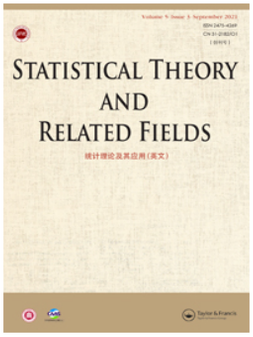

Statistical Theory and Related Fields

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tstf20

Personalized treatment selection via the covariate-specific treatment effect curve for longitudinal data

Yanghui Liu, Riquan Zhang, Shujie Ma & Xiuzhen Zhang

To cite this article: Yanghui Liu, Riquan Zhang, Shujie Ma & Xiuzhen Zhang (2021) Personalized treatment selection via the covariate-specific treatment effect curve for longitudinal data, Statistical Theory and Related Fields, 5:3, 253-264, DOI: 10.1080/24754269.2020.1762059

To link to this article: https://doi.org/10.1080/24754269.2020.1762059

Published online: 14 May 2020.

STATISTICAL THEORY AND RELATED FIELDS
2021, VOL. 5, NO. 3, 253–264
https://doi.org/10.1080/24754269.2020.1762059

Personalized treatment selection via the covariate-specific treatment effect curve for longitudinal data

Yanghui Liu[a], Riquan Zhang[a], Shujie Ma[b] and Xiuzhen Zhang[a]

[a]School of Statistics, East China Normal University, Shanghai, People's Republic of China; [b]Department of Statistics, University of California, Riverside, CA, USA

ABSTRACT

Treatment selection based on patient characteristics has been widely recognised in modern medicine. In this paper, we propose a generalised partially linear single-index mixed-effects modelling strategy for treatment selection and heterogeneous treatment effect estimation in longitudinal clinical and observational studies. We model the treatment effect as an unknown functional curve of a weighted linear combination of time-dependent covariates. This method enables us to investigate covariate-specific treatment effects and make personalised treatment selection in a flexible fashion. We develop a method that combines local linear regression and penalised quasi-likelihood to estimate the weight for each covariate, the unknown treatment effect curve and the parameters for mixed-effects. Based on pointwise confidence intervals for the treatment effect curve, we can make individualised treatment decisions from the information of patient characteristics. A simulation study is conducted to evaluate finite sample performance of the proposed method. We also illustrate the method via analysis of a real data example.

ARTICLE HISTORY
Received 13 November 2019
Revised 21 April 2020
Accepted 25 April 2020

KEYWORDS
personalized medicine; treatment selection; semiparametric model; longitudinal data

1. Introduction

Personalized or precision medicine, which aims to provide treatment strategies according to the characteristics of individuals or subgroups of the population, has gained much attention from biomedical researchers. The main goal of personalised medicine is to investigate covariate-specific (heterogeneous) effects of treatment, based on which individualised clinical decision can be better made to patients. In recent years, there has been an increasing amount of literature on heterogeneous treatment effect (HTE) estimation. A typical way of exploring heterogeneous treatment effects is to examine patient outcomes in mutually exclusive subgroups defined by observable patient characteristics, see Berger et al. (2014), Ciampi et al. (1995), Negassa et al. (2005), Foster et al. (2011), Su et al. (2008) and Wang et al. (2007). Bonetti and Gelber (2004) introduced a subpopulation treatment effect pattern plot (STEPP), which characterises treatment effects across potentially overlapping intervals of a continuous covariate. Bonetti et al. (2009) indicated that this method is effective only for large sample sizes, and proposed a permutation-based method for inference and achieved better performance for smaller sample sizes. The major limitation of such subgroup approach is that dichotomisation of continuous covariates can be artificial, and thus it may lose important information from the data.

In the literature, some other works apply nonparametric and semiparametric modelling methods to study HTE. These methods impose no or very relaxed assumptions on the model structure, and thus can explore HTE in a flexible fashion. For continuous response, Foster et al. (2015) proposed a two-stage procedure. They obtain nonparametric estimates of treatment effects for each subject in stage 1, and these estimates are used to identify optimal subset for a treatment, thus determine a treatment regime in stage 2. However, this method only applies to situations that the optimal subset is contiguous. For time-to-event data, Ma and Zhou (2014) defined a covariate-specific treatment effect (CSTE) curve, which is used to represent clinical utility of a continuous biomarker. They derived estimate of CSTE curve, and constructed pointwise confidence interval to select the optimal treatment for individual patient, as well as simultaneous confidence band to identify subpopulation who respond well to a treatment. Han et al. (2017) extended the method of Ma and Zhou (2014) to the case of binary response. Both of them only considered a single biomarker. In order to incorporate multivariate or even high-dimensional covariates, Guo et al. (2018) proposed a sparse logistic single-index coefficient model for optimal treatment selection using the CSTE curve. This method offers a flexible way for studying the CSTE curve without a restrictive assumption on the structure of the curve

CONTACT Riquan Zhang rqzhang@stat.ecnu.edu.cn School of Statistics, East China Normal University, Shanghai 200062, People's Republic of China

while achieving great dimension reduction of the high-dimensional covariates.

All the above methods were proposed for cross-sectional data with responses measured at one time point. In practice, the patient outcomes are often collected at multiple follow-up times in order to better evaluate the effectiveness of treatments. In this paper, we extend the model considered in Guo et al. (2018) to the longitudinal clinical and observational studies, and consider a generalised partially linear single-index coefficient mixed-effects model (GPLSIMM) for our longitudinal setting. Similar as Guo et al. (2018), we treat the treatment effect as an unknown functional curve of a weighted linear combination of time-varying covariates. The weights for covariates account for their different contributions to the treatment effect, and they are estimated from the data. In longitudinal studies, the repeated measures are correlated within subjects, and thus the estimating method considered in Guo et al. (2018) is not applicable to our proposed GPLSIMM. In our model, we need to estimate the unknown functional curve, the weight of each covariate and the parameters for the mixed-effects. Estimation of generalised semiparametric mixed-effects models has been considered in some existing works. Liang (2009) proposed a local linear regression and penalised quasi-likelihood method for estimation of a generalised partially linear mixed-effects model. Pang and Xue (2012) considered a local linear regression with GEE method for a single-index mixed effects model. Xu and Zhu (2012) and Chen et al. (2014) developed a kernel and a P-spline estimation method, respectively, together with quasi-lilikelhood for longitudinal generalised single-index models.

Based on the methods considered in these works, we see that penalised quasi-likelihood is a commonly used method for estimation of generalised mixed-effects models. It circumvents the calculation of high-dimensional integral in likelihood function (Breslow & Clayton, 1993). In our proposed GPLSIMM, we approximate the unknown treatment effect curve by the local linear method and estimate the parameters for the parametric and nonparametric parts through alternatively optimising the local and global penalised quasi-likelihood functions. We then select optimal treatment for a future patient based on pointwise confidence intervals for the CSTE curve.

The rest of the paper is organised as follows. In Section 2, we introduce the proposed model, the CSTE curve and the estimation method. Section 3 gives asymptotic properties of parametric and nonparametric estimates. Section 4 provides the algorithm for model estimation. In Section 5 we evaluate the finite sample properties of the proposed method via simulation studies, while Section 6 illustrates the application of the proposed method in a real data set. All technical proofs are relegated to Appendix.

2. Model and estimation

2.1. Model

Suppose our data are obtained from n independent subjects, and observations of the ith subject is $\{(Y_{ij}, X_{ij}, Z_{ij}, A_{ij}, D_{ij}),\ j = 1, \ldots, n_i\}$. Y_{ij} is the response, X_{ij}, Z_{ij} and A_{ij} are covariates of dimension p, q and s, and $D_{ij} \in \{0, 1\}$ is an indicator of exposure to treatment. We assume the relationship of response and covariates is specified by the following GPLSIMM:

$$\begin{aligned} &\mathrm{E}(Y_{ij}|X_{ij}, Z_{ij}, A_{ij}, D_{ij}, \boldsymbol{\gamma}_i) \\ &\quad = g^{-1}\{\boldsymbol{\alpha}^{\mathrm{T}} X_{ij} + f(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij}) D_{ij} + A_{ij}^{\mathrm{T}} \boldsymbol{\gamma}_i\}, \\ &\mathrm{var}(Y_{ij}|X_{ij}, Z_{ij}, A_{ij}, D_{ij}, \boldsymbol{\gamma}_i) \\ &\quad = w_{ij}^{-1} \phi V(\mu_{ij}), \quad i = 1, \ldots, n, j = 1, \ldots, n_i, \end{aligned} \tag{1}$$

where $\mu_{ij} = \mathrm{E}(Y_{ij}|X_{ij}, Z_{ij}, A_{ij}, D_{ij}, \boldsymbol{\gamma}_i)$, $g(\cdot)$ and $V(\cdot)$ are known functions, ϕ is a scale parameter and w_{ij} is a pre-determined weight for the jth observation of the ith subject. $f(\cdot)$ is an unknown function, and $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are unknown parameter vectors. For model identifiability, we assume that $\|\boldsymbol{\beta}\| = 1$ and $\beta_1 > 0$. We incorporate random effect $A_{ij}^{\mathrm{T}} \boldsymbol{\gamma}_i$ to account for the within-subject correlation, where $\boldsymbol{\gamma}_i$, $i = 1, \ldots, n$ are s-dimensional random effect vectors. We assume that $\boldsymbol{\gamma}_i \sim N(\mathbf{0}, \boldsymbol{G})$.

In this model, we characterise the treatment effect with a single-index term $f(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij})$, and the relationship of response and covariates for control group with a generalised linear mixed effect model. When response Y is binary, we can easily see that

$$\begin{aligned} f(\boldsymbol{\beta}^{\mathrm{T}} Z) = &\ \mathrm{logit}\{\mathrm{E}(Y|X, Z, A, \boldsymbol{\gamma}, D = 1)\} \\ &- \mathrm{logit}\{\mathrm{E}(Y|X, Z, A, \boldsymbol{\gamma}, D = 0)\}. \end{aligned} \tag{2}$$

When Y is a count variable and follows Poisson distribution,

$$\begin{aligned} f(\boldsymbol{\beta}^{\mathrm{T}} Z) = &\ \log\{\mathrm{E}(Y|X, Z, A, \boldsymbol{\gamma}, D = 1)\} \\ &- \log\{\mathrm{E}(Y|X, Z, A, \boldsymbol{\gamma}, D = 0)\}. \end{aligned} \tag{3}$$

Assume that $D = 0$ and $D = 1$ represents standard care and treatment respectively, X and Z are patient characteristics, Y indicates outcome of the study, and the better outcome corresponds to larger value of Y. From (2) and (3), $f(\boldsymbol{\beta}^{\mathrm{T}} Z)$ could be regarded as a measure of treatment effect. We define $f(\cdot)$ as covariate-specific treatment effect (CSTE) curve under this model.

Given the estimate and confidence interval of $f(\cdot)$, we could suggest the optimal treatment for a future patient based on his or her personal characteristics. Taking binary response as an example, we assume that $Y = 1$ represent a disease being cured, while $Y = 0$ indicate uncured. Let u_0 be the estimated value of $\boldsymbol{\theta}^{\mathrm{T}} Z$ calculated from personal characteristics of a patient. If the lower bound of confidence interval for $f(u_0)$ is greater than 0, the treatment is more effective than standard care for the patient. On the other hand, if the upper

bound of confidence interval is smaller than 0, standard care is more effective. In this case, we could not recommend the treatment to the patient. If neither of the above cases happen, that is, 0 is contained in the confidence interval, we would draw the conclusion that there's no significant difference between treatment and standard care.

2.2. Estimation

Denote $\boldsymbol{\Gamma} = (\gamma_1^{\mathrm{T}}, \ldots, \gamma_n^{\mathrm{T}})^{\mathrm{T}}$. Let $\boldsymbol{\beta}^{(1)} = (\beta_2, \ldots, \beta_q)^{\mathrm{T}}$ be the vector after deleting the first element from $\boldsymbol{\beta}$. Then $\boldsymbol{\beta} = (\sqrt{1 - \|\boldsymbol{\beta}^{(1)}\|^2}, \boldsymbol{\beta}^{(1)\mathrm{T}})^{\mathrm{T}}$, and we define the Jacobian matrix

$$\boldsymbol{J} = \boldsymbol{J}_{\boldsymbol{\beta}^{(1)}} = \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\beta}^{(1)\mathrm{T}}} = \begin{pmatrix} -\dfrac{\boldsymbol{\beta}^{(1)\mathrm{T}}}{\sqrt{1 - \|\boldsymbol{\beta}^{(1)}\|^2}} \\ \boldsymbol{I}_{q-1} \end{pmatrix}.$$

Our primary interest is to estimate $f(\cdot), \boldsymbol{\alpha}, \boldsymbol{\beta}$ and $\boldsymbol{G}$.

Suppose that $\boldsymbol{G}$ is known. Given $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, we combine penalised quasi-likelihood with local linear technique to obtain estimates of $f(\cdot)$ and $f'(\cdot)$. If u is in a neighbourhood of $\boldsymbol{\beta}^{\mathrm{T}} Z_{ij}$, $g(\mu_{ij})$ can be approximated by

$$g(\widetilde{\mu}_{ij}) = \boldsymbol{\alpha}^{\mathrm{T}} X_{ij} + \{f(u) + f'(u)(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij} - u)\} D_{ij} + A_{ij}^{\mathrm{T}} \boldsymbol{\gamma}_i.$$

Let $\mathrm{d}(y, \mu) = -2 \int_y^{\mu} ((y - u)/V(u))\, \mathrm{d}u$, $K_h(\cdot) = h^{-1} K(\cdot/h)$, where h is bandwidth, $K(\cdot)$ is a zero-mean symmetric density function. We maximise the following local penalised quasi-likelihood

$$-\frac{1}{2} \sum_{i=1}^{n} \left\{ \sum_{j=1}^{n_i} K_h(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij} - u) w_{ij} \phi^{-1} \mathrm{d}\big(Y_{ij}, g^{-1}[\boldsymbol{\alpha}^{\mathrm{T}} X_{ij} + \{a + b(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij} - u)\} D_{ij} + A_{ij}^{\mathrm{T}} \boldsymbol{\gamma}_i]\big) + \boldsymbol{\gamma}_i^{\mathrm{T}} \boldsymbol{G}^{-1} \boldsymbol{\gamma}_i \right\} \tag{4}$$

with respect to a, b and $\boldsymbol{\Gamma}$, and obtain estimates $\widehat{f}(u) = \widehat{a}$ and $\widehat{f}'(u) = \widehat{b}$. Taking derivative on (4) with respect to (a, b) and $\boldsymbol{\Gamma}$ yields the following estimating equations:

$$\sum_{i=1}^{n} \sum_{j=1}^{n_i} K_h(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij} - u) \{g'(\widetilde{\mu}_{ij})\}^{-1} w_{ij} \phi^{-1} V^{-1}(\widetilde{\mu}_{ij}) \times \Big(Y_{ij} - g^{-1}[\boldsymbol{\alpha}^{\mathrm{T}} X_{ij} + \{a + b(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij} - u)\} D_{ij} + A_{ij}^{\mathrm{T}} \boldsymbol{\gamma}_i]\Big) D_{ij} \begin{pmatrix} 1 \\ \boldsymbol{\beta}^{\mathrm{T}} Z_{ij} - u \end{pmatrix} = \boldsymbol{0}, \tag{5}$$

and for each i,

$$\sum_{j=1}^{n_i} K_h(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij} - u) A_{ij} \{g'(\widetilde{\mu}_{ij})\}^{-1} w_{ij} \phi^{-1} V^{-1}(\widetilde{\mu}_{ij}) \times \Big(Y_{ij} - g^{-1}[\boldsymbol{\alpha}^{\mathrm{T}} X_{ij} + \{a + b(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij} - u)\} D_{ij} + A_{ij}^{\mathrm{T}} \boldsymbol{\gamma}_i]\Big) - \boldsymbol{G}^{-1} \boldsymbol{\gamma}_i = \boldsymbol{0}. \tag{6}$$

When $f(\cdot)$ is known, we obtain estimate of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\Gamma}$ by maximising the global penalised quasi-likelihood

$$-\frac{1}{2} \sum_{i=1}^{n} \left(\sum_{j=1}^{n_i} w_{ij} \phi^{-1} d\big[Y_{ij}, g^{-1}\{\boldsymbol{\alpha}^{\mathrm{T}} X_{ij} + f(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij}) D_{ij} + A_{ij}^{\mathrm{T}} \boldsymbol{\gamma}_i\}\big] + \boldsymbol{\gamma}_i^{\mathrm{T}} \boldsymbol{G}^{-1} \boldsymbol{\gamma}_i \right), \tag{7}$$

with respect to $\boldsymbol{\alpha}$, $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\Gamma}$. The corresponding estimating equations are as follows:

$$\begin{aligned} &\boldsymbol{S}_{01}(\boldsymbol{\alpha}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\Gamma}) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n_i} X_{ij} \{g'(\mu_{ij})\}^{-1} w_{ij} \phi^{-1} V^{-1}(\mu_{ij}) \\ &\quad \times \Big[Y_{ij} - g^{-1}\{\boldsymbol{\alpha}^{\mathrm{T}} X_{ij} + f(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij}) D_{ij} + A_{ij}^{\mathrm{T}} \boldsymbol{\gamma}_i\}\Big] \\ &= \boldsymbol{0}, \end{aligned} \tag{8}$$

$$\begin{aligned} &\boldsymbol{S}_{02}(\boldsymbol{\alpha}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\Gamma}) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n_i} \boldsymbol{J}_{\boldsymbol{\beta}^{(1)}}^{\mathrm{T}} Z_{ij} D_{ij} f'(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij}) \{g'(\mu_{ij})\}^{-1} w_{ij} \phi^{-1} \\ &\quad + V^{-1}(\mu_{ij}) \Big[Y_{ij} - g^{-1}\{\boldsymbol{\alpha}^{\mathrm{T}} X_{ij} f(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij}) D_{ij} + A_{ij}^{\mathrm{T}} \boldsymbol{\gamma}_i\}\Big] \\ &= \boldsymbol{0}, \end{aligned} \tag{9}$$

and for each $i = 1, \ldots, n$,

$$\begin{aligned} &\boldsymbol{S}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\gamma}_i) \\ &= \sum_{j=1}^{n_i} A_{ij} \{g'(\mu_{ij})\}^{-1} w_{ij} \phi^{-1} V^{-1}(\mu_{ij}) \\ &\quad \times \Big[Y_{ij} - g^{-1}\{\boldsymbol{\alpha}^{\mathrm{T}} X_{ij} + f(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij}) D_{ij} + A_{ij}^{\mathrm{T}} \boldsymbol{\gamma}_i\}\Big] \\ &\quad - \boldsymbol{G}^{-1} \boldsymbol{\gamma}_i \\ &= \boldsymbol{0}. \end{aligned} \tag{10}$$

We obtain the parametric and nonparametric estimates by iteratively maximising quasi-likelihoods (4) and (7). The corresponding algorithm is summarised in Section 4 for practical implementation.

3. Asymptotic properties

Denote $q_k(t, y) = -\frac{1}{2}(\partial^k d\{y, g^{-1}(t)\}/\partial t^k)$, and $\rho_k(t) = \{\mathrm{d}g^{-1}(t)/\mathrm{d}t\}^k V^{-1}\{g^{-1}(t)\}$, $k = 1, 2, 3$. Let $\kappa_j = \int u^j K(u)\, \mathrm{d}u$, $\mu_j = \int u^j K^2(u)\, \mathrm{d}u$ for $j = 1, 2$, $N = \sum_{i=1}^{n} n_i$ and $N_1 = \sum_{i=1}^{n} n_i(n_i - 1)$. We assume n tends to infinity and n_i's are bounded. Denote $\eta_{ij} = \boldsymbol{\alpha}^{\mathrm{T}} X_{ij} + f(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij}) D_{ij} + A_{ij}^{\mathrm{T}} \boldsymbol{\gamma}_i$, and $\eta = \boldsymbol{\alpha}^{\mathrm{T}} X + f(\boldsymbol{\beta}^{\mathrm{T}} Z) D + A^{\mathrm{T}} \boldsymbol{\gamma}$.

In order to obtain asymptotic behaviours of the proposed parametric and nonparametric estimates, we assume the following regularity conditions:

(C1) The marginal density of Z is positive and uniformly continuous with a compact support $\mathcal{Z} \subseteq \mathbb{R}^q$. $f_{\mathcal{U}}(\cdot)$, density function of $U = \boldsymbol{\beta}^{\mathrm{T}} Z$, is twice continuously differentiable for $U \in \mathcal{U}$, where $\mathcal{U} = [b_1, b_2] = \{\boldsymbol{\beta}^{\mathrm{T}} Z, Z \in \mathcal{Z}\}$ is a compact interval.

(C2) For all x, $\rho_2(x) > 0$.

(C3) The second derivative of $g(\cdot)$, $V(\cdot)$ and $f(\cdot)$ are bounded and continuous.

(C4) The kernel $K(\cdot)$ is a bounded and symmetric probability density function with bounded support $[-c_0, c_0]$, and satisfies $\int u^2 K(u)\mathrm{d}u \neq 0$ and $\int u^2 K(u)\mathrm{d}u < \infty$. $K(\cdot)$ satisfies the Lipschitz condition of order 1.

(C5) $\mathrm{E}\{\rho_2(\eta)D^2|\boldsymbol{\beta}^{\mathrm{T}} Z = u\}$, $\mathrm{E}\{\rho_2(\eta)DX|\boldsymbol{\beta}^{\mathrm{T}} Z = u\}$ and $\mathrm{E}\{\rho_2(\eta)D^2 Z|\boldsymbol{\beta}^{\mathrm{T}} Z = u\}$ are twice differentiable on u. $\mathrm{E}\{q_1(\eta_{11}, Y_{11})q_1(\eta_{12}, Y_{12})D_{11}^2 D_{12}^2|\boldsymbol{\beta}^{\mathrm{T}} Z_{11} = u_1, \boldsymbol{\beta}^{\mathrm{T}} Z_{12} = u_2\}$ is continuous on u_1 and u_2.

(C6) As $n \to \infty$, h satisfies $nh^4 \to 0$ and $nh^2/\log(1/h) \to \infty$.

Denote

$$\widetilde{X} = X - \left[\mathrm{E}\{\rho_2(\eta)D^2|\boldsymbol{\beta}^{\mathrm{T}} Z\}\right]^{-1}\left[\mathrm{E}\{\rho_2(\eta)DX|\boldsymbol{\beta}^{\mathrm{T}} Z\}\right],$$

$$\widetilde{Z} = f'(\boldsymbol{\beta}^{\mathrm{T}} Z)\boldsymbol{J}^{\mathrm{T}}\left(Z - \left[\mathrm{E}\{\rho_2(\eta)D^2|\boldsymbol{\beta}^{\mathrm{T}} Z\}\right]^{-1} \times \left[\mathrm{E}\{\rho_2(\eta)D^2 Z|\boldsymbol{\beta}^{\mathrm{T}} Z\}\right]\right).$$

Let $\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}$ be the final estimator of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively. Theorem 3.1 gives the asymptotic distributions of $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$.

Theorem 3.1: *Under conditions* (C1)–(C6),

$$\sqrt{N}\begin{pmatrix}\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 \\ \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\end{pmatrix} \to N(0, \boldsymbol{\Sigma}), \tag{11}$$

where

$$\boldsymbol{\Sigma} = \mathcal{D}\left(\mathcal{A}^{-1} + \frac{N_1}{N}\mathcal{A}^{-1}\mathcal{C}\mathcal{A}^{-1}\right)\mathcal{D}^{\mathrm{T}},$$

with

$$\mathcal{A} = \mathrm{E}\left\{\rho_2(\eta)\begin{pmatrix}\widetilde{X} \\ \widetilde{Z}D\end{pmatrix}^{\otimes 2}\right\},$$

$$\mathcal{C} = \mathrm{cov}\left\{\begin{pmatrix}\widetilde{X}_{11} \\ \widetilde{Z}_{11}D_{11}\end{pmatrix} q_1(\eta_{11}, Y_{11}), \begin{pmatrix}\widetilde{X}_{12} \\ \widetilde{Z}_{12}D_{12}\end{pmatrix} q_1(\eta_{12}, Y_{12})\right\},$$

and

$$\mathcal{D} = \begin{pmatrix}\boldsymbol{I}_p & \boldsymbol{0}_{p\times(q-1)} \\ \boldsymbol{0}_{q\times p} & \boldsymbol{J}_{\boldsymbol{\beta}_0^{(1)}}\end{pmatrix}.$$

Theorem 3.1 indicates that in order to obtain $\sqrt{n}-$consistent estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, we need to undersmooth the nonparametric function $f(\cdot)$. For any $z \in \mathcal{Z}$, we let $\widehat{f}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}} z; \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$ be the estimate of $f(\boldsymbol{\beta}^{\mathrm{T}} z)$ given $\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}$ and h_1 be the corresponding bandwidth. Theorem 3.2 presents the asymptotic distribution of $\widehat{f}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}} z; \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$.

Theorem 3.2: *Under conditions* (C1)–(C5), *as* $n \to \infty$, $h_1 \to 0$ *and* $Nh_1 \to \infty$,

$$\sqrt{Nh}\{\widehat{f}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}} z; \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}) - f(\boldsymbol{\beta}_0^{\mathrm{T}} z) - \frac{1}{2}h^2\kappa_2 f''(\boldsymbol{\beta}_0^{\mathrm{T}} z)\} \to N(0, \sigma_f^2), \tag{12}$$

where $\sigma_f^2 = \mu_0 f_{\mathcal{U}}^{-1}(\boldsymbol{\beta}_0^{\mathrm{T}} z)[\mathrm{E}\{\rho_2(\eta)D^2|\boldsymbol{\beta}_0^{\mathrm{T}} Z = \boldsymbol{\beta}_0^{\mathrm{T}} z\}]^{-1}$.

4. Computation

Given $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, we solve estimating Equations (5) and (6) by Fisher's scoring algorithm (see Wu & Zhang, 2006, Section 10.4), and obtain estimates $\widehat{f}(\cdot)$ and $\widehat{f}'(\cdot)$.

When $f(\cdot)$ is given, we obtain estimates of $\boldsymbol{\alpha}, \boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\Gamma}$ by solving Equations (8), (9) and (10). Here we apply the quasi-Fisher scoring algorithm to this problem.

Denote

$$\boldsymbol{\theta} = (\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}, \quad \boldsymbol{\Theta} = (\boldsymbol{\theta}^{\mathrm{T}}, \boldsymbol{\Gamma}^{\mathrm{T}})^{\mathrm{T}},$$

$$\boldsymbol{\theta}^{(1)} = (\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\beta}^{(1)\mathrm{T}})^{\mathrm{T}}, \quad \boldsymbol{\Theta}^{(1)} = (\boldsymbol{\theta}^{(1)\mathrm{T}}, \boldsymbol{\Gamma}^{\mathrm{T}})^{\mathrm{T}},$$

$$\boldsymbol{S}(\boldsymbol{\Theta}^{(1)}) = \{\boldsymbol{S}_{01}^{\mathrm{T}}(\boldsymbol{\alpha}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\Gamma}), \boldsymbol{S}_{02}^{\mathrm{T}}(\boldsymbol{\alpha}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\Gamma}), \times \boldsymbol{S}_1^{\mathrm{T}}(\boldsymbol{\alpha}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\gamma}_1), \ldots, \boldsymbol{S}_n^{\mathrm{T}}(\boldsymbol{\alpha}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\gamma}_n)\}^{\mathrm{T}}.$$

Using the quasi-Fisher scoring algorithm, estimate of $\boldsymbol{\Theta}^{(1)}$ is updated by

$$\boldsymbol{\Theta}_{\mathrm{new}}^{(1)} = \boldsymbol{\Theta}_{\mathrm{old}}^{(1)} + \dot{\boldsymbol{S}}^{-1}(\boldsymbol{\Theta}_{\mathrm{old}}^{(1)})\boldsymbol{S}(\boldsymbol{\Theta}_{\mathrm{old}}^{(1)}), \tag{13}$$

where

$$\dot{\boldsymbol{S}}^{-1}(\boldsymbol{\Theta}_1^{(1)}) = \boldsymbol{Q} = \begin{pmatrix}\boldsymbol{Q}_{11} & \boldsymbol{Q}_{12} & \boldsymbol{Q}_{13} \\ \boldsymbol{Q}_{12}^{\mathrm{T}} & \boldsymbol{Q}_{22} & \boldsymbol{Q}_{23} \\ \boldsymbol{Q}_{13}^{\mathrm{T}} & \boldsymbol{Q}_{23}^{\mathrm{T}} & \boldsymbol{Q}_{33}\end{pmatrix}$$

with

$$\boldsymbol{Q}_{11} = \sum_{i=1}^{n}\sum_{j=1}^{n_i} X_{ij}\{g'(\mu_{ij})\}^{-2} w_{ij}\phi^{-1}V^{-1}(\mu_{ij})X_{ij}^{\mathrm{T}},$$

$$\boldsymbol{Q}_{12} = \sum_{i=1}^{n}\sum_{j=1}^{n_i} X_{ij}\{g'(\mu_{ij})\}^{-2} w_{ij}\phi^{-1}V^{-1}(\mu_{ij}) \times f'(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij})Z_{ij}^{\mathrm{T}}\boldsymbol{J}_{\boldsymbol{\beta}^{(1)}},$$

$$\boldsymbol{Q}_{22} = \sum_{i=1}^{n}\sum_{j=1}^{n_i} \boldsymbol{J}_{\boldsymbol{\beta}^{(1)}}^{\mathrm{T}} Z_{ij}D_{ij}^2\{f'(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij})\}^2\{g'(\mu_{ij})\}^{-2} \times w_{ij}\phi^{-1}V^{-1}(\mu_{ij})Z_{ij}^{\mathrm{T}}\boldsymbol{J}_{\boldsymbol{\beta}^{(1)}},$$

$$\boldsymbol{Q}_{13} = \left(\boldsymbol{Q}_{13}^{[1]}, \ldots, \boldsymbol{Q}_{13}^{[n]}\right), \quad \boldsymbol{Q}_{23} = \left(\boldsymbol{Q}_{23}^{[1]}, \ldots, \boldsymbol{Q}_{23}^{[n]}\right),$$
$$\boldsymbol{Q}_{33} = \operatorname{diag}\left(\boldsymbol{Q}_{33}^{[1]}, \ldots, \boldsymbol{Q}_{33}^{[n]}\right),$$

and for $i = 1, \ldots, n$,

$$\boldsymbol{Q}_{13}^{[i]} = \sum_{j=1}^{n_i} X_{ij}\{g'(\mu_{ij})\}^{-2} w_{ij}\phi^{-1} V^{-1}(\mu_{ij}) A_{ij}^{\mathrm{T}},$$
$$\boldsymbol{Q}_{23}^{[i]} = \sum_{j=1}^{n_i} \boldsymbol{J}_{\boldsymbol{\beta}^{(1)}}^{\mathrm{T}} Z_{ij} D_{ij} f'(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij})\{g'(\mu_{ij})\}^{-2} \times w_{ij}\phi^{-1} V^{-1}(\mu_{ij}) A_{ij}^{\mathrm{T}},$$
$$\boldsymbol{Q}_{33}^{[i]} = \sum_{j=1}^{n_i} A_{ij}\{g'(\mu_{ij})\}^{-2} w_{ij}\phi^{-1} V^{-1}(\mu_{ij}) A_{ij}^{\mathrm{T}} + G^{-1}.$$

Let

$$\boldsymbol{P}_i = \begin{pmatrix} \{g'(\mu_{i1})\}^{-1} X_{i1}^{\mathrm{T}} & \{'(\mu_{i1})\}^{-1} f'(\boldsymbol{\beta}^{\mathrm{T}} Z_{i1}) D_{i1} Z_{i1}^{\mathrm{T}} \boldsymbol{J}_{\boldsymbol{\beta}^{(1)}} \\ \vdots & \vdots \\ \{g'(\mu_{in_i})\}^{-1} X_{in_i}^{\mathrm{T}} & \{g'(\mu_{in_i})\}^{-1} f'(\boldsymbol{\beta}^{\mathrm{T}} Z_{in_i}) D_{in_i} Z_{in_i}^{\mathrm{T}} \boldsymbol{J}_{\boldsymbol{\beta}^{(1)}} \end{pmatrix},$$
$$\boldsymbol{A}_i = \begin{pmatrix} \{g'(\mu_{i1})\}^{-1} A_{i1}^{\mathrm{T}} \\ \vdots \\ \{g'(\mu_{in_i})\}^{-1} A_{in_i}^{\mathrm{T}} \end{pmatrix},$$
$$\boldsymbol{R}_i = \operatorname{diag}\left(w_{i1}^{-1}\phi V(\mu_{i1}), \ldots, w_{in_i}^{-1}\phi V(\mu_{in_i})\right),$$
$$\boldsymbol{\varepsilon}_i = \left(Y_{i1} - \mu_{i1}, \ldots, Y_{in_i} - \mu_{in_i}\right)^{\mathrm{T}}.$$

We further denote

$$\boldsymbol{P} = \left(\boldsymbol{P}_1^{\mathrm{T}}, \ldots, \boldsymbol{P}_n^{\mathrm{T}}\right)^{\mathrm{T}}, \quad \boldsymbol{A} = \operatorname{diag}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n),$$
$$\boldsymbol{\varepsilon} = \left(\boldsymbol{\varepsilon}_1^{\mathrm{T}}, \ldots, \boldsymbol{\varepsilon}_n^{\mathrm{T}}\right)^{\mathrm{T}},$$
$$\boldsymbol{R} = \operatorname{diag}(\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n), \quad \widetilde{\boldsymbol{G}} = \operatorname{diag}(\boldsymbol{G}, \ldots, \boldsymbol{G}).$$

Equation (13) can be rewritten as

$$\boldsymbol{\Theta}_{\text{new}}^{(1)} - \boldsymbol{\Theta}_{\text{old}}^{(1)} = \begin{pmatrix} \boldsymbol{P}_{\text{old}}^{\mathrm{T}} \mathbf{R}_{\text{old}}^{-1} \boldsymbol{P}_{\text{old}} & \boldsymbol{P}_{\text{old}}^{\mathrm{T}} \mathbf{R}_{\text{old}}^{-1} \boldsymbol{A}_{\text{old}} \\ \boldsymbol{A}_{\text{old}}^{\mathrm{T}} \mathbf{R}_{\text{old}}^{-1} \boldsymbol{P}_{\text{old}} & \boldsymbol{A}_{\text{old}}^{\mathrm{T}} \mathbf{R}_{\text{old}}^{-1} \boldsymbol{A}_{\text{old}} + \widetilde{\boldsymbol{G}}^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{P}_{\text{old}}^{\mathrm{T}} \mathbf{R}_{\text{old}}^{-1} \boldsymbol{\varepsilon}_{\text{old}} \\ \boldsymbol{A}_{\text{old}}^{\mathrm{T}} \mathbf{R}_{\text{old}}^{-1} \boldsymbol{\varepsilon}_{\text{old}} - \widetilde{\boldsymbol{G}}^{-1} \boldsymbol{\Gamma}_{\text{old}} \end{pmatrix}.$$

Note that $\boldsymbol{P}_{\text{old}}, \boldsymbol{A}_{\text{old}}, \mathbf{R}_{\text{old}}$ and $\boldsymbol{\varepsilon}_{\text{old}}$ on the right hand of above equation are calculated from $\boldsymbol{\Theta}_{\text{old}}$. For simplicity we suppress their subscripts hereinafter. Denote $\boldsymbol{H}_i = \boldsymbol{A}_i \boldsymbol{G} \boldsymbol{A}_i^{\mathrm{T}} + \mathbf{R}_i, \boldsymbol{H} = \operatorname{diag}(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_n)$. By matrix algebra, we can obtain

$$\begin{aligned} \boldsymbol{\theta}_{\text{new}}^{(1)} &= \boldsymbol{\theta}_{\text{old}}^{(1)} + (\boldsymbol{P}^{\mathrm{T}} \boldsymbol{H}^{-1} \boldsymbol{P})^{-1} \boldsymbol{P}^{\mathrm{T}} \boldsymbol{H}^{-1} (\boldsymbol{\varepsilon} + \boldsymbol{A} \boldsymbol{\Gamma}_{\text{old}}), \\ \boldsymbol{\gamma}_{i,\text{new}} &= \boldsymbol{G} \boldsymbol{A}_i^{\mathrm{T}} \boldsymbol{H}_i^{-1} \left\{ \boldsymbol{\varepsilon}_i + \boldsymbol{A}_i \boldsymbol{\gamma}_{i,\text{old}} - \boldsymbol{P}_i (\boldsymbol{\theta}_{\text{new}}^{(1)} - \boldsymbol{\theta}_{\text{old}}^{(1)}) \right\}, \\ & \qquad i = 1, \ldots, n. \end{aligned} \tag{14}$$

Based on $\|\boldsymbol{\beta}_{\text{new}}\| = 1$ and $\|\boldsymbol{\beta}_{\text{old}}\| = 1$, we can show that

$$\boldsymbol{\theta}_{\text{new}} - \boldsymbol{\theta}_{\text{old}} = \begin{pmatrix} \boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{J}_{\boldsymbol{\beta}_{\text{old}}^{(1)}} \end{pmatrix} (\boldsymbol{\theta}_{\text{new}}^{(1)} - \boldsymbol{\theta}_{\text{old}}^{(1)})\{1 + o_p(1)\}.$$

Thus, we update $\boldsymbol{\theta}_{\text{old}}$ by $\boldsymbol{\theta}_{\text{new}} = (\boldsymbol{\alpha}_{\text{new}}^{*\mathrm{T}}, \boldsymbol{\beta}_{\text{new}}^{*\mathrm{T}} / \|\boldsymbol{\beta}_{\text{new}}^{*}\|)^{\mathrm{T}}$, with $\boldsymbol{\theta}_{\text{new}}^{*} = (\boldsymbol{\alpha}_{\text{new}}^{*\mathrm{T}}, \boldsymbol{\beta}_{\text{new}}^{*\mathrm{T}})^{\mathrm{T}}$, and

$$\boldsymbol{\theta}_{\text{new}}^{*} = \boldsymbol{\theta}_{\text{old}} + \begin{pmatrix} \boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{J}_{\boldsymbol{\beta}_{\text{old}}^{(1)}} \end{pmatrix} (\boldsymbol{P}^{\mathrm{T}} \boldsymbol{H}^{-1} \boldsymbol{P})^{-1} \times \boldsymbol{P}^{\mathrm{T}} \boldsymbol{H}^{-1} (\boldsymbol{\varepsilon} + \boldsymbol{A} \boldsymbol{\Gamma}_{\text{old}}).$$

Our algorithm can be summarised as following:

Step 1. Obtain initial estimate $\widehat{\boldsymbol{\theta}}_{\text{init}}$ by solving generalised linear mixed effect model

$$\begin{aligned} \mathrm{E}(Y_{ij} | X_{ij}, Z_{ij}, D_{ij}) &= g^{-1}(\boldsymbol{\alpha}^{\mathrm{T}} X_{ij} + \boldsymbol{\beta}^{\mathrm{T}} Z_{ij} D_{ij} + A_{ij}^{\mathrm{T}} \boldsymbol{\gamma}_i) \\ \operatorname{var}(Y_{ij} | X_{ij}, Z_{ij}, D_{ij}) &= w_{ij}^{-1} \phi V(\mu_{ij}). \end{aligned}$$

Step 2. Given parametric estimate $\widehat{\boldsymbol{\theta}}_{\text{old}}$, derive estimates $\widehat{f}(\widehat{\boldsymbol{\beta}}_{\text{old}}^{\mathrm{T}} Z_{ij})$ and $\widehat{f}'(\widehat{\boldsymbol{\beta}}_{\text{old}}^{\mathrm{T}} Z_{ij})$ by solving estimating Equations (5) and (6).

Step 3. Obtain $\widehat{\boldsymbol{\theta}}_{\text{new}}$ and $\widehat{\boldsymbol{\Gamma}}_{\text{new}}$ using Equation (14).

Step 4. Iterate between Steps 2 and 3 until convergence, and obtain the final estimate $\widehat{\boldsymbol{\theta}}$, $\widehat{\boldsymbol{\Gamma}}$ and $\widehat{f}(\cdot)$. We further calculate the final estimate $\widehat{f}(\cdot)$ by Step 2.

Note that matrix $\boldsymbol{G}$ is still unknown. To obtain estimate of $\boldsymbol{G}$, we apply the maximum likelihood method under the normality assumption, and implement the method by EM algorithm (see Laird & Ware, 1982). To be specific, we set the initial value of $\boldsymbol{G}$ to be the identity matrix. After obtaining the estimates of $\boldsymbol{\theta}, \boldsymbol{\Gamma}$ and $f(\cdot)$ by Steps 1–4, we have an updated estimate of $\boldsymbol{G}$. We repeat this procedure until convergence, and obtain the final estimate $\widehat{\boldsymbol{G}}$. Given the estimate $\widehat{\boldsymbol{G}}$, we derive the estimates of $\boldsymbol{\theta}, \boldsymbol{\Gamma}$ and $f(\cdot)$ by substitution of $\widehat{\boldsymbol{G}}$ for $\boldsymbol{G}$ in Steps 2 and 3.

During the process of implementation, bandwidth h in Step 2 also needs to be selected. Theorem 3.1 indicates that undersmoothing the nonparametric part is necessary to guarantee $\sqrt{n}$ consistency of parametric estimates. We adopt the ad hoc method in Carroll et al. (1997) to select appropriate bandwidth for Step 2.

Table 1. Simulation results for parametric estimates with binary data.

			GPLSIMM			GPLSIM		
n	Parameters	True value	Mean	Sd	MSE $\times$ 10	Mean	Sd	MSE $\times$ 10
100	α_1	0.500	0.438	0.124	0.192	0.442	0.123	0.184
	α_2	1.000	0.894	0.126	0.271	0.884	0.128	0.297
	α_3	−0.500	−0.439	0.119	0.178	−0.425	0.117	0.193
	β_1	0.707	0.694	0.092	0.086	0.696	0.107	0.116
	β_2	0.707	0.709	0.089	0.080	0.702	0.106	0.113
300	α_1	0.500	0.447	0.067	0.074	0.430	0.069	0.096
	α_2	1.000	0.892	0.076	0.175	0.859	0.075	0.255
	α_3	−0.500	−0.450	0.067	0.070	−0.430	0.066	0.093
	β_1	0.707	0.702	0.051	0.027	0.701	0.055	0.031
	β_2	0.707	0.708	0.051	0.026	0.709	0.054	0.029
500	α_1	0.500	0.447	0.055	0.058	0.429	0.050	0.076
	α_2	1.000	0.892	0.059	0.151	0.857	0.058	0.239
	α_3	−0.500	−0.446	0.052	0.056	−0.427	0.051	0.080
	β_1	0.707	0.707	0.042	0.018	0.706	0.044	0.020
	β_2	0.707	0.705	0.042	0.018	0.706	0.045	0.020

Table 2. Simulation results for parametric estimates with count data.

			GPLSIMM			GPLSIM		
n	Parameters	True value	Mean	Sd	MSE $\times$ 10^2	Mean	Sd	MSE $\times$ 10^2
100	α_1	0.500	0.508	0.048	0.236	0.505	0.129	1.671
	α_2	1.000	1.020	0.047	0.262	1.004	0.138	1.908
	α_3	−0.500	−0.508	0.054	0.302	−0.511	0.134	1.800
	β_1	0.707	0.703	0.063	0.395	0.706	0.082	0.666
	β_2	0.707	0.703	0.089	0.790	0.685	0.164	2.757
300	α_1	0.500	0.512	0.021	0.060	0.504	0.093	0.860
	α_2	1.000	1.023	0.024	0.108	1.016	0.096	0.954
	α_3	−0.500	−0.511	0.023	0.068	−0.505	0.090	0.815
	β_1	0.707	0.708	0.017	0.030	0.706	0.048	0.234
	β_2	0.707	0.706	0.021	0.046	0.705	0.049	0.242
500	α_1	0.500	0.512	0.015	0.037	0.516	0.073	0.560
	α_2	1.000	1.028	0.018	0.108	1.015	0.079	0.642
	β_1	−0.500	−0.512	0.016	0.041	−0.515	0.076	0.596
	β_2	0.707	0.706	0.011	0.013	0.708	0.039	0.156
	β_3	0.707	0.708	0.011	0.012	0.704	0.039	0.153

5. Simulation

In this section, we assess the finite sample performance of the proposed method via two Monte Carlo simulations with binary and count responses respectively.

In these simulations,

$$\eta_{ij} = \boldsymbol{\alpha}^{\mathrm{T}} X_{ij} + f(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij}) D_{ij} + A_{ij}^{\mathrm{T}} \boldsymbol{\gamma}_i,$$
$$i = 1, \ldots, n, j = 1, \ldots, n_i,$$

where $\boldsymbol{\alpha} = (0.5, 1, -0.5)^{\mathrm{T}}$, $\boldsymbol{\beta} = (1, 1)^{\mathrm{T}}/\sqrt{2}$ and $f(u) = 3\sin\{\pi(u - c_1)/(c_2 - c_1)\}$, with $c_1 = \frac{\sqrt{2}}{2} - \frac{1.645}{\sqrt{12}}$, and $c_2 = \frac{\sqrt{2}}{2} + \frac{1.645}{\sqrt{12}}$. We assume the random effect be random intercept, that is, $A \equiv 1$ and $\boldsymbol{\gamma}$ is one-dimensional. We let $\boldsymbol{\gamma}$ follow $N(0, 1)$. For the covariates, X is distributed as $X = (X_1, X_2, X_3)^{\mathrm{T}}$, where X_1, X_2 and X_3 follows $N(0, 1)$, $Z = (Z_1, Z_2)^{\mathrm{T}}$, where Z_1, Z_2 are uniform $U(0, 1)$ variables, and the exposure indicator D follows Bernoulli(0.5). We assume n_i, the number of observations for each subject, is 5, and set the number of subjects n to be 100, 300 and 500. For the response variable, we consider the following two cases:

Case 1 (Binary data):

$$Y_{ij} \sim \text{Bernoulli}(\mu_{ij}), \quad \text{where} \quad \text{logit}(\mu_{ij}) = \eta_{ij};$$

Case 2 (Count data):

$$Y_{ij} \sim \text{Poisson}(\mu_{ij}), \quad \text{where} \quad \log(\mu_{ij}) = \eta_{ij}.$$

All simulations are repeated for 500 times. For comparison, we ignore the random effects and apply generalised partially linear single-index model (GPLSIM) to the simulated data. The corresponding estimates are obtained using the method in Xu and Zhu (2012). The parametric estimates for binary and count data are summarised in Tables 1 and 2 respectively. We assess the performance of nonparametric estimator $\widehat{f}(\cdot)$ by mean integrated squared error (MISE) defined as

$$\text{MISE}(\widehat{f}) = \text{E} \int \{\widehat{f}(u) - f(u)\}^2 \, \mathrm{d}u.$$

Tables 3 and 4 shows mean and standard deviation of MISE for nonparametric function.

From Tables 1 and 2, MSE of the parametric estimates under our model are generally smaller than that of ignoring the random effects. As the sample size increases, the performance of parametric estimates improves. Similar results hold for nonparametric estimates. The MISE of nonparametric estimates using our method is smaller than those derived under GPLSIM. MISE decreases as the sample size increases.

Table 3. MISE for nonparametric estimate with binary data.

	GPLSIMM		GPLSIM	
n	Mean	Sd	Mean	Sd
100	0.287	0.182	0.291	0.186
300	0.155	0.085	0.163	0.085
500	0.106	0.052	0.112	0.052

Table 4. MISE for nonparametric estimate with count data.

	GPLSIMM		GPLSIM	
n	Mean	Sd	Mean	Sd
100	0.191	0.135	0.243	0.172
300	0.169	0.090	0.193	0.111
500	0.168	0.062	0.191	0.082

Table 5. Estimated $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

Estimated $\boldsymbol{\alpha}$	Estimate	Sd
Intercept	6.450	2.754
Visit year	0.029	0.026
Age	0.029	0.019
EDUC	−0.012	0.037
DIABETES	0.062	0.416
DEP	0.454	0.179
MMSE	−0.384	0.095
Estimated $\boldsymbol{\beta}$	Estimate	Sd
MMSE	0.842	0.115
Age	−0.540	0.304

6. Real data analysis

We apply the proposed method to the US National Alzhemer's Coordinating Center Uniform Data Set (https://www.alz.washington.edu). Our goal is to investigate the effect of heredity on development of Alzheimer's disease (AD) among women. We take the diagnosis of AD in each observation of patients (yes/no) as response. The covariates that may influence the occurrence of AD include age, visit year, years of education (EDUC), indicator of first-degree family member with cognitive impairment (yes/no, FAM), depression (yes/no, DEP), diabetes (yes/no, DIABETES) and minimental state exam (MMSE) score. To avoid large computational burden including all the observations, we randomly select 500 subjects with at least 2 follow-up visits from the original data set. Our final sample includes 2491 observations and the median follow-up is 3. The visit year is between 2005 and 2019. The proportion of occurence of AD is 28.3% among all observations, and the proportion of subjects whose family member has cognitive impairment is 64.2%. Note that we repeated the sample selection for several times, and achieved consistent results.

Taking FAM as indicator of treatment, we apply the proposed model and method to the final data set. We include logarithm of age and MMSE score into Z, and intercept, age, visit year, EDUC, DEP, DIABETES and MMSE score into X. The estimate of treatment effect reflects the influence of family heredity on development of AD. Note that in this data set, FAM is not a real treatment indicator, thus the corresponding CSTE curve does not indicate a real treatment effect. This data set is only used to for illustrative purpose.

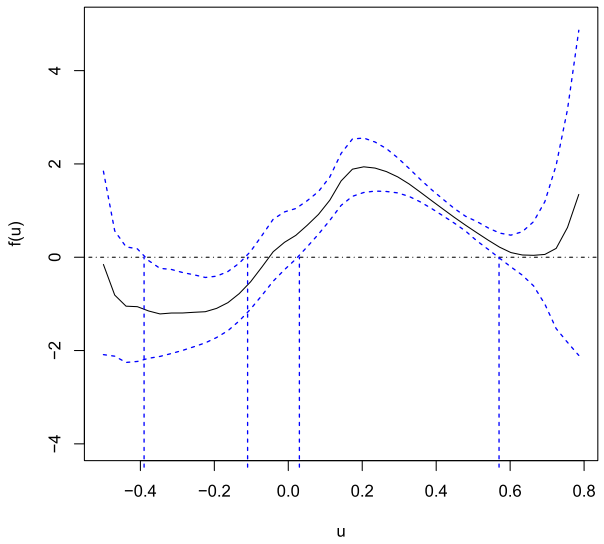

Figure 1. Estimate of nonparametric function $f(\cdot)$ (real line) and its 95% pointwise confidence interval (dashed line).

Table 5 shows the parametric estimates. The standard deviations of parameters are calculated from 500 bootstrap samples. Figure 1 presents estimate of nonparametric function and the corresponding 95% pointwise confidence interval. We construct the confidence interval based on result of Theorem 3.2, and apply the method of Zhang and Peng (2010) to estimate the bias and variance. Figure 2 displays the curve of treatment effect versus MMSE score with age fixed on its mean value, as well as the curve of treatment effect versus age with MMSE score fixed on the mean value.

Chen and Zhou (2011) used a generalised linear model to investigate risk factors that influence the occurrence of AD. From their results, age, DEP and MMSE score are significant under four estimation methods. They found that age and DEP have positive associations with the occurence of AD, and MMSE score is negatively correlated with the development of AD. From Table 5, our result of parametric estimates is consistent with these findings. From Figure 1, we can clearly see that if the estimated index of a patient is between 0.03 and 0.57, family inheritance of congnitive impairment increases the risk of getting AD. However, if the estimated index is between −0.39 and −0.11, family heredity decreases the risk of AD. Figure 2 shows that the effect of family heredity on occurence of AD has a bell shape association with MMSE score, that is, heredity has lower or even negative influence on risk of AD when the value of MMSE score is small or large. Also, the effect has an increasing trend with age. The effect of family heredity on occurrence of AD is stronger for elderly people.

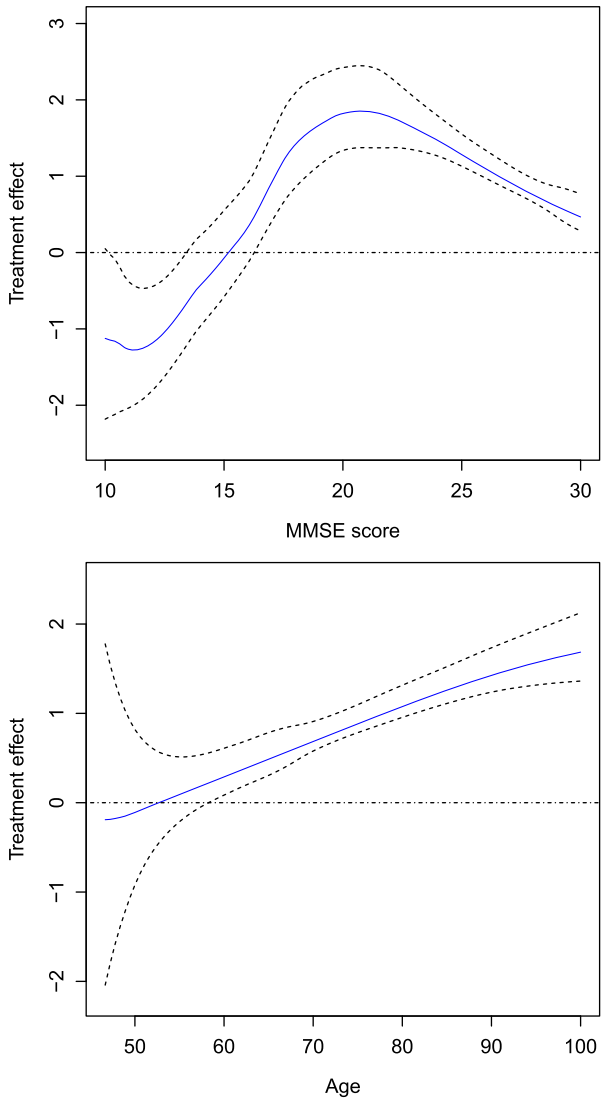

Figure 2. Relationship of treatment effect with MMSE score and Age: the estimated curve represented by real line and the corresponding 95% pointwise confidence interval by dashed line.

7. Discussion

This paper focuses on a generalised partially linear single-index mixed effects model for personalised treatment effect estimation and treatment selection in longitudinal studies. In our model, the treatment effect is described as a function of a linear combination of covariates. We develop a method combining local linear regression and penalised quasi-likelihood to estimate the coefficients for each covariate, the treatment effect curve and the parameters for mixed effects. Based on the pointwise confidence intervals for treatment effect curve, we can make individualised treatment decisions from the information of patient characteristics. Our simulation study and real data analysis illustrate effectiveness of the proposed method.

Nonparametric and semiparametric methods provide a flexible way to explore HTE. The previous research in this area mostly focus on cross-sectional data. Our work fills in the gap of semiparametric modelling of HTE with longitudinal data. On the other hand, we develop a new estimation method combining local linear technique and penalised quasi-likelihood, for generalised partially linear single-index model. Pointwise confidence interval can be directly constructed for the estimated treatment effect curve based on its asymptotic normality. The theory of simultaneous confidence band for treatment effect curve can also be established accordingly, and we leave this for future work.

There are still some limitations in our work. We directly apply the method of Zhang and Peng (2010) to estimate bias and variance for the treatment effect curve. It would be more rigorous that the performance of this method in our context be validated via simulation studies. We will include this in our future research. Another limitation of our work is that, based on the pointwise confidence interval of treatment effect curve, we could only make treatment decision for a future patient. To identify subgroup of patients that benefit from each treatment, it is necessary to construct simultaneous confidence band for treatment effect curve. Some possible extensions of our work could also be considered in future research. Our model could be extended to high-dimensional covariates to cope with longitudinal studies in which large number of patient characteristics are recorded. It is also of interest to consider robust regression to limit the impact of outlying observations. An even further extension is a survival model for time-to-event response.

Acknowledgments

We are grateful to the Editor-in-Chief, Prof. Jun Shao, an Associate Editor and anonymous reviewers for their thorough reading of our manuscript and insightful comments that have led to significant improvement of this work. We also thank the US National Alzheimer's Coordinating Center for providing the data.

Funding

This work was supported by Natural Science Foundation of Shanghai (17ZR1409000). The NACC database is funded by NIA/NIH Grant U01 AG016976. NACC data are contributed by the NIA -funded ADCs: P30 AG019610 (PI Eric Reiman, MD), P30 AG013846 (PI Neil Kowall, MD), P30 AG062428-01 (PI James Leverenz, MD) P50 AG008702 (PI Scott Small, MD), P50 AG025688 (PI Allan Levey, MD, PhD), P50 AG047266 (PI Todd Golde, MD, PhD), P30 AG010133 (PI Andrew Saykin, PsyD), P50 AG005146 (PI Marilyn Albert, PhD), P30 AG062421-01 (PI Bradley Hyman, MD, PhD), P30 AG062422-01 (PI Ronald Petersen, MD, PhD), P50 AG005138 (PI Mary Sano, PhD), P30 AG008051 (PI Thomas Wisniewski, MD), P30 AG013854 (PI Robert Vassar, PhD), P30 AG008017 (PI Jeffrey Kaye, MD), P30 AG010161 (PI David Bennett, MD), P50 AG047366 (PI Victor Henderson, MD, MS), P30 AG010129 (PI Charles DeCarli, MD), P50 AG016573 (PI Frank LaFerla, PhD), P30 AG062429-01 (PI James Brewer, MD, PhD), P50 AG023501

(PI Bruce Miller, MD), P30 AG035982 (PI Russell Swerdlow, MD), P30 AG028383 (PI Linda Van Eldik, PhD), P30 AG053760 (PI Henry Paulson, MD, PhD), P30 AG010124 (PI John Trojanowski, MD, PhD), P50 AG005133 (PI Oscar Lopez, MD), P50 AG005142 (PI Helena Chui, MD), P30 AG012300 (PI Roger Rosenberg, MD), P30 AG049638 (PI Suzanne Craft, PhD), P50 AG005136 (PI Thomas Grabowski, MD), P30 AG062715-01 (PI Sanjay Asthana, MD, FRCP), P50 AG005681 (PI John Morris, MD), P50 AG047270 (PI Stephen Strittmatter, MD, PhD) .

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Yanghui Liu is a Ph.D. candidate in Statistics at East China Normal University.

Riquan Zhang is a Professor in School of Statistics at East China Normal University.

Shujie Ma is an Associate Professor in Department of Statistics at University of California, Riverside.

Xiuzhen Zhang is a Ph.D. candidate in Statistics at East China Normal University.

References

Berger, J. O., Wang, X., & Shen, L. (2014). A Bayesian approach to subgroup identification. *Journal of Biopharmaceutical Statistics*, *24*(1), 110–129. https://doi.org/10.1080/10543406.2013.856026

Bonetti, M., & Gelber, R. (2004). Patterns of treatment eects in subsets of patients in clinical trials. *Biostatistics*, *5*(3), 465–481. https://doi.org/10.1093/biostatistics/kxh002

Bonetti, M., Zahrieh, D., Cole, B. F., & Gelber, R. D. (2009). A small sample study of the stepp approach to assessing treatment-covariate interactions in survival data. *Statistics in Medicine*, *28*(8), 1255–1268. https://doi.org/10.1002/sim.v28:8

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, *88*(21), 9–25. https://doi.org/10.1080/01621459.1993.10594284.

Carroll, R. J., Fan, J., Gijbels, I., & Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, *92*(438), 477–489. https://doi.org/10.1080/01621459.1997.10474001

Chen, J., Kim, I., Terrell, G. R., & Liu, L. (2014). Generalized partially linear single-index mixed model for repeated measures data. *Journal of Nonparametric Statistics*, *26*(2), 291–303. https://doi.org/10.1080/10485252.2014.891029

Chen, B., & Zhou, X. (2011). Doubly robust estimates for binary longitudinal data analysis with missing response and missing covariates. *Biometrics*, *67*(3), 830–842. https://doi.org/10.1111/biom.2011.67.issue-3

Ciampi, A., Negassa, A., & Lou, Z. (1995). Tree-structured prediction for censored survival data and the cox model. *Journal of Clinical Epidemiology*, *48*(5), 675–689. https://doi.org/10.1016/0895-4356(94)00164-L

Cui, X., Härdle, W. K., & Zhu, L. (2011). The EFM approach for single-index models. *The Annals of Statistics*, *39*(3), 1658–2688. https://projecteuclid.org/euclid.aos/1311600279

Foster, J. C., Taylor, J. M. G., Kaciroti, N., & Nan, B. (2015). Simple subgroup approximations to optimal treatment regimes from randomized clinical trial data. *Biostatistics*, *16*(2), 368–382. https://doi.org/10.1093/biostatistics/kxu049

Foster, J. C., Taylor, J. M.G., & Ruberg, S. J. (2011). Subgroup identification from randomized clinical trial data. *Statistics in Medicine*, *30*(24), 2867–2880. https://doi.org/10.1002/sim.v30.24

Guo, W., Zhou, X., & Ma, S. (2018). Optimal treatment selection using the covariate-specific treatment effect curve with high-dimensional covariates, arXiv:1812.10018.

Han, K., Zhou, X., & Liu, B. (2017). CSTE curve for selection the optimal treatment when outcome is binary. *Scientia Sinica (Mathematica)*, *47*(4), 497–514. https://doi.org/10.1360/SCM-2015-0595

Laird, N. M., & Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics*, *38*(4), 963–974. https://doi.org/10.2307/2529876

Liang, H. (2009). Generalized partially linear mixed-effects models incorporating mismeasured covariates. *Annals of the Institute of Statistical Mathematics*, *61*(1), 27–46. https://doi.org/10.1007/s10463-007-0146-0

Ma, Y., & Zhou, X. (2014). Treatment selection in a randomized clinical trial via covariate-specific treatment effect curves. *Statistical Methods in Medical Research*, *26*(1), 124–141. https://doi.org/10.1177/0962280214541724

Negassa, A., Ciampi, A., Abrahamowicz, M., Shapiro, S., & Boivin, J. (2005). Tree-structured subgroup analysis for censored survival data: validation of computationally inexpensive model selection criteria. *Statistics and Computing*, *15*(3), 231–239. https://doi.org/10.1007/s11222-005-1311-z

Pang, Z., & Xue, L. (2012). Estimation for the single-index models with random effects. *Computational Statistics & Data Analysis*, *56*(6), 1837–1853. https://doi.org/10.1016/j.csda.2011.11.007

Su, X., Tsai, C., Wang, H., Nickerson, D., & Li, B. (2008). Subgroup analysis via recursive partitioning. *Journal of Machine Learning Research*, *10*, 141–158. https://dl.acm.org/doi/10.5555/1577069.1577074.

Wang, R., Lagakos, S. W., Ware, J. H., Hunter, D. J., & Drazen, J. M. (2007). Statistics in medicine – Reporting of subgroup analyses in clinical trials. *The New England Journal of Medicine*, *357*(21), 2189–2194. https://doi.org/10.1056/NEJMsr077003

Wu, H., & Zhang, J. T. (2006). *Nonparametric regression methods for longitudinal data analysis*. John Wiley & Sons.

Xu, P., & Zhu, L. X. (2012). Estimation for a marginal generalized single-index longitudinal model. *Journal of Multivariate Analysis*, *105*(1), 285–299. https://doi.org/10.1016/j.jmva.2011.10.004

Zhang, W., & Peng, H. (2010). Simultaneous confidence band and hypothesis test in generalised varying-coefficient models. *Journal of Multivariate Analysis*, *101*(7), 1656–1680. https://doi.org/10.1016/j.jmva.2010.03.003

Appendix

Proof of Theorem 3.1: We use a similar proof strategy with Xu and Zhu (2012). Our proof is divided to three steps.

Step 1: Given $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$, we derive the asymptotic expansion of $\widehat{f}(u) - f(u)$.

Let $c_n = (Nh)^{-1/2}, \delta_n = c_n^2 \log^{1/2}(1/h)$,

$$\Lambda = \begin{pmatrix} c_n^{-1}\{a - f(u)\} \\ c_n^{-1} h\{b - f'(u)\} \end{pmatrix}, \quad \boldsymbol{F}_{ij} = \begin{pmatrix} D_{ij} \\ h^{-1}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_{ij} - u)D_{ij} \end{pmatrix},$$

$$\widehat{\Lambda} = \begin{pmatrix} c_n^{-1}\{\widehat{f}(u) - f(u)\} \\ c_n^{-1}h\{\widehat{f}'(u) - f(u)\} \end{pmatrix}.$$

Denote $\bar{\eta}_{ij} = \boldsymbol{\alpha}^{\mathrm{T}}X_{ij} + \{f(u) + f'(u)(\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - u)\}D_{ij} + A_{ij}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_i$. By (4), $\widehat{\Lambda}$ minimises

$$L(\Lambda) = h\sum_{i=1}^{n}\sum_{j=1}^{n_i} K_h(\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - u) \times \left[d\{Y_{ij}, g^{-1}(\bar{\eta}_{ij} + c_n\boldsymbol{F}_{ij}^{\mathrm{T}}\Lambda)\} - d\{Y_{ij}, g^{-1}(\bar{\eta}_{ij})\}\right]$$

with respect to Λ. Using Taylor expansion,

$$\begin{aligned} -\frac{1}{2}L(\Lambda) &= c_n h\sum_{i=1}^{n}\sum_{j=1}^{n_i} K_h(\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - u)q_1(\bar{\eta}_{ij}, Y_{ij})\boldsymbol{F}_{ij}^{\mathrm{T}}\Lambda \\ &\quad + c_n^2 h\Lambda^{\mathrm{T}}\left\{\sum_{i=1}^{n}\sum_{j=1}^{n_i} K_h(\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - u)q_2(\bar{\eta}_{ij}, Y_{ij})\boldsymbol{F}_{ij}\boldsymbol{F}_{ij}^{\mathrm{T}}\right\} \\ &\quad \times \Lambda\{1 + o_p(1)\} \\ &\triangleq \boldsymbol{T}_n^{\mathrm{T}}\Lambda + \Lambda^{\mathrm{T}}\boldsymbol{W}_n\Lambda\{1 + o_p(1)\}. \end{aligned}$$

Let $\widetilde{\eta}_{ij} = \boldsymbol{\alpha}^{\mathrm{T}}X_{ij} + \{f(u) + f'(u)(\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - u)\}D_{ij} + A_{ij}^{\mathrm{T}}\boldsymbol{\gamma}_i$. Then $\bar{\eta}_{ij} - \widetilde{\eta}_{ij} = A_{ij}^{\mathrm{T}}(\widehat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i)$. Using similar arguments as Section A.1 in Liang (2009),

$$\boldsymbol{W}_n = hc_n^2\sum_{i=1}^{n}\sum_{j=1}^{n_i} K_h(\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - u)q_2(\widetilde{\eta}_{ij}, Y_{ij})\boldsymbol{F}_{ij}\boldsymbol{F}_{ij}^{\mathrm{T}} + o(h),$$

$$\boldsymbol{T}_n = hc_n\sum_{i=1}^{n}\sum_{j=1}^{n_i} K_h(\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - u)q_1(\widetilde{\eta}_{ij}, Y_{ij})\boldsymbol{F}_{ij} + o(c_n^{-1}h^2).$$

It is easy to see that $q_2(t, y) = \{y - g^{-1}(t)\}\rho_1'(t) - \rho_2(t)$. Denote the first element of $\boldsymbol{T}_n$ by $T_n^{[1]}$. Using Taylor expansion,

$$\begin{aligned} T_n^{[1]} &= hc_n\sum_{i=1}^{n}\sum_{j=1}^{n_i} K_h(\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - u)q_1(\eta_{ij})D_{ij} \\ &\quad - hc_n\sum_{i=1}^{n}\sum_{j=1}^{n_i} K_h(\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - u)q_2(\eta_{ij}) \\ &\quad \times (\eta_{ij} - \widetilde{\eta}_{ij})D_{ij}\{1 + o_p(1)\} \\ &= hc_n\sum_{i=1}^{n}\sum_{j=1}^{n_i} K_h(\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - u)q_1(\eta_{ij})D_{ij} \\ &\quad - \frac{1}{2}hc_n f''(u)\sum_{i=1}^{n}\sum_{j=1}^{n_i} K_h(\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - u)q_2(\eta_{ij}) \\ &\quad \times (\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - u)^2 D_{ij}^2\{1 + o_p(1)\} \\ &= hc_n\sum_{i=1}^{n}\sum_{j=1}^{n_i} K_h(\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - u)q_1(\eta_{ij}, Y_{ij})D_{ij} + \frac{1}{2}c_n^{-1} \\ &\quad \times f''(u)h^2\kappa_2 f_{\mathcal{U}}(u)\mathrm{E}\{\rho_2(\eta)D^2|\boldsymbol{\beta}^{\mathrm{T}}Z = u\} + O_p(\delta_n). \end{aligned}$$

Using similar arguments, we also have

$$\boldsymbol{W}_n = -f_{\mathcal{U}}(u)\mathrm{E}\{\rho_2(\eta)\Xi|\boldsymbol{\beta}^{\mathrm{T}}Z = u\} + O_p(h^2 + \delta_n),$$

where $\Xi = D^2\mathrm{diag}(1, \kappa_2)$. By the concavity of function $L(\Lambda)$, we obtain that $\widehat{\Lambda} = -\boldsymbol{W}_n^{-1}\boldsymbol{T}_n + o_p(1)$. Therefore,

$$\begin{aligned} \widehat{f}(u) - f(u) &= \frac{1}{2}f''(u)h^2\kappa_2 + \left[f_{\mathcal{U}}(u)\mathrm{E}\{\rho_2(\eta)D^2|\beta^{\mathrm{T}}Z = u\}\right]^{-1} \\ &\quad \times N^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n_i} K_h(\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - u)q_1(\eta_{ij}, Y_{ij})D_{ij} \\ &\quad + O_p(\delta_n). \end{aligned}$$

Step 2: Derive expression of

$$\left.\frac{\partial\widehat{f}(\boldsymbol{\beta}^{\mathrm{T}}z;\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial\boldsymbol{\alpha}}\right|_{(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0^{(1)})} \quad \text{and} \quad \left.\frac{\partial\widehat{f}(\boldsymbol{\beta}^{\mathrm{T}}z;\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^{(1)}}\right|_{(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0^{(1)})}.$$

From Step 1, $\widehat{f}(\boldsymbol{\beta}^{\mathrm{T}}z)$, $\widehat{f}'(\boldsymbol{\beta}^{\mathrm{T}}z)$ satisfy the following equation:

$$\begin{aligned} &N^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n_i} K_h(\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - \boldsymbol{\beta}^{\mathrm{T}}z)q_1 \\ &\quad \times \left[\boldsymbol{\alpha}^{\mathrm{T}}X_{ij} + \{\widehat{a} + \widehat{b}(\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - \boldsymbol{\beta}^{\mathrm{T}}z)\}D_{ij} + A_{ij}^{\mathrm{T}}\boldsymbol{\gamma}_i, Y_{ij}\right] \\ &\quad \times D_{ij}\{1 + o_p(1)\} = 0, \end{aligned} \tag{A1}$$

where $\widehat{a} = \widehat{f}(\boldsymbol{\beta}^{\mathrm{T}}z)$ and $\widehat{b} = \widehat{f}'(\boldsymbol{\beta}^{\mathrm{T}}z)$. Taking derivative with respect to $\boldsymbol{\alpha}$ yields:

$$\begin{aligned} &N^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n_i} K_h(\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - \boldsymbol{\beta}^{\mathrm{T}}z)q_2 \\ &\quad \times \left[\boldsymbol{\alpha}^{\mathrm{T}}X_{ij} + \{\widehat{a} + \widehat{b}(\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - \boldsymbol{\beta}^{\mathrm{T}}z)\}D_{ij} + A_{ij}^{\mathrm{T}}\boldsymbol{\gamma}_i, Y_{ij}\right]D_{ij} \\ &\quad \times \left[X_{ij} + \left\{\frac{\partial\widehat{a}}{\partial\boldsymbol{\alpha}} + \frac{\partial\widehat{b}}{\partial\boldsymbol{\alpha}}(\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - \boldsymbol{\beta}^{\mathrm{T}}z)\right\}D_{ij}\right] = 0. \end{aligned}$$

From this equation we obtain

$$\frac{\partial\widehat{a}}{\partial\boldsymbol{\alpha}} = -A_n^{-1}(B_{1n} + B_{2n}),$$

with

$$\begin{aligned} A_n &= N^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n_i} K_h(\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - \boldsymbol{\beta}^{\mathrm{T}}z)q_2 \\ &\quad \times \left[\boldsymbol{\alpha}^{\mathrm{T}}X_{ij} + \{\widehat{a} + \widehat{b}(\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - \boldsymbol{\beta}^{\mathrm{T}}z)\}D_{ij} + A_{ij}^{\mathrm{T}}\boldsymbol{\gamma}_i, Y_{ij}\right]D_{ij}^2, \\ B_{1n} &= N^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n_i} K_h(\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - \boldsymbol{\beta}^{\mathrm{T}}z)q_2 \\ &\quad \times \left[\boldsymbol{\alpha}^{\mathrm{T}}X_{ij} + \{\widehat{a} + \widehat{b}(\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - \boldsymbol{\beta}^{\mathrm{T}}z)\}D_{ij} + A_{ij}^{\mathrm{T}}\boldsymbol{\gamma}_i, Y_{ij}\right] \\ &\quad \times D_{ij}X_{ij}, \\ B_{2n} &= N^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n_i} K_h(\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - \boldsymbol{\beta}^{\mathrm{T}}z)q_2 \\ &\quad \times \left[\boldsymbol{\alpha}^{\mathrm{T}}X_{ij} + \{\widehat{a} + \widehat{b}(\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - \boldsymbol{\beta}^{\mathrm{T}}z)\}D_{ij} + A_{ij}^{\mathrm{T}}\boldsymbol{\gamma}_i, Y_{ij}\right]D_{ij}^2 \\ &\quad \times (\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - \boldsymbol{\beta}^{\mathrm{T}}z)\frac{\partial\widehat{b}}{\partial\boldsymbol{\alpha}}. \end{aligned}$$

Taking derivative on (A1) with respect to $\boldsymbol{\beta}^{(1)}$, we get

$$N^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n_i} h^{-1}K_h'(\boldsymbol{\beta}^{\mathrm{T}}Z_{ij} - \boldsymbol{\beta}^{\mathrm{T}}z)\boldsymbol{J}^{\mathrm{T}}(Z_{ij} - z)q_1$$

$$\times \left[\boldsymbol{\alpha}^{\mathrm{T}} X_{ij} + \{\widehat{a} + \widehat{b}(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij} - \boldsymbol{\beta}^{\mathrm{T}} z)\} D_{ij} + A_{ij}^{\mathrm{T}} \boldsymbol{\gamma}_i, Y_{ij}\right] D_{ij}$$
$$+ \sum_{i=1}^{n} \sum_{j=1}^{n_i} K_h(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij} - \boldsymbol{\beta}^{\mathrm{T}} z) q_2$$
$$\times \left[\boldsymbol{\alpha}^{\mathrm{T}} X_{ij} + \{\widehat{a} + \widehat{b}(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij} - \boldsymbol{\beta}^{\mathrm{T}} z)\} + A_{ij}^{\mathrm{T}} \boldsymbol{\gamma}_i, Y_{ij}\right] D_{ij}^2$$
$$\times \left\{ \frac{\partial \widehat{a}}{\partial \boldsymbol{\beta}^{(1)}} + \frac{\partial \widehat{b}}{\partial \boldsymbol{\beta}^{(1)}} (\boldsymbol{\beta}^{\mathrm{T}} Z_{ij} - \boldsymbol{\beta}^{\mathrm{T}} z) + \widehat{b} \boldsymbol{J}^{\mathrm{T}} (Z_{ij} - z) \right\} = 0.$$

It follows that

$$\frac{\partial \widehat{a}}{\partial \boldsymbol{\beta}^{(1)}} = -A_n^{-1}(C_{1n} + C_{2n} + C_{3n}),$$

where

$$C_{1n} = N^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n_i} K_h(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij} - \boldsymbol{\beta}^{\mathrm{T}} z) q_2$$
$$\times \left[\boldsymbol{\alpha}^{\mathrm{T}} X_{ij} + \{\widehat{a} + \widehat{b}(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij} - \boldsymbol{\beta}^{\mathrm{T}} z)\} D_{ij} + A_{ij}^{\mathrm{T}} \boldsymbol{\gamma}_i, Y_{ij}\right]$$
$$\times D_{ij}^2 \widehat{b} \boldsymbol{J}^{\mathrm{T}} (Z_{ij} - z),$$
$$C_{2n} = N^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n_i} K_h(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij} - \boldsymbol{\beta}^{\mathrm{T}} z) q_2$$
$$\times \left[\boldsymbol{\alpha}^{\mathrm{T}} X_{ij} + \{\widehat{a} + \widehat{b}(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij} - \boldsymbol{\beta}^{\mathrm{T}} z)\} D_{ij} + A_{ij}^{\mathrm{T}} \boldsymbol{\gamma}_i, Y_{ij}\right]$$
$$\times D_{ij}^2, (\boldsymbol{\beta}^{\mathrm{T}} Z_{ij} - \boldsymbol{\beta}^{\mathrm{T}} z) \frac{\partial \widehat{b}}{\partial \boldsymbol{\beta}^{(1)}},$$
$$C_{3n} = N^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n_i} h^{-1} K_h'(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij} - \boldsymbol{\beta}^{\mathrm{T}} z) \boldsymbol{J}^{\mathrm{T}} (Z_{ij} - z) q_1$$
$$\times \left[\boldsymbol{\alpha}^{\mathrm{T}} Z_{ij} + \{\widehat{a} + \widehat{b}(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij} - \boldsymbol{\beta}^{\mathrm{T}} z)\} D_{ij} + A_{ij}^{\mathrm{T}} \boldsymbol{\gamma}_i, Y_{ij}\right]$$
$$\times D_{ij}.$$

By similar calculations with Cui et al. (2011),

$$A_n = -f_{\mathcal{U}}(\boldsymbol{\beta}^{\mathrm{T}} z) \mathrm{E}\left\{\rho_2(\eta) D^2 | \boldsymbol{\beta}^{\mathrm{T}} Z = \boldsymbol{\beta}^{\mathrm{T}} z\right\}$$
$$+ O_p(h^2 + n^{-/2} h^{-1/2}),$$
$$B_{1n} = -f_{\mathcal{U}}(\boldsymbol{\beta}^{\mathrm{T}} z) \mathrm{E}\left\{\rho_2(\eta) DX | \boldsymbol{\beta}^{\mathrm{T}} Z = \boldsymbol{\beta}^{\mathrm{T}} z\right\}$$
$$+ O_p(h^2 + n^{-1/2} h^{-1/2}),$$
$$B_{2n} = O_p(h^2 + n^{-1/2} h^{1/2}),$$
$$C_{1n} = -f_{\mathcal{U}}(\boldsymbol{\beta}^{\mathrm{T}} z) f'(\boldsymbol{\beta}^{\mathrm{T}} z) \boldsymbol{J}^{\mathrm{T}} \left[\mathrm{E}\{\rho_2(\eta) D^2 Z | \boldsymbol{\beta}^{\mathrm{T}} Z = \boldsymbol{\beta}^{\mathrm{T}} z\}\right.$$
$$\left. - \mathrm{E}\{\rho_2(\eta) D^2 | \boldsymbol{\beta}^{\mathrm{T}} Z = \boldsymbol{\beta}^{\mathrm{T}} z\} z\right]$$
$$+ O_p(h^2 + n^{-1/2} h^{-1/2}),$$
$$C_{2n} = O_p(h^2 + n^{-1/2} h^{1/2}),$$
$$C_{3n} = O_p(h^2 + n^{-1/2} h^{-3/2}).$$

Therefore,

$$\frac{\partial \widehat{a}}{\partial \boldsymbol{\alpha}} = -\left[\mathrm{E}\{\rho_2(\eta) D^2 | \boldsymbol{\beta}^{\mathrm{T}} Z = \boldsymbol{\beta}^{\mathrm{T}} z\}\right]^{-1}$$
$$\times \left[\mathrm{E}\{\rho_2(\eta) DX | \boldsymbol{\beta}^{\mathrm{T}} Z = \boldsymbol{\beta}^{\mathrm{T}} z\}\right]$$
$$+ O(h^2 + n^{-1/2} h^{1/2}),$$

$$\frac{\partial \widehat{a}}{\partial \boldsymbol{\beta}^{(1)}} = f'(\boldsymbol{\beta}^{\mathrm{T}} z) \boldsymbol{J}^{\mathrm{T}} \left(z - \left[\mathrm{E}\{\rho_2(\eta) D^2 | \boldsymbol{\beta}^{\mathrm{T}} Z = \boldsymbol{\beta}^{\mathrm{T}} z\}\right]^{-1} \right.$$
$$\left. \times \left[\mathrm{E}\{\rho_2(\eta) D^2 Z | \boldsymbol{\beta}^{\mathrm{T}} Z = \boldsymbol{\beta}^{\mathrm{T}} z\}\right] \right)$$
$$+ O(h^2 + n^{-1/2} h^{1/2}).$$

We further denote

$$\widetilde{X} = X - \left[\mathrm{E}\{\rho_2(\eta) D^2 | \boldsymbol{\beta}^{\mathrm{T}} Z\}\right]^{-1} \left[\mathrm{E}\{\rho_2(\eta) DX | \boldsymbol{\beta}^{\mathrm{T}} Z\}\right],$$
$$\widetilde{Z} = f'(\boldsymbol{\beta}^{\mathrm{T}} Z) \boldsymbol{J}^{\mathrm{T}} \left(Z - \left[\mathrm{E}\{\rho_2(\eta) D^2 | \boldsymbol{\beta}^{\mathrm{T}} Z\}\right]^{-1} \right.$$
$$\left. \times \left[\mathrm{E}\{\rho_2(\eta) D^2 Z | \boldsymbol{\beta}^{\mathrm{T}} Z\}\right] \right).$$

It is easy to see that

$$\mathrm{E}\{\rho_2(\eta) D\widetilde{X} | \boldsymbol{\beta}^{\mathrm{T}} Z\} = 0, \quad \mathrm{E}\{\rho_2(\eta) D^2 \widetilde{Z} | \boldsymbol{\beta}^{\mathrm{T}} Z\} = 0. \quad \text{(A2)}$$

Step 3: The asymptotic normality of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.
From estimating Equations (8) and (9),

$$\sqrt{N} \begin{pmatrix} \widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 \\ \widehat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}_0^{(1)} \end{pmatrix} = \mathcal{A}_n^{-1} \sqrt{N} \mathcal{B}_n + o_p(1),$$

where

$$\mathcal{A}_n = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \rho_2(\eta_{ij}) \begin{pmatrix} X_{ij} \\ f'(\boldsymbol{\beta}_0^{\mathrm{T}} Z_{ij}) \boldsymbol{J}^{\mathrm{T}} Z_{ij} D_{ij} \end{pmatrix} \begin{pmatrix} \widetilde{X}_{ij} \\ \widetilde{Z}_{ij} D_{ij} \end{pmatrix}^{\mathrm{T}},$$

$$\mathcal{B}_n = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \begin{pmatrix} X_{ij} \\ f'(\boldsymbol{\beta}_0^{\mathrm{T}} Z_{ij}) \boldsymbol{J}^{\mathrm{T}} Z_{ij} D_{ij} \end{pmatrix} \{g'(\mu_{ij})\}^{-1} V^{-1}(\mu_{ij})$$
$$\times \left[Y_{ij} - g^{-1}\{\boldsymbol{\alpha}_0^{\mathrm{T}} X_{ij} + \widehat{f}(\boldsymbol{\beta}_0^{\mathrm{T}} Z_{ij}) D_{ij} + A_{ij}^{\mathrm{T}} \boldsymbol{\gamma}_i \right].$$

Based on (A2), we have

$$\mathcal{A}_n = \mathrm{E}\left(\rho_2(\eta) \begin{pmatrix} \widetilde{X} \\ \widetilde{Z} D \end{pmatrix}^{\otimes 2} \right) + o_p(1) \triangleq \mathcal{A} + o_p(1).$$

By Taylor expansion,

$$\mathcal{B}_n = \mathcal{B}_{1n} + \mathcal{B}_{2n}\{1 + o_p(1)\},$$

where

$$\mathcal{B}_{1n} = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \begin{pmatrix} X_{ij} \\ f'(\boldsymbol{\beta}_0^{\mathrm{T}} Z_{ij}) \boldsymbol{J}^{\mathrm{T}} Z_{ij} D_{ij} \end{pmatrix} q_1(\eta_{ij}, Y_{ij}),$$

$$\mathcal{B}_{2n} = -\frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \begin{pmatrix} X_{ij} \\ f'(\boldsymbol{\beta}_0^{\mathrm{T}} Z_{ij}) \boldsymbol{J}^{\mathrm{T}} Z_{ij} D_{ij} \end{pmatrix} \rho_2(\eta_{ij}) \{\widehat{f}(\boldsymbol{\beta}_0^{\mathrm{T}} Z_{ij})$$
$$- f(\boldsymbol{\beta}_0^{\mathrm{T}} Z_{ij})\} D_{ij}.$$

From the result of Step 1,

$$\sqrt{N} \mathcal{B}_{2n} = -\frac{1}{\sqrt{N}} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \begin{pmatrix} X_{ij} \\ f'(\boldsymbol{\beta}_0^{\mathrm{T}} Z_{ij}) \boldsymbol{J}^{\mathrm{T}} Z_{ij} D_{ij} \end{pmatrix} \rho_2(\eta_{ij}) D_{ij}$$
$$\times \left[f_{\mathcal{U}}(\boldsymbol{\beta}_0^{\mathrm{T}} Z_{ij}) \mathrm{E}\{\rho_2(\eta) D^2 | \boldsymbol{\beta}_0^{\mathrm{T}} Z = \boldsymbol{\beta}_0^{\mathrm{T}} Z_{ij}\} \right]^{-1}$$
$$\times \frac{1}{N} \sum_{k=1}^{n} \sum_{l=1}^{n_i} K_h(\boldsymbol{\beta}_0^{\mathrm{T}} Z_{kl} - \boldsymbol{\beta}_0^{\mathrm{T}} Z_{ij}) q_1(\eta_{kl}, Y_{kl}) D_{kl}$$
$$- \frac{1}{2} h^2 \kappa_2 \frac{1}{\sqrt{N}} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \begin{pmatrix} X_{ij} \\ f'(\boldsymbol{\beta}_0^{\mathrm{T}} Z_{ij}) \boldsymbol{J}^{\mathrm{T}} Z_{ij} D_{ij} \end{pmatrix}$$
$$\times \rho_2(\eta_{ij}) D_{ij} f''(\boldsymbol{\beta}^{\mathrm{T}} Z_{ij})$$

$$
\begin{aligned}
&= -\frac{1}{\sqrt{N}} \sum_{k=1}^{n} \sum_{l=1}^{n_i} q_1(\eta_{kl}, Y_{kl}) D_{kl} \\
&\quad \times \sum_{i=1}^{n} \sum_{j=1}^{n_i} K_h(\boldsymbol{\beta}_0^{\mathrm{T}} Z_{ij} - \boldsymbol{\beta}_0^{\mathrm{T}} Z_{kl}) \\
&\quad \times \begin{pmatrix} X_{ij} \\ f'(\boldsymbol{\beta}_0^{\mathrm{T}} Z_{ij}) \boldsymbol{J}^{\mathrm{T}} Z_{ij} D_{ij} \end{pmatrix} \rho_2(\eta_{ij}) D_{ij} \\
&\quad \times \left[f_{\mathcal{U}}(\boldsymbol{\beta}_0^{\mathrm{T}} Z_{ij}) \mathrm{E}\{\rho_2(\eta) D^2 | \boldsymbol{\beta}_0^{\mathrm{T}} Z = \boldsymbol{\beta}_0^{\mathrm{T}} Z_{ij}\} \right]^{-1} \\
&\quad + O_p(n^{1/2} h^2) \\
&= -\frac{1}{\sqrt{N}} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left[\mathrm{E}\{\rho_2(\eta) D^2 | \boldsymbol{\beta}_0^{\mathrm{T}} Z = \boldsymbol{\beta}_0^{\mathrm{T}} Z_{ij}\} \right]^{-1} \\
&\quad \times \begin{bmatrix} \mathrm{E}\left\{ \rho_2(\eta) D X | \boldsymbol{\beta}_0^{\mathrm{T}} Z = \boldsymbol{\beta}_0^{\mathrm{T}} Z_{ij} \right\} \\ f'(\boldsymbol{\beta}_0^{\mathrm{T}} Z_{ij}) \boldsymbol{J}^{\mathrm{T}} \mathrm{E}\{\rho_2(\eta) D^2 Z | \boldsymbol{\beta}_0^{\mathrm{T}} Z = \boldsymbol{\beta}_0^{\mathrm{T}} Z_{ij}\} \end{bmatrix} \\
&\quad \times D_{ij} q_1(\eta_{ij}, Y_{ij}) + o_p(1).
\end{aligned}
$$

It is easy to see that

$$
\sqrt{N} \mathcal{B}_n = \frac{1}{\sqrt{N}} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \begin{pmatrix} \widetilde{X}_{ij} \\ \widetilde{Z}_{ij} D_{ij} \end{pmatrix} q_1(\eta_{ij}, Y_{ij}).
$$

By CLT, $\sqrt{N}\mathcal{B}_n$ has asymptotic normal distribution with zero-mean. Denote $s_{ij} = \begin{pmatrix} \widetilde{X}_{ij} \\ \widetilde{Z}_{ij} D_{ij} \end{pmatrix}$, we have

$$
\begin{aligned}
\mathrm{var}(\sqrt{N}\mathcal{B}_n) = &\frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \mathrm{var}\{s_{ij} q_1(\eta_{ij}, Y_{ij})\} \\
&+ \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \sum_{\substack{k=1 \\ k \neq j}}^{n_i} \mathrm{cov}\{s_{ij} q_1(\eta_{ij}, Y_{ij}), \\
&s_{ik} q_1(\eta_{ik}, Y_{ik})\}.
\end{aligned}
$$

Using simple calculation,

$$
\mathrm{var}\{s_{ij} q_1(\eta_{ij}, Y_{ij})\} = \mathrm{E}\{\rho_2(\eta) s s^{\mathrm{T}}\} = \mathcal{A}.
$$

Denote $\mathcal{C} = \mathrm{cov}\{s_{11} q_1(\eta_{11}, Y_{11}), s_{12} q_1(\eta_{12}, Y_{12})\}$,

$$
\mathrm{var}(\sqrt{N}\mathcal{B}_n) = \mathcal{A} + \frac{N_1}{N} \mathcal{C}.
$$

Therefore, when $n \to \infty$,

$$
\sqrt{N} \begin{pmatrix} \widehat{\boldsymbol{\alpha}} - \alpha_0 \\ \widehat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}_0^{(1)} \end{pmatrix} \to N(0, \mathcal{A}^{-1} + \frac{N_1}{N} \mathcal{A}^{-1} \mathcal{C} \mathcal{A}^{-1}).
$$

Denote $\mathcal{D} = \begin{pmatrix} \boldsymbol{I}_p & \boldsymbol{0}_{p \times (q-1)} \\ \boldsymbol{0}_{q \times p} & \boldsymbol{J}_{\boldsymbol{\beta}_0^{(1)}} \end{pmatrix}$. An application of multivariate delta-method yields

$$
\sqrt{N} \begin{pmatrix} \widehat{\boldsymbol{\alpha}} - \alpha_0 \\ \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \end{pmatrix} \to N(0, \boldsymbol{\Sigma}),
$$

where $\boldsymbol{\Sigma} = \mathcal{D}(\mathcal{A}^{-1} + \frac{N_1}{N} \mathcal{A}^{-1} \mathcal{C} \mathcal{A}^{-1}) \mathcal{D}^{\mathrm{T}}$. ■

Proof of Theorem 3.2: From Theorem 3.1, $\sqrt{N} \begin{pmatrix} \widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 \\ \widehat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}_0^{(1)} \end{pmatrix} = O_p(1)$. We write

$$
\begin{aligned}
&\widehat{f}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}} z; \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}) - f(\boldsymbol{\beta}_0^{\mathrm{T}} z) \\
&\quad = \widehat{f}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}} z; \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}) - \widehat{f}(\boldsymbol{\beta}_0^{\mathrm{T}} z) + \widehat{f}(\boldsymbol{\beta}_0^{\mathrm{T}} z) - f(\boldsymbol{\beta}_0^{\mathrm{T}} z) \\
&\quad = \frac{\partial \widehat{f}(\boldsymbol{\beta}^{\mathrm{T}} z; \boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial (\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\beta}^{(1)\mathrm{T}})} \Big|_{(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0^{(1)})} \begin{pmatrix} \widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 \\ \widehat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}_0^{(1)} \end{pmatrix} \{1 + o_p(1)\} \\
&\quad + \left\{ \widehat{f}(\boldsymbol{\beta}_0^{\mathrm{T}} z) - f(\boldsymbol{\beta}_0^{\mathrm{T}} z) \right\} \\
&\quad = \widehat{f}(\boldsymbol{\beta}_0^{\mathrm{T}} z) - f(\boldsymbol{\beta}_0^{\mathrm{T}} z) + O_p(n^{-1/2}).
\end{aligned}
$$

Hence the asymptotic distribution of $\widehat{f}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}} z; \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$ is the same as that of $\widehat{f}(\boldsymbol{\beta}_0^{\mathrm{T}} z)$. Based on the result of Step 1, we now derive the asymptotic variance of $\widehat{f}(\boldsymbol{\beta}_0^{\mathrm{T}} z)$. Denote $r_{ij} = K_h(\boldsymbol{\beta}_0^{\mathrm{T}} Z_{ij} - \boldsymbol{\beta}_0^{\mathrm{T}} z) q_1(\eta_{ij}, Y_{ij}) D_{ij}$, we have

$$
\begin{aligned}
\mathrm{var}\left\{ \sum_{i=1}^{n} \sum_{j=1}^{n_i} r_{ij} \right\} = &\sum_{i=1}^{n} \sum_{j=1}^{n_i} \mathrm{var}(r_{ij}) \\
&+ \sum_{i=1}^{n} \sum_{j=1}^{n_i} \sum_{\substack{k=1 \\ k \neq j}}^{n_i} \mathrm{cov}(r_{ij}, r_{ik}),
\end{aligned}
$$

where

$$
\begin{aligned}
\mathrm{var}(r_{ij}) &= \mathrm{E} K_h^2(\boldsymbol{\beta}_0^{\mathrm{T}} Z_{ij} - \boldsymbol{\beta}_0^{\mathrm{T}} z) q_1^2(\eta_{ij}, Y_{ij}) D_{ij}^2 \\
&= h^{-1} \mu_0 f_{\mathcal{U}}(\boldsymbol{\beta}_0^{\mathrm{T}} z) \mathrm{E}\{\rho_2(\eta) D^2 | \boldsymbol{\beta}_0^{\mathrm{T}} Z = \boldsymbol{\beta}_0^{\mathrm{T}} z\}, \\
\mathrm{cov}(r_{ij}, r_{ik}) &= f_{\mathcal{U}}^2(\boldsymbol{\beta}_0^{\mathrm{T}} z) \mathrm{E}\left\{ q_1(\eta_{11}, Y_{11}) q_1(\eta_{12}, Y_{12}) D_{11}^2 \right. \\
&\quad \left. D_{12}^2 | \boldsymbol{\beta}_0^{\mathrm{T}} Z_{11} = \boldsymbol{\beta}_0^{\mathrm{T}} z, \boldsymbol{\beta}_0^{\mathrm{T}} Z_{12} = \boldsymbol{\beta}_0^{\mathrm{T}} z \right\}.
\end{aligned}
$$

Therefore, we have

$$
\sqrt{Nh} \{ \widehat{f}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}} z; \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}) - f(\boldsymbol{\beta}_0^{\mathrm{T}} z) - \frac{1}{2} h^2 \kappa_2 f''(\boldsymbol{\beta}_0^{\mathrm{T}} z) \} \to N(0, \sigma_f^2),
$$

where $\sigma_f^2 = \mu_0 f_{\mathcal{U}}^{-1}(\boldsymbol{\beta}_0^{\mathrm{T}} z) [\mathrm{E}\{\rho_2(\eta) D^2 | \boldsymbol{\beta}_0^{\mathrm{T}} Z = \boldsymbol{\beta}_0^{\mathrm{T}} z\}]^{-1}$. ■