Z}}_1^{\mathrm{T}}, \ldots, \underline{\mathbf{Z}}_n^{\mathrm{T}})^{\mathrm{T}}$. Let Δ_i be the diagonal matrix with the diagonal elements being the first derivative of $\mu(\cdot)$ evaluated at $\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_0 + \sum_{l=1}^{d_2}\theta_{l0}(Z_{ijl})$, $j = 1, \ldots, m_i$, and $\mathbf{V}_i = \mathbf{A}_i^{1/2}\mathbf{R}_i(\boldsymbol{\alpha})\mathbf{A}_i^{1/2}$ with $\mathbf{A}_i$ being the marginal variance of $\underline{\mathbf{Y}}_i$ evaluated at the true parameters and additive functions. To make $\boldsymbol{\beta}$ estimable, we need a condition to ensure $\mathcal{X}$ and $\mathcal{Z}$ not functionally related, which is similar to the condition given in [21]. Define the Hilbert space $\mathcal{H} = \{\psi(\mathbf{z}) = \sum_{l=1}^{d_2}\psi_l(z_l), E\psi_l(z_l) = 0, \|\psi_l\|_2 < \infty\}$ of theoretically centered L_2 additive functions on $[0, 1]^{d_2}$, where $\|\psi_l\|_2 = \{\int_0^1 \psi_l^2(z_l)\,dz_l\}^{1/2}$. Let ψ_k^* be the function $\psi \in \mathcal{H}$ that minimizes $\sum_{i=1}^n E[\{\underline{\mathbf{X}}_i^{(k)} - \psi(\underline{\mathbf{Z}}_i)\}^{\mathrm{T}}\Delta_i\mathbf{V}_i^{-1}\Delta_i\{\underline{\mathbf{X}}_i^{(k)} - \psi(\underline{\mathbf{Z}}_i)\}]$, where $\underline{\mathbf{X}}_i^{(k)} = (X_{i1k}, \ldots, X_{im_ik})^{\mathrm{T}}$, $1 \le k \le d_1$. Some other assumptions needed are given as follows.

(A3) Given $1 \le k \le d_1$, $\psi_k^{*(p-1)} \in C^{0,1}[0, 1]$, for $1 \le p \le q + 1$.

The order requirements of the number of interior knots N and $N^{\mathcal{S}}$ in steps I and II are given in the following assumption:

(A4) (i) $\sqrt{(\log n_{\mathrm{T}})N^{\mathcal{S}}/n_{\mathrm{T}}}(\tau_n^{\max}/\lambda_n^{\min})^{1/2} = o(1)$, $(N^{\mathcal{S}})^{-p-1/2}n_{\mathrm{T}}^{1/2}(\lambda_n^{\max}/\lambda_n^{\min})(\lambda_n^{\max}/\tau_n^{\min})^{1/2} = O(1)$, and (ii) $(\lambda_n^{\max}/\tau_n^{\min})^{1/2}(\lambda_n^{\max}/\lambda_n^{\min})^2(n_{\mathrm{T}}/N^{\mathcal{S}})^{1/2} \times N^{-p} = o(1)$, $(\lambda_n^{\max}/\tau_n^{\min})^{1/2}(\lambda_n^{\max}/\lambda_n^{\min})^2(\log n_{\mathrm{T}}/N^{\mathcal{S}})^{1/2} = o(1)$, $(N_n^{\mathcal{S}}N_n \times \log n_{\mathrm{T}})^{1/2}n_{\mathrm{T}}^{-1} = o(1)$.

Since $\lambda_n^{\min} \le \tau_n^{\min} \le \tau_n^{\max} \le m_{(n)}\lambda_n^{\max}$, condition (A4) is implied by a stronger condition as below:

(A4*) (i) $\sqrt{(\log n_{\mathrm{T}})N^{\mathcal{S}}/n_{\mathrm{T}}}m_{(n)}^{1/2}(\lambda_n^{\max}/\lambda_n^{\min})^{1/2} = o(1)$, $(N^{\mathcal{S}})^{-p-1/2} \times n_{\mathrm{T}}^{1/2}(\lambda_n^{\max}/\lambda_n^{\min})^{3/2} = O(1)$, and (ii) $(\lambda_n^{\max}/\lambda_n^{\min})^{5/2}(n_{\mathrm{T}}/N^{\mathcal{S}})^{1/2}N^{-p} = o(1)$, $(\lambda_n^{\max}/\lambda_n^{\min})^{5/2}(\log n_{\mathrm{T}}/N^{\mathcal{S}})^{1/2} = o(1)$, $(N_n^{\mathcal{S}}N_n \log n_{\mathrm{T}})^{1/2}n_{\mathrm{T}}^{-1} = o(1)$.

Condition (A3) is weaker than the second part of condition (A5) in [21]. Condition (A4*) does not depend on the true correlation matrices $\overline{\mathbf{R}}_i$, which are not specified. It is clear that the first conditions in (A4) and (A4*) ensure conditions (A1) and (A1*), respectively.

REMARK 1. (A4)(i) lists the order requirements for $N^{\mathcal{S}}$ to obtain the asymptotic results of the oracle estimator in Theorem 3. (A4)(ii) ensures the uniform oracle efficiency of the two-step spline estimator. It will be shown in Theorem 4 that the difference between the two-step spline and the oracle estimators is of uniform order $O_P\{(\lambda_n^{\max}/\lambda_n^{\min})^2(J_n^{-p} + \sqrt{\log n_{\mathrm{T}}/n_{\mathrm{T}}})\}$ with $O_P\{(\lambda_n^{\max}/\lambda_n^{\min})^2\sqrt{\log n_{\mathrm{T}}/n_{\mathrm{T}}}\}$ and $O_P\{(\lambda_n^{\max}/\lambda_n^{\min})^2 J_n^{-p}\}$ caused by the noise and bias terms, respectively, in the first step spline estimation. The inverse of the asymptotic standard deviation of the oracle estimator is of order

$O\{\sqrt{n_{\mathrm{T}}/J_n^{\mathcal{S}}}(\lambda_n^{\max}/\tau_n^{\min})^{1/2}\}$. The first two conditions of (A4)(ii) ensure that the difference is asymptotically uniformly negligible. If we let N have the order $n_{\mathrm{T}}^{1/(2p)}$, then the difference is of uniform order $O_P\{(\lambda_n^{\max}/\lambda_n^{\min})^2\sqrt{\log n_{\mathrm{T}}/n_{\mathrm{T}}}\}$. Therefore, an undersmoothing procedure is applied in the first step to reduce the bias. When $\lambda_n^{\min}$, $\lambda_n^{\max}$, $\tau_n^{\min}$ and $\tau_n^{\max}$ are finite numbers, (A4)(i) becomes $\sqrt{(\log n_{\mathrm{T}})N^{\mathcal{S}}/n_{\mathrm{T}}} = o(1)$ and $(N^{\mathcal{S}})^{-p-1/2}n_{\mathrm{T}}^{1/2} = O(1)$. The optimal order of $N^{\mathcal{S}}$ is $n_{\mathrm{T}}^{1/(2p+1)}$. Define

$$\widetilde{\mathbf{X}}_{ik} = \underline{\mathbf{X}}_i^{(k)} - \psi_k^*(\underline{\mathbf{Z}}_i), \qquad 1 \le k \le d_1, \qquad \underline{\widetilde{\mathbf{X}}}_i = (\widetilde{\mathbf{X}}_{i1}, \ldots, \widetilde{\mathbf{X}}_{id_1})_{m_i \times d_1}.$$

Define $\widetilde{\Psi}_n = \sum_{i=1}^n \underline{\widetilde{\mathbf{X}}}_i^{\mathrm{T}} \Delta_i \mathbf{V}_i^{-1} \Delta_i \underline{\widetilde{\mathbf{X}}}_i$, $\widetilde{\Phi}_n = \sum_{i=1}^n \underline{\widetilde{\mathbf{X}}}_i^{\mathrm{T}} \Delta_i \mathbf{V}_i^{-1} \Sigma_i \mathbf{V}_i^{-1} \Delta_i \underline{\widetilde{\mathbf{X}}}$, and

$$\widetilde{\Xi}_n = \{E(\widetilde{\Psi}_n)\}^{-1} E(\widetilde{\Phi}_n) \{E(\widetilde{\Psi}_n)\}^{-1}. \tag{9}$$

The following result gives the asymptotic distribution and consistency rate of $\widehat{\boldsymbol{\beta}}_n$ for general working covariance matrices.

THEOREM 2. *Under conditions* (A2)–(A4), *as* $n_{\mathrm{T}} \to \infty$, $\widetilde{\Xi}_n^{-1/2}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \to \mathrm{Normal}(0, \mathbf{I}_{d_1})$, *and* $\|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\| = O_p\{n_{\mathrm{T}}^{-1/2}(\tau_n^{\max})^{1/2}(\lambda_n^{\min})^{-1/2}\}$. *If condition* (A4) *is replaced by* (A4*), *then*

$$\|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\| = O_p\{n_{\mathrm{T}}^{-1/2} m_{(n)}^{1/2} (\lambda_n^{\max})^{1/2} (\lambda_n^{\min})^{-1/2}\}.$$

REMARK 2. It is easy to show that the covariance $\widetilde{\Xi}_n$ in (9) is minimized when the working covariance matrices are equal to the true covariance matrices such that $\mathbf{V}_i = \Sigma_i$ for all $1 \le i \le n$, and in this case equal to $\{E(\widetilde{\Psi}_n)\}^{-1}$. To construct the confidence sets for $\boldsymbol{\beta}$, $\widetilde{\Xi}_n$ is consistently estimated by $\widehat{\Xi}_n = \widehat{\Psi}_n^{-1}\widehat{\Phi}_n\widehat{\Psi}_n^{-1}$, where $\widehat{\Psi}_n = \sum_{i=1}^n \underline{\widehat{\mathbf{X}}}_i^{\mathrm{T}} \Delta_i \mathbf{V}_i^{-1} \Delta_i \underline{\widehat{\mathbf{X}}}_i$, $\widehat{\Phi}_n = \sum_{i=1}^n \underline{\widehat{\mathbf{X}}}_i^{\mathrm{T}} \Delta_i \mathbf{V}_i^{-1} \widehat{\boldsymbol{\Sigma}}_i \mathbf{V}_i^{-1} \Delta_i \underline{\widehat{\mathbf{X}}}_i$, and $\underline{\widehat{\mathbf{X}}}_i = \underline{\mathbf{X}}_i - \mathrm{Proj}_{G_n^*} \underline{\mathbf{X}}_i$, $i = 1, \ldots, n$, in which $\mathrm{Proj}_{G_n^*}$ is the projection onto the empirically centered spline inner product space and $\widehat{\boldsymbol{\Sigma}}_i$ is a consistent estimator of $\boldsymbol{\Sigma}_i$.

For $1 \le l \le d_2$, let $\boldsymbol{\gamma}_{l,0}^{\mathcal{S}} = (\gamma_{sl,0})_{s=1}^{J_n^{\mathcal{S}}}$, with $\gamma_{sl,0}$ defined in the same fashion as given in (8), and $\boldsymbol{\theta}_{-l0} = \{\theta_{l'0}(\cdot), 1 \le l' \le d_2, l' \ne l\}$. Define

$$\mathcal{D}_{n,l}^*(\boldsymbol{\gamma}_l^{\mathcal{S}}) = \{-\partial \mathbf{g}_{n,l}^*(\boldsymbol{\gamma}_l^{\mathcal{S}})/\partial(\boldsymbol{\gamma}_l^{\mathcal{S}})^{\mathrm{T}}\}_{J_n^{\mathcal{S}} \times J_n^{\mathcal{S}}},$$

$$\Psi_{n,l}^*(\boldsymbol{\gamma}_{l,0}^{\mathcal{S}}) = \left\{\sum_{i=1}^n (\mathbf{B}_{i\cdot l}^{\mathcal{S}})^{\mathrm{T}} \Delta_i(\boldsymbol{\beta}_0, \boldsymbol{\theta}_{-l0}, \boldsymbol{\gamma}_{l,0}^{\mathcal{S}}) \mathbf{V}_i^{-1}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_{-l0}, \boldsymbol{\gamma}_{l,0}^{\mathcal{S}}) \times \Delta_i(\boldsymbol{\beta}_0, \boldsymbol{\theta}_{-l0}, \boldsymbol{\gamma}_{l,0}^{\mathcal{S}}) \mathbf{B}_{i\cdot l}^{\mathcal{S}}\right\}_{J_n^{\mathcal{S}} \times J_n^{\mathcal{S}}}.$$

In order to ensure the existence and uniformly weak convergence of the oracle estimator $\widehat{\theta}_{n,l}^{\mathcal{S}}(z_l, \boldsymbol{\beta}_0, \boldsymbol{\theta}_{-l0})$, we need the following conditions:

(A5) For $1 \le l \le d_2$, there is a constant $c_l > 0$, for any $r > 0$, such that $P\{\mathcal{D}^*_{n,l}(\boldsymbol{\gamma}^{\mathcal{S}}_l) \ge c_l \Psi^*_{n,l}(\boldsymbol{\gamma}^{\mathcal{S}}_{l,0})$ and $\mathcal{D}^*_{n,l}(\boldsymbol{\gamma}^{\mathcal{S}}_l)$ is nonsingular, for all $\boldsymbol{\gamma}^{\mathcal{S}}_l \in \xi_n(r)\} \to 1$, where $\xi_n(r) = \{\boldsymbol{\gamma}^{\mathcal{S}}_l : \|\{\Psi^*_{n,l}(\boldsymbol{\gamma}^{\mathcal{S}}_{l,0})\}^{1/2}(\boldsymbol{\gamma}^{\mathcal{S}}_l - \boldsymbol{\gamma}^{\mathcal{S}}_{l,0})\| \le (\tau^{\max}_n)^{1/2} r\}$.

For $1 \le l \le d_2$, define $\Xi^*_{n,l} = \{E(\Psi^*_{n,l})\}^{-1} E(\Phi^*_{n,l})\{E(\Psi^*_{n,l})\}^{-1}$, where

$$\Phi^*_{n,l} = \sum_{i=1}^{n} (\mathbf{B}^{\mathcal{S}}_{i\cdot l})^{\mathrm{T}} \Delta_i \mathbf{V}_i^{-1} \Sigma_i \mathbf{V}_i^{-1} \Delta_i \mathbf{B}^{\mathcal{S}}_{i\cdot l}, \qquad \Psi^*_{n,l} = \sum_{i=1}^{n} (\mathbf{B}^{\mathcal{S}}_{i\cdot l})^{\mathrm{T}} \Delta_i \mathbf{V}_i^{-1} \Delta_i \mathbf{B}^{\mathcal{S}}_{i\cdot l}.$$

THEOREM 3. *Let* $\theta^*_{l0}(z_l) = E\{\widehat{\theta}^{\mathcal{S}}_{n,l}(z_l, \boldsymbol{\beta}_0, \boldsymbol{\theta}_{-l0}) | \mathcal{X}, \mathcal{Z}\}$. *Under conditions* (A3), (A4)(i) *and* (A5), *for* $1 \le l \le d_2$ *and* $z_l \in [0, 1]$, *as* $n_{\mathrm{T}} \to \infty$,

$$\big(\mathbf{B}^{\mathcal{S}}_l(z_l)^{\mathrm{T}} \Xi^*_{n,l} \mathbf{B}^{\mathcal{S}}_l(z_l)\big)^{-1/2} \{\widehat{\theta}^{\mathcal{S}}_{n,l}(z_l, \boldsymbol{\beta}_0, \boldsymbol{\theta}_{-l0}) - \theta^*_{l0}(z_l)\} \longrightarrow N(0, 1),$$

$$\sup_{z_l \in [0,1]} \big|\widehat{\theta}^{\mathcal{S}}_{n,l}(z_l, \boldsymbol{\beta}_0, \boldsymbol{\theta}_{-l0}) - \theta^*_{l0}(z_l)\big| = O_P\Big\{\sqrt{(\log n_{\mathrm{T}}) J^{\mathcal{S}}_n / n_{\mathrm{T}}} \big(\tau^{\max}_n / \lambda^{\min}_n\big)^{1/2}\Big\},$$

$$\sup_{z_l \in [0,1]} \big|\theta^*_{l0}(z_l) - \theta_{l0}(z_l)\big| = O_P\{(\lambda^{\max}_n / \lambda^{\min}_n)(J^{\mathcal{S}}_n)^{-p}\}, \tag{10}$$

and there are constants $0 < c_{l,\Xi} \le C_{l,\Xi} < \infty$, *such that for all* $z_l \in [0, 1]$,

$$\{\mathbf{B}^{\mathcal{S}}_l(z_l)^{\mathrm{T}} \Xi^*_{n,l} \mathbf{B}^{\mathcal{S}}_l(z_l)\}^{1/2} \ge c_{l,\Xi} \sqrt{J^{\mathcal{S}}_n / n_{\mathrm{T}}} \big(\tau^{\min}_n / \lambda^{\max}_n\big)^{1/2},$$

$$\{\mathbf{B}^{\mathcal{S}}_l(z_l)^{\mathrm{T}} \Xi^*_{n,l} \mathbf{B}^{\mathcal{S}}_l(z_l)\}^{1/2} \le C_{l,\Xi} \sqrt{J^{\mathcal{S}}_n / n_{\mathrm{T}}} \big(\tau^{\max}_n / \lambda^{\min}_n\big)^{1/2}. \tag{11}$$

Replacing (A4)(i) *by* (A4*)(i), *one has* $\sup_{z_l \in [0,1]} |\widehat{\theta}^{\mathcal{S}}_{n,l}(z_l, \boldsymbol{\beta}_0, \boldsymbol{\theta}_{-l0}) - \theta^*_{l0}(z_l)| = O_P\{\sqrt{(\log n_{\mathrm{T}}) J^{\mathcal{S}}_n m_{(n)} / n_{\mathrm{T}}} (\lambda^{\max}_n / \lambda^{\min}_n)^{1/2}\}$.

REMARK 3. Pointwise confidence intervals for $\theta_{l0}(z_l)$ can be constructed based on the results in Theorem 3. By (10) and (11), the bias term in (10) is asymptotically uniformly negligible through undersmoothing if $(N^{\mathcal{S}})^{-p-1/2} n_{\mathrm{T}}^{1/2} (\lambda^{\max}_n / \lambda^{\min}_n)(\lambda^{\max}_n / \tau^{\min}_n)^{1/2} = o(1)$. Thus, $N^{\mathcal{S}}$ is of the form $[(\lambda^{\max}_n / \lambda^{\min}_n)^2 (\lambda^{\max}_n / \tau^{\min}_n) \times n_{\mathrm{T}}]^{1/(2p+1)} N^*$, where the sequence N^* satisfies $N^* \to \infty$ and $n_{\mathrm{T}}^{-\tau} N^* \to 0$ for any $\tau > 0$. Under (A4*)(i), $N^{\mathcal{S}}$ is of the form $[(\lambda^{\max}_n / \lambda^{\min}_n)^3 n_{\mathrm{T}}]^{1/(2p+1)} N^*$.

Theorem 3 presents asymptotic normality and uniform convergence rate of the oracle estimator $\widehat{\theta}^{\mathcal{S}}_{n,l}(z_l, \boldsymbol{\beta}_0, \boldsymbol{\theta}_{-l0})$. The oracle estimator achieves the convergence rate of univariate spline regression function estimation. References [35] and [13] studied asymptotic normality of spline estimators for nonparametric regression functions with i.i.d. data. Reference [14] established the asymptotic distribution for the univariate spline estimator in partially linear models for clustered data with $m_{(n)} < \infty$. Reference [13] discussed the difficulty of obtaining asymptotic normality of spline estimators for additive models. Reference [21] studied convergence rate of the one-step additive spline estimator for clustered data with $m_{(n)} < \infty$, but it lacks the limiting distribution. The next theorem will present the uniform convergence rate of the two-step spline estimator $\widehat{\theta}^{\mathcal{S}}_{n,l}(z_l, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}_{n,-l})$ to the oracle estimator $\widehat{\theta}^{\mathcal{S}}_{n,l}(z_l, \boldsymbol{\beta}_0, \boldsymbol{\theta}_{-l0})$, and establish the asymptotic normality of $\widehat{\theta}^{\mathcal{S}}_{n,l}(z_l, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}_{n,-l})$.

THEOREM 4. *Under conditions* (A2)–(A5), *for* $1 \le l \le d_2$,

$$\begin{aligned}
&\sup_{z_l \in [0,1]} \left|\widehat{\theta}^{\mathcal{S}}_{n,l}(z_l, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}_{n,-l}) - \widehat{\theta}^{\mathcal{S}}_{n,l}(z_l, \boldsymbol{\beta}_0, \boldsymbol{\theta}_{-l0})\right| \\
&\quad = O_p\left\{\left(\lambda_n^{\max}/\lambda_n^{\min}\right)^2\left(\sqrt{\log n_{\mathrm{T}}/n_{\mathrm{T}}} + J_n^{-p}\right)\right\} \qquad (12)\\
&\quad = o_p\left\{\left(J_n^{\mathcal{S}}/n_{\mathrm{T}}\right)^{1/2}\left(\tau_n^{\min}/\lambda_n^{\max}\right)^{1/2}\right\}
\end{aligned}$$

and replacing (A4) *by* (A4*),

$$\begin{aligned}
&\sup_{z_l \in [0,1]} \left|\widehat{\theta}^{\mathcal{S}}_{n,l}(z_l, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}_{n,-l}) - \widehat{\theta}^{\mathcal{S}}_{n,l}(z_l, \boldsymbol{\beta}_0, \boldsymbol{\theta}_{-l0})\right| \\
&\quad = o_p\left\{\left(J_n^{\mathcal{S}}/n_{\mathrm{T}}\right)^{1/2}\left(\lambda_n^{\min}/\lambda_n^{\max}\right)^{1/2}\right\}.
\end{aligned}$$

Hence, for $1 \le l \le d_2$ *and* $z_l \in [0, 1]$, *as* $n_{\mathrm{T}} \to \infty$,

$$\left(\mathbf{B}_l^{\mathcal{S}}(z_l)^{\mathrm{T}} \Xi^*_{n,l} \mathbf{B}_l^{\mathcal{S}}(z_l)\right)^{-1/2} \left\{\widehat{\theta}^{\mathcal{S}}_{n,l}(z_l, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}_{n,-l}) - \theta^*_{l0}(z_l)\right\} \longrightarrow N(0, 1).$$

REMARK 4. Similarly as $\widetilde{\Xi}_n$ in (9), $\Xi^*_{n,l}$ is minimized when $\mathbf{V}_i = \Sigma_i$ for all $1 \le i \le n$, and in this case is equal to $\{E(\Psi^*_{n,l})\}^{-1}$. To construct a pointwise confidence interval for $\theta_{l0}(z_l)$ at $z_l \in [0, 1]$, $\Xi^*_{n,l}$ is consistently estimated by $\widehat{\Xi}^*_{n,l} = \widehat{\Psi}^{*-1}_{n,l} \widehat{\Phi}^*_{n,l} \widehat{\Psi}^{*-1}_{n,l}$, where $\widehat{\Psi}^*_{n,l} = \sum_{i=1}^n (\mathbf{B}^{\mathcal{S}}_{i\cdot l})^{\mathrm{T}} \Delta_i \mathbf{V}_i^{-1} \Delta_i \mathbf{B}^{\mathcal{S}}_{i\cdot l}$ and $\widehat{\Phi}^*_{n,l} = \sum_{i=1}^n (\mathbf{B}^{\mathcal{S}}_{i\cdot l})^{\mathrm{T}} \Delta_i \mathbf{V}_i^{-1} \widehat{\boldsymbol{\Sigma}}_i \mathbf{V}_i^{-1} \Delta_i \mathbf{B}^{\mathcal{S}}_{i\cdot l}$. Then under the assumption given in Remark 3, for any $\alpha \in (0, 1)$, an asymptotic $100(1 - \alpha)\%$ pointwise confidence interval for $\theta_{l0}(z_l)$ is

$$\widehat{\theta}^{\mathcal{S}}_{n,l}(z_l, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}_{n,-l}) \pm z_{\alpha/2}\left(\mathbf{B}_l^{\mathcal{S}}(z_l)^{\mathrm{T}} \Xi^*_{n,l} \mathbf{B}_l^{\mathcal{S}}(z_l)\right)^{1/2}. \qquad (13)$$

REMARK 5. By letting N have order $n_{\mathrm{T}}^{1/(2p)}$, the difference in (12) is of uniform order $O_P\{(\lambda_n^{\max}/\lambda_n^{\min})^2\sqrt{\log n_{\mathrm{T}}/n_{\mathrm{T}}}\}$. So undersmoothing is applied to reduce the approximation error caused by the bias in the first step.

4. Simulation. In this section we conduct simulations to illustrate the finite-sample behavior of the proposed GEE estimators for both normal and binary responses. For each procedure, we consider three different working correlation structures: independence (IND), exchangeable (EX) and first order autocorrelation (AR(1)). For notation simplicity, denote the two-step spline estimator $\widehat{\theta}^{\mathcal{S}}_{n,l}(z_l,\widehat{\boldsymbol{\beta}}_n,\widehat{\boldsymbol{\theta}}_{n,-l})$ defined in (7) as $\widehat{\theta}^{\mathcal{SS}}_{n,l}(z_l)=\mathbf{B}^{\mathcal{S}}_l(z_l)^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}^{\mathcal{SS}}_{n,l}$, and the oracle estimator $\widehat{\theta}^{\mathcal{S}}_{n,l}(z_l,\boldsymbol{\beta},\boldsymbol{\theta}_{-l})$ in (5) as $\widehat{\theta}^{\mathrm{OR}}_{n,l}(z_l)=\mathbf{B}^{\mathcal{S}}_l(z_l)^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}^{\mathrm{OR}}_{n,l}$. In the first step, the pilot estimators are obtained by an undersmoothed spline procedure to reduce bias. By the order requirements of the number of interior knots, we select a relatively large N by letting $N=[2n_{\mathrm{T}}^{1/(2p)}]$, where $[a]$ denotes the nearest integer to a. In the second step, $N^{\mathcal{S}}$ is selected from the interval $I_{N^{\mathcal{S}}}=[[a_n],[5a_n]]$, $a_n=(n_{\mathrm{T}}\log n_{\mathrm{T}})^{1/(2p+1)}$, minimizing the BIC criterion

$$\mathrm{BIC}(N^{\mathcal{S}})=\log\{2Q^*_{n,l}(\widehat{\boldsymbol{\gamma}}^{\mathcal{S}}_{n,l})/n\}+J^{\mathcal{S}}_n\log(n)/n, \tag{14}$$

where $Q^*_{n,l}(\widehat{\boldsymbol{\gamma}}^{\mathcal{S}}_{n,l})=2^{-1}\sum_{i=1}^n(\underline{\mathbf{Y}}_i-\widehat{\underline{\boldsymbol{\mu}}}_i)^{\mathrm{T}}\mathbf{V}_i^{-1}(\widehat{\boldsymbol{\beta}}_n,\widehat{\boldsymbol{\theta}}_{n,-l},\widehat{\boldsymbol{\gamma}}^{\mathcal{S}}_{n,l})(\underline{\mathbf{Y}}_i-\widehat{\underline{\boldsymbol{\mu}}}_i)$ with $\widehat{\underline{\boldsymbol{\mu}}}_i=\mu(\underline{\mathbf{X}}_i\widehat{\boldsymbol{\beta}}_n+\sum_{l'=1,l'\neq l}^{d_2}\widehat{\theta}_{n,l'}(\mathbf{Z}_{il'})+(\mathbf{B}^{\mathcal{S}}_{i\cdot l})^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}^{\mathcal{S}}_{n,l})$. The optimal number of interior knots $N^{\mathcal{S}}$ is chosen as $\widehat{N}^{\mathcal{S}}=\arg\min_{N^{\mathcal{S}}\in I_{N^{\mathcal{S}}}}\mathrm{BIC}(N^{\mathcal{S}})$. We use cubic B-splines ($q=3$) to estimate the additive nonparametric functions. We generate $\mathrm{nsim}=500$ replications for each simulation study.

Given $1\le l\le d_2$, to compare the performance of the two-step estimator $\widehat{\theta}^{\mathcal{SS}}_{n,l}(z_l)$ with the pilot spline estimator $\widehat{\theta}_{n,l}(z_l)$ and the oracle estimator $\widehat{\theta}^{\mathrm{OR}}_{n,l}(z_l)$, we define the mean integrated squared error (MISE) for $\widehat{\theta}^{\mathcal{SS}}_{n,l}(z_l)$ as $\mathrm{MISE}(\widehat{\theta}^{\mathcal{SS}}_{n,l})=\frac{1}{\mathrm{nsim}}\sum_{\alpha=1}^{\mathrm{nsim}}\mathrm{ISE}(\widehat{\theta}^{\mathcal{SS}}_{n,l,\alpha})$, where $\mathrm{ISE}(\widehat{\theta}^{\mathcal{SS}}_{n,l,\alpha})=n_{\mathrm{T}}^{-1}\sum_{i=1}^n\sum_{j=1}^{m_i}(\widehat{\theta}^{\mathcal{SS}}_{n,l,\alpha}(Z_{ijl,\alpha})-\theta_l(Z_{ijl,\alpha}))^2$, and $\widehat{\theta}^{\mathcal{SS}}_{n,l,\alpha}$ is the estimator of θ_l and $Z_{ijl,\alpha}$ is the observation of Z_{ijl} in the αth sample. The MISEs for $\widehat{\theta}_{n,l}(z_l)$ and $\widehat{\theta}^{\mathrm{OR}}_{n,l}(z_l)$ denoted as $\mathrm{MISE}(\widehat{\theta}_{n,l})$ and $\mathrm{MISE}(\widehat{\theta}^{\mathrm{OR}}_{n,l})$ are defined in the same way. The empirical relative efficiency for the two-step estimator in the αth sample is defined as $\mathrm{eff}_{l,\alpha}=\{\mathrm{ISE}(\widehat{\theta}^{\mathcal{SS}}_{n,l,\alpha})/\mathrm{ISE}(\widehat{\theta}^{\mathrm{OR}}_{n,l,\alpha})\}^{1/2}$. To construct confidence intervals for coefficient parameters $(\beta_{0,0},\dots,\beta_{0,(d_1-1)})$ by using the first result in Theorem 2 and to construct pointwise confidence intervals for the lth nonparametric function $\theta_{l0}(z_l)$ given in (13), the true correlation matrix $\overline{\mathbf{R}}$ is consistently estimated by

$$\begin{aligned}\widehat{\mathbf{R}}=n^{-1}\sum_{i=1}^n\mathbf{A}_i^{-1/2}(\widehat{\boldsymbol{\beta}}_n,\widehat{\boldsymbol{\gamma}}_n)[\underline{\mathbf{Y}}_i-\mu\{\widetilde{\underline{\boldsymbol{\eta}}}_i(\widehat{\boldsymbol{\beta}}_n,\widehat{\boldsymbol{\gamma}}_n)\}]\\ \times[\underline{\mathbf{Y}}_i-\mu\{\widetilde{\underline{\boldsymbol{\eta}}}_i(\widehat{\boldsymbol{\beta}}_n,\widehat{\boldsymbol{\gamma}}_n)\}]^{\mathrm{T}}\mathbf{A}_i^{-1/2}(\widehat{\boldsymbol{\beta}}_n,\widehat{\boldsymbol{\gamma}}_n).\end{aligned}$$

And the covariance matrix Σ_i is estimated by $\widehat{\Sigma}_i=\mathbf{A}_i^{1/2}\widehat{\mathbf{R}}\mathbf{A}_i^{1/2}$. Let $\boldsymbol{\beta}_0=(\beta_{0,k})_{k=0}^{(d_1-1)}$ and $\widehat{\boldsymbol{\beta}}_n=(\widehat{\beta}_{n,k})_{k=0}^{(d_1-1)}$. For evaluating estimation accuracy of each coefficient parameter, we report the root mean squared error (RMSE) defined as $\{\sum_{\alpha=1}^{\mathrm{nsim}}(\widehat{\beta}^{\alpha}_{n,k}-\beta_{0,k})^2/\mathrm{nsim}\}^{1/2}$, for $0\le k\le d_1-1$, where $\widehat{\beta}^{\alpha}_{n,k}$ is the estimate of $\beta_{0,k}$ obtained from the αth sample.

EXAMPLE 1 (Continuous response). The correlated normal responses are generated from the model $Y_{ij} = \mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \theta_1(Z_{ij1}) + \theta_2(Z_{ij2}) + \theta_3(Z_{ij3}) + \varepsilon_{ij}$, where $\boldsymbol{\beta} = (1, -1, 0.5)$, $\mathbf{X}_{ij} = (X_{ij,1}, X_{ij,2}, X_{ij,3})^{\mathrm{T}}$, $\theta_l(Z_l) = \sin(2\pi Z_l)$, $1 \le l \le 3$. For the covariates, let $Z_{ijl} = \Phi(Z_{ijl}^*)$, $1 \le l \le 3$, with $\mathbf{Z}_{ij}^* = (Z_{ij1}^*, Z_{ij2}^*, Z_{ij3}^*)^{\mathrm{T}}$ generated from the multivariate normal distribution with mean 0 and an AR(1) covariance with marginal variance 1 and autocorrelation coefficient 0.5, $X_{ij,1} = \pm 1/2$ with probability $1/2$, and $(X_{ij,2}, X_{ij,3})^{\mathrm{T}} \sim \mathrm{N}[(0,0)^{\mathrm{T}}, \mathrm{diag}(a(Z_{ij1}), a(Z_{ij2}))]$ with $a(z) = \frac{5-0.5\sin(2\pi z)}{5+0.5\sin(2\pi z)}$. The error term $\underline{\boldsymbol{\varepsilon}}_i = (\varepsilon_{i1}, \dots, \varepsilon_{im_i})^{\mathrm{T}}$ is generated from the multivariate normal distribution with mean 0, marginal variance 1 and an exchangeable correlation matrix with parameter $\rho = 0.5$. We let $n = 250$ and cluster size $m_i = m = 20, 50, 100$, respectively. For computational simplicity, we choose the same cluster size for each subject. The computational algorithm can be easily extended to the case with varying cluster sizes. Table 1 lists the empirical coverage rates of the 95% confidence intervals of the estimators $(\widehat{\beta}_{n,k})_{k=1}^3$ for coefficients $(\beta_{0,k})_{k=1}^3$, the RMSE and the absolute value of the empirical bias denoted as Bias for IND, EX and AR(1) and $m = 20, 50, 100$.

The empirical coverage rates are close to the nominal coverage probabilities 95% for all cases. The results are confirmative to Theorem 2. EX has the smallest RMSE, since it is the true correlation structure, which leads to the most efficient estimators (Remark 2). The RMSEs decrease as cluster size increases for all three working correlation structures. The last three columns show that the empirical biases are close to zero for all cases.

Table 2 shows the MISE($\times 10^{-3}$) for the two-step spline estimator $\widehat{\theta}_{n,l}^{SS}(\cdot)$, the pilot estimator $\widehat{\theta}_{n,l}(\cdot)$ and the oracle estimator $\widehat{\theta}_{n,l}^{\mathrm{OR}}(\cdot)$, $l = 1, 2, 3$, for IND, EX and

TABLE 1

The empirical coverage rates of the 95% *confidence intervals for* $(\beta_{0,k})_{k=1}^3$, *the RMSE and Bias for the IND, EX and AR(1) working correlation structures with* $m = 20, 50, 100$

		Coverage frequency			**RMSE**			**Bias**		
m		$\boldsymbol{\beta_{0,1}}$	$\boldsymbol{\beta_{0,2}}$	$\boldsymbol{\beta_{0,3}}$	$\boldsymbol{\beta_{0,1}}$	$\boldsymbol{\beta_{0,2}}$	$\boldsymbol{\beta_{0,3}}$	$\boldsymbol{\beta_{0,1}}$	$\boldsymbol{\beta_{0,2}}$	$\boldsymbol{\beta_{0,3}}$
20	IND	0.948	0.956	0.950	0.0279	0.0137	0.0137	0.0050	0.0002	0.0008
	EX	0.954	0.950	0.948	0.0196	0.0098	0.0108	0.0018	0.0000	0.0006
	AR(1)	0.936	0.954	0.956	0.0260	0.0123	0.0121	0.0026	0.0003	0.0011
50	IND	0.948	0.952	0.948	0.0177	0.0092	0.0091	0.0006	0.0001	0.0009
	EX	0.946	0.950	0.948	0.0126	0.0063	0.0066	0.0002	0.0001	0.0002
	AR(1)	0.944	0.956	0.948	0.0157	0.0079	0.0081	0.0003	0.0002	0.0003
100	IND	0.948	0.956	0.958	0.0126	0.0063	0.0064	0.0001	0.0003	0.0002
	EX	0.950	0.954	0.948	0.0084	0.0044	0.0045	0.0001	0.0002	0.0001
	AR(1)	0.946	0.954	0.956	0.0111	0.0056	0.0055	0.0001	0.0004	0.0001

TABLE 2
The MISE($\times 10^{-3}$) *for* $\widehat{\theta}^{SS}_{n,l}(\cdot)$, $\widehat{\theta}_{n,l}(\cdot)$ *and* $\widehat{\theta}^{\mathrm{OR}}_{n,l}(\cdot)$, $l = 1, 2, 3$, *for the IND, EX and AR*(1) *working correlation structures with* $m = 20, 50, 100$

m		$\widehat{\theta}^{SS}_{n,1}$	$\widehat{\theta}_{n,1}$	$\widehat{\theta}^{\mathrm{OR}}_{n,1}$	$\widehat{\theta}^{SS}_{n,2}$	$\widehat{\theta}_{n,2}$	$\widehat{\theta}^{\mathrm{OR}}_{n,2}$	$\widehat{\theta}^{SS}_{n,3}$	$\widehat{\theta}_{n,3}$	$\widehat{\theta}^{\mathrm{OR}}_{n,3}$
20	IND	1.678	2.231	1.588	1.659	2.278	1.517	1.516	2.118	1.448
	EX	0.883	1.228	0.836	0.943	1.232	0.848	0.849	1.167	0.811
	AR(1)	1.249	1.710	1.186	1.324	1.790	1.205	1.252	1.713	1.182
50	IND	0.633	0.862	0.601	0.677	0.927	0.608	0.631	0.881	0.601
	EX	0.342	0.463	0.328	0.348	0.475	0.321	0.353	0.465	0.335
	AR(1)	0.473	0.664	0.459	0.513	0.690	0.478	0.486	0.679	0.464
100	IND	0.319	0.440	0.306	0.346	0.461	0.317	0.315	0.436	0.299
	EX	0.173	0.234	0.166	0.176	0.237	0.162	0.172	0.227	0.164
	AR(1)	0.247	0.333	0.235	0.252	0.348	0.230	0.244	0.338	0.232

AR(1) structures and cluster size $m = 20, 50, 100$. $\widehat{\theta}^{SS}_{n,l}(\cdot)$ and $\widehat{\theta}^{\mathrm{OR}}_{n,l}(\cdot)$ have similar MISE values, while $\widehat{\theta}_{n,l}(\cdot)$ has the largest MISE value. The EX structure has the smallest MISEs, and the MISEs decrease as the cluster size increases.

We plotted the kernel density estimates in Figure 1 of 500 empirical efficiencies $\mathrm{eff}_{l,\alpha}$ for the estimators of the first function $\theta_1(\cdot)$ for IND (dashed lines), EX (thick lines) and AR(1) (thin lines) structures with $m = 20, 50$ and $n = 250$. The vertical line at efficiency $= 1$ is the standard line for the comparison of the two-step estimator (7) and the oracle estimator (5). The centers of density distributions are close to 1 for all working correlation structures, and EX has the narrowest distribution.

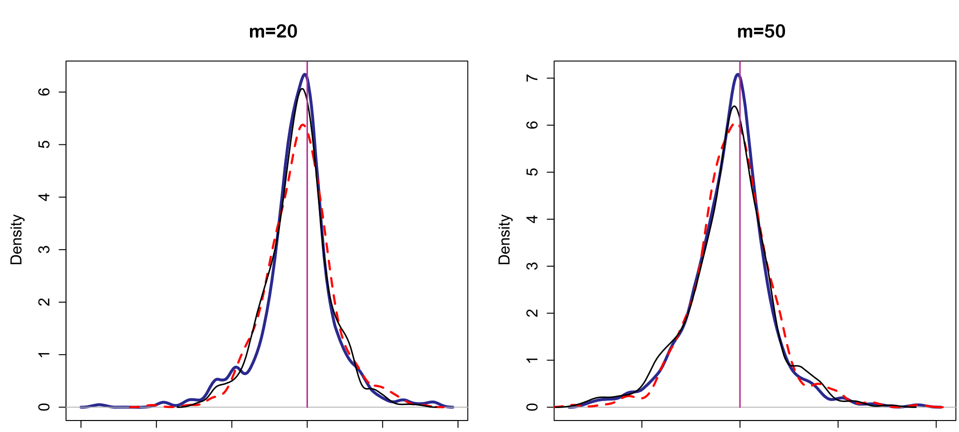

FIG. 1. *Kernel density plots of the* 500 *empirical efficiencies of the two-step estimator to the oracle estimator of the first function* $\theta_1(\cdot)$ *for the IND* (*dashed lines*), *EX* (*thick lines*) *and AR*(1) (*thin lines*) *working correlation structures with* $m = 20, 50$.

EXAMPLE 2 (Binary response). The correlated binary responses $\{Y_{ij}\}$ are generated from a marginal logit model

$$\operatorname{logit} P(Y_{ij} = 1|\mathbf{X}_{ij}, \mathbf{Z}_{ij}) = \mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \theta_1(Z_{ij1}) + \theta_2(Z_{ij2}),$$

where $\boldsymbol{\beta} = (0.5, -0.3, 0.3)$, $\mathbf{X}_{ij} = (1, X_{ij,1}, X_{ij,2})^{\mathrm{T}}$, $\theta_1(Z_1) = 0.5 \times \sin(2\pi Z_1)$, and $\theta_2(Z_2) = -0.5 \times \{Z_2 - 0.5 + \sin(2\pi Z_2)\}$. For the covariates, we generate X_{ijk} and Z_{ijl} independently from standard normal and uniform distributions, respectively, such that $X_{ijk} \sim \mathrm{N}(0, 1)$ and $Z_{ijl} \sim \mathrm{Uniform}[0, 1]$. We use the R package "mvtBinaryEP" to generate the correlated binary responses with exchangeable correlation structure with a correlation parameter of 0.1 within each cluster. We let the number of clusters be $n = 100, 200, 500$, respectively, and let the cluster size be equal and increase with n, such that $m_{(n)} = m_i = \lfloor 2n^{1/2} \rfloor$, for $1 \le i \le n$, where $\lfloor a \rfloor$ denotes the largest integer no greater than a. So $m = 20, 28, 44$ for $n = 100, 200, 500$, respectively. Table 3 shows the empirical coverage rates of the 95% confidence intervals of the estimators $(\widehat{\beta}_{n,k})_{k=0}^2$ for the coefficients $(\beta_{0,k})_{k=0}^2$ and the RMSEs for IND, EX and AR(1) and $n = 100, 200, 500$. Table 4 shows that the empirical coverage rates are close to the nominal coverage probabilities 95% for all cases. EX has the smallest RMSE values, and the RMSEs decrease as n increases.

Table 4 shows the MISE for the two-step spline estimator $\widehat{\theta}_{n,l}^{SS}(\cdot)$, the pilot estimator $\widehat{\theta}_{n,l}(\cdot)$ and the oracle estimator $\widehat{\theta}_{n,l}^{\mathrm{OR}}(\cdot)$, $l = 1, 2$, for the IND, EX and AR(1) structures and $n = 100, 200, 500$. The MISE values for $\widehat{\theta}_{n,l}^{SS}(\cdot)$ and $\widehat{\theta}_{n,l}^{\mathrm{OR}}(\cdot)$ are close and $\widehat{\theta}_{n,l}(\cdot)$ has the largest MISE values. EX has the smallest MISEs among the three working correlation structures, and the MISEs decrease as n increases.

TABLE 3
The empirical coverage rates of the 95% confidence intervals for $(\beta_{0,k})_{k=0}^2$ and the estimated MSE for the IND, EX and AR(1) working correlation structures with $n = 100, 200, 500$

		Coverage frequency			RMSE		
		$\boldsymbol{\beta_{0,0}}$	$\boldsymbol{\beta_{0,1}}$	$\boldsymbol{\beta_{0,2}}$	$\boldsymbol{\beta_{0,0}}$	$\boldsymbol{\beta_{0,1}}$	$\boldsymbol{\beta_{0,2}}$
$n = 100, m = 20$	IND	0.960	0.946	0.940	0.0821	0.0549	0.0506
	EX	0.940	0.946	0.946	0.0763	0.0469	0.0454
	AR(1)	0.966	0.930	0.940	0.0773	0.0540	0.0488
$n = 200, m = 28$	IND	0.944	0.946	0.940	0.0559	0.0299	0.0328
	EX	0.948	0.952	0.942	0.0554	0.0289	0.0310
	AR(1)	0.940	0.950	0.940	0.0556	0.0291	0.0325
$n = 500, m = 44$	IND	0.952	0.946	0.942	0.0340	0.0157	0.0154
	EX	0.948	0.952	0.946	0.0336	0.0136	0.0142
	AR(1)	0.952	0.952	0.942	0.0340	0.0153	0.0153

TABLE 4
The MISE for $\widehat{\theta}^{SS}_{n,l}(\cdot)$, $\widehat{\theta}_{n,l}(\cdot)$ *and* $\widehat{\theta}^{\mathrm{OR}}_{n,l}(\cdot)$, $l=1,2$, *for the IND, EX and AR*(1) *working correlation structures with* $n = 100, 200, 500$

n		$\widehat{\theta}^{SS}_{n,1}$	$\widehat{\theta}_{n,1}$	$\widehat{\theta}^{\mathrm{OR}}_{n,1}$	$\widehat{\theta}^{SS}_{n,2}$	$\widehat{\theta}_{n,2}$	$\widehat{\theta}^{\mathrm{OR}}_{n,2}$
100	IND	0.0172	0.0243	0.0174	0.0158	0.0222	0.0159
	EX	0.0148	0.0223	0.0148	0.0139	0.0204	0.0137
	AR(1)	0.0178	0.0265	0.0176	0.0161	0.0234	0.0163
200	IND	0.0059	0.0086	0.0059	0.0056	0.0082	0.0056
	EX	0.0048	0.0069	0.0048	0.0054	0.0075	0.0053
	AR(1)	0.0058	0.0085	0.0058	0.0056	0.0081	0.0056
500	IND	0.0015	0.0022	0.0015	0.0015	0.0021	0.0015
	EX	0.0013	0.0019	0.0013	0.0014	0.0019	0.0013
	AR(1)	0.0015	0.0022	0.0015	0.0015	0.0020	0.0014

For visualization of the actual function estimates, in Figure 2 we plotted the oracle estimator given in (5) (dashed curve), the two-step estimator given in (7) (thick curve) and the 95% pointwise confidence intervals constructed in (13) (upper and lower curves) of $\theta_1(\cdot)$ (thin curve) for $n = 200$ based on one simulated sample. The proposed two-step estimator seems satisfactory.

5. Application. In this section we apply the proposed estimation procedure to analyze unemployment-economic growth and employment relationship at the U.S. state level for the 1970–1986 period. Reference [2] has first studied the effect of economic growth on unemployment rate by establishing a paramet-

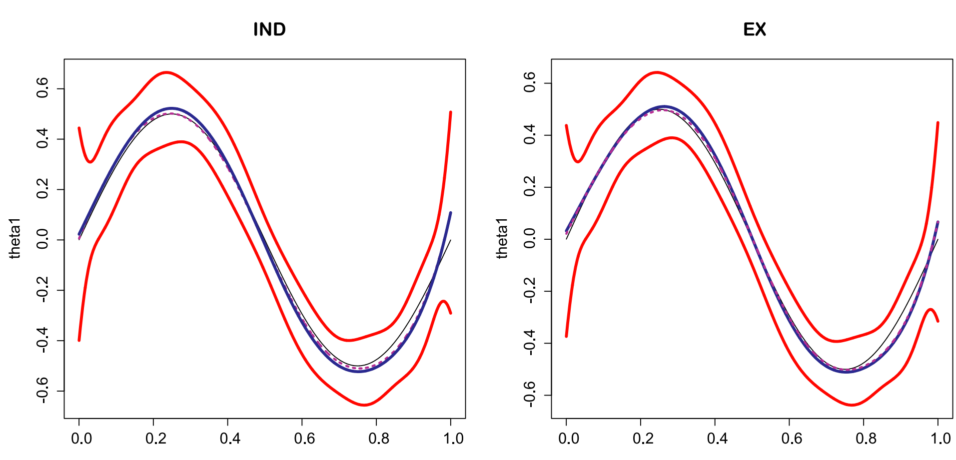

FIG. 2. *Plots of oracle estimator* (*dashed curve*), *the two-step estimator* (*thick curve*) *and the* 95% *pointwise confidence intervals* (*upper and lower curves*) *of* $\theta_1(\cdot)$ (*thin curve*) *for* $n = 200$.

ric unemployment-growth model. They concluded that relatively high economic growth is more likely to show reduced unemployment rates when compared to states in which the economy is growing more slowly by obtaining a negative coefficient for growth. Reference [27] demonstrated a strong negative correlation between the change of unemployment rate and employment. We restudy their relationship by considering possible nonlinear relations of the unemployment rate with economic growth and time. The economic growth rate is calculated from the logarithm difference of the gross state product (GSP). The data for the unemployment rate, gross state product and employment are available for the U.S. 48 contiguous states over the period 1970–1986. Details on this data set can be found in [24]. The number of time periods for each state in estimation is $m = 16$, since the year 1970 is taken as the initial observation. We consider the following GAPLM:

$$U_{ij} = \beta_0 + \beta_1 E_{ij} + \theta_1(T_{ij}) + \theta_2(G_{ij}) + \varepsilon_{ij}, \qquad j = 2, \dots, 17, i = 1, \dots, 48,$$

where U_{ij} is the change in the unemployment rate for the jth year in the ith state, E_{ij} is the empirically centered value of the relative change in employment, G_{ij} is the GSP growth, and T_{ij} is time. $\theta_1(\cdot)$ and $\theta_2(\cdot)$ are nonparametric functions of time and GSP growth, respectively.

To test whether $\theta_l(\cdot)$, $l = 1, 2$, has a specific parametric form, we construct simultaneous confidence bands according to Theorem 2 of [28]. For any $\alpha \in (0, 1)$, an asymptotic $100(1 - \alpha)\%$ conservative confidence band for $\theta_{l0}(z_l)$ over the domain of z_l is given as

$$\widehat{\theta}^{S}_{n,l}(z_l, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}_{n,-l}) \pm \{2\log(N^s + 1) - 2\log\alpha\}^{1/2}\big(\mathbf{B}^{S}_{l}(z_l)^{\mathrm{T}}\,\Xi^{*}_{n,l}\mathbf{B}^{S}_{l}(z_l)\big)^{1/2}$$

with $\widehat{\theta}^{S}_{n,l}$ obtained by linear splines with degree $q = 1$. We use linear splines in both steps of estimation.

We use three working correlation structures to analyze this data set, including the working independence $\mathbf{R}_i(\alpha) = \mathbf{I}_m$, where $\mathbf{I}_m$ is an $m \times m$ identity matrix, the exchangeable $\mathbf{R}_i(\alpha) = \alpha \times 1_m 1_m^{\mathrm{T}} + (1 - \alpha)\mathbf{I}_m$, where 1_m is the m-dimensional vector with 1's, and the AR(1) $\mathbf{R}_i(\alpha) = (R_{ijj'})^{m}_{j,j'=1}$ with $R_{ijj'} = \alpha^{|j-j'|}$. The parameter α is estimated by the R package geepack from the first spline estimation step. We obtain the estimated values for α which are $\widehat{\alpha} = 0.088$ for the EX structure and $\widehat{\alpha} = -0.199$ for the AR(1) structure, respectively. Table 5 shows the estimated values $\widehat{\beta}_0$ and $\widehat{\beta}_1$ of β_0 and β_1 and the corresponding standard errors $\mathrm{SE}(\widehat{\beta}_0)$ and $\mathrm{SE}(\widehat{\beta}_1)$ for the three working correlation structures. The estimation results are very similar for the three structures. The negative values of $\widehat{\beta}_1$ imply a negative relationship between U_{ij} and E_{ij}, confirmative to the result in [27]. Both of the estimators are significant with p-values close to 0 for the three different working correlation structures. The correlation coefficient $r = 0.785$, 0.822 and 0.762 for the IND, EX and AR(1) structures, respectively.

TABLE 5
The estimated values $\widehat{\beta}_0$ and $\widehat{\beta}_1$ of β_0 and β_1 and the standard errors SE($\widehat{\beta}_0$) and SE($\widehat{\beta}_1$) for the IND, EX and AR(1) working correlation structures

	$\widehat{\beta}_0$	SE($\widehat{\beta}_0$)	$\widehat{\beta}_1$	SE($\widehat{\beta}_1$)
IND	0.127	0.0417	−0.219	0.0230
EX	0.127	0.0494	−0.249	0.0220
AR(1)	0.127	0.0484	−0.250	0.0216

Figure 3 displays the two-step spline estimators $\widehat{\theta}_{n,1}^{SS}(\cdot)$ (dashed lines) and $\widehat{\theta}_{n,2}^{SS}(\cdot)$ (dashed lines) of $\theta_1(\cdot)$ and $\theta_2(\cdot)$ and the corresponding 95% pointwise confidence intervals (thin lines) and simultaneous confidence bands (thick lines) for the three working structures. Figure 3 shows that the change patterns of U_{ij} with T_{ij} and G_{ij} are very similar for the three working structures. In the upper panel of Figure 3, we can observe a declining trend for $\widehat{\theta}_{n,1}^{SS}(\cdot)$ in general. The values of $\widehat{\theta}_{n,1}^{SS}(\cdot)$ were all positive before the year 1976, which means that the unemployment rate was increasing with time during that period. The increasing unemployment rate was caused by a severe economic recession that happened in the years 1973–1975. A local peak of $\widehat{\theta}_{n,1}^{SS}(\cdot)$ is observed around 1980, when another recession happened.

In order to test the linearity of the nonparametric function θ_2, we plotted straight solid lines in the lower panel of Figure 3, which are the regression lines obtained by solving the GEE in (6) by assuming that $\theta_1(\cdot)$ is a linear function of GSP growth. All the three plots in the lower panel of Figure 3 show that the confidence bands with 95% confidence level do not totally cover the straight regression lines, that is, the linearity of the component function for GSP growth is rejected at the significance level 0.05. The lower panel of Figure 3 indicates a general negative relation between the GSP growth and the change in unemployment rate.

6. Discussion. In this paper we propose a two-step spline estimating equations procedure for generalized additive partially linear models with large cluster sizes. We develop asymptotic distributions and consistency properties for the two-step estimators of the additive functions and the one-step estimator of the parametric vector. We establish the oracle properties of the two-step estimators. Because the two-step estimator is a mixture of two different spline bases, and an infinite number of observations within clusters are correlated in complex ways, we encountered challenging tasks when developing the theories. We demonstrate our proposed method by two simulated examples and one real data example. Our proposed method can be extended to generalized additive models and generalized additive coefficient models, and it provides a useful tool for studying clustered

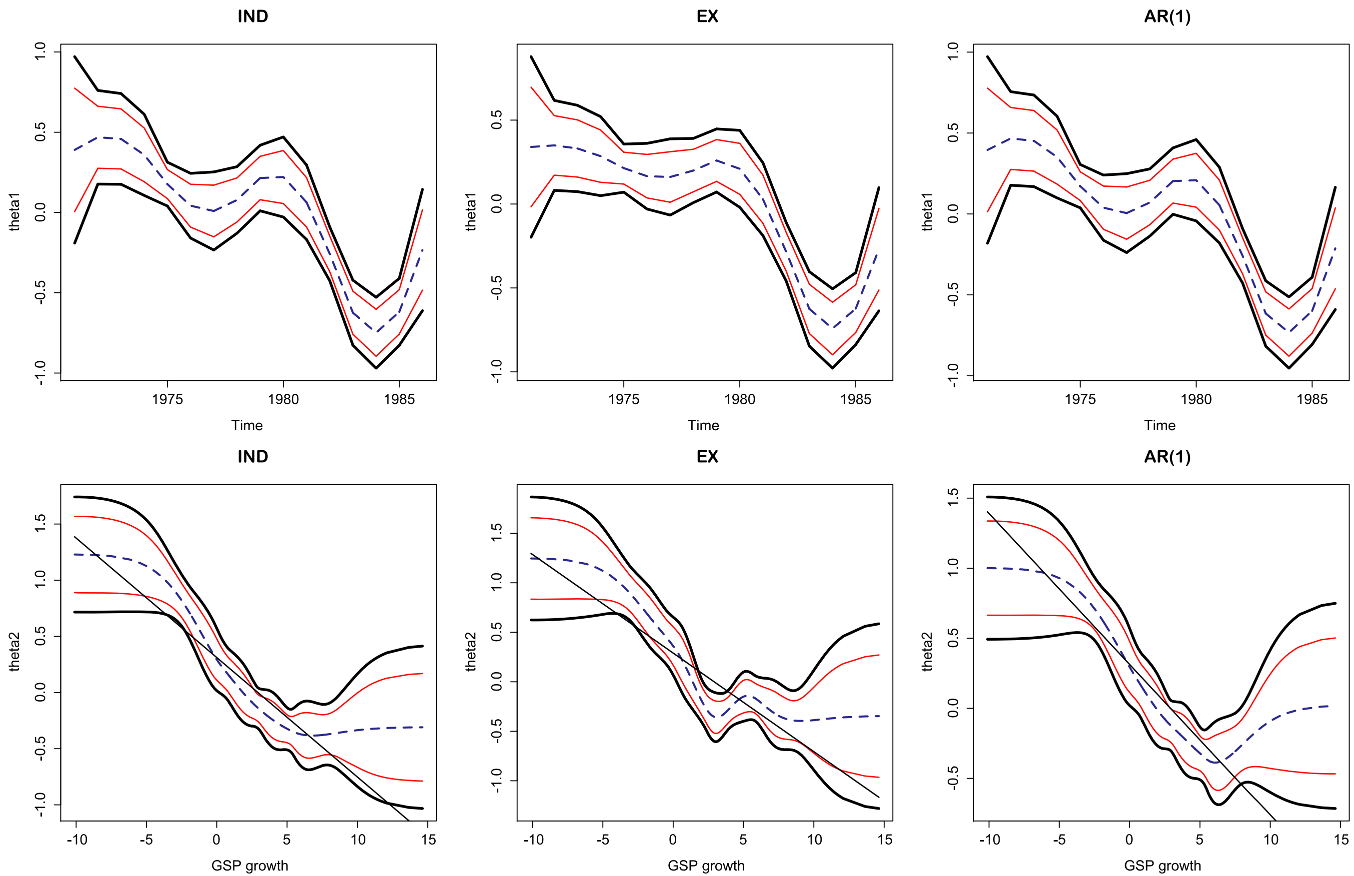

FIG. 3. *Plots of the two-step spline estimated functions* (*dashed line*), *the* 95% *pointwise confidence intervals* (*thin lines*) *and the* 95% *confidence bands* (*thick lines*) *for* $\theta_1(\cdot)$ (*upper panel*) *and* $\theta_2(\cdot)$ (*lower panel*), *and the GEE estimator of* $\theta_2(\cdot)$ *by assuming linearity* (*straight solid line*).

data. The theoretical development in this paper helps us further investigate semiparametric models with clustered data. In the real data example, we constructed confidence bands to test the linearity of the nonparametric function. To establish confidence bands with rigorous theoretical proofs will be our future work.

In this paper we focus on the two-step spline estimation procedure, which is computationally expedient and theoretically reliable. As mentioned in Section 2, that kernel smoothing method can be applied to the second step. Let $K_h(\cdot)$ be a kernel weight function, where $K_h(z) = h^{-1}K(z/h)$ with bandwidth h. Let $G_1(z_l) = (1, z_l)^{\mathrm{T}}$. If we use local linear kernel estimation, then by assuming that $\boldsymbol{\beta}$ and $\boldsymbol{\theta}_{-l}$ are known, $\theta_l(\cdot)$ is estimated by the oracle estimator $\widehat{\theta}_l^{\mathrm{OR}}(Z_l) = G_1(Z_l - z_l)^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_l^{\mathrm{OR}}$ at any given point z_l, where $\widehat{\boldsymbol{\gamma}}_l^{\mathrm{OR}} = (\widehat{\gamma}_{l0}^{\mathrm{OR}}, \widehat{\gamma}_{l1}^{\mathrm{OR}})^{\mathrm{T}}$ with $\widehat{\boldsymbol{\gamma}}_l^{\mathrm{OR}}$ solving the kernel estimating equations

$$\sum_{i=1}^{n} G_{i1}(z_l)^{\mathrm{T}} \Delta_i(\boldsymbol{\beta}, \boldsymbol{\theta}_{-l}, \widehat{\boldsymbol{\gamma}}_l^{\mathrm{OR}}) \mathbf{V}_i^{-1}(\boldsymbol{\beta}, \boldsymbol{\theta}_{-l}, \widehat{\boldsymbol{\gamma}}_l^{\mathrm{OR}}) \mathbf{K}_{ih}(z_l) \times \left\{ \underline{\mathbf{Y}}_i - \mu\left(\underline{\mathbf{X}}_i \boldsymbol{\beta} + \sum_{l'=1, l' \neq l}^{d_2} \theta_{l'}(\mathbf{Z}_{il'}) + G_{i1}(z_l) \widehat{\boldsymbol{\gamma}}_l^{\mathrm{OR}} \right) \right\} = 0,$$

where $\mathbf{K}_{ih}(z_l) = \mathrm{diag}\{K_h(Z_{ijl} - z_l)\}$ and $G_{i1}(z_l) = \{G_1(Z_{i1l} - z_l), \ldots, G_1(Z_{im_il} - z_l)\}^{\mathrm{T}}$. Then $\theta_l(z_l)$ is estimated by $\widehat{\theta}_l^{\mathrm{OR}}(z_l) = \widehat{\gamma}_{l0}^{\mathrm{OR}}$. The two-step spline backfitted kernel (SBK) estimator $\widehat{\theta}_l^{\mathrm{SBK}}(z_l)$ is obtained by replacing $\boldsymbol{\beta}$ and $\boldsymbol{\theta}_{-l}$ with the pilot estimators $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{\boldsymbol{\theta}}_{n,-l}$ from step I. The asymptotic normality of the oracle estimator $\widehat{\theta}_l^{\mathrm{OR}}(z_l)$ which is a pure local linear kernel estimator of $\theta_l(z_l)$ by GEE can be obtained following the same idea in the proofs for Theorem 3 and the results in [16] for kernel estimators using GEE. The uniform oracle efficiency of the SBK estimator $\widehat{\theta}_l^{\mathrm{SBK}}(z_l)$ is achievable by following the same procedure as the proofs for Theorem 4 and by studying the properties of spline-kernel combination. See [20, 29] and [22] for the oracle properties of the SBK estimators in additive models, additive coefficient models and additive partially linear models with weekly-dependent data and a continuous response variable. The asymptotic distributions and the oracle properties of the SBK estimators for GAPLMs with large cluster sizes still need us to explore as future work.

APPENDIX

We denote by the same letters c, C, any positive constants without distinction. For any $s \times s'$ matrix $\mathbf{M}$, let $\|\mathbf{M}\|_\infty = \max_{1 \le i \le s} \sum_{j=1}^{s'} |M_{ij}|$. For any vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_s)^{\mathrm{T}}$, denote $\|\boldsymbol{\alpha}\|_\infty = \max_{1 \le i \le s} |\alpha_i|$ as the maximum norm. Let $\mathbf{I}_s$ be the $s \times s$ identity matrix. Let $\widehat{\Pi}_n$, Π_n denote, respectively, the projection onto G_n^0 relative to the empirical and the theoretical inner products. For any function ϕ, define the empirical norm as $\|\phi\|_{n_{\mathrm{T}}}^2 = n_{\mathrm{T}}^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \phi(X_{ij}, Z_{ij})^2$. For positive numbers a_n and b_n, let $a_n \asymp b_n$ denote that $\lim_{n\to\infty} a_n/b_n = c$, where c is some nonzero constant.

A.1. Proof of Theorem 1. It can be proved following the similar reasoning as in [21] that under condition (A1) with $n_{\mathrm{T}} \to \infty$, $J_n \to \infty$, and $J_n n^{-1} = o(1)$, there exist constants $0 < c' < C' < \infty$, such that with probability 1, for n_{T} sufficiently large,

$$c' n_{\mathrm{T}} \le \lambda_{\min}\left(\sum_{i=1}^{n} \underline{\mathbf{B}}_i^{\mathrm{T}} \underline{\mathbf{B}}_i\right) \le \lambda_{\max}\left(\sum_{i=1}^{n} \underline{\mathbf{B}}_i^{\mathrm{T}} \underline{\mathbf{B}}_i\right) \le C' n_{\mathrm{T}}$$

and $\|\sum_{i=1}^{n} \underline{\mathbf{X}}_i^{\mathrm{T}} \underline{\mathbf{B}}_i\|_\infty = O_{\mathrm{a.s.}}\{(n_{\mathrm{T}} \log n_{\mathrm{T}})^{1/2}\}$. By these results together with condition (C4), one has with probability 1,

$$c'' n_{\mathrm{T}} \le \lambda_{\min}\left(\sum_{i=1}^{n} \underline{\mathbf{D}}_i^{\mathrm{T}} \underline{\mathbf{D}}_i\right) \le \lambda_{\max}\left(\sum_{i=1}^{n} \underline{\mathbf{D}}_i^{\mathrm{T}} \underline{\mathbf{D}}_i\right) \le C'' n_{\mathrm{T}} \tag{15}$$

for some constants $0 < c'' < C'' < \infty$. Then by condition (A2),

$$(\tau_n^{\max})^{-1} \lambda_{\min}\{\Psi_n(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)\} \ge c c'' (\tau_n^{\max})^{-1} \lambda_n^{\min} n_{\mathrm{T}} \to \infty.$$

Results in Theorem 1 can be proved similarly as Theorems 1 and 2 in [33] with $r = \sqrt{2(d_1 + d_2 J_n)/c_0 \varepsilon}$ for any given $\varepsilon > 0$.

A.2. Proof of Theorem 2. By Taylor's expansion, one has

$$\mathbf{g}_n(\widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\gamma}}_n) - \mathbf{g}_n(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0) = -\mathcal{D}_n(\boldsymbol{\beta}_n^*, \boldsymbol{\gamma}_n^*) \begin{pmatrix} \widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0 \\ \widehat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_0 \end{pmatrix}, \tag{16}$$

where $\boldsymbol{\beta}_n^* = t_1 \widehat{\boldsymbol{\beta}}_n + (1 - t_1)\boldsymbol{\beta}_0$, and $\boldsymbol{\gamma}_n^* = t_2 \widehat{\boldsymbol{\gamma}}_n + (1 - t_2)\boldsymbol{\gamma}_0$ for some $t_1, t_2 \in (0, 1)$. Let $\Pi_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \Delta_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) \mathbf{V}_i^{-1}(\boldsymbol{\beta}, \boldsymbol{\gamma})$, for $1 \le i \le n$. Then

$$\mathcal{D}_n(\boldsymbol{\beta}_n^*, \boldsymbol{\gamma}_n^*) = \Psi_n(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0) + \Pi_{n,1}(\boldsymbol{\beta}_n^*, \boldsymbol{\gamma}_n^*) + \Pi_{n,2}(\boldsymbol{\beta}_n^*, \boldsymbol{\gamma}_n^*) + \Pi_{n,3} + O(n_{\mathrm{T}} J_n^{-p}),$$

where $\Pi_{n,1}(\boldsymbol{\beta}_n^*, \boldsymbol{\gamma}_n^*) = -\sum_{i=1}^{n} \underline{\mathbf{D}}_i^{\mathrm{T}} \dot{\Pi}_i(\boldsymbol{\beta}_n^*, \boldsymbol{\gamma}_n^*) \underline{\boldsymbol{\varepsilon}}_i$,

$$\Pi_{n,2}(\boldsymbol{\beta}_n^*, \boldsymbol{\gamma}_n^*) = \sum_{i=1}^{n} \underline{\mathbf{D}}_i^{\mathrm{T}} \dot{\Pi}_i(\boldsymbol{\beta}_n^*, \boldsymbol{\gamma}_n^*) \Delta_i(\boldsymbol{\beta}_n^{**}, \boldsymbol{\gamma}_n^{**}) \underline{\mathbf{D}}_i \begin{pmatrix} \boldsymbol{\beta}_n^* - \boldsymbol{\beta}_0 \\ \boldsymbol{\gamma}_n^* - \boldsymbol{\gamma}_0 \end{pmatrix},$$

$\Pi_{n,3} = \Psi_n(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0) - \Psi_n(\boldsymbol{\beta}_n^*, \boldsymbol{\gamma}_n^*)$, where $\dot{\Pi}_i(\boldsymbol{\beta}_n^*, \boldsymbol{\gamma}_n^*)$ is the first order derivative of $\Pi_i(\boldsymbol{\beta}, \boldsymbol{\gamma})$ evaluated at $(\boldsymbol{\beta}_n^{*\mathrm{T}}, \boldsymbol{\gamma}_n^{*\mathrm{T}})^{\mathrm{T}}$, which is a $m_i \times m_i \times (d_1 + d_2 J_n)$-dimensional array, $\boldsymbol{\beta}_n^{**}$ is between $\boldsymbol{\beta}_n^*$ and $\boldsymbol{\beta}_0$, and $\boldsymbol{\gamma}_n^{**}$ is between $\boldsymbol{\gamma}_n^*$ and $\boldsymbol{\gamma}_0$. By conditions (C3) and (C4) and (15), for any given vector $\alpha_n \in R^{(d_1 + d_2 J_n)}$ with $\|\alpha_n\| = 1$, there exists a constant $0 < c < \infty$, such that with probability approaching 1, $\alpha_n^{\mathrm{T}} \Psi_n(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0) \alpha_n \ge c n_{\mathrm{T}} \lambda_n^{\min}$. By Theorem 1 and (15), $\alpha_n^{\mathrm{T}} \Pi_{n,2}(\boldsymbol{\beta}_n^*, \boldsymbol{\gamma}_n^*) \alpha_n = o_p(\lambda_n^{\max})$. Since $E\{\Pi_{n,1}(\boldsymbol{\beta}_n^*, \boldsymbol{\gamma}_n^*) | \mathcal{X}, \mathcal{Z}\} = 0$, it can be proved by Bernstein's inequality of [1] $\alpha_n^{\mathrm{T}} \Pi_{n,1}(\boldsymbol{\beta}_n^*, \boldsymbol{\gamma}_n^*) \alpha_n = O_P\{(n_{\mathrm{T}} \log n_{\mathrm{T}})^{1/2}\}$. By condition (C1), $\lambda_n^{\max} = O(\tau_n^{\max}) = o(n_{\mathrm{T}} \lambda_n^{\min} J_n^{-1/2})$. Therefore, $\Psi_n(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$

dominates $\Pi_{n,1}(\boldsymbol{\beta}_n^*, \boldsymbol{\gamma}_n^*)$ and $\Pi_{n,2}(\boldsymbol{\beta}_n^*, \boldsymbol{\gamma}_n^*)$, and by Theorem 1, $\Psi_n(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$ dominates $\Pi_{n,3}(\boldsymbol{\beta}_n^*, \boldsymbol{\gamma}_n^*)$. Thus, from (16), one has

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0 \\ \widehat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_0 \end{pmatrix} = \Psi_n(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)^{-1} \mathbf{g}_n(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)\{1 + o_p(1)\}. \tag{17}$$

Let $\Delta_{i0} = \Delta_i(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$ and $\mathbf{V}_{i0} = \mathbf{V}_i(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$. To obtain the closed-form expression of $\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0$, we need the following block form of the inverse of $\sum_{i=1}^n \underline{\mathbf{D}}_i^{\mathrm{T}} \Delta_{i0} \mathbf{V}_{i0}^{-1} \Delta_{i0} \underline{\mathbf{D}}_i$:

$$\begin{aligned} &\begin{pmatrix} \sum_{i=1}^n \underline{\mathbf{X}}_i^{\mathrm{T}} \Delta_{i0} \mathbf{V}_{i0}^{-1} \Delta_{i0} \underline{\mathbf{X}}_i & \sum_{i=1}^n \underline{\mathbf{X}}_i^{\mathrm{T}} \Delta_{i0} \mathbf{V}_{i0}^{-1} \Delta_{i0} \underline{\mathbf{B}}_i \\ \sum_{i=1}^n \underline{\mathbf{B}}_i^{\mathrm{T}} \Delta_{i0} \mathbf{V}_{i0}^{-1} \Delta_{i0} \underline{\mathbf{X}}_i & \sum_{i=1}^n \underline{\mathbf{B}}_i^{\mathrm{T}} \Delta_{i0} \mathbf{V}_{i0}^{-1} \Delta_{i0} \underline{\mathbf{B}}_i \end{pmatrix}^{-1} \\ &\qquad = \begin{pmatrix} \mathbf{H}_{\mathbf{XX}} & \mathbf{H}_{\mathbf{XB}} \\ \mathbf{H}_{\mathbf{BX}} & \mathbf{H}_{\mathbf{BB}} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{H}^{11} & \mathbf{H}^{12} \\ \mathbf{H}^{21} & \mathbf{H}^{22} \end{pmatrix}, \end{aligned} \tag{18}$$

where $\mathbf{H}^{11} = (\mathbf{H}_{\mathbf{XX}} - \mathbf{H}_{\mathbf{XB}} \mathbf{H}_{\mathbf{BB}}^{-1} \mathbf{H}_{\mathbf{BX}})^{-1}$, $\mathbf{H}^{22} = (\mathbf{H}_{\mathbf{BB}} - \mathbf{H}_{\mathbf{BX}} \mathbf{H}_{\mathbf{XX}}^{-1} \mathbf{H}_{\mathbf{XB}})^{-1}$, $\mathbf{H}^{12} = -\mathbf{H}^{11} \mathbf{H}_{\mathbf{XB}} \mathbf{H}_{\mathbf{BB}}^{-1}$, and $\mathbf{H}^{21} = -\mathbf{H}^{22} \mathbf{H}_{\mathbf{BX}} \mathbf{H}_{\mathbf{XX}}^{-1}$. Consequently, $\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0 = (\widetilde{\boldsymbol{\beta}}_{n,e} + \widetilde{\boldsymbol{\beta}}_{n,\mu})\{1 + o_p(1)\}$, in which

$$\widetilde{\boldsymbol{\beta}}_{n,e} = \mathbf{H}^{11} \left\{ \sum_{i=1}^n \underline{\mathbf{X}}_i^{\mathrm{T}} \Delta_{i0} \mathbf{V}_{i0}^{-1} \underline{\boldsymbol{\varepsilon}}_i - \mathbf{H}_{\mathbf{XB}} \mathbf{H}_{\mathbf{BB}}^{-1} \sum_{i=1}^n \underline{\mathbf{B}}_i^{\mathrm{T}} \Delta_{i0} \mathbf{V}_{i0}^{-1} \underline{\boldsymbol{\varepsilon}}_i \right\},$$

$$\begin{aligned} \widetilde{\boldsymbol{\beta}}_{n,\mu} = \mathbf{H}^{11} \Bigg[&\sum_{i=1}^n \underline{\mathbf{X}}_i^{\mathrm{T}} \Delta_{i0} \mathbf{V}_{i0}^{-1} \left\{ \mu\left(\underline{\mathbf{X}}_i \boldsymbol{\beta}_0 + \sum_{l=1}^{d_2} \theta_{l0}(\mathbf{Z}_{il}) \right) - \mu(\underline{\mathbf{X}}_i \boldsymbol{\beta}_0 + \underline{\mathbf{B}}_i \boldsymbol{\gamma}_0) \right\} \\ &- \mathbf{H}_{\mathbf{XB}} \mathbf{H}_{\mathbf{BB}}^{-1} \sum_{i=1}^n \underline{\mathbf{B}}_i^{\mathrm{T}} \Delta_{i0} \mathbf{V}_{i0}^{-1} \left\{ \mu\left(\underline{\mathbf{X}}_i \boldsymbol{\beta}_0 + \sum_{l=1}^{d_2} \theta_{l0}(\mathbf{Z}_{il}) \right) \right. \\ &\left. \left. - \mu(\underline{\mathbf{X}}_i \boldsymbol{\beta}_0 + \underline{\mathbf{B}}_i \boldsymbol{\gamma}_0) \right\} \right]. \end{aligned}$$

LEMMA 1. *Under condition* (A4), *there are constants* $0 < c_{H_1} < C_{H_1} < \infty$, *such that with probability approaching* 1, *for* n_{T} *sufficiently large*, $c_{H_1}(\lambda_n^{\max} \times n_{\mathrm{T}})^{-1} \mathbf{I}_{d_1} \le \mathbf{H}^{11} \le C_{H_1}(\lambda_n^{\max} n_{\mathrm{T}})^{-1} \mathbf{I}_{d_1}$ *with* $\mathbf{H}^{11}$ *in* (18).

PROOF. The proof of Lemma 1 follows the same fashion as the proof of Lemma A.4 in [21], and is hence omitted. □

LEMMA 2. *Under conditions* (A2) *and* (A4), $\|\widetilde{\boldsymbol{\beta}}_{n,\mu}\| = O_P\{(\lambda_n^{\max}/\lambda_n^{\min}) \times J_n^{-2p}\}$.

PROOF. Let $\Delta\mu(\underline{\boldsymbol{\eta}}_i) = \mu(\underline{\mathbf{X}}_i\boldsymbol{\beta}_0 + \sum_{l=1}^{d_2}\theta_{l0}(\mathbf{Z}_{il})) - \mu(\underline{\mathbf{X}}_i\boldsymbol{\beta}_0 + \underline{\mathbf{B}}_i\boldsymbol{\gamma}_0) = \{\Delta\mu(\eta_{ij})\}_{j=1}^{m_i}$, then

$$\begin{aligned}\widetilde{\boldsymbol{\beta}}_{n,\mu} &= \mathbf{H}^{11}\left[\sum_{i=1}^n \underline{\mathbf{X}}_i^{\mathrm{T}}\Delta_{i0}\mathbf{V}_{i0}^{-1}\{\Delta\mu(\underline{\boldsymbol{\eta}}_i)\} - \mathbf{H}_{\mathbf{XB}}\mathbf{H}_{\mathbf{BB}}^{-1}\sum_{i=1}^n \underline{\mathbf{B}}_i^{\mathrm{T}}\Delta_{i0}\mathbf{V}_{i0}^{-1}\{\Delta\mu(\underline{\boldsymbol{\eta}}_i)\}\right] \\ &= \mathbf{H}^{11}\sum_{i=1}^n \underline{\mathbf{X}}_i^{\mathrm{T}}\Delta_{i0}\mathbf{V}_{i0}^{-1}[\{\Delta\mu(\underline{\boldsymbol{\eta}}_i)\} - \widehat{\Pi}_n\{\Delta\mu(\underline{\boldsymbol{\eta}}_i)\}] = n_{\mathrm{T}}\mathbf{H}^{11}\mathbf{W},\end{aligned}$$

where $\mathbf{W} = (W_1, \dots, W_{d_1})$, with

$$\begin{aligned}|W_k| &= n_{\mathrm{T}}^{-1}\left|\sum_{i=1}^n (\underline{\mathbf{X}}_i^{(k)})^{\mathrm{T}}\Delta_{i0}\mathbf{V}_{i0}^{-1}[\{\Delta\mu(\underline{\boldsymbol{\eta}}_i)\} - \widehat{\Pi}_n\{\Delta\mu(\underline{\boldsymbol{\eta}}_i)\}]\right| \\ &\le C\lambda_n^{\max} n_{\mathrm{T}}^{-1}\sum_{i=1}^n\sum_{j=1}^{m_i}\left|X_{ijk}\{\Delta\mu(\eta_{ij})\} - \widehat{\Pi}_n\{\Delta\mu(\eta_{ij})\}\right|.\end{aligned}$$

Following similar reasoning as in the proof of Lemma A.5 in [21], it can be proved that $n_{\mathrm{T}}^{-1}\sum_{i=1}^n\sum_{j=1}^{m_i}|X_{ijk}\{\Delta\mu(\eta_{ij})\} - \widehat{\Pi}_n\{\Delta\mu(\eta_{ij})\}| = O_P(J_n^{-2p})$. Therefore, $|W_k| = O_P(\lambda_n^{\max}J_n^{-2p})$. By the above result and Lemma 1, one has $\|\widetilde{\boldsymbol{\beta}}_{n,\mu}\| = O_P\{(\lambda_n^{\min})^{-1}\lambda_n^{\max}J_n^{-2p}\}$. □

LEMMA 3. *Under conditions* (A2)–(A4), *as* $n_{\mathrm{T}} \to \infty$, $\widetilde{\Xi}_n^{-1/2}(\widetilde{\boldsymbol{\beta}}_{n,e}) \longrightarrow N(0, \mathbf{I}_{d_1})$, *where* $\widetilde{\Xi}_n$ *is defined in* (9).

PROOF. Lemma 3 can be proved by using the Linderberg–Feller CLT and similar techniques for the proofs of Lemmas A.6 and A.7 in [21]. □

LEMMA 4. *Under conditions* (A2) *and* (A4), *there exist constants* $0 < c_\Xi \le C_\Xi < \infty$, *such that*

$$c_\Xi n_{\mathrm{T}}^{-1}(\lambda_n^{\max})^{-1}\tau_n^{\min}\mathbf{I}_{d_1} \le \widetilde{\Xi}_n \le C_\Xi n_{\mathrm{T}}^{-1}\tau_n^{\max}(\lambda_n^{\min})^{-1}\mathbf{I}_{d_1}$$

and $\|\widetilde{\boldsymbol{\beta}}_{n,e}\| = O_p\{n_{\mathrm{T}}^{-1/2}(\tau_n^{\max})^{1/2}(\lambda_n^{\min})^{-1/2}\}$.

PROOF. For any vector $a \in R^{d_1}$ with $\|a\| = 1$, one has

$$a^{\mathrm{T}}\widetilde{\Xi}_n a \le \tau_n^{\max}a^{\mathrm{T}}\left\{E\left(\sum_{i=1}^n \underline{\widetilde{\mathbf{X}}}_i^{\mathrm{T}}\Delta_{i0}\mathbf{V}_{i0}^{-1}\Delta_{i0}\underline{\widetilde{\mathbf{X}}}_i\right)\right\}^{-1} a \le C_\Xi n_{\mathrm{T}}^{-1}\tau_n^{\max}(\lambda_n^{\min})^{-1},$$

$$a^{\mathrm{T}}\widetilde{\Xi}_n a \ge \left\{E\left(\sum_{i=1}^n \underline{\widetilde{\mathbf{X}}}_i^{\mathrm{T}}\Delta_{i0}\mathbf{V}_{i0}^{-1}\Delta_{i0}\underline{\widetilde{\mathbf{X}}}_i\right)\right\}^{-1}\tau_n^{\min} \ge c_\Xi n_{\mathrm{T}}^{-1}(\lambda_n^{\max})^{-1}\tau_n^{\min},$$

and the second result in Lemma 4 follows from Chebyshev's inequality. □

PROOF OF THEOREM 2. By Lemmas 2 and 4, for any vector $a \in R^{d_1}$ with $\|a\| = 1$, one has

$$a^{\mathrm{T}} \widetilde{\Xi}_n^{-1/2} \widetilde{\boldsymbol{\beta}}_{n,\mu} a \le c_{\Xi}^{-1/2} n_{\mathrm{T}}^{1/2} (\lambda_n^{\max})^{1/2} (\tau_n^{\min})^{-1/2} O_P\{(\lambda_n^{\min})^{-1} \lambda_n^{\max} J_n^{-2p}\}$$
$$= O_P\{n_{\mathrm{T}}^{1/2} J_n^{-2p} (\lambda_n^{\max})^{3/2} (\lambda_n^{\min})^{-1} (\tau_n^{\min})^{-1/2}\} = o_p(1).$$

Therefore, Theorem 2 follows from Lemma 3, the above result and Slutsky's theorem. □

A.3. Proof of Theorem 3. Following the same reasoning as deriving (17), it can be proved that

$$\begin{aligned} \widehat{\boldsymbol{\gamma}}_{n,l}^{\mathcal{S}}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_{-l0}) - \boldsymbol{\gamma}_{l,0}^{\mathcal{S}} &= \Psi_{n,l}^*(\boldsymbol{\gamma}_{l,0}^{\mathcal{S}})^{-1} \mathbf{g}_{n,l}^*(\boldsymbol{\gamma}_{l,0})(1 + o_p(1)) \\ &= (\widetilde{\boldsymbol{\gamma}}_{n,e,l}^{\mathcal{S}} + \widetilde{\boldsymbol{\gamma}}_{n,\mu,l}^{\mathcal{S}})(1 + o_p(1)), \end{aligned} \tag{19}$$

where

$$\begin{aligned} \widetilde{\boldsymbol{\gamma}}_{n,e,l}^{\mathcal{S}} &= \widetilde{\boldsymbol{\gamma}}_{n,e,l}^{\mathcal{S}}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_{-l0}) \\ &= \Psi_{n,l}^*(\boldsymbol{\gamma}_{l,0}^{\mathcal{S}})^{-1} \sum_{i=1}^{n} (\mathbf{B}_{i\cdot l}^{\mathcal{S}})^{\mathrm{T}} \Delta_i(\boldsymbol{\beta}_0, \boldsymbol{\theta}_{-l0}, \boldsymbol{\gamma}_{l,0}^{\mathcal{S}}) \mathbf{V}_i^{-1}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_{-l0}, \boldsymbol{\gamma}_{l,0}^{\mathcal{S}}) \underline{\boldsymbol{\varepsilon}}_i, \\ \widetilde{\boldsymbol{\gamma}}_{n,\mu,l}^{\mathcal{S}} &= \widetilde{\boldsymbol{\gamma}}_{n,\mu,l}^{\mathcal{S}}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_{-l0}) = (\widetilde{\gamma}_{n,\mu,sl}^{\mathcal{S}})_{s=1}^{J_n^{\mathcal{S}}} \\ &= \Psi_{n,l}^*(\boldsymbol{\gamma}_{l,0}^{\mathcal{S}})^{-1} \sum_{i=1}^{n} (\mathbf{B}_{i\cdot l}^{\mathcal{S}})^{\mathrm{T}} \Delta_i(\boldsymbol{\beta}_0, \boldsymbol{\theta}_{-l0}, \boldsymbol{\gamma}_{l,0}^{\mathcal{S}}) \mathbf{V}_i^{-1}(\boldsymbol{\beta}_0, \boldsymbol{\theta}_{-l0}, \boldsymbol{\gamma}_{l,0}^{\mathcal{S}}) \\ &\qquad \times \Big\{\mu\Big(\underline{\mathbf{X}}_i \boldsymbol{\beta}_0 + \sum_{l' \ne l} \theta_{l'0}(\mathbf{Z}_{il'}) + \theta_{l0}(\mathbf{Z}_{il})\Big) \\ &\qquad - \mu\Big(\underline{\mathbf{X}}_i \boldsymbol{\beta}_0 + \sum_{l' \ne l} \theta_{l'0}(\mathbf{Z}_{il'}) + \mathbf{B}_{i\cdot l}^{\mathcal{S}} \boldsymbol{\gamma}_{l,0}^{\mathcal{S}}\Big)\Big\}. \end{aligned}$$

By the decomposition in (19),

$$\begin{aligned} \widehat{\theta}_{n,l}^{\mathcal{S}}(z_l, \boldsymbol{\beta}_0, \boldsymbol{\theta}_{-l0}) - \theta_{l0}^*(z_l) &= \mathbf{B}_l^{\mathcal{S}}(z_l)^{\mathrm{T}} \widetilde{\boldsymbol{\gamma}}_{n,e,l}^{\mathcal{S}}(1 + o_p(1)), \\ \theta_{l0}^*(z_l) - \theta_{l0}(z_l) &= \{\mathbf{B}_l^{\mathcal{S}}(z_l)^{\mathrm{T}} \widetilde{\boldsymbol{\gamma}}_{n,\mu,l}^{\mathcal{S}} + \mathbf{B}_l^{\mathcal{S}}(z_l)^{\mathrm{T}} \boldsymbol{\gamma}_{l,0}^{\mathcal{S}} - \theta_{l0}(z_l)\} \\ &\qquad \times (1 + o_p(1)). \end{aligned}$$

It can be proved by the Linderberg–Feller CLT that as $n_{\mathrm{T}} \to \infty$,

$$(\mathbf{B}_l^{\mathcal{S}}(z_l)^{\mathrm{T}} \Xi_{n,l}^* \mathbf{B}_l^{\mathcal{S}}(z_l))^{-1/2} (\mathbf{B}_l^{\mathcal{S}}(z_l)^{\mathrm{T}} \widetilde{\boldsymbol{\gamma}}_{n,e,l}^{\mathcal{S}}) \longrightarrow N(0, 1).$$

Following similar reasoning as in the proofs in Lemma 5, it can be proved

$$\sup_{1\le s\le J_n^{\mathcal{S}}} |\widetilde{\gamma}_{n,\mu,sl}^{\mathcal{S}}| = O_P\{(\lambda_n^{\min})^{-1}\lambda_n^{\max}(J_n^{\mathcal{S}})^{-p-1/2}\}$$

and

$$\|\widetilde{\boldsymbol{\gamma}}_{n,\varepsilon,l}^{\mathcal{S}}\|_\infty = O_P\{(\log n_{\mathrm{T}}/n_{\mathrm{T}})^{1/2}(\tau_n^{\max})^{1/2}(\lambda_n^{\min})^{-1/2}\}.$$

By B-spline properties, $\sup_{z_l\in[0,1]}|\mathbf{B}_l^{\mathcal{S}}(z_l)^{\mathrm{T}}\widetilde{\boldsymbol{\gamma}}_{n,\mu,l}^{\mathcal{S}}| = O_P\{(\lambda_n^{\max}/\lambda_n^{\min})(J_n^{\mathcal{S}})^{-p}\}$, and $\sup_{z_l\in[0,1]}|\mathbf{B}_l^{\mathcal{S}}(z_l)^{\mathrm{T}}\widetilde{\boldsymbol{\gamma}}_{n,\varepsilon,l}^{\mathcal{S}}| = O_P\{\sqrt{(\log n_{\mathrm{T}})J_n^{\mathcal{S}}/n_{\mathrm{T}}}(\tau_n^{\max}/\lambda_n^{\min})^{1/2}\}$, so

$$\begin{aligned}\sup_{z_l\in[0,1]}|\theta_{l0}^*(z_l)-\theta_{l0}(z_l)| &\le \sup_{z_l\in[0,1]}|\mathbf{B}_l^{\mathcal{S}}(z_l)^{\mathrm{T}}\widetilde{\boldsymbol{\gamma}}_{n,\mu,l}^{\mathcal{S}}| \\ &\quad + \sup_{z_l\in[0,1]}|\mathbf{B}_l^{\mathcal{S}}(z_l)^{\mathrm{T}}\boldsymbol{\gamma}_{l,0}^{\mathcal{S}}-\theta_{l0}(z_l)| \\ &= O_P\{(\lambda_n^{\min})^{-1}\lambda_n^{\max}(J_n^{\mathcal{S}})^{-p}\},\end{aligned}$$

$\sup_{z_l\in[0,1]}|\widehat{\theta}_{n,l}^{\mathcal{S}}(z_l,\boldsymbol{\beta}_0,\boldsymbol{\theta}_{-l0})-\theta_{l0}^*(z_l)| = O_P\{\sqrt{(\log n_{\mathrm{T}})J_n^{\mathcal{S}}/n_{\mathrm{T}}}(\tau_n^{\max}/\lambda_n^{\min})^{1/2}\}$.

A.4. Proof of Theorem 4.

LEMMA 5. *Under conditions* (A2)–(A4),

$$\|\widehat{\boldsymbol{\gamma}}_n-\boldsymbol{\gamma}_0\| = O_P\{J_n^{1/2}n_{\mathrm{T}}^{-1/2}(\tau_n^{\max}/\lambda_n^{\min})^{1/2}+(\lambda_n^{\max}/\lambda_n^{\min})J_n^{-p}\},$$

$$\|\widehat{\boldsymbol{\gamma}}_n-\boldsymbol{\gamma}_0\|_\infty = O_P\{(\log n_{\mathrm{T}}/n_{\mathrm{T}})^{1/2}(\tau_n^{\max}/\lambda_n^{\min})^{1/2}+(\lambda_n^{\max}/\lambda_n^{\min})J_n^{-p-1/2}\}.$$

PROOF. From (17) and (18), one obtains $\widehat{\boldsymbol{\gamma}}_n-\boldsymbol{\gamma}_0 = (\widetilde{\boldsymbol{\gamma}}_{n,e}+\widetilde{\boldsymbol{\gamma}}_{n,\mu})(1+o_p(1))$, where

$$\widetilde{\boldsymbol{\gamma}}_{n,e} = \mathbf{H}^{22}\left\{\sum_{i=1}^n \underline{\mathbf{B}}_i^{\mathrm{T}}\Delta_{i0}\mathbf{V}_{i0}^{-1}\underline{\boldsymbol{\varepsilon}}_i - \mathbf{H}_{\mathbf{BX}}\mathbf{H}_{\mathbf{XX}}^{-1}\sum_{i=1}^n \underline{\mathbf{X}}_i^{\mathrm{T}}\Delta_{i0}\mathbf{V}_{i0}^{-1}\underline{\boldsymbol{\varepsilon}}_i\right\},$$

$$\begin{aligned}\widetilde{\boldsymbol{\gamma}}_{n,\mu} = \mathbf{H}^{22}\Bigg[&\sum_{i=1}^n \underline{\mathbf{B}}_i^{\mathrm{T}}\Delta_{i0}\mathbf{V}_{i0}^{-1}\left\{\mu\left(\underline{\mathbf{X}}_i\boldsymbol{\beta}_0+\sum_{l=1}^{d_2}\theta_{l0}(\mathbf{Z}_{il})\right)-\mu(\underline{\mathbf{X}}_i\boldsymbol{\beta}_0+\underline{\mathbf{B}}_i\boldsymbol{\gamma}_0)\right\} \\ &- \mathbf{H}_{\mathbf{BX}}\mathbf{H}_{\mathbf{XX}}^{-1}\sum_{i=1}^n \underline{\mathbf{X}}_i^{\mathrm{T}}\Delta_{i0}\mathbf{V}_{i0}^{-1}\bigg\{\mu\left(\underline{\mathbf{X}}_i\boldsymbol{\beta}_0+\sum_{l=1}^{d_2}\theta_{l0}(\mathbf{Z}_{il})\right) \\ &\qquad -\mu(\underline{\mathbf{X}}_i\boldsymbol{\beta}_0+\underline{\mathbf{B}}_i\boldsymbol{\gamma}_0)\bigg\}\Bigg].\end{aligned}$$

It can be proved that there exist constants $0<c_{H_2}<C_{H_2}<\infty$, such that with probability approaching 1, for n_{T} sufficiently large,

$$c_{H_2}(\lambda_n^{\max})^{-1}n_{\mathrm{T}}^{-1}\mathbf{I}_{d_1} \le \mathbf{H}^{22} \le C_{H_2}(\lambda_n^{\min})^{-1}n_{\mathrm{T}}^{-1}\mathbf{I}_{d_1}.$$

Letting $\widehat{\Pi}_{n,\mathbf{X}}$ be the projection on $\{\underline{\mathbf{X}}_i\}_{i=1}^n$ to the empirical inner product,

$$\widetilde{\boldsymbol{\gamma}}_{n,\mu} = \mathbf{H}^{22} \sum_{i=1}^n \underline{\mathbf{B}}_i^{\mathrm{T}} \Delta_{i0} \mathbf{V}_{i0}^{-1} [\{\Delta\mu(\underline{\boldsymbol{\eta}}_i)\} - \widehat{\Pi}_{n,\mathbf{X}}\{\Delta\mu(\underline{\boldsymbol{\eta}}_i)\}] = n_{\mathrm{T}} \mathbf{H}^{22} \mathbf{W},$$

where $\mathbf{W} = (W_1, \dots, W_{J_n d_2})$, with

$$W_{s,l} = n_{\mathrm{T}}^{-1} \sum_{i=1}^n (\mathbf{B}_i^{(s,l)})^{\mathrm{T}} \Delta_{i0} \mathbf{V}_{i0}^{-1} [\{\Delta\mu(\underline{\boldsymbol{\eta}}_i)\} - \widehat{\Pi}_{n,\mathbf{X}}\{\Delta\mu(\underline{\boldsymbol{\eta}}_i)\}],$$

$\mathbf{B}_i^{(s,l)} = [\{B_{s,l}(Z_{i1l}), \dots, B_{s,l}(Z_{im_il})\}^{\mathrm{T}}]$. The Cauchy–Schwarz inequality implies

$$\begin{aligned}
|W_{s,l}| &\le C\lambda_n^{\max} n_{\mathrm{T}}^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} |B_{s,l}(Z_{ijl})\{\Delta\mu(\eta_{ij})\} - \widehat{\Pi}_{n,\mathbf{X}}\{\Delta\mu(\eta_{ij})\}| \\
&\le C\lambda_n^{\max} \|B_{s,l}\|_{n_{\mathrm{T}}} \|\Delta\mu - \widehat{\Pi}_{n,\mathbf{X}}(\Delta\mu)\|_{n_{\mathrm{T}}} = O_P(\lambda_n^{\max} J_n^{-p-1/2}),
\end{aligned}$$

thus, $\|\widetilde{\boldsymbol{\gamma}}_{n,\mu}\| = O_P\{(\lambda_n^{\max}/\lambda_n^{\min}) J_n^{-p}\}$, $\|\widetilde{\boldsymbol{\gamma}}_{n,\mu}\|_\infty = O_P\{(\lambda_n^{\max}/\lambda_n^{\min}) J_n^{-p-1/2}\}$. For any $\omega \in \mathcal{R}^{J_n d_2}$ with $\|\omega\| = 1$, it can be proved that $\mathrm{Var}(\omega^{\mathrm{T}} \widetilde{\boldsymbol{\gamma}}_{n,e} | \mathcal{X}, \mathcal{Z}) \le O_P\{n_{\mathrm{T}}^{-1}(\tau_n^{\max}/\lambda_n^{\min})\}$, thus, $\omega^{\mathrm{T}} \widetilde{\boldsymbol{\gamma}}_{n,e} = O_P\{n_{\mathrm{T}}^{-1/2}(\tau_n^{\max}/\lambda_n^{\min})^{1/2}\}$. Therefore, $\|\widetilde{\boldsymbol{\gamma}}_{n,e}\| \le J_n^{1/2} |\omega^{\mathrm{T}} \widetilde{\boldsymbol{\gamma}}_{n,e}| = O_P\{J_n^{1/2} n_{\mathrm{T}}^{-1/2} (\tau_n^{\max}/\lambda_n^{\min})^{1/2}\}$, and by Bernstein's inequality of [1] that $\|\widetilde{\boldsymbol{\gamma}}_{n,e}\|_\infty = O_P\{(\log n_{\mathrm{T}}/n_{\mathrm{T}})^{1/2} (\tau_n^{\max}/\lambda_n^{\min})^{1/2}\}$. □

LEMMA 6. *Under conditions* (A2)–(A4),

$$\|\widehat{\boldsymbol{\gamma}}_{n,l}^{\mathcal{SS}} - \widehat{\boldsymbol{\gamma}}_{n,l}^{\mathrm{OR}}\|_\infty = O_p\Big\{(\lambda_n^{\max}/\lambda_n^{\min})^2 \Big(\sqrt{\log n_{\mathrm{T}}/(J_n^{\mathcal{S}} n_{\mathrm{T}})} + (J_n^{\mathcal{S}})^{-1/2} J_n^{-p}\Big)\Big\}.$$

PROOF. Let $\widetilde{\boldsymbol{\theta}}_{-l0} = \{\widetilde{\theta}_{l'0}(\cdot), l' \ne l\}$, where $\widetilde{\theta}_{l'0}(\cdot)$ is defined in (8). Let $\widehat{\boldsymbol{\gamma}}_{n,-l} = (\widehat{\gamma}_{n,sl'} : 1 \le s \le J_n, l' \ne l)^{\mathrm{T}}$ and $\boldsymbol{\gamma}_{-l0} = (\gamma_{sl',0} : 1 \le s \le J_n, l' \ne l)^{\mathrm{T}}$. By the Taylor expansion, $\mathbf{g}_{n,l}^{\mathcal{S}}(\widehat{\boldsymbol{\gamma}}_{n,l}^{\mathrm{OR}}, \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\theta}}_{n,-l}) - \mathbf{g}_{n,l}^{\mathcal{S}}(\widehat{\boldsymbol{\gamma}}_{n,l}^{\mathrm{OR}}, \widehat{\boldsymbol{\beta}}_n, \widetilde{\boldsymbol{\theta}}_{-l0}) = \{\partial \mathbf{g}_{n,l}^{\mathcal{S}}(\widehat{\boldsymbol{\gamma}}_{n,l}^{\mathrm{OR}}, \widehat{\boldsymbol{\beta}}_n, \widetilde{\boldsymbol{\theta}}_{-l}) / \partial \widetilde{\boldsymbol{\gamma}}_{-l}^{\mathrm{T}}\}(\widehat{\boldsymbol{\gamma}}_{n,-l} - \boldsymbol{\gamma}_{-l0})$, where $\widetilde{\boldsymbol{\gamma}}_{-l} = t\boldsymbol{\gamma}_{-l0} + (1-t)\widehat{\boldsymbol{\gamma}}_{n,-l}$ for $t \in (0,1)$. Let $\widehat{\Delta}_i = \Delta_i(\widehat{\boldsymbol{\beta}}_n, \widetilde{\boldsymbol{\theta}}_{-l}, \widehat{\boldsymbol{\gamma}}_{n,l}^{\mathrm{OR}})$, $\widehat{\mathbf{V}}_i = \mathbf{V}_i(\widehat{\boldsymbol{\beta}}_n, \widetilde{\boldsymbol{\theta}}_{-l}, \widehat{\boldsymbol{\gamma}}_{n,l}^{\mathrm{OR}})$, $\widetilde{\underline{\boldsymbol{\varepsilon}}}_i = \underline{\boldsymbol{\varepsilon}}_i - \widehat{\Pi}_{n,\mathbf{X}}(\underline{\boldsymbol{\varepsilon}}_i)$, $\widetilde{\Delta\mu}(\underline{\boldsymbol{\eta}}_i) = \Delta\mu(\underline{\boldsymbol{\eta}}_i) - \widehat{\Pi}_{n,\mathbf{X}}\{\Delta\mu(\underline{\boldsymbol{\eta}}_i)\}$, $\mathbf{B}_{ij,-l} = \{(\mathbf{B}_{ijl'}^{\mathrm{T}}, l' \ne l)^{\mathrm{T}}\}_{(d_2-1)J_n \times 1}$, $\mathbf{B}_{i,-l} = \{(\mathbf{B}_{i1,-l}, \dots, \mathbf{B}_{im_i,-l})^{\mathrm{T}}\}_{m_i \times (d_2-1)J_n}$. Thus, by (6) and the proofs for Lemma 5, with probability approaching 1, there are constants $0 < C_1, C_2 < \infty$ such that

$$\begin{aligned}
&\|\mathbf{g}_{n,l}^{\mathcal{S}}(\widehat{\boldsymbol{\gamma}}_{n,l}^{\mathrm{OR}}, \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\theta}}_{n,-l}) - \mathbf{g}_{n,l}^{\mathcal{S}}(\widehat{\boldsymbol{\gamma}}_{n,l}^{\mathrm{OR}}, \widehat{\boldsymbol{\beta}}_n, \widetilde{\boldsymbol{\theta}}_{-l0})\|_\infty \\
&\quad \le C_1 (\lambda_n^{\min})^{-1} n_{\mathrm{T}}^{-1} \\
&\qquad \times \left\| \left(\sum_{i=1}^n (\mathbf{B}_{i\cdot l}^{\mathcal{S}})^{\mathrm{T}} \widehat{\Delta}_i \widehat{\mathbf{V}}_i^{-1} \mathbf{B}_{i,-l} \right) \left\{ \sum_{i=1}^n \mathbf{B}_{i,-l}^{\mathrm{T}} \Delta_{i0} \mathbf{V}_{i0}^{-1} (\widetilde{\underline{\boldsymbol{\varepsilon}}}_i + \widetilde{\Delta\mu}(\underline{\boldsymbol{\eta}}_i)) \right\} \right\|_\infty \\
&\quad \le C_2 (\lambda_n^{\min})^{-1} (\|\zeta_1\|_\infty + \|\zeta_2\|_\infty),
\end{aligned}$$

where $\zeta_1 = n_T^{-1}\{\sum_{i=1}^n (\mathbf{B}_{i\cdot l}^{\mathcal{S}})^T \Delta_{i0}\mathbf{V}_{i0}^{-1}\mathbf{B}_{i,-l}\}\{\sum_{i=1}^n \mathbf{B}_{i,-l}^T \Delta_{i0}\mathbf{V}_{i0}^{-1}(\widetilde{\Delta\mu}(\underline{\boldsymbol{\eta}}_i))\}$, $\zeta_2 = n_T^{-1}\{\sum_{i=1}^n (\mathbf{B}_{i\cdot l}^{\mathcal{S}})^T \Delta_{i0}\mathbf{V}_{i0}^{-1}\mathbf{B}_{i,-l}\}(\sum_{i=1}^n \mathbf{B}_{i,-l}^T \Delta_{i0}\mathbf{V}_{i0}^{-1}\underline{\widetilde{\boldsymbol{\varepsilon}}}_i)$, and then $\|\zeta_1\|_\infty \le (\lambda_n^{\max})^2\|\zeta_3\|_\infty O(J_n^{-p})$, where $\zeta_3 = \Delta_1 + \Delta_2 + \Delta_3$, $\Delta_1 = (\delta_{1s})_{s=1}^{J_n^{\mathcal{S}}}$, $\Delta_2 = (\delta_{2s})_{s=1}^{J_n^{\mathcal{S}}}$ and $\Delta_3 = (\delta_{3s})_{s=1}^{J_n^{\mathcal{S}}}$ with $\delta_{1s} = n_T^{-1}\sum_{i=1}^n \delta_{1s,i}$, $\delta_{2s} = n_T^{-1}\sum_{i=1}^n \delta_{2s,i}$ and $\delta_{3s} = n_T^{-1}\sum_{i=1}^n \delta_{3s,i}$,

$$\delta_{1s,i} = \sum_{j=1}^{m_i}\sum_{l'=1,l'\ne l}^{d_2}\sum_{s'=1}^{J_n}|B_{s,l}^{\mathcal{S}}(Z_{ijl})||B_{s',l'}(Z_{ijl'})|^2,$$

$$\delta_{2s,i} = \sum_{j=1}^{m_i}\sum_{j'\ne j}\sum_{l'\ne l}\sum_{s'=1}^{J_n}|B_{s,l}^{\mathcal{S}}(Z_{ijl})||B_{s',l'}(Z_{ijl'})||B_{s',l'}(Z_{ij'l'})|,$$

$$\delta_{3s,i} = \sum_{j=1}^{m_i}\sum_{i'\ne i}\sum_{j'}\sum_{l'\ne l}\sum_{s'=1}^{J_n}|B_{s,l}^{\mathcal{S}}(Z_{ijl})||B_{s',l'}(Z_{ijl'})||B_{s',l'}(Z_{i'j'l'})|.$$

Let $\delta_{1s,i}^* = \delta_{1s,i} - E(\delta_{1s,i})$. It can be proved by B-spline properties that $E(\delta_{1s,i}) \asymp m_i J_n/\sqrt{J_n^{\mathcal{S}}}$, $E(\delta_{1s,i}^*) = 0$, $E(\delta_{1s,i}^*)^2 \asymp m_i J_n^2 + m_i^2 J_n^2 (J_n^{\mathcal{S}})^{-1}$, and $E(|\delta_{1s,i}^*|^k) \le C\{m_i J_n^k (J_n^{\mathcal{S}})^{k/2-1} + m_i^2 J_n^k (J_n^{\mathcal{S}})^{k/2-2}\}$ for $k \ge 3$ and some constant $C > 0$. Thus, $E(|\delta_{1s,i}^*|^k) \le (C'(J_n^{\mathcal{S}})^{1/2}J_n)^{k-2}k!E(\delta_{1s,ijl's'}^2)$ with $C' = C^{1/(k-2)}$. By Bernstein's inequality in [1],

$$P\left(\left|\sum_{i=1}^n \delta_{1s,i}\right| \ge t\right) \le 2\exp\left\{-\frac{t^2}{4\sum_{i=1}^n E(\delta_{1s,i}^*)^2 + 2C'(J_n^{\mathcal{S}})^{1/2}J_n t}\right\}.$$

Let $t = c\{\{n_T J_n^2 + (\sum_{i=1}^n m_i^2)J_n^2(J_n^{\mathcal{S}})^{-1}\}\log n_T\}^{1/2}$ for a large constant $0 < c < \infty$. There is a constant $0 < c' < \infty$ such that $E(\delta_{1s,i}^*)^2 \le c'\{m_i J_n^2 + m_i^2 J_n^2 (J_n^{\mathcal{S}})^{-1}\}$. For $J_n^{\mathcal{S}} = O((\log n_T)^{-1} n_T^{1/2} m_{(n)}^{1/2})$, one has $P(|\sum_{i=1}^n \delta_{1s,i}| \ge t) \le 2n_T^{-c^2/(4c')}$. By the Borel–Cantelli lemma,

$$\max_{1\le s\le J_n^{\mathcal{S}}}|\delta_{1s} - E(\delta_{1s})| = O_{\text{a.s.}}\{n_T^{-1/2}J_n(1 + m_{(n)}/J_n^{\mathcal{S}})^{1/2}(\log n_T)^{1/2}\}.$$

Since $E(\delta_{1s}) \asymp J_n/\sqrt{J_n^{\mathcal{S}}}$, one has $\|\Delta_1\|_\infty = O_{\text{a.s.}}(J_n/\sqrt{J_n^{\mathcal{S}}})$. Since $E(\delta_{2s}) \asymp n_T^{-1}(\sum_{i=1}^n m_i^2)/\sqrt{J_n^{\mathcal{S}}}$ and $E(\delta_{3s}) \asymp n_T/\sqrt{J_n^{\mathcal{S}}}$, similarly it can be proved that $\|\Delta_2\|_\infty = O_{\text{a.s.}}(m_{(n)}/\sqrt{J_n^{\mathcal{S}}})$ and $\|\Delta_3\|_\infty = O_{\text{a.s.}}(n_T/\sqrt{J_n^{\mathcal{S}}})$. Therefore, $\|\zeta_1\|_\infty = O_{\text{a.s.}}\{(\lambda_n^{\max})^2 n_T (J_n^{\mathcal{S}})^{-1/2} J_n^{-p}\}$. Following similar reasoning, by Bernstein's inequality one can prove $\|\zeta_2\|_\infty = O_{\text{a.s.}}((\lambda_n^{\max})^2\sqrt{n_T \log n_T/J_n^{\mathcal{S}}})$. Thus,

$$\|\mathbf{g}_{n,l}^{\mathcal{S}}(\widehat{\boldsymbol{\gamma}}_{n,l}^{\text{OR}}, \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\theta}}_{n,-l}) - \mathbf{g}_{n,l}^{\mathcal{S}}(\widehat{\boldsymbol{\gamma}}_{n,l}^{\text{OR}}, \widehat{\boldsymbol{\beta}}_n, \widetilde{\boldsymbol{\theta}}_{-l0})\|_\infty = O_p(a_n + b_n),$$

where $a_n = c_n(n_{\mathrm{T}} \log n_{\mathrm{T}}/J_n^{\mathcal{S}})^{1/2}$ and $b_n = c_n n_{\mathrm{T}}(J_n^{\mathcal{S}})^{-1/2} J_n^{-p}$ with $c_n = (\lambda_n^{\min})^{-1}(\lambda_n^{\max})^2$. Following similar reasoning, one can prove that $\|\mathbf{g}_{n,l}^{\mathcal{S}}(\widehat{\boldsymbol{\gamma}}_{n,l}^{\mathrm{OR}}, \widehat{\boldsymbol{\beta}}_n, \widetilde{\boldsymbol{\theta}}_{-l0}) - \mathbf{g}_{n,l}^{\mathcal{S}}(\widehat{\boldsymbol{\gamma}}_{n,l}^{\mathrm{OR}}, \boldsymbol{\beta}, \widetilde{\boldsymbol{\theta}}_{-l0})\|_\infty = O_p(a_n + d_n)$, where $d_n = c_n n_{\mathrm{T}}(J_n^{\mathcal{S}})^{-1/2} J_n^{-2p}$, $\|\mathbf{g}_{n,l}^{\mathcal{S}}(\widehat{\boldsymbol{\gamma}}_{n,l}^{\mathrm{OR}}, \boldsymbol{\beta}, \widetilde{\boldsymbol{\theta}}_{-l0}) - \mathbf{g}_{n,l}^{\mathcal{S}}(\widehat{\boldsymbol{\gamma}}_{n,l}^{\mathrm{OR}}, \boldsymbol{\beta}, \boldsymbol{\theta}_{-l})\|_\infty = O_p(b_n)$, where $\mathbf{g}_{n,l}^{\mathcal{S}}(\widehat{\boldsymbol{\gamma}}_{n,l}^{\mathrm{OR}}, \boldsymbol{\beta}, \boldsymbol{\theta}_{-l}) = \mathbf{0}$. Thus, $\|\mathbf{g}_{n,l}^{\mathcal{S}}(\widehat{\boldsymbol{\gamma}}_{n,l}^{\mathrm{OR}}, \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\theta}}_{n,-l})\|_\infty = O_p(a_n + b_n)$. By the Taylor expansion, there is $t \in (0,1)$ such that $\widetilde{\boldsymbol{\gamma}}_{n,l} = t\widehat{\boldsymbol{\gamma}}_{n,l}^{\mathrm{OR}} + (1-t)\widehat{\boldsymbol{\gamma}}_{n,l}^{\mathcal{SS}}$,

$$\widehat{\boldsymbol{\gamma}}_{n,l}^{\mathcal{SS}} - \widehat{\boldsymbol{\gamma}}_{n,l}^{\mathrm{OR}} = -\{\partial \mathbf{g}_{n,l}^{\mathcal{S}}(\widetilde{\boldsymbol{\gamma}}_{n,l}, \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\theta}}_{n,-l})/\partial \widetilde{\boldsymbol{\gamma}}_{n,l}^{\mathrm{T}}\}^{-1} \mathbf{g}_{n,l}^{\mathcal{S}}(\widehat{\boldsymbol{\gamma}}_{n,l}^{\mathrm{OR}}, \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\theta}}_{n,-l}).$$

$\partial \mathbf{g}_{n,l}^{\mathcal{S}}(\widetilde{\boldsymbol{\gamma}}_{n,l}, \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\theta}}_{n,-l})/\partial \widetilde{\boldsymbol{\gamma}}_{n,l}^{\mathrm{T}} = \Lambda_n(1 + o_p(1))$, with $\Lambda_n = \sum_{i=1}^n (\mathbf{B}_{i\cdot l}^{\mathcal{S}})^{\mathrm{T}} \widetilde{\Delta}_i \widetilde{\mathbf{V}}_i^{-1} \times \widetilde{\Delta}_i \mathbf{B}_{i\cdot l}^{\mathcal{S}}$, $\widetilde{\Delta}_i = \Delta_i(\widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\theta}}_{n,-l}, \widetilde{\boldsymbol{\gamma}}_{n,l})$ and $\widetilde{\mathbf{V}}_i = \mathbf{V}_i(\widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\theta}}_{n,-l}, \widetilde{\boldsymbol{\gamma}}_{n,l})$. There exist constants $0 < c_3 < C_3 < \infty$, such that with probability 1, for n_{T} sufficiently large, $c_3 \lambda_n^{\min} n_{\mathrm{T}} \le \lambda_{\min}(\Lambda_n) \le \lambda_{\max}(\Lambda_n) \le C_3 \lambda_n^{\max} n_{\mathrm{T}}$. By Theorem 13.4.3 of [4], one has $\|\Lambda_n^{-1}\|_\infty = O_{\mathrm{a.s.}}\{(\lambda_n^{\min} n_{\mathrm{T}})^{-1}\}$. Therefore,

$$\begin{aligned}\|\widehat{\boldsymbol{\gamma}}_{n,l}^{\mathcal{SS}} - \widehat{\boldsymbol{\gamma}}_{n,l}^{\mathrm{OR}}\|_\infty &\le \|\{\partial \mathbf{g}_{n,l}^{\mathcal{S}}(\widetilde{\boldsymbol{\gamma}}_{n,l}, \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\theta}}_{n,-l})/\partial \widetilde{\boldsymbol{\gamma}}_{n,l}^{\mathrm{T}}\}^{-1}\|_\infty \|\mathbf{g}_{n,l}^{\mathcal{S}}(\widehat{\boldsymbol{\gamma}}_{n,l}^{\mathrm{OR}}, \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\theta}}_{n,-l})\|_\infty \\ &= O_p\{(\lambda_n^{\max}/\lambda_n^{\min})^2(\sqrt{\log n_{\mathrm{T}}/(J_n^{\mathcal{S}} n_{\mathrm{T}})} + (J_n^{\mathcal{S}})^{-1/2} J_n^{-p})\}. \qquad \square\end{aligned}$$

PROOF OF THEOREM 4. By Lemma 6,

$$\begin{aligned}&\sup_{z_l \in [0,1]} |\widehat{\theta}_{n,l}^{\mathcal{S}}(z_l, \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\theta}}_{n,-l}) - \widehat{\theta}_{n,l}^{\mathcal{S}}(z_l, \boldsymbol{\beta}_0, \boldsymbol{\theta}_{-l0})| \\ &\qquad \le \sum_{s=1}^{J_n^{\mathcal{S}}} |B_{s,l}(z_l)| \|\widehat{\boldsymbol{\gamma}}_{n,l}^{\mathcal{SS}} - \widehat{\boldsymbol{\gamma}}_{n,l}^{\mathrm{OR}}\|_\infty \\ &\qquad = O_p\{(\lambda_n^{\max}/\lambda_n^{\min})^2(\sqrt{\log n_{\mathrm{T}}/n_{\mathrm{T}}} + J_n^{-p})\}.\end{aligned}$$

By the above result and (11),

$$\sup_{z_l \in [0,1]} |(\mathbf{B}_l^{\mathcal{S}}(z_l)^{\mathrm{T}} \Xi_{n,l}^* \mathbf{B}_l^{\mathcal{S}}(z_l))^{-1/2} \{\widehat{\theta}_{n,l}^{\mathcal{S}}(z_l, \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\theta}}_{n,-l}) - \widehat{\theta}_{n,l}^{\mathcal{S}}(z_l, \boldsymbol{\beta}_0, \boldsymbol{\theta}_{-l0})\}| = o_p(1).$$

Thus, the asymptotic normality of $\widehat{\theta}_{n,l}^{\mathcal{S}}(z_l, \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\theta}}_{n,-l})$ follows from Theorem 3, the above result and Slutsky's theorem. $\square$

Acknowledgments. The author is grateful for the insightful comments from the Editor, an Associate Editor and anonymous referees.

REFERENCES

[1] BOSQ, D. (1998). *Nonparametric Statistics for Stochastic Processes*: *Estimation and Prediction*, 2nd ed. *Lecture Notes in Statistics* **110**. Springer, New York. MR1640691

[2] BUN, M. J. G. and CARREE, M. A. (2005). Bias-corrected estimation in dynamic panel data models. *J. Bus. Econom. Statist.* **23** 200–210. MR2157271

[3] DE BOOR, C. (2001). *A Practical Guide to Splines*, revised ed. *Applied Mathematical Sciences* **27**. Springer, New York. MR1900298

[4] DEVORE, R. A. and LORENTZ, G. G. (1993). *Constructive Approximation*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **303**. Springer, Berlin. MR1261635

[5] HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models*. *Monographs on Statistics and Applied Probability* **43**. Chapman & Hall, London. MR1082147

[6] HE, X., FUNG, W. K. and ZHU, Z. (2005). Robust estimation in generalized partial linear models for clustered data. *J. Amer. Statist. Assoc.* **100** 1176–1184. MR2236433

[7] HE, X. and SHI, P. (1996). Bivariate tensor-product B-splines in a partly linear model. *J. Multivariate Anal.* **58** 162–181. MR1405586

[8] HECKMAN, N. E. (1986). Spline smoothing in a partly linear model. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **48** 244–248. MR0868002

[9] HOOVER, D. R., RICE, J. A., WU, C. O. and YANG, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85** 809–822. MR1666699

[10] HOROWITZ, J., KLEMELÄ, J. and MAMMEN, E. (2006). Optimal estimation in additive regression models. *Bernoulli* **12** 271–298. MR2218556

[11] HOROWITZ, J. L. and LEE, S. (2005). Nonparametric estimation of an additive quantile regression model. *J. Amer. Statist. Assoc.* **100** 1238–1249. MR2236438

[12] HOROWITZ, J. L. and MAMMEN, E. (2004). Nonparametric estimation of an additive model with a link function. *Ann. Statist.* **32** 2412–2443. MR2153990

[13] HUANG, J. Z. (2003). Local asymptotics for polynomial spline regression. *Ann. Statist.* **31** 1600–1635. MR2012827

[14] HUANG, J. Z., ZHANG, L. and ZHOU, L. (2007). Efficient estimation in marginal partially linear models for longitudinal/clustered data using splines. *Scand. J. Stat.* **34** 451–477. MR2368793

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

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

[17] LIN, X. and CARROLL, R. J. (2001). Semiparametric regression for clustered data. *Biometrika* **88** 1179–1185. MR1872228

[18] LIN, X., WANG, N., WELSH, A. H. and CARROLL, R. J. (2004). Equivalent kernels of smoothing splines in nonparametric regression for clustered/longitudinal data. *Biometrika* **91** 177–193. MR2050468

[19] LINTON, O. B. (2000). Efficient estimation of generalized additive nonparametric regression models. *Econometric Theory* **16** 502–523. MR1790289

[20] LIU, R. and YANG, L. (2010). Spline-backfitted kernel smoothing of additive coefficient model. *Econometric Theory* **26** 29–59. MR2587102

[21] MA, S., SONG, Q. and WANG, L. (2013). Simultaneous variable selection and estimation in semiparametric modeling of longitudinal/clustered data. *Bernoulli* **19** 252–274.

[22] MA, S. and YANG, L. (2011). Spline-backfitted kernel smoothing of partially linear additive model. *J. Statist. Plann. Inference* **141** 204–219. MR2719488

[23] MAMMEN, E., LINTON, O. and NIELSEN, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Ann. Statist.* **27** 1443–1490. MR1742496

[24] MUNNELL, A. H. (1990). How does public infrastructure affect regional economic performance. *New England Econ. Rev.* Sep. 11–33.

[25] OPSOMER, J. D. and RUPPERT, D. (1997). Fitting a bivariate additive model by local polynomial regression. *Ann. Statist.* **25** 186–211. MR1429922

[26] SONG, Q. and YANG, L. (2010). Oracally efficient spline smoothing of nonlinear additive autoregression models with simultaneous confidence band. *J. Multivariate Anal.* **101** 2008–2025. MR2671198

[27] WALTERSKIRCHEN, E. (1999). The relationship between growth, employment and unemployment in the EU. European economists for an alternative economic policy, Workshop in Barcelona.

[28] WANG, J. and YANG, L. (2009). Polynomial spline confidence bands for regression curves. *Statist. Sinica* **19** 325–342. MR2487893

[29] WANG, L. and YANG, L. (2007). Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *Ann. Statist.* **35** 2474–2503. MR2382655

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

[31] WELSH, A. H., LIN, X. and CARROLL, R. J. (2002). Marginal longitudinal nonparametric regression: Locality and efficiency of spline and kernel methods. *J. Amer. Statist. Assoc.* **97** 482–493. MR1941465

[32] WILD, C. J. and YEE, T. W. (1996). Additive extensions to generalized estimating equation methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **58** 711–725. MR1410186

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

[34] XUE, L. and YANG, L. (2006). Additive coefficient modeling via polynomial spline. *Statist. Sinica* **16** 1423–1446. MR2327498

[35] ZHOU, S., SHEN, X. and WOLFE, D. A. (1998). Local asymptotics for regression splines and confidence regions. *Ann. Statist.* **26** 1760–1782. MR1673277

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, RIVERSIDE
RIVERSIDE, CALIFORNIA 92521
USA
E-MAIL: shujie.ma@ucr.edu
URL: http://faculty.ucr.edu/~shujiema/