

# Efficient estimation for generalized partially linear single-index models

LI WANG<sup>1</sup> and GUANQUN CAO<sup>2</sup>

<sup>1</sup>*Department of Statistics and the Statistical Laboratory, Iowa State University, Ames, IA 50011, USA.  
E-mail: lilywang@iastate.edu*

<sup>2</sup>*Department of Mathematics and Statistics, Auburn University, Auburn, AL 36849, USA.  
E-mail: gzc0009@auburn.edu*

In this paper, we study the estimation for generalized partially linear single-index models, where the systematic component in the model has a flexible semi-parametric form with a general link function. We propose an efficient and practical approach to estimate the single-index link function, single-index coefficients as well as the coefficients in the linear component of the model. The estimation procedure is developed by applying quasi-likelihood and polynomial spline smoothing. We derive large sample properties of the estimators and show the convergence rate of each component of the model. Asymptotic normality and semiparametric efficiency are established for the coefficients in both the single-index and linear components. By making use of spline basis approximation and Fisher score iteration, our approach has numerical advantages in terms of computing efficiency and stability in practice. Both simulated and real data examples are used to illustrate our proposed methodology.

*Keywords:* asymptotic normality; generalized linear model; polynomial splines; quasi-likelihood; semi-parametric regression; single-index model

## 1. Introduction

Generalized linear models (GLM [23]) provide a unified likelihood framework for parametric regression analysis. They extend linear models to allow for parametrically modeling the relation between a transformation of the mean response and some covariates. However, the parametric approach may not always be desirable since the appropriate function form may not be known in advance and the response may depend on the covariate in a very complicated manner. Different approaches have been proposed to face this lack of linearity and to solve problems arising when a misspecified model is fitted. This recognition has motivated the development of non-/semi-parametric regression techniques with far reaching applications.

On the other hand, another problem when confronted with multiple covariates is the well-known “curse of dimensionality”. For example, models with one multivariate function can only accommodate low dimensional covariates. A lot of effort has been devoted to the circumventing of this difficulty. There are essentially two possible approaches: function approximation and dimension reduction [34]. A favorite function approximation technique is the generalized additive model (GAM) advocated by [10], see also, for example, [12,22,29,35] and the recent comprehensive book by [31] for additional references. GAM replaces the multivariate nonparametric component by a sum of univariate functions of the covariates. However, the model with a summand

of univariate functions does not take into account the interaction effects among covariates. An attractive dimension reduction method is the single-index model (SIM); see [6,7,14] and [3] for detailed discussions and illustrations of the usefulness of this model. Over the last two decades, many authors have devised various intelligent methods to estimate the SIM, for instance, [8,9,11,15,16,24,33] and [25].

Semiparametric SIMs present an appealing and effective statistical concept to model the relationship between the response variable and multivariate covariates, since it achieves dimension reduction and relaxes the restrictive parametric assumptions. In this paper, we consider the generalized partially linear single-index model (GPLSIM) proposed by [2], a realistic and parsimonious semiparametric model. The GPLSIM as a natural extension to the SIM allows discrete explanatory variables to be modeled in the linear part. Various methods have been studied in depth for the partially linear single-index model (PLSIM), a special case of GPLSIM with an identity link function. For instance, [26,36] and [20] investigate the PLSIM under independent setting, while [4] and [21] study the estimation of the PLSIM for longitudinal data case.

Estimation of the single-index link function requires a degree of statistical smoothing. Disregarding the great flexibility and easy interpretability, the GPLSIMs haven't been widely used in practical data analysis, due to the scarcity of smoothing tools that are not only computationally expedient but also theoretically reliable, which has motivated the proposed procedures of this paper.

Two classes of smoothing method which have received much attention in the semiparametric SIM literature are kernel smoothing and penalized spline smoothing. For example, the kernel type of smoothing is adopted in [2] and [19], which can become computationally expensive when extensive iterations are required to update the estimator. The heavy computation burden is also noticed by a sequence of papers; see, for example, [26] and [32]. Piecewise polynomial functions, or splines, have been proven to be an extremely powerful method in the statistical literature on smoothing. Splines are simple, yet eminently practical tool with computational tractability and statistical efficiency. Yu and Ruppert [36] has presented a thorough account of penalized spline smoothing methodology for PLSIMs, in which the underlying regression function is assumed to be a spline function with a fixed number of knots. However, the bias caused by the spline approximation in developing the asymptotic theory is ignored. In addition, it has been shown in many contexts of function estimation that, by letting the number of knots increase with the sample size at an appropriate rate, the spline estimate of the unknown function can achieve the optimal nonparametric rate of convergence.

In our paper, the estimation is carried out using the regression splines to save the computational cost. We allow a general link function and the number of knots to increase with the sample size, which makes our approach fundamentally different from [36]. Our method can be applied for continuous as well as categorical responses. The application of polynomial spline smoothing makes our approach more stable and computationally expedient. Once the knots sequence is chosen, we can estimate the parameters and the nonparametric single-index link function simultaneously by maximizing the spline approximation of the quasi-likelihood. We make use of the Fisher scoring algorithm to solve the maximization problem iteratively. We also propose a generalized cross-validation method to select the number of knots in the spline approximation.

Under regularity conditions, we show that the maximum quasi-likelihood estimators of the parameters are root- $n$  consistent, asymptotically normal and achieve the information bound. We

also provide the convergence rate of the nonparametric component estimator. The establishment of the asymptotic normal distribution of the estimators for the parameters in the GPLSIM is quite challenging as our estimator is motivated from an approximated model, the standard theory of the parametric quasi-likelihood estimation does not carry over.

The rest of the paper is structured as the following. In Section 2, we present the generalized partially linear single-index model. In Section 3, we describe the quasi-likelihood estimation procedures via polynomial spline smoothing. In Section 4, we state the asymptotic properties of the proposed estimators. The computing algorithm and its implementations are given in Section 5. In Section 6, we evaluate the estimation procedure on simulated datasets. We also apply the model and estimation procedure to analyzing the datasets from the Framingham Heart Study and the Munich Dust Study. Concluding remarks are given in Section 7. We present the regularity conditions and technical proofs in Appendix and Supplementary Material [27].

## 2. Generalized partially linear single-index models

Let  $Y$  be the response variable and  $(\mathbf{X}, \mathbf{Z}) \in R^{d_1} \times R^{d_2}$  be the predictor vector with  $\mathbf{X} = (X_1, \dots, X_{d_1})^\top$  and  $\mathbf{Z} = (Z_1, \dots, Z_{d_2})^\top$ . We assume the conditional density of  $Y$  given  $(\mathbf{X}, \mathbf{Z}) = (\mathbf{x}, \mathbf{z})$  belongs to a canonical exponential family

$$f_{Y|\mathbf{X},\mathbf{Z}}(y|\mathbf{x}, \mathbf{z}) = \exp[y\xi(\mathbf{x}, \mathbf{z}) - \mathcal{B}\{\xi(\mathbf{x}, \mathbf{z})\} + \mathcal{C}(y)], \quad (2.1)$$

for known functions  $\mathcal{B}$  and  $\mathcal{C}$ , and  $\xi$  is the natural parameter in parametric GLMs, which is related to the unknown mean response  $\mu(\mathbf{x}, \mathbf{z}) = E(Y|\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}) = \mathcal{B}'\{\xi(\mathbf{x}, \mathbf{z})\}$ . In parametric GLMs, the mean function  $\mu$  is defined via a known link function  $g$  by  $g\{\mu(\mathbf{x}, \mathbf{z})\} = \boldsymbol{\alpha}_0^\top \mathbf{x} + \boldsymbol{\beta}_0^\top \mathbf{z}$  with coefficients  $\boldsymbol{\alpha}_0$  and  $\boldsymbol{\beta}_0$ . In this paper, to enhance the model flexibility, we model  $g(\mu)$  as a PLSIM:

$$g\{\mu(\mathbf{x}, \mathbf{z})\} = \eta_0(\boldsymbol{\alpha}_0^\top \mathbf{x}) + \boldsymbol{\beta}_0^\top \mathbf{z}, \quad (2.2)$$

where  $\boldsymbol{\alpha}_0$  is the single-index coefficient vector with dimension  $d_1$ ,  $\boldsymbol{\beta}_0$  is the coefficient vector in the linear component with dimension  $d_2$ , and  $\eta_0$  is some unknown and smooth function, referred to as the single-index link function. For model identifiability, we assume  $\|\boldsymbol{\alpha}_0\| = 1$  and the first element of  $\boldsymbol{\alpha}_0$  is positive. The GPLSIM includes as special cases the linear model, the SIM, as well as the partially linear model (PLM) and the PLSIM.

If the conditional variance function  $\text{Var}(Y|\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}) = \sigma^2 V\{\mu(\mathbf{x}, \mathbf{z})\}$  for some known positive function  $V$ , then estimation of the mean can be achieved by replacing the conditional log-likelihood function  $f_{Y|\mathbf{X},\mathbf{Z}}(y|\mathbf{x}, \mathbf{z})$  in (2.1) by a quasi-likelihood function  $Q(m, y)$ , which satisfies

$$\frac{\partial}{\partial m} Q(m, y) = \frac{y - m}{\sigma^2 V(m)}.$$

## 3. Estimation method

Let  $(Y_i, \mathbf{X}_i, \mathbf{Z}_i)$ ,  $i = 1, \dots, n$ , be independent copies of  $(Y, \mathbf{X}, \mathbf{Z})$ . We assume  $\mathbf{X}$  is supported on a  $d_1$ -dimensional ball. Under some smoothness assumptions,  $\eta_0(u)$ ,  $u \in [a, b]$ , can be well

approximated by a spline function. Let  $\mathcal{S}_n$  be the space of polynomial splines on  $[a, b]$  of order  $r \geq 1$ . We introduce a knot sequence with  $J$  interior knots

$$k_{-r+1} = \dots = k_{-1} = k_0 = a < k_1 < \dots < k_J < b = k_{J+1} = \dots = k_{J+r},$$

where  $J \equiv J_n$  increases when sample size  $n$  increases, and the precise order is given in Condition (C5) in the Appendix. Then  $\mathcal{S}_n$  consists of functions  $s$  satisfying:

- (i)  $s$  is a polynomial of degree  $r - 1$  on each of the subintervals  $I_j = [k_j, k_{j+1})$ ,  $j = 0, \dots, J_n - 1$ ,  $I_{J_n} = [k_{J_n}, b]$ ;
- (ii) for  $r \geq 2$ ,  $s$  is  $r - 2$  continuously differentiable on  $[a, b]$ .

We seek a function  $\eta \in \mathcal{S}_n$  along with a value of  $\alpha$  and a value of  $\beta$  that maximize the following quasi-likelihood function

$$L(\eta, \alpha, \beta) = \frac{1}{n} \sum_{i=1}^n Q[g^{-1}\{\eta(\alpha^\top \mathbf{X}_i) + \beta^\top \mathbf{Z}_i\}, Y_i], \quad \eta \in \mathcal{S}_n. \tag{3.1}$$

Let  $N_n = J_n + r$ , and let  $b_{j,r}(u)$ ,  $j = 1, \dots, N_n$ , be the B-spline basis functions of order  $r$ . For any  $\eta \in \mathcal{S}_n$ , one can write  $\eta(u) = \boldsymbol{\gamma}^\top \mathbf{B}(u)$ , where  $\mathbf{B}(u) = \{b_{j,r}(u)\}_{j=1}^{N_n}$  is the spline basis, and  $\boldsymbol{\gamma} = \{\gamma_j\}_{j=1}^{N_n}$  is the spline coefficient vector.

Denote  $\boldsymbol{\theta} = (\alpha^\top, \beta^\top, \boldsymbol{\gamma}^\top)^\top$ , then the maximization problem in (3.1) is equivalent to find a value of  $\boldsymbol{\theta}$  to maximize

$$\ell(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n Q[g^{-1}\{\boldsymbol{\gamma}^\top \mathbf{B}(\alpha^\top \mathbf{X}_i) + \beta^\top \mathbf{Z}_i\}, Y_i], \tag{3.2}$$

and we denote the maximizer as  $\widehat{\boldsymbol{\theta}} = (\widehat{\alpha}^\top, \widehat{\beta}^\top, \widehat{\boldsymbol{\gamma}}^\top)^\top$ . Then the spline estimator of  $\eta_0(u)$  is  $\widehat{\eta}(u) = \widehat{\boldsymbol{\gamma}}^\top \mathbf{B}(u)$ .

### 4. Asymptotic properties of the estimators

In this section, we derive the asymptotic properties of the estimators for the parametric and nonparametric components.

#### 4.1. Estimation of the nonparametric component

For any measurable functions  $\varphi_1, \varphi_2$  on  $[a, b]$ , define the empirical inner product and the corresponding norm as  $\langle \varphi_1, \varphi_2 \rangle_n = n^{-1} \sum_{i=1}^n \{\varphi_1(U_i)\varphi_2(U_i)\}$ ,  $\|\varphi\|_n^2 = n^{-1} \sum_{i=1}^n \varphi^2(U_i)$ .

If  $\varphi_1$  and  $\varphi_2$  are  $L^2$ -integrable, define the theoretical inner product and the corresponding norm as  $\langle \varphi_1, \varphi_2 \rangle = E\{\varphi_1(U)\varphi_2(U)\}$ ,  $\|\varphi\|_2^2 = E\varphi^2(U) = \int_a^b \varphi^2(u)f(u)du$ .

Let  $h = (b - a)/(J_n + 1)$  be the distance between neighboring knots. The following theorem shows the convergence rate of the proposed estimator  $\widehat{\eta}$ .

**Theorem 1.** Under Conditions (C1)–(C5) in Appendix A.1,

$$\begin{aligned} \|\widehat{\eta} - \eta_0\|_2 &= O_P\{N_n^{1/2}(h^p + n^{-1/2}h^{-1})\}, \\ \|\widehat{\eta} - \eta_0\|_n &= O_P\{N_n^{1/2}(h^p + n^{-1/2}h^{-1})\}. \end{aligned}$$

### 4.2. Estimation of the parametric components

The next theorem shows that the maximum quasi-likelihood estimator is root- $n$  consistent and asymptotically normal though the convergence rate of the nonparametric component  $\widehat{\eta}$  is slower than root- $n$ . For simplicity of notation, let  $\mathbf{T} = (\mathbf{X}^\top, \mathbf{Z}^\top)^\top$ ,  $m_0(\mathbf{T}) = \eta_0(U_0) + \beta_0^\top \mathbf{Z}$ . Denote  $U_{\tau,0} = \boldsymbol{\alpha}^\top(\boldsymbol{\tau}_0)\mathbf{X}$  the true single-index variable. For  $l = 1, 2$ , let  $\rho_l(m) = \{dg^{-1}(m)/dm\}^l / (\sigma^2 V\{g^{-1}(m)\})$ . Further let

$$\begin{aligned} \Upsilon(u_{\tau,0}) &= \frac{E[\mathbf{X}\rho_2\{m_0(\mathbf{T})\}|U_{\tau,0} = u_{\tau,0}]}{E[\rho_2\{m_0(\mathbf{T})\}|U_{\tau,0} = u_{\tau,0}]}, \\ \Gamma(u_{\tau,0}) &= \frac{E[\mathbf{Z}\rho_2\{m_0(\mathbf{T})\}|U_{\tau,0} = u_{\tau,0}]}{E[\rho_2\{m_0(\mathbf{T})\}|U_{\tau,0} = u_{\tau,0}]}, \\ \Phi(\mathbf{x}) &\equiv \Phi(u_{\tau,0}, \mathbf{x}) = \mathbf{x} - \Upsilon(u_{\tau,0}), \quad \Psi(\mathbf{z}) \equiv \Psi(u_{\tau,0}, \mathbf{z}) = \mathbf{z} - \Gamma(u_{\tau,0}). \end{aligned}$$

Recall for model identifiability, we assume  $\|\boldsymbol{\alpha}\| = 1$  and the first element of  $\boldsymbol{\alpha}$  is positive. Following [36], we re-parameterize  $\boldsymbol{\alpha}$  using the “delete-one-component” to handle this constrain. Let  $\boldsymbol{\tau} = (\tau_1, \tau_2, \dots, \tau_{d_1-1})^\top$  and  $\boldsymbol{\alpha}(\boldsymbol{\tau}) = (\sqrt{1 - \|\boldsymbol{\tau}\|^2}, \boldsymbol{\tau}^\top)^\top$ . Note that  $\boldsymbol{\tau}$  is unconstrained and one dimension lower than  $\boldsymbol{\alpha}$ . The Jacobian matrix of  $\boldsymbol{\alpha}(\boldsymbol{\tau})$  of dimension  $d_1 \times (d_1 - 1)$  is

$$\mathbf{J}(\boldsymbol{\tau}) = \begin{bmatrix} -(1 - \|\boldsymbol{\tau}\|^2)^{-1/2} \boldsymbol{\tau}^\top \\ \mathbf{I}_{(d_1-1) \times (d_1-1)} \end{bmatrix}. \tag{4.1}$$

Another typical “delete-one-component” re-parameterization is  $\boldsymbol{\alpha}(\boldsymbol{\tau}) = \frac{(1, \boldsymbol{\tau}^\top)^\top}{\sqrt{1 + \|\boldsymbol{\tau}\|^2}}$  with the Jacobian matrix

$$\mathbf{J}(\boldsymbol{\tau}) = \begin{pmatrix} -(1 + \|\boldsymbol{\tau}\|^2)^{-3/2} \boldsymbol{\tau}^\top \\ (1 + \|\boldsymbol{\tau}\|^2)^{-1/2} \mathbf{I}_{(d_1-1) \times (d_1-1)} - (1 + \|\boldsymbol{\tau}\|^2)^{-3/2} \boldsymbol{\tau} \boldsymbol{\tau}^\top \end{pmatrix}. \tag{4.2}$$

Finally, let

$$\mathbf{R}(\boldsymbol{\tau}) = \begin{bmatrix} \mathbf{J}(\boldsymbol{\tau}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{d_2 \times d_2} \end{bmatrix} \tag{4.3}$$

be the Jacobian matrix of  $(\boldsymbol{\alpha}^\top(\boldsymbol{\tau}), \boldsymbol{\beta}^\top)^\top$ , which is of dimension  $(d_1 + d_2) \times (d_1 + d_2 - 1)$ .

**Theorem 2.** Under Conditions (C1)–(C8) in Appendix A.1, the constrained quasi-likelihood estimators  $\widehat{\boldsymbol{\alpha}}$  and  $\widehat{\boldsymbol{\beta}}$  with  $\|\widehat{\boldsymbol{\alpha}}\| = 1$  is asymptotically normally distributed, that is,

$$\sqrt{n} \begin{pmatrix} \widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 \\ \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \end{pmatrix} \longrightarrow N\{0, \mathbf{R}(\boldsymbol{\tau}_0)\mathbf{D}^{-1}\mathbf{R}^\top(\boldsymbol{\tau}_0)\},$$

where  $\mathbf{R}(\cdot)$  is given in (4.3) and

$$\mathbf{D} = E \left[ \rho_2\{m_0(\mathbf{T})\} \begin{pmatrix} \eta'_0(U_{\boldsymbol{\tau},0})\mathbf{J}^\top(\boldsymbol{\tau}_0)\boldsymbol{\Phi}(\mathbf{X}) \\ \boldsymbol{\Psi}(\mathbf{Z}) \end{pmatrix} \begin{pmatrix} \eta'_0(U_{\boldsymbol{\tau},0})\mathbf{J}^\top(\boldsymbol{\tau}_0)\boldsymbol{\Phi}(\mathbf{X}) \\ \boldsymbol{\Psi}(\mathbf{Z}) \end{pmatrix}^\top \right]. \quad (4.4)$$

### 4.3. Semiparametric efficient score and information bound

In this section, we derive the semiparametric efficient score and the information bound for the semiparametric model (2.1) and (2.2).

**Theorem 3.** We assume that the joint density of the random vector  $\mathbf{T} = (\mathbf{X}^\top, \mathbf{Z}^\top)^\top$  exists and is completely unknown. Under Conditions (C1)–(C8) in Appendix A.1, the efficient score for estimating  $\boldsymbol{\alpha}_0$  and  $\boldsymbol{\beta}_0$  is given by

$$i^* = \{Y - \mu(\mathbf{X}, \mathbf{Z})\} \rho_1\{m_0(\mathbf{T})\} \begin{pmatrix} \eta'_0(U_{\boldsymbol{\tau},0})\mathbf{J}^\top(\boldsymbol{\tau}_0)\boldsymbol{\Phi}(\mathbf{X}) \\ \boldsymbol{\Psi}(\mathbf{Z}) \end{pmatrix}.$$

Thus, the information bound for  $\boldsymbol{\alpha}_0$  and  $\boldsymbol{\beta}_0$  equals  $\mathbf{D}$  in (4.4).

## 5. Computational algorithm

The above estimation approach can be easily implemented with the existing GLM in any statistics software. To impose the constraints  $\|\boldsymbol{\alpha}\|$  and  $\alpha_1 > 0$ , we consider the re-parametrization of  $\boldsymbol{\alpha}$ ,  $\boldsymbol{\alpha}(\boldsymbol{\tau})$ , given in Section 4.2. Let  $\boldsymbol{\theta}_\tau = (\boldsymbol{\tau}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$ . Let  $\mathbf{H}(\boldsymbol{\theta}_\tau) = \frac{\partial^2 \ell(\boldsymbol{\theta}_\tau)}{\partial \boldsymbol{\theta}_\tau \partial \boldsymbol{\theta}_\tau^\top}$  be the Hessian (second derivative) matrix of the quasi-likelihood in (3.2) and let  $\mathbf{S}(\boldsymbol{\theta}_\tau) = \frac{\partial \ell(\boldsymbol{\theta}_\tau)}{\partial \boldsymbol{\theta}_\tau}$  be the corresponding gradient vector.

Denote

$$\boldsymbol{\xi}_i(\boldsymbol{\tau}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \begin{pmatrix} \boldsymbol{\gamma}^\top \mathbf{B}'(\boldsymbol{\alpha}(\boldsymbol{\tau})^\top \mathbf{X}_i) \mathbf{J}^\top(\boldsymbol{\tau}) \mathbf{X}_i \\ \mathbf{Z}_i \\ \mathbf{B}(\boldsymbol{\alpha}(\boldsymbol{\tau})^\top \mathbf{X}_i) \end{pmatrix},$$

where  $\mathbf{J}(\boldsymbol{\tau})$  is the Jacobian matrix of  $\boldsymbol{\alpha}(\boldsymbol{\tau})$  of dimension  $d_1 \times (d_1 - 1)$ ; see (4.1) and (4.1), for example. Then the gradient vector

$$\mathbf{S}(\boldsymbol{\theta}_\tau) = \frac{1}{n} \sum_{i=1}^n \{Y_i - g^{-1}(m_i)\} \rho_1(m_i, Y_i) \boldsymbol{\xi}_i(\boldsymbol{\tau}, \boldsymbol{\beta}, \boldsymbol{\gamma}),$$

where  $m_i = \boldsymbol{\gamma}^\top \mathbf{B}(\boldsymbol{\alpha}^\top \mathbf{X}_i) + \boldsymbol{\beta}^\top \mathbf{Z}_i$ , and the Hessian matrix

$$\mathbf{H}(\boldsymbol{\theta}_\tau) = -\frac{1}{n} \sum_{i=1}^n \rho_2(m_i, Y_i) \boldsymbol{\xi}_i(\boldsymbol{\tau}, \boldsymbol{\beta}, \boldsymbol{\gamma}) \boldsymbol{\xi}_i^\top(\boldsymbol{\tau}, \boldsymbol{\beta}, \boldsymbol{\gamma}).$$

The Fisher scoring update equations become

$$\begin{aligned} \boldsymbol{\theta}_\tau^{(k+1)} &= \boldsymbol{\theta}_\tau^{(k)} + \left\{ \sum_{i=1}^n \rho_2^{(k)} \boldsymbol{\xi}_i(\boldsymbol{\tau}^{(k)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)}) \boldsymbol{\xi}_i^\top(\boldsymbol{\tau}^{(k)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)}) \right\}^{-1} \\ &\times \sum_{i=1}^n \rho_1^{(k)} (Y_i - \mu_i^{(k)}) \boldsymbol{\xi}_i(\boldsymbol{\tau}^{(k)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{\gamma}^{(k)}). \end{aligned} \tag{5.1}$$

### 5.1. Initial values selection

Unlike [2], we do not require root- $n$  consistent pilot estimators in our procedure. Our algorithm usually converges within a few steps for reasonable initial values. We recommend repeating the algorithm by using a variety of initial values.

To obtain a reasonable initial value of  $\boldsymbol{\theta}_\tau^{(0)}$ , we generate i.i.d. standard normal random vectors  $\{\boldsymbol{\alpha}_b, 1 \leq b \leq B\}$  (say  $B = 200$ ). Normalize  $\boldsymbol{\alpha}_b$  such that  $\|\boldsymbol{\alpha}_b\| = 1$ ,  $b = 1, 2, \dots, B$ , and impose the constraint that its first element positive. Obtain  $\boldsymbol{\beta}_b$  and  $\boldsymbol{\gamma}_b$  by maximizing the quasi-likelihood  $Q_{n,b}(\boldsymbol{\beta}_b, \boldsymbol{\gamma}_b; U_b)$  with index values  $\mathbf{U}_b = \{U_{i,b}\}_{i=1}^n = \{\boldsymbol{\alpha}_b^\top \mathbf{X}_i\}_{i=1}^n$ . Let  $(\boldsymbol{\alpha}_{b^0}, \boldsymbol{\beta}_{b^0}, \boldsymbol{\gamma}_{b^0}) = \operatorname{argmax}_{b: 1 \leq b \leq B} Q_{n,b}(\boldsymbol{\beta}_b, \boldsymbol{\gamma}_b; \mathbf{U}_b)$ , then the initial value is set as  $\boldsymbol{\theta}_\tau^{(0)} = (\boldsymbol{\tau}_{b^0}^\top, \boldsymbol{\beta}_{b^0}^\top, \boldsymbol{\gamma}_{b^0}^\top)^\top$ , where  $\boldsymbol{\tau}_{b^0} = \boldsymbol{\tau}(\boldsymbol{\alpha}_{b^0})$ , and  $\boldsymbol{\tau}(\cdot)$  is the inverse map of  $\boldsymbol{\alpha}(\boldsymbol{\tau})$ .

### 5.2. Knots selection

The spline approximation for the function  $\eta(\cdot)$  requires an appropriate selection of the knot sequences. Note that the distribution of  $\boldsymbol{\alpha}^\top \mathbf{X}$  is unknown, deciding where to place knots can become problematic. To solve the problem, we first transform the index variable,  $\boldsymbol{\alpha}^\top \mathbf{X}$  using a re-scaled centered Beta $\{(d_1 + 1)/2, (d_1 + 1)/2\}$  cumulative distribution function:

$$F^*(u) = \int_{-1}^{u/\chi} \frac{\Gamma(d_1 + 1)}{\Gamma\{(d_1 + 1)/2\}^2 2^{d_1}} (1 - t^2)^{(d_1-1)/2} dt,$$

for  $u = \boldsymbol{\alpha}^\top \mathbf{x}$  with  $\|\mathbf{x}\| \leq \chi$ . Wang and Yang [30] has shown that the probability density function of the transformed index variable is bounded below and above uniformly for all  $\boldsymbol{\alpha}$ . After the above transformation, one can simply adopt the equally-spaced knots when applying spline smoothing.

Next, we choose the number of knots,  $J \equiv J_n$ , by minimizing the following generalized cross-validation (GCV) criterion

$$\operatorname{GCV}(J) = \frac{n \sum_{i=1}^n D(Y_i, \hat{\mu}_i; J)}{\{n - \operatorname{tr}(\mathbf{S}(J))\}^2}, \tag{5.2}$$

where  $D(Y, \mu; J)$  is the deviance of  $Y$  corresponding to fitting with  $J$  knots, and

$$\mathbf{S}(J) = \hat{\boldsymbol{\xi}} (\hat{\boldsymbol{\xi}}^\top \mathbf{W} \hat{\boldsymbol{\xi}})^{-1} \hat{\boldsymbol{\xi}}^\top \mathbf{W}$$

is the influence (hat) matrix with  $\hat{\boldsymbol{\xi}} = (\hat{\boldsymbol{\xi}}_1, \dots, \hat{\boldsymbol{\xi}}_n)^\top$ , and  $\mathbf{W}$  is a diagonal matrix such that  $W_{ii} = \rho_{2i}$ . Based on the study in [36] and [29], 2 to 12 knots seem to be a good range to evaluate the above GCV, and the number of interior knots  $J_n$  selected is the one with minimum GCV value.

We also investigate another user friendly knots selection method where the number of interior knots follows

$$J_n = Cn^{1/(2r)} \log(n), \quad (5.3)$$

where  $r$  is the order of B-spline basis function and  $C \in [0.3, 1]$  is a tuning constant whose default value is 0.6 in our simulation.

Our simulation shows that the estimation results using (5.3) are similar to those using the GCV method. In the supplementary material, additional simulation works demonstrate that the estimation outcomes are not very sensitive to the constant  $C$  in the range from  $[0.3, 1]$ .

## 6. Numerical examples

In this section, we illustrate our method by evaluating its performance on the simulated data and in two real data analysis.

### 6.1. Simulation

We first conducted some simulation studies to study the finite sample behavior of the proposed estimators for GPLSIM using a design similar to [2]. We examined the performance with two popular link functions: identity link and logit link.

*Case I: Identity link function.* Following [2], we generate the data according to the ‘‘sine-bump’’ model  $Y_i = \sin\{\pi(\boldsymbol{\alpha}^\top \mathbf{X}_i - A)/(B - A)\} + \beta Z_i + \varepsilon_i$ , where  $\mathbf{X}_i$ 's are trivariate with independent uniform  $(-0.5, 0.5)$  components,  $Z_i = 0$  for  $i$  odd and  $Z_i = 1$  for  $i$  even, and  $\varepsilon_i$ 's are normally distributed with mean 0 and variance 0.01. The parameters are set as  $\boldsymbol{\alpha} = (1, 1, 1)^\top / \sqrt{3}$  and  $\beta = 0.3$ . We take  $A = \sqrt{3}/2 - 1.645/\sqrt{12}$  and  $B = \sqrt{3}/2 + 1.645/\sqrt{12}$  to ensure that the design was relatively thick in the tails. We run 200 replications with sample sizes  $n = 100, 200$  and 500.

*Case II: Logit link function.* The data are generated from model  $\text{logit}\{P(Y_i = 1 | \mathbf{X}_i, Z_i)\} = \sin\{\pi(\boldsymbol{\alpha}^\top \mathbf{X}_i - A)/(B - A)\} + \beta Z_i + \varepsilon_i$ , where all the parameters and variables are defined in the same way as in Case I. To mimic the Framingham Data and Munich Data, we try larger sample size,  $n = 1000, 1500$  and 2000, in the example. For each sample size, we run 200 replications.

We apply our proposed method to estimate the GPLSIM with two knots selection mechanisms: formula (5.3) and GCV (methods GPLSIM1 and GPLSIM2, respectively). We compare the performances of the proposed method with the penalized spline method in [36] (method GPLSIM3), and the generalized linear model (method GLM). For a fair comparison, in GLM



we set  $g\{\mu(\mathbf{x}, \mathbf{z})\} = c\boldsymbol{\alpha}^\top \mathbf{x} + \boldsymbol{\beta}^\top \mathbf{z}$  with the constraints  $\|\boldsymbol{\alpha}\| = 1$  and  $\alpha_1 > 0$ . For GPLSIM3, the values of the tuning parameter are selected by the GCV as suggested in [36].

To assess the estimation accuracy, we summarize the results using the estimated bias (Bias), standard errors (SD) and mean squared error (MSE) for both cases. Tables 1 and 2 report the estimation results for cases I and II with four different methods. For the logit case, as the penalized spline method in [36] is only developed for the identity link, so we are not able to obtain the GPLSIM3 in this case. Both tables correspondingly indicate the consistency of  $\hat{\boldsymbol{\alpha}}$  and  $\hat{\boldsymbol{\beta}}$  as the bias, SD and MSE decrease as the sample size increasing. Three algorithms for GPLSIM have comparable performance. The knots selection with formula  $Cn^{1/(2r)} \log(n)$  by using  $C \in [0.3, 1]$  is close to the knots selection using the GCV in the range of 2 to 10. The simulation results of GPLSIM1 presented in Tables 1 and 2 are based on  $C = 0.6$ . Additional simulation studies with  $C = 0.4$  and  $0.8$  are given in supplementary material; see Tables S.1 and S.2. Compared with GLM, the GPLSIM estimators significantly over-perform the GLM regardless of sample sizes and link functions. The poor performance of GLM is because it assumes the single-index function is linear, thus, completely misspecifies the model.

Figure 1 depicts the estimated mean functions of  $\eta$  in one replication with  $n = 200$  for case I. One can see that the algorithm described in Section 3 works very well in fitting the data as the fit is very close to the true mean function.

Finally, we evaluate the computing cost of different methods. The last columns in Tables 1 and 2 report the computing time of one simulation example on an ordinary x64 PC with Intel Dual Core i5. If using formula (5.3) to select the knots, it takes about one second to compute the GPLSIM1 in one run of simulation with 200 observations. It takes longer if using the GCV to select knots, but usually one is able to obtain the result within one minute even for a sample size of 2000. Overall, the proposed algorithm is fast to compute and usually converges in 5 iterations.

## 6.2. Example: Framingham data

In this example, we consider the Framingham dataset studied by [2] to illustrate an application of our proposed method. Cardiovascular disease (CVD) is the leading cause of death and serious illness in the United States. The objective of the Framingham Heart Study aims at identifying the common factors that contribute to CVD by following its development over a long period of time in a large group of participants who had not yet developed overt symptoms of CVD or suffered a heart attack or stroke. For the purpose of illustration, we use Exam #3 as the baseline [2]. The dataset includes 1615 men age 31–65, with the outcome indicating the occurrence of coronary heart disease (CHD).

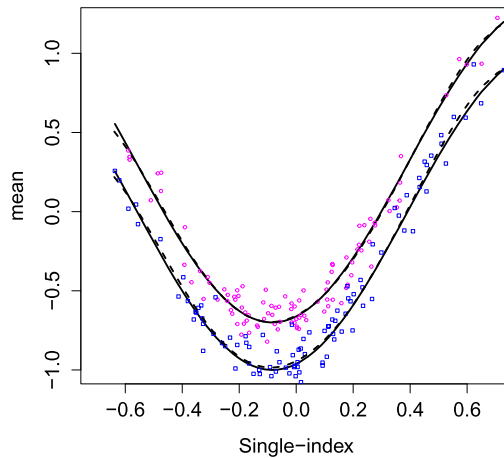
Let the response  $Y$  be the incidence of CHD. Predictors used in this example are patient's age, systolic blood pressure (SBP), serum cholesterol level, and smoking status. Each variable is scaled to lie between 0 and 1. Following [2], we fit the data with GPLSIM by using the continuous variables as the single index components, including  $X_1 = \text{age}$ ,  $X_2 = \text{trblood}$  (transformed systolic blood pressure), and  $X_3 = \text{logchol}$  (log of serum cholesterol). The dummy variable, "smoker" (smoking status), enters the model as the linear component naturally. We consider the

**Table 1.** GPLSIM results of simulation study: Case I (identity link function)

$n$		$\alpha_1$			$\alpha_2$			$\alpha_3$			$\beta$			Time
		Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE	
100	GPLSIM1	0.0028	0.0160	0.0003	0.0021	0.0171	0.0003	0.0001	0.0150	0.0002	-0.0016	0.0241	0.0006	0.41
	GPLSIM2	-0.0034	0.0159	0.0003	0.0004	0.0165	0.0003	-0.0022	0.0169	0.0003	0.0007	0.0216	0.0005	4.98
	GPLSIM3	-0.0016	0.0158	0.0003	0.0002	0.0154	0.0002	0.0007	0.0168	0.0003	-0.0015	0.0213	0.0005	4.13
	GLM	-0.0417	0.0570	0.0263	0.1385	0.0861	0.0266	-0.2518	0.1649	0.0904	0.0005	0.0892	0.0079	0.00
200	GPLSIM1	-0.0003	0.0104	0.0001	-0.0007	0.0104	0.0001	0.0007	0.0102	0.0001	0.0004	0.0159	0.0003	0.64
	GPLSIM2	-0.0006	0.0100	0.0001	-0.0003	0.0112	0.0001	0.0006	0.0101	0.0001	0.0007	0.0165	0.0003	6.94
	GPLSIM3	-0.0003	0.0103	0.0001	0.0000	0.0094	0.0001	0.0000	0.0089	0.0001	0.0018	0.0147	0.0002	6.68
	GLM	0.0183	0.0984	0.0100	0.1026	0.0614	0.0143	-0.2160	0.1137	0.0595	0.0019	0.0571	0.0032	0.00
500	GPLSIM1	0.0001	0.0064	0.0000	-0.0004	0.0068	0.0000	0.0002	0.0070	0.0000	-0.0001	0.0087	0.0001	1.48
	GPLSIM2	-0.0011	0.0066	0.0000	0.0004	0.0065	0.0000	0.0005	0.0068	0.0000	-0.0006	0.0090	0.0001	13.69
	GPLSIM3	0.0001	0.0062	0.0000	-0.0008	0.0062	0.0000	0.0000	0.0057	0.0000	-0.0003	0.0092	0.0001	14.48
	GLM	-0.0021	0.0270	0.0007	0.0674	0.0367	0.0059	-0.2233	0.0765	0.0557	0.0060	0.0366	0.0014	0.00

**Table 2.** GPLSIM results of simulation study: Case II (logit link function)

$n$		$\alpha_1$			$\alpha_2$			$\alpha_3$			$\beta$			Time
		Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE	
1000	GPLSIM1	-0.0587	0.1350	0.0216	-0.0329	0.1334	0.0188	0.0258	0.1875	0.0357	0.0190	0.1397	0.0198	3.72
	GPLSIM2	-0.0179	0.1224	0.0152	-0.0237	0.1164	0.0140	0.0029	0.1243	0.0154	-0.0013	0.1229	0.0150	24.99
	GLM	-0.1404	0.2862	0.1012	0.2141	0.0998	0.0557	-0.3361	0.2215	0.1618	0.0486	0.1196	0.0166	0.00
1500	GPLSIM1	-0.0315	0.1038	0.0117	-0.0116	0.1007	0.0102	0.0154	0.0989	0.0100	0.0001	0.1091	0.0119	5.83
	GPLSIM2	-0.0122	0.0944	0.0090	-0.0093	0.0948	0.0090	-0.0019	0.0945	0.0089	-0.0031	0.1139	0.0129	36.46
	GLM	-0.1156	0.2569	0.0790	-0.2102	0.0934	0.0529	-0.3739	0.1836	0.1733	0.0444	0.1065	0.0133	0.01
2000	GPLSIM1	-0.0262	0.0953	0.0097	-0.0028	0.0748	0.0056	0.0084	0.0923	0.0085	0.0078	0.0988	0.0098	8.02
	GPLSIM2	-0.0115	0.0754	0.0058	-0.0012	0.0837	0.0070	-0.0042	0.0821	0.0067	0.0094	0.0877	0.0077	50.37
	GLM	-0.0684	0.1550	0.0286	0.2025	0.0910	0.0492	-0.3364	0.1630	0.1369	0.0335	0.0832	0.0080	0.01



**Figure 1.** Curve estimates for a single replication of the sine-bump function. The solid curves are the true functions. The dashed curves are the estimated mean functions from GPLSIM for  $Z = 0$  and  $Z = 1$ . The data are shown by squares:  $Z = 0$  and circles:  $Z = 1$ .

following GPLSIM for the mean response

$$\text{logit}\{P(\text{CHD}|\mathbf{X}, Z)\} = \eta_0(\boldsymbol{\alpha}_0^T \mathbf{X}) + \beta_0 Z,$$

specified as

$$\begin{aligned} \text{logit}\{P(\text{disease} = 1 | \text{age, trblood, logchol, smoker})\} \\ = \eta_0(\alpha_{01}(\text{age}) + \alpha_{02}(\text{trblood}) + \alpha_{03}(\text{logchol})) + \beta_0(\text{smoker}). \end{aligned}$$

We apply our proposed polynomial spline GPLSIM method to the data and compare it with the local polynomial quasi-likelihood GPLSIM method in [2]. As suggested in [2], to avoid problems with sparse data near the boundaries, we use only those data with a single-index value in range  $[0.4, 1.2]$  for estimation. The resulted data set is slightly different from that in [2], where 45 of the 1615 observations are excluded.

For our method, we fit the model using cubic splines and 3 interior knots selected according to the GCV criterion in (5.2). We also compare the GPLSIM with the GLM and the generalized additive model (GAM). In GLM,  $\eta_0$  is assumed to be a constant multiple of identity function, with the constant determined by the constraints  $\|\boldsymbol{\alpha}\| = 1$  and  $\alpha_1 > 0$  to make the estimation  $\hat{\boldsymbol{\alpha}}$  comparable to estimates from a single-index model. In GAM, the smoking status enters as a linear term and  $X_1, X_2, X_3$  enter as nonparametric additive terms.

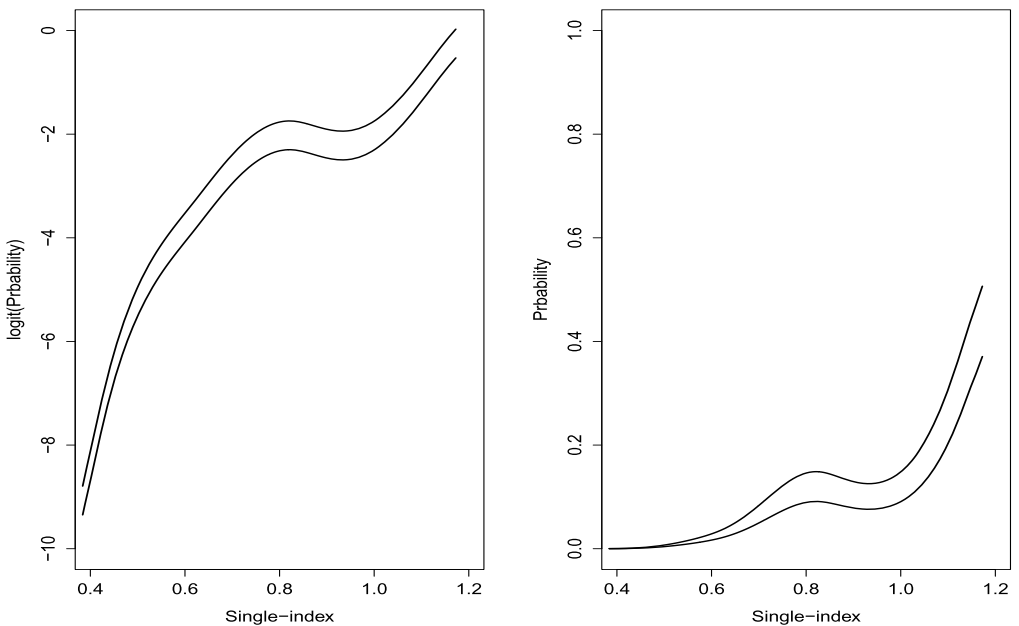
Table 3 presents the estimation results of each coefficients and their standard errors. The results for the GPLSIM-local polynomial method are adapted from [2]. From Table 3, we observe that the polynomial spline estimates for GPLSIM are fairly similar to the local polynomial GPLSIM estimates. To evaluate the model fits, we also provide the generalized cross validation deviance

**Table 3.** Framingham Heart Study Data

Model	Method		age	trblood	logchol	smoker	GCV
GPLSIM	Local Polynomial	Estimate	0.370	0.650	0.660	0.590	0.513
		SD	0.086	0.110	0.120	0.240	
	Polynomial Spline	Estimate	0.370	0.660	0.654	0.553	
		SD	0.058	0.094	0.045	0.252	
GLM		Estimate	0.416	0.597	0.686	0.594	0.548
		SD	0.087	0.150	0.181	0.253	
GAM		Estimate	–	–	–	0.570	0.524
		SD	–	–	–	0.325	

(GCV) for each method in Table 3. In terms of GCV, both the GPLSIM and GAM over-perform GLM.

Figure 2 displays the estimates of  $\eta_0$  and the conditional probability of heart disease for both smokers and nonsmokers. An interesting feature of this figure is the curvature of the  $\eta_0(\cdot)$



**Figure 2.** Curve estimates for the Framingham Heart Study Data. Left: Curve estimates of the logit  $P(\text{heart disease})$  for the smokers (upper curve) and nonsmokers (lower curve) against the estimated single-index described in the text. Right: Upper (lower) curve correspond to the estimates of  $P(\text{heart disease})$  for smokers (nonsmokers) against the estimated single-index described in the text.

and logit function when the single index is larger than 0.8. Similar findings are also discussed in [2].

### 6.3. Example: Munich dust study data

In this example, we use the Munich Dust data analyzed in [2] for another illustration of our method. It is important to assess the health hazard of specific harmful substances in occupational medicine. In this data analysis, we consider a specific problem of modeling risk of bronchitis in a dust burdened mechanical engineering plant in Munich [2].

The regressor variable  $X_1$  (trdust) is the logarithm of 1.0 plus the average dust concentration in the working area over the period of time in question;  $X_2$  (duration) is the duration of exposure; and  $Z$  (smoker) is the smoking status. There are 1246 observations: 23% of the workers reported chronic bronchitis and 74% are smokers. The data were analyzed by [17] as an example of segmented GLM with additive measurement error and were further explored by [2]. Following [2], we conduct our analysis using the following GPLSIM

$$\begin{aligned} & \text{logit}\{P(\text{bronchitis} = 1 \mid \text{trdust}, \text{duration}, \text{smoker})\} \\ & = \eta(\alpha_{01}(\text{trdust}) + \alpha_{02}(\text{duration})) + \beta_0(\text{smoker}), \end{aligned}$$

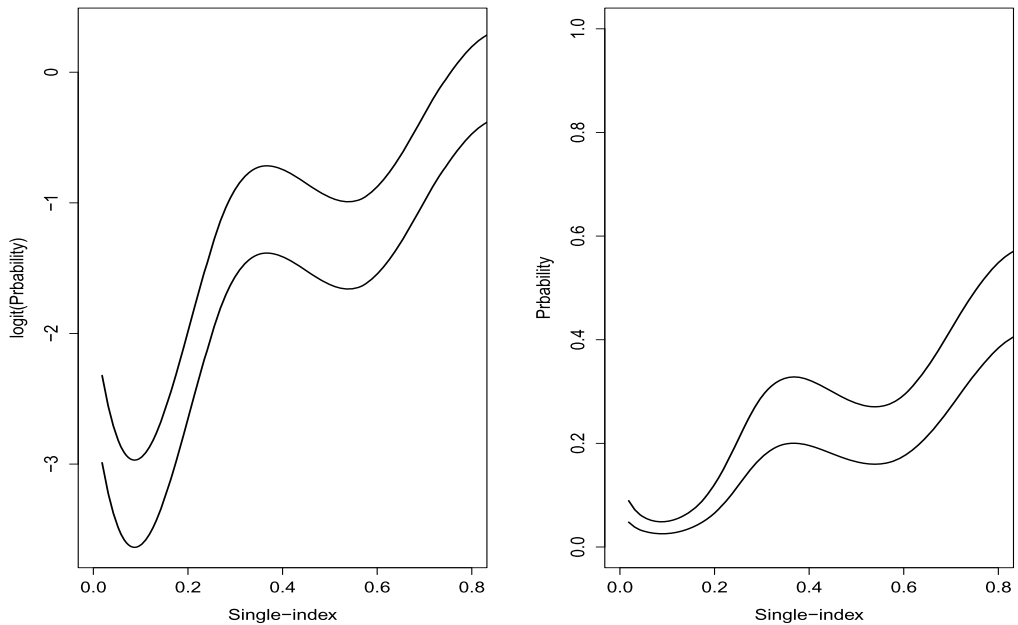
with cubic splines and 4 interior knots selected according to the GCV criterion in (5.2).

Table 4 presents the model fitting results of GPLSIM, GLM and GAM. In GAM, the smoking status enters as a linear term and  $X_1, X_2$  enter as nonparametric additive terms. The results for the local polynomial quasi-likelihood method are adapted from [2]. We conclude from Table 4 that the GPLSIM and GAM fit the data better than the GLM in terms of GCV. In addition, we observe that the polynomial spline GPLSIM fit is virtually similar to the local linear GPLSIM fit in [2].

Figure 3 shows the logit and probability of bronchitis for smokers and nonsmokers, and [2] gives analogous results. Two important curvatures in these data are depicted in Figure 3, which can not be well fitted by GLM. Küchenhoff and Carroll [17] suggested that the second curvature

**Table 4.** Munich Dust Study Data

Model	Method		trdust	duration	smoker	GCV
GPLSIM	Local Polynomial	Estimate	0.222	0.975	0.668	1.020
		SD	0.089	0.021	0.178	
	Polynomial Spline	Estimate	0.207	0.978	0.673	1.020
		SD	0.030	0.006	0.176	
GLM		Estimate	0.397	0.918	0.682	1.052
		SD	0.104	0.142	0.174	
GAM		Estimate	–	–	0.714	1.021
		SD	–	–	0.180	



**Figure 3.** Curve estimates for the Munich Dust Study Data. Left: Curve estimates of the  $\text{logit}P(\text{bronchitis})$  for the smokers (upper curve) and nonsmokers (lower curve) against the estimated single-index described in the text. Right: Upper (lower) curve correspond to the estimates of  $P(\text{bronchitis})$  for smokers (nonsmokers) against the estimated single-index described in the text.

around single index value 0.6 may reflect a threshold effect a threshold concentration, under which there is no risk due to the substance.

## 7. Conclusions and discussions

In this paper, we have proposed an effective estimation method for the GPLSIM based on a combination of the polynomial spline smoothing and quasi-likelihood. The contributions we made to the existing literature can be summarized in three ways: (i) the procedures are computationally efficient through and stable through “delete-one-component” re-parameterization, spline approximation and Fisher scoring algorithm; (ii) the estimators of the coefficients in the single-index component and linear component are asymptotically normal and achieve the semiparametric information bound under some regularity conditions; and (iii) all the parameters in the model can be estimated simultaneously, and we do not require any root- $n$  consistent pilot estimator for our estimation. The proposed method greatly enhances the application of GPLSIMs to practical data analysis. Both the simulation study and empirical data analysis show that the proposed methods work well for finite samples.

Recently, extensive literatures have developed new semiparametric estimation methods for longitudinal regression models. For example, [18] considers the PLSIM with longitudinal data

and proposes the bias-corrected quadratic inference function method. Chen *et al.* [4] studies the same model and proposes a combination of the local linear smoothing and generalized estimation equations under a general framework which includes both the sparse and dense longitudinal data cases. These methods take into account the within-subject correlation information and thus generally improves the asymptotic estimation efficiency. Our methods can be extended to longitudinal data settings through marginal models or mixed-effects models. More work, however, is needed to understand the properties of the estimators in such new settings. Some associated theoretical properties with respect to model selection and estimation as well as inference need to be carefully investigated. We will leave this issue for future research.

## Appendix

Throughout the article, let  $\|\cdot\|$  be the Euclidean norm and  $\|\varphi\|_\infty = \sup_{a \leq m \leq b} |\varphi(m)|$  be the supremum norm of a function  $\varphi$  on  $[a, b]$ . For any matrix  $\mathbf{A}$ , denote its  $L_2$  norm as  $\|\mathbf{A}\|_2 = \sup_{\|\mathbf{x}\| \neq 0} \frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|}$ . Denote  $q_l(m, y) = \partial^l / \partial m^l Q\{g^{-1}(m), y\}$ , then

$$\begin{aligned} q_1(m, y) &= \partial / \partial m Q\{g^{-1}(m), y\} = \{y - g^{-1}(m)\} \rho_1(m), \\ q_2(m, y) &= \partial^2 / \partial m^2 Q\{g^{-1}(m), y\} = \{y - g^{-1}(m)\} \rho'_1(m) - \rho_2(m), \end{aligned}$$

where  $\rho_l(m) = \{dg^{-1}(m)/dm\}^l / (\sigma^2 V\{g^{-1}(m)\})$ . For easy asymptotic analysis, we adopt the normalized B-spline space  $S_n^0$  introduced in [35] with the following normalized basis

$$B_{j,r}(u) = \sqrt{N_n} \left\{ b_{j+1,r}(u) - \frac{E(b_{j+1,r})}{E(b_{1,r})} b_{1,r}(u) \right\}, \quad 0 \leq j \leq N_n - 1. \quad (\text{A.1})$$

### A.1. Assumptions

The following are some conditions to obtain Theorems 1 and 2. Let  $v$  be a positive integer, let  $\alpha \in (0, 1]$  be such that  $p = v + \alpha > 1.5$ . Let  $\mathcal{H}(p)$  be the collection of functions  $g$  on  $[a, b]$  whose  $v$ th derivative,  $g^{(v)}$ , exists and satisfies the Lipschitz condition of order  $\alpha$ ,

$$|g^{(v)}(m') - g^{(v)}(m)| \leq C |m' - m|^\alpha, \quad \text{for } a \leq m', m \leq b.$$

Let  $\varepsilon = Y - g^{-1}(m_0(\mathbf{T}))$ .

- (C1) The function  $\eta_0 \in \mathcal{H}(p)$ .
- (C2) For  $m \in R$  and  $y$  in the range of the response variable, the function  $q_2(m, y) < 0$ , and  $c_q < |q_2^k(m, y)| < C_q$  ( $k = 1, 2$ ), for some positive constants  $c_q$  and  $C_q$ .
- (C3) The  $v$ th order partial derivatives of the joint density function of  $\mathbf{X}$  satisfy the Lipschitz condition of order  $\alpha$  ( $\alpha \in (0, 1]$ ). The marginal density function of  $\alpha^\top \mathbf{X}$  is continuous and bounded away from zero and infinity on its support  $[a, b]$ .

(C4) For any  $\tau$ , there exist positive constants  $c_\tau$  and  $C_\tau$  such that

$$c_\tau \mathbf{I}_{d \times d} \leq E \left\{ \begin{pmatrix} 1 \\ \mathbf{T} \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{T} \end{pmatrix}^\top \middle| \boldsymbol{\alpha}^\top(\tau) \mathbf{X} = \boldsymbol{\alpha}^\top(\tau) \mathbf{x} \right\} \leq C_\tau \mathbf{I}_{d \times d},$$

where  $d = 1 + d_1 + d_2$ .

(C5) The number of knots  $n^{1/2(p+1)} \ll N_n \ll n^{1/8}$  ( $p > 3$ ).

(C6) For some constants  $0 < C_\rho, C_\rho^*, M_0 < \infty$ ,

$$|\rho_1(m_0)| \leq C_\rho \quad \text{and} \quad |\rho_1(m) - \rho_1(m_0)| \leq C_\rho^* |m - m_0| \quad \text{for all } |m - m_0| \leq M_0.$$

(C7) For some constants  $0 < C_g, C_g^*, M_1 < \infty$ , the link function  $g$  in model (2.2) satisfies that  $|\frac{d}{dm} g(m)|_{m=m_0} \leq C_g$  and

$$\left| \frac{d}{dm} g^{-1}(m) - \left. \frac{d}{dm} g^{-1}(m) \right|_{m=m_0} \right| \leq C_g^* |m - m_0| \quad \text{for all } |m - m_0| \leq M_1.$$

(C8) There exists a positive constant  $C_0$ , such that  $E(\varepsilon^2 | U_{\tau,0}) \leq C_0$ , almost surely.

The smoothness condition in (C1) describes a requirement on the best rate of convergence that the single-index function  $\eta_0(\cdot)$  can be approximated by functions in  $\mathcal{S}_n^0$ , where  $\mathcal{S}_n^0$  is B-spline space with the normalized basis (A.1). Condition (C2) is imposed to ensure the uniqueness of the solution; see, for example, Condition 1a of [2] and Condition (C2) of [28]. Condition (C3) requires a smoothness condition on the joint and marginal density functions of the covariates, which is often assumed in asymptotic analysis of nonparametric regression problems; see Assumption A2 in [30]. Condition (C4) implies that the eigenvalues of

$$E \left\{ \begin{pmatrix} 1 \\ \mathbf{T} \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{T} \end{pmatrix}^\top \middle| \boldsymbol{\alpha}^\top(\tau) \mathbf{X} = \boldsymbol{\alpha}^\top(\tau) \mathbf{x} \right\}$$

are bounded away from 0 and  $\infty$ . Condition (C5) gives the rate of growth of the dimension of the spline spaces relative to the sample size. Conditions (C6) and (C7) describe some requirements on the link function  $g$ .

In Section A.2 below, we define the best spline approximation estimator and state its asymptotic distribution. In Sections A.3 and A.4, we prove Theorems 1 and 2, respectively. Lemmas S.1 to S.7 used in the proofs below are presented in the Supplementary Material.

### A.2. Best spline approximation estimator

According to the result of [5] (on 149 page), for any function  $g \in \mathcal{H}(p)$  and  $n \geq 1$ , there exists a function  $\tilde{g} \in \mathcal{S}_n$ , such that  $\|\tilde{g} - g\|_\infty \leq Ch^p$ , where  $C$  is some fixed positive constant. For  $\eta_0$  satisfying Condition (C1), we can find  $\tilde{\boldsymbol{\gamma}} = \{\tilde{\gamma}_j\}_{j=1}^N$  and a spline function  $\tilde{\eta} = \tilde{\boldsymbol{\gamma}}^\top \mathbf{B}(u) \in \mathcal{S}_n^0$ , such that  $\|\tilde{\eta} - \eta_0\|_\infty = O(h^p)$ .

In the following, denote  $\tilde{m}_0(\mathbf{t}) = \tilde{\eta}(\boldsymbol{\alpha}_0^\top \mathbf{x}) + \boldsymbol{\beta}_0^\top \mathbf{z}$ ,  $\tilde{m}_{0i} \equiv \tilde{m}_0(\mathbf{T}_i) = \tilde{\eta}(\boldsymbol{\alpha}_0^\top \mathbf{X}_i) + \boldsymbol{\beta}_0^\top \mathbf{Z}_i$ . Let  $(\tilde{\boldsymbol{\alpha}}^\top, \tilde{\boldsymbol{\beta}}^\top)^\top = \operatorname{argmax}_{\|\boldsymbol{\alpha}\|=1, \boldsymbol{\beta}} \frac{1}{n} \sum_{i=1}^n Q[g^{-1}\{\tilde{\eta}(\boldsymbol{\alpha}^\top \mathbf{X}_i) + \boldsymbol{\beta}^\top \mathbf{Z}_i\}, Y_i]$ .



**Lemma A.1.** Under Conditions (C1)–(C5),

$$\sqrt{n} \begin{pmatrix} \tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 \\ \tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \end{pmatrix} \rightarrow N(\mathbf{0}, \mathbf{R}(\boldsymbol{\tau}_0) \mathbf{A}^{-1} \boldsymbol{\Sigma}_1 \mathbf{A}^{-1} \mathbf{R}^\top(\boldsymbol{\tau}_0)),$$

where  $\mathbf{R}(\cdot)$  is given in (4.3),

$$\boldsymbol{\Sigma}_1 = E \left[ q_1^2 \{m_0(\mathbf{T})\} \begin{pmatrix} \eta'_0(U_{\boldsymbol{\tau},0}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \mathbf{X} \\ \mathbf{Z} \end{pmatrix} \begin{pmatrix} \eta'_0(U_{\boldsymbol{\tau},0}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \mathbf{X} \\ \mathbf{Z} \end{pmatrix}^\top \right],$$

and  $\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{12}^\top & \mathbf{A}_{22} \end{pmatrix}$  with

$$\mathbf{A}_{11} = E[\rho_2 \{m_0(\mathbf{T})\} \{\eta'_0(U_{\boldsymbol{\tau},0})\}^2 \mathbf{J}^\top(\boldsymbol{\tau}_0) \mathbf{X} \mathbf{X}^\top \mathbf{J}(\boldsymbol{\tau}_0)],$$

$$\mathbf{A}_{12} = E[\rho_2 \{m_0(\mathbf{T})\} \eta'_0(U_{\boldsymbol{\tau},0}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \mathbf{X} \mathbf{Z}^\top],$$

$$\mathbf{A}_{22} = E[\rho_2 \{m_0(\mathbf{T})\} \mathbf{Z} \mathbf{Z}^\top].$$

The proof of Lemma A.1 is given in the Supplementary Material.

### A.3. Proof of Theorem 1

According to Lemma S.5 in the Supplementary Material,

$$\begin{aligned} \|\hat{\eta}(u; \boldsymbol{\alpha}(\hat{\boldsymbol{\tau}})) - \tilde{\eta}(u)\|_2^2 &= \|(\hat{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}})^\top \mathbf{B}(u)\|_2^2 \\ &= (\hat{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}})^\top E \left\{ \frac{1}{n} \sum_{i=1}^n \mathbf{B}(u) \mathbf{B}^\top(u) \right\} (\hat{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}}) \\ &\leq C \|\hat{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}}\|_2^2, \end{aligned}$$

thus  $\|\hat{\eta} - \tilde{\eta}\|_2 = O_P\{N_n^{1/2}(h^p + n^{-1/2}h^{-1})\}$  and

$$\begin{aligned} \|\hat{\eta}(u; \boldsymbol{\alpha}(\hat{\boldsymbol{\tau}})) - \eta_0(u)\|_2 &\leq \|\hat{\eta}(u; \boldsymbol{\alpha}(\hat{\boldsymbol{\tau}})) - \tilde{\eta}(u; \boldsymbol{\alpha}(\hat{\boldsymbol{\tau}}))\|_2 + \|\tilde{\eta} - \eta_0\|_2 \\ &= O_P\{N_n^{1/2}(h^p + n^{-1/2}h^{-1})\} + O_P(h^p) \\ &= O_P\{N_n^{1/2}(h^p + n^{-1/2}h^{-1})\}. \end{aligned}$$

According to Lemma S.3 in the Supplementary Material, one has

$$\sup_{\eta_1, \eta_2 \in \mathcal{G}_n} \left| \frac{\langle \eta_1, \eta_2 \rangle_n - \langle \eta_1, \eta_2 \rangle}{\|\eta_1\|_2 \|\eta_2\|_2} \right| = O_P \left\{ \left( \frac{\log n}{nh} \right)^{1/2} \right\}.$$

Thus,  $\|\widehat{\eta} - \widetilde{\eta}\|_n = O_P\{N_n^{1/2}(h^p + n^{-1/2}h^{-1})\}$ . Then

$$\begin{aligned} \|\widehat{\eta} - \eta_0\|_n &\leq \|\widehat{\eta} - \widetilde{\eta}\|_n + \|\widetilde{\eta} - \eta_0\|_n \\ &= O_P\{N_n^{1/2}(h^p + n^{-1/2}h^{-1})\} + O_P(h^p) \\ &= O_P\{N_n^{1/2}(h^p + n^{-1/2}h^{-1})\}. \end{aligned}$$

### A.4. Proof of Theorem 2

We use the notation  $\widehat{U}_\tau = \alpha^\top(\widehat{\tau})\mathbf{X}$ . For any  $\mathbf{v} = (\mathbf{v}_1^\top, \mathbf{v}_2^\top)^\top$  with  $\mathbf{v}_1 \in R^{d_1}$  and  $\mathbf{v}_2 \in R^{d_2}$  define

$$\widehat{m}(\mathbf{v}; \alpha(\widehat{\tau}), \widehat{\beta}) = \widehat{\eta}\{\widehat{u}_\tau + \mathbf{v}_1 \mathbf{J}^\top(\widehat{\tau})\Phi(\mathbf{x}); \alpha(\widehat{\tau}), \widehat{\beta}\} + \mathbf{v}_2^\top \Psi(\mathbf{z}).$$

Define

$$\mathcal{M}_n = \{m(\mathbf{x}, \mathbf{z}) = \eta(\alpha^\top(\tau)\mathbf{x}) + \beta^\top \mathbf{z} : \eta \in \mathcal{H}(p)\}.$$

Note that  $\widehat{m}(\mathbf{v}; \alpha(\widehat{\tau}), \widehat{\beta})$  maximizes  $l(m) = \frac{1}{n} \sum_{i=1}^n Q[g^{-1}\{m(\mathbf{X}_i, \mathbf{Z}_i)\}, Y_i]$  for all  $m \in \mathcal{M}_n$  when  $\mathbf{v} = \mathbf{0}$ , thus  $\frac{\partial}{\partial \mathbf{v}} l(\widehat{m})|_{\mathbf{v}=\mathbf{0}} = \mathbf{0}$ , i.e.

$$\frac{\partial}{\partial \mathbf{v}_1} l(\widehat{m}) \Big|_{\mathbf{v}=\mathbf{0}} = \mathbf{0}, \quad \frac{\partial}{\partial \mathbf{v}_2} l(\widehat{m}) \Big|_{\mathbf{v}=\mathbf{0}} = \mathbf{0}.$$

Denote by  $\widehat{m}_i = \widehat{\eta}(\widehat{U}_{\tau,i}; \alpha(\widehat{\tau}), \widehat{\beta}) + \widehat{\beta}^\top \mathbf{Z}_i = \widehat{\gamma}^\top \mathbf{B}(\widehat{U}_{\tau,i}) + \widehat{\beta}^\top \mathbf{Z}_i$ . First,

$$\begin{aligned} \mathbf{0} &\equiv \frac{\partial}{\partial \mathbf{v}_1} l(\widehat{m}) \Big|_{\mathbf{v}_1=\mathbf{0}} \\ &= \frac{1}{n} \sum_{i=1}^n [Y_i - g^{-1}(\widehat{m}_i)] \rho_1(\widehat{m}_i) \widehat{\eta}'(\widehat{U}_{\tau,i}; \alpha(\widehat{\tau}), \widehat{\beta}) \mathbf{J}^\top(\widehat{\tau}) \Phi(\mathbf{X}_i) \\ &= \frac{1}{n} \sum_{i=1}^n [Y_i - g^{-1}(\widehat{m}_i)] \rho_1(\widehat{m}_i) \eta'_0(U_{\tau,0i}) \mathbf{J}^\top(\widehat{\tau}) \Phi(\mathbf{X}_i) \\ &\quad + \frac{1}{n} \sum_{i=1}^n [Y_i - g^{-1}(\widehat{m}_i)] \rho_1(\widehat{m}_i) [\widehat{\eta}'(\widehat{U}_{\tau,i}; \alpha(\widehat{\tau}), \widehat{\beta}) - \eta'_0(U_{\tau,0i})] \mathbf{J}^\top(\widehat{\tau}) \Phi(\mathbf{X}_i). \end{aligned}$$

Using the following expansion

$$\begin{aligned} &\widehat{\eta}'(\widehat{U}_{\tau,i}; \alpha(\widehat{\tau}), \widehat{\beta}) - \eta'_0(U_{\tau,0i}) \\ &= \widehat{\eta}'(\widehat{U}_{\tau,i}; \alpha(\widehat{\tau}), \widehat{\beta}) - \widehat{\eta}'(\alpha^\top(\tau_0)\mathbf{X}_i; \alpha(\widehat{\tau}), \widehat{\beta}) + \widehat{\eta}'(U_{\tau,0i}; \alpha(\widehat{\tau}), \widehat{\beta}) - \eta'_0(U_{\tau,0i}) \\ &= \widehat{\eta}''(\alpha^\top(\tau_0)\mathbf{X}_i; \alpha(\widehat{\tau}), \widehat{\beta}) \{\alpha(\widehat{\tau}) - \alpha(\tau_0)\}^\top \mathbf{X}_i + \widehat{\eta}'(U_{\tau,0i}; \alpha(\widehat{\tau}), \widehat{\beta}) - \eta'_0(U_{\tau,0i}) \\ &\quad + o_P(n^{-1/2}), \end{aligned} \tag{A.2}$$

and Conditions (C3), (C4) and (C7), we have

$$\begin{aligned}
 & \frac{1}{n} \sum_{i=1}^n [Y_i - g^{-1}(\widehat{m}_i)] \rho_1(\widehat{m}_i) \{ \widehat{\eta}'(\widehat{U}_{\tau,i}; \boldsymbol{\alpha}(\widehat{\boldsymbol{\tau}}), \widehat{\boldsymbol{\beta}}) - \eta'_0(U_{\tau,0i}) \} \mathbf{J}^\top(\widehat{\boldsymbol{\tau}}) \boldsymbol{\Phi}(\mathbf{X}_i) \\
 &= \frac{1}{n} \sum_{i=1}^n \varepsilon_i \rho_1(\widehat{m}_i) \widehat{\eta}''(U_{\tau,0i}; \boldsymbol{\alpha}(\widehat{\boldsymbol{\tau}}), \widehat{\boldsymbol{\beta}}) \mathbf{J}^\top(\widehat{\boldsymbol{\tau}}) \boldsymbol{\Phi}(\mathbf{X}_i) \mathbf{X}_i^\top \mathbf{J}(\boldsymbol{\tau}_0) (\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) \\
 & \quad + \frac{1}{n} \sum_{i=1}^n \varepsilon_i \rho_1(\widehat{m}_i) \{ \widehat{\eta}'(U_{\tau,0i}; \boldsymbol{\alpha}(\widehat{\boldsymbol{\tau}}), \widehat{\boldsymbol{\beta}}) - \eta'_0(U_{\tau,0i}) \} \mathbf{J}^\top(\widehat{\boldsymbol{\tau}}) \boldsymbol{\Phi}(\mathbf{X}_i) \\
 & \quad - \frac{1}{n} \sum_{i=1}^n [g^{-1}(\widehat{m}_i) - g^{-1}(m_{0i})] \rho_1(\widehat{m}_i) \widehat{\eta}''(U_{\tau,0i}; \boldsymbol{\alpha}(\widehat{\boldsymbol{\tau}}), \widehat{\boldsymbol{\beta}}) \mathbf{J}^\top(\widehat{\boldsymbol{\tau}}) \boldsymbol{\Phi}(\mathbf{X}_i) \mathbf{X}_i^\top \mathbf{J}(\boldsymbol{\tau}_0) (\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) \\
 & \quad - \frac{1}{n} \sum_{i=1}^n [g^{-1}(\widehat{m}_i) - g^{-1}(m_{0i})] \rho_1(\widehat{m}_i) \{ \widehat{\eta}'(U_{\tau,0i}; \boldsymbol{\alpha}(\widehat{\boldsymbol{\tau}}), \widehat{\boldsymbol{\beta}}) - \eta'_0(U_{\tau,0i}) \} \mathbf{J}^\top(\widehat{\boldsymbol{\tau}}) \boldsymbol{\Phi}(\mathbf{X}_i) \\
 &= o_P(n^{-1/2}).
 \end{aligned}$$

Thus,

$$\begin{aligned}
 \mathbf{0} &= \frac{1}{n} \sum_{i=1}^n \varepsilon_i \rho_1(m_{0i}) \eta'_0(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}_i) \\
 & \quad + \frac{1}{n} \sum_{i=1}^n \varepsilon_i \{ \rho_1(\widehat{m}_i) - \rho_1(m_{0i}) \} \eta'_0(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}_i) \\
 & \quad - \frac{1}{n} \sum_{i=1}^n [g^{-1}(\widehat{m}_i) - g^{-1}(m_{0i})] \rho_1(\widehat{m}_i) \eta'_0(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}_i) + o_P(n^{-1/2}) \\
 &\equiv \text{I} + \text{II} - \text{III} + o_P(n^{-1/2}).
 \end{aligned} \tag{A.3}$$

By (S.10) in the supplement,  $\|\widehat{\boldsymbol{m}} - \boldsymbol{m}_0\|_\infty = O_P\{N_n(h^p + n^{-1/2}h^{-1})\}$ , so

$$\text{II} = \frac{1}{n} \sum_{i=1}^n \varepsilon_i \rho'_1(m_{0i}) (\widehat{m}_i - m_{0i}) \eta'_0(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}_i) + o_P(n^{-1/2}) = \text{II}^* + o_P(n^{-1/2}),$$

by Condition (C8). Note that the expectation of the square of the  $k$ th column of  $\text{II}^*$  is

$$E \left\{ \frac{1}{n} \sum_{i=1}^n \varepsilon_i \rho'_1(m_{0i}) (\widehat{m}_i - m_{0i}) \eta'_0(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}_i) \mathbf{e}_k \right\}^2 = o(n^{-1}).$$

By Markov's inequality,

$$\frac{1}{n} \sum_{i=1}^n \varepsilon_i \rho_1'(m_{0i}) (\widehat{m}_i - m_{0i}) \eta_0'(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}_i) = o_P(n^{-1/2}).$$

Therefore,

$$\mathbf{\Pi} = o_P(n^{-1/2}). \tag{A.4}$$

For the third term, we have

$$\begin{aligned} \mathbf{III} &= \frac{1}{n} \sum_{i=1}^n [g^{-1}(\widehat{m}_i) - g^{-1}(m_{0i})] \rho_1(\widehat{m}_i) \eta_0'(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}_i) \\ &= \frac{1}{n} \sum_{i=1}^n (\widehat{m}_i - m_{0i}) \rho_2(m_{0i}) \eta_0'(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}_i) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \left. \frac{dg^{-2}(m)}{dm^2} \right|_{m=\bar{m}_i} (\widehat{m}_i - m_{0i})^2 \rho_1(m_{0i}) \eta_0'(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}_i) \\ &\quad + \frac{1}{n} \sum_{i=1}^n [g^{-1}(\widehat{m}_i) - g^{-1}(m_{0i})] [\rho_1(\widehat{m}_i) - \rho_1(m_{0i})] \eta_0'(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}_i) \\ &= \mathbf{III}_1 + \mathbf{III}_2 + \mathbf{III}_3, \end{aligned}$$

where  $\bar{m}_i$  is between  $m_{0i}$  and  $\widehat{m}_i$ .

Similar to (A.2), we have

$$\begin{aligned} \widehat{m}_i - m_{0i} &= \widehat{\eta}(\widehat{U}_{\tau,i}; \boldsymbol{\alpha}(\widehat{\boldsymbol{\tau}}), \widehat{\boldsymbol{\beta}}) - \eta_0(U_{\tau,0i}) + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top \mathbf{Z}_i \\ &= \eta'(U_{\tau,0i}) \{ \boldsymbol{\alpha}(\widehat{\boldsymbol{\tau}}) - \boldsymbol{\alpha}(\boldsymbol{\tau}_0) \}^\top \boldsymbol{\Phi}(\mathbf{X}_i) + \widehat{\eta}(U_{\tau,0i}; \boldsymbol{\alpha}(\widehat{\boldsymbol{\tau}}), \widehat{\boldsymbol{\beta}}) - \eta_0(U_{\tau,0i}) \\ &\quad + \eta'(U_{\tau,0i}) \{ \boldsymbol{\alpha}(\widehat{\boldsymbol{\tau}}) - \boldsymbol{\alpha}(\boldsymbol{\tau}_0) \}^\top \boldsymbol{\Upsilon}(U_{\tau,0i}) + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top \boldsymbol{\Psi}(\mathbf{T}_i) \\ &\quad + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\top \boldsymbol{\Gamma}(U_{\tau,0i}) + o_P(n^{-1/2}), \end{aligned}$$

so

$$\begin{aligned} \mathbf{III}_1 &= \frac{1}{n} \sum_{i=1}^n \rho_2(m_{0i}) \{ \eta'(U_{\tau,0i}) \}^2 \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}_i) \boldsymbol{\Phi}^\top(\mathbf{X}_i) \mathbf{J}(\boldsymbol{\tau}_0) (\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \rho_2(m_{0i}) \eta_0'(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}_i) \boldsymbol{\Psi}^\top(\mathbf{Z}_i) (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \rho_2(m_{0i}) \{ \widehat{\eta}(U_{\tau,0i}; \boldsymbol{\alpha}(\widehat{\boldsymbol{\tau}}), \widehat{\boldsymbol{\beta}}) - \eta_0(U_{\tau,0i}) \} \eta_0'(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}_i) \end{aligned}$$

$$\begin{aligned}
 & + \frac{1}{n} \sum_{i=1}^n \rho_2(m_{0i}) \{\eta'(U_{\tau,0i})\}^2 \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}_i) \boldsymbol{\Upsilon}^\top(U_{\tau,0i}) \mathbf{J}(\boldsymbol{\tau}_0) (\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) \\
 & + \frac{1}{n} \sum_{i=1}^n \rho_2(m_{0i}) \eta'_0(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}_i) \boldsymbol{\Gamma}^\top(U_{\tau,0i}) (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_P(n^{-1/2}).
 \end{aligned}$$

By Lemma S.6 in the Supplementary Material,

$$\begin{aligned}
 & \frac{1}{n} \sum_{i=1}^n \rho_2(m_{0i}) \{\widehat{\eta}(U_{\tau,0i}; \boldsymbol{\alpha}(\widehat{\boldsymbol{\tau}}), \widehat{\boldsymbol{\beta}}) - \eta_0(U_{\tau,0i})\} \eta'_0(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}_i) = o_P(n^{-1/2}), \\
 & \frac{1}{n} \sum_{i=1}^n \rho_2(m_{0i}) \{\eta'(U_{\tau,0i})\}^2 \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}_i) \boldsymbol{\Upsilon}^\top(U_{\tau,0i}) \mathbf{J}(\boldsymbol{\tau}_0) (\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) = o_P(n^{-1/2}), \\
 & \frac{1}{n} \sum_{i=1}^n \rho_2(m_{0i}) \eta'_0(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}_i) \boldsymbol{\Gamma}^\top(U_{\tau,0i}) (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = o_P(n^{-1/2}).
 \end{aligned}$$

Thus,

$$\begin{aligned}
 \text{III}_1 & = \frac{1}{n} \sum_{i=1}^n \rho_2(m_{0i}) \{\eta'(U_{\tau,0i})\}^2 \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}_i) \boldsymbol{\Phi}^\top(\mathbf{X}_i) \mathbf{J}(\boldsymbol{\tau}_0) (\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) + o_P(n^{-1/2}) \\
 & + \frac{1}{n} \sum_{i=1}^n \rho_2(m_{0i}) \eta'_0(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}_i) \boldsymbol{\Psi}^\top(\mathbf{Z}_i) (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0).
 \end{aligned}$$

By Conditions (C5)–(C7)

$$\begin{aligned}
 \text{III}_2 & = \frac{1}{n} \sum_{i=1}^n \left. \frac{dg^{-2}(m)}{dm^2} \right|_{m=\widehat{m}_i} (\widehat{m}_i - m_{0i})^2 \rho_1(m_{0i}) \eta'_0(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}_i) \\
 & \leq C \|\widehat{m} - m_0\|_\infty^2 = O_P\{N_n^2(h^p + n^{-1/2}h^{-1})^2\} \\
 & = o_P(n^{-1/2}),
 \end{aligned}$$

and similarly

$$\begin{aligned}
 \text{III}_3 & = \frac{1}{n} \sum_{i=1}^n [g^{-1}(\widehat{m}_i) - g^{-1}(m_{0i})][\rho_1(\widehat{m}_i) - \rho_1(m_{0i})] \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}_i) \\
 & \leq C \|\widehat{m} - m_0\|_\infty^2 = o_P(n^{-1/2}).
 \end{aligned}$$

Therefore,

$$\begin{aligned} \text{III} &= \frac{1}{n} \sum_{i=1}^n \rho_2(m_{0i}) \{ \eta'(U_{\tau,0i}) \}^2 \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}_i) \boldsymbol{\Phi}^\top(\mathbf{X}_i) \mathbf{J}(\boldsymbol{\tau}_0) (\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) + o_P(n^{-1/2}) \\ &+ \frac{1}{n} \sum_{i=1}^n \rho_2(m_{0i}) \eta'_0(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}_i) \boldsymbol{\Psi}^\top(\mathbf{Z}_i) (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0). \end{aligned} \tag{A.5}$$

Next,

$$\begin{aligned} \mathbf{0} &\equiv \frac{\partial}{\partial \mathbf{v}_2} l(\hat{\boldsymbol{m}}) \Big|_{\mathbf{v}_2=\mathbf{0}} = \frac{1}{n} \sum_{i=1}^n [Y_i - g^{-1}(\hat{m}_i)] \rho_1(\hat{m}_i) \boldsymbol{\Psi}(\mathbf{Z}_i) \\ &= \frac{1}{n} \sum_{i=1}^n \varepsilon_i \rho_1(m_{0i}) \boldsymbol{\Psi}(\mathbf{Z}_i) + \frac{1}{n} \sum_{i=1}^n \varepsilon_i \{ \rho_1(\hat{m}_i) - \rho_1(m_{0i}) \} \boldsymbol{\Psi}(\mathbf{Z}_i) \\ &\quad - \frac{1}{n} \sum_{i=1}^n [g^{-1}(\hat{m}_i) - g^{-1}(m_{0i})] \rho_1(\hat{m}_i) \boldsymbol{\Psi}(\mathbf{Z}_i) = \text{IV} + \text{V} - \text{VI}. \end{aligned} \tag{A.6}$$

Similar to (A.4),

$$\text{V} = \frac{1}{n} \sum_{i=1}^n \varepsilon_i [ \rho_1(\hat{m}_i) - \rho_1(m_{0i}) ] \boldsymbol{\Psi}(\mathbf{Z}_i) = o_P(n^{-1/2}). \tag{A.7}$$

For term VI, we have

$$\begin{aligned} \text{VI} &= \frac{1}{n} \sum_{i=1}^n [g^{-1}(\hat{m}_i) - g^{-1}(m_{0i})] \rho_1(\hat{m}_i) \boldsymbol{\Psi}(\mathbf{Z}_i) \\ &= \frac{1}{n} \sum_{i=1}^n (\hat{m}_i - m_{0i}) \rho_2(m_{0i}) \boldsymbol{\Psi}(\mathbf{Z}_i) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \frac{dg^{-2}(m)}{dm^2} \Big|_{m=\bar{m}_i} (\hat{m}_i - m_{0i})^2 \rho_1(m_{0i}) \boldsymbol{\Psi}(\mathbf{Z}_i) \\ &\quad + \frac{1}{n} \sum_{i=1}^n [g^{-1}(\hat{m}_i) - g^{-1}(m_{0i})] [ \rho_1(\hat{m}_i) - \rho_1(m_{0i}) ] \boldsymbol{\Psi}(\mathbf{Z}_i) \\ &= \text{VI}_1 + \text{VI}_2 + \text{VI}_3, \end{aligned}$$

where  $\bar{m}_i$  is between  $m_{0i}$  and  $\hat{m}_i$ . Note that

$$\text{VI}_1 = \frac{1}{n} \sum_{i=1}^n (\hat{m}_i - m_{0i}) \rho_2(m_{0i}) \boldsymbol{\Psi}(\mathbf{Z}_i)$$

$$\begin{aligned}
 &= \frac{1}{n} \sum_{i=1}^n \eta'_0(U_{\tau,0i}) \rho_2(m_{0i}) \Psi(\mathbf{Z}_i) \Phi^\top(\mathbf{X}_i) \mathbf{J}(\boldsymbol{\tau}_0) (\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) + o_P(n^{-1/2}) \\
 &\quad + \left\{ \frac{1}{n} \sum_{i=1}^n \rho_2(m_{0i}) \Psi(\mathbf{T}_i) (\Psi^\top(\mathbf{Z}_i) + \boldsymbol{\Gamma}^\top(U_{\tau,0i})) \right\} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
 &\quad + \frac{1}{n} \sum_{i=1}^n \{ \widehat{\eta}(U_{\tau,0i}) - \eta_0(U_{\tau,0i}) \} \rho_2(m_{0i}) \Psi(\mathbf{Z}_i) \\
 &\quad + \frac{1}{n} \sum_{i=1}^n \eta'_0(U_{\tau,0i}) \rho_2(m_{0i}) \Psi(\mathbf{Z}_i) \boldsymbol{\Upsilon}^\top(U_{\tau,0i}) \mathbf{J}(\boldsymbol{\tau}_0) (\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0).
 \end{aligned}$$

By Lemma S.7 in the Supplementary Material, one has

$$\begin{aligned}
 &\frac{1}{n} \sum_{i=1}^n \{ \widehat{\eta}(U_{\tau,0i}) - \eta_0(U_{\tau,0i}) \} \rho_2(m_{0i}) \Psi(\mathbf{Z}_i) = o_P(n^{-1/2}), \\
 &\frac{1}{n} \sum_{i=1}^n \eta'_0(U_{\tau,0i}) \rho_2(m_{0i}) \Psi(\mathbf{Z}_i) \boldsymbol{\Upsilon}^\top(U_{\tau,0i}) \mathbf{J}(\boldsymbol{\tau}_0) (\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) = o_P(n^{-1/2}), \\
 &\frac{1}{n} \sum_{i=1}^n \rho_2(m_{0i}) \Psi(\mathbf{Z}_i) \boldsymbol{\Gamma}(U_{\tau,0i})^\top (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = o_P(n^{-1/2}).
 \end{aligned}$$

Hence,

$$\begin{aligned}
 \text{VI}_1 &= \left\{ \frac{1}{n} \sum_{i=1}^n \rho_2(m_{0i}) \Psi(\mathbf{Z}_i) \{ \Psi(\mathbf{Z}_i) \}^\top \right\} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
 &\quad + \frac{1}{n} \sum_{i=1}^n \eta'_0(U_{\tau,0i}) \rho_2(m_{0i}) \Psi(\mathbf{Z}_i) \Phi^\top(\mathbf{X}_i) \mathbf{J}(\boldsymbol{\tau}_0) (\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) + o_P(n^{-1/2}),
 \end{aligned}$$

and using similar arguments about III<sub>2</sub> and III<sub>3</sub>, we can show  $|\text{VI}_2| = o_P(n^{-1/2})$ ,  $|\text{VI}_3| = o_P(n^{-1/2})$ . Therefore,

$$\begin{aligned}
 \text{VI} &= \left\{ \frac{1}{n} \sum_{i=1}^n \rho_2(m_{0i}) \Psi(\mathbf{Z}_i) \{ \Psi(\mathbf{Z}_i) \}^\top \right\} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_P(n^{-1/2}) \\
 &\quad + \frac{1}{n} \sum_{i=1}^n \eta'_0(U_{\tau,0i}) \rho_2(m_{0i}) \Psi(\mathbf{Z}_i) \Phi^\top(\mathbf{X}_i) \mathbf{J}(\boldsymbol{\tau}_0) (\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0).
 \end{aligned} \tag{A.8}$$

Combining (A.3)–(A.8), we have

$$\begin{aligned} \mathbf{0} &= \frac{1}{n} \sum_{i=1}^n q_1(m_{0i}, Y_i) \begin{pmatrix} \eta'_0(U_{\tau,0i}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}_i) \\ \boldsymbol{\Psi}(\mathbf{Z}_i) \end{pmatrix} \\ &+ \left\{ E \left[ \rho_2\{m_0(\mathbf{T})\} \begin{pmatrix} \eta'_0(U_{\tau,0}) \mathbf{J}^\top(\boldsymbol{\tau}_0) \boldsymbol{\Phi}(\mathbf{X}) \\ \boldsymbol{\Psi}(\mathbf{Z}) \end{pmatrix} \right. \right. \\ &\left. \left. \times \begin{pmatrix} \eta'_0(U_{\tau,0}) \boldsymbol{\Phi}^\top(\mathbf{X}) \mathbf{J}(\boldsymbol{\tau}_0) & \{\boldsymbol{\Psi}(\mathbf{Z})\}^\top \end{pmatrix} \right] + o_P(1) \right\} \begin{pmatrix} \widehat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0 \\ \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \end{pmatrix} + o_P(n^{-1/2}). \end{aligned}$$

Applying the central limit theorem and the delta method, we obtain the desired distribution of  $\begin{pmatrix} \widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 \\ \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \end{pmatrix}$  in the original parameter space.

### A.5. Proof of Theorem 3

Let  $\pi(\mathbf{x}, \mathbf{z})$  be the joint density of  $\mathbf{T} = (\mathbf{X}^\top, \mathbf{Z}^\top)^\top$ . Then under the semiparametric model (2.1) and (2.2), the joint density of  $(\mathbf{X}, Y, \mathbf{Z})$  is given by

$$f(\mathbf{x}, y, \mathbf{z}) = \exp[y\xi(\mathbf{x}, \mathbf{z}) - \mathcal{B}\{\xi(\mathbf{x}, \mathbf{z})\} + \mathcal{C}(y)]\pi(\mathbf{x}, \mathbf{z}), \tag{A.9}$$

where  $\xi(\mathbf{x}, \mathbf{z}) = g_0 \circ g^{-1}\{\eta_0(\boldsymbol{\alpha}_0^\top \mathbf{X}) + \boldsymbol{\beta}_0^\top \mathbf{Z}\}$  with  $\|\boldsymbol{\alpha}_0\| = 1$  and  $g_0$  as the canonical link function. Define the following submodels:

- $P_1 = \{\text{Model (A.9) with given } \eta_0(\cdot)\},$
- $P_2 = \{\text{Model (A.9) with given } \boldsymbol{\alpha}_0 \text{ and } \boldsymbol{\beta}_0\},$
- $P_3 = \{\text{Model (A.9) with given } \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0 \text{ and } \eta_0(\cdot)\},$

and let  $\dot{P}^\perp$  denote the orthogonal complement of  $\dot{P}$ . The score function for  $\boldsymbol{\alpha}_0$  and  $\boldsymbol{\beta}_0$  in the parametric model  $P_1$  is

$$\dot{l} = \{Y - \mu(\mathbf{X}, \mathbf{Z})\} g'_1\{\eta_0(\boldsymbol{\alpha}_0^\top \mathbf{X}) + \boldsymbol{\beta}_0^\top \mathbf{Z}\} \begin{pmatrix} \eta'_0(U) \mathbf{X} \\ \mathbf{Z} \end{pmatrix}.$$

According to Theorem 3.4.1 of [1], the efficient score for  $\boldsymbol{\alpha}_0$  and  $\boldsymbol{\beta}_0$  in model (A.9) is given by

$$\dot{l}^* = \dot{l} - \Pi(\dot{l} | \dot{P}_2 + \dot{P}_3),$$

where  $\Pi(\dot{l} | \dot{P}_2 + \dot{P}_3)$  is the projection of  $\dot{l}$  into the linear space  $\dot{P}_2 + \dot{P}_3$  and

$$\begin{aligned} \dot{l} - \Pi(\dot{l} | \dot{P}_2 + \dot{P}_3) &= \dot{l} - \Pi(\dot{l} | \dot{P}_3) - \Pi(\dot{l} | \Pi_{\dot{P}_3^\perp} \dot{P}_2) \\ &= \Pi(\dot{l} | \dot{P}_3^\perp) - \Pi\{\Pi(\dot{l} | \dot{P}_3^\perp) | \Pi_{\dot{P}_3^\perp} \dot{P}_2\}. \end{aligned}$$



By Lemma A4 in [13],

$$\Pi(\dot{i}|\dot{P}_3^\perp) = \{Y - \mu(\mathbf{X}, \mathbf{Z})\} g_1' \{ \eta_0(\alpha_0^\top \mathbf{X}) + \beta_0^\top \mathbf{Z} \} \begin{pmatrix} \eta_0'(U)\mathbf{X} \\ \mathbf{Z} \end{pmatrix},$$

where  $U = \alpha_0^\top \mathbf{X}$ . Next, we can show that

$$\Pi_{\dot{P}_3^\perp} \dot{P}_2 = \{ \{Y - \mu(\mathbf{X}, \mathbf{Z})\} g_1' \{ \eta_0(\alpha_0^\top \mathbf{X}) + \beta_0^\top \mathbf{Z} \} \psi(U), \psi(\cdot) \in L_2 \}.$$

Therefore,

$$i^* = \{Y - \mu(\mathbf{X}, \mathbf{Z})\} g_1' \{ \eta_0(\alpha_0^\top \mathbf{X}) + \beta_0^\top \mathbf{Z} \} \begin{pmatrix} \eta_0'(U) \{ \mathbf{X} - \Upsilon^*(U) \} \\ \mathbf{Z} - \Gamma^*(U) \end{pmatrix},$$

where  $\Upsilon^*(\cdot)$  and  $\Gamma^*(U)$  satisfies that

$$E \left[ \rho_2(\cdot) \begin{pmatrix} \eta_0'(U) \{ \mathbf{X} - \Upsilon^*(U) \} \\ \mathbf{Z} - \Gamma^*(U) \end{pmatrix} \psi(U) \right] = 0, \quad \psi(U) \in L_2.$$

So

$$\Upsilon^*(U) = \frac{E\{\mathbf{X}\rho_2(\cdot)|U\}}{E\{\rho_2(\cdot)|U\}}, \quad \Gamma^*(U) = \frac{E\{\mathbf{Z}\rho_2(\cdot)|U\}}{E\{\rho_2(\cdot)|U\}}.$$

Hence,

$$i^* = \{Y - \mu(\mathbf{X}, \mathbf{Z})\} g_1'(\cdot) \begin{pmatrix} \eta_0'(U) [\mathbf{X} - E\{\mathbf{X}\rho_2(\cdot)|U\}] / E\{\rho_2(\cdot)|U\} \\ \mathbf{Z} - E\{\mathbf{Z}\rho_2(\cdot)|U\} / E\{\rho_2(\cdot)|U\} \end{pmatrix}.$$

It is easy to verify that  $\mathbf{D} = E\{i^*(i^*)^\top\}$ .

## Acknowledgements

Wang’s research is supported in part by NSF awards DMS 1106816 and 1542332. Cao’s research was supported in part by the Simons Foundation under grant #354917. We sincerely thank the Editor, the Associate Editor and the anonymous reviewers for their insightful comments that have lead to significant improvements on the paper.

## Supplementary Material

Supplement to “Efficient estimation for generalized partially linear single-index models” (DOI: 10.3150/16-BEJ873SUPP; .pdf). We provide detailed proof for Lemma A.1, the statements and/or proofs of other technical lemmas and additional simulation results.

## References

- [1] Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J.A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins Series in the Mathematical Sciences. Baltimore, MD: Johns Hopkins Univ. Press. [MR1245941](#)
- [2] Carroll, R.J., Fan, J., Gijbels, I. and Wand, M.P. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.* **92** 477–489. [MR1467842](#)
- [3] Chen, H. (1991). Estimation of a projection-pursuit type regression model. *Ann. Statist.* **19** 142–157. [MR1091843](#)
- [4] Chen, J., Li, D., Liang, H. and Wang, S. (2015). Semiparametric GEE analysis in partially linear single-index models for longitudinal data. *Ann. Statist.* **43** 1682–1715. [MR3357875](#)
- [5] de Boor, C. (2001). *A Practical Guide to Splines*, Revised ed. *Applied Mathematical Sciences* **27**. New York: Springer. [MR1900298](#)
- [6] Friedman, J.H. and Stuetzle, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76** 817–823. [MR0650892](#)
- [7] Hall, P. (1989). On projection pursuit regression. *Ann. Statist.* **17** 573–588. [MR0994251](#)
- [8] Härdle, W., Hall, P. and Ichimura, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21** 157–178. [MR1212171](#)
- [9] Härdle, W. and Stoker, T.M. (1989). Investigating smooth multiple regression by the method of average derivatives. *J. Amer. Statist. Assoc.* **84** 986–995. [MR1134488](#)
- [10] Hastie, T.J. and Tibshirani, R.J. (1990). *Generalized Additive Models*. Boca Raton: CRC Press.
- [11] Hristache, M., Juditsky, A. and Spokoiny, V. (2001). Direct estimation of the index coefficient in a single-index model. *Ann. Statist.* **29** 595–623. [MR1865333](#)
- [12] Huang, J.Z. and Yang, L. (2004). Identification of non-linear additive autoregressive models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 463–477. [MR2062388](#)
- [13] 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](#)
- [14] Huber, P.J. (1985). Projection pursuit. *Ann. Statist.* **13** 435–525. [MR0790553](#)
- [15] Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J. Econometrics* **58** 71–120. [MR1230981](#)
- [16] Klein, R.W. and Spady, R.H. (1993). An efficient semiparametric estimator for binary response models. *Econometrica* **61** 387–421. [MR1209737](#)
- [17] Küchenhoff, H. and Carroll, R.J. (1997). Segmented regression with errors in predictors: Semiparametric and parametric methods. *Stat. Med.* **16** 169–188.
- [18] Lai, P., Li, G. and Lian, H. (2013). Quadratic inference functions for partially linear single-index models with longitudinal data. *J. Multivariate Anal.* **118** 115–127. [MR3054094](#)
- [19] Lai, P., Tian, Y. and Lian, H. (2014). Estimation and variable selection for generalised partially linear single-index models. *J. Nonparametr. Stat.* **26** 171–185. [MR3174315](#)
- [20] Liang, H., Liu, X., Li, R. and Tsai, C.-L. (2010). Estimation and testing for partially linear single-index models. *Ann. Statist.* **38** 3811–3836. [MR2766869](#)
- [21] Ma, S., Liang, H. and Tsai, C.-L. (2014). Partially linear single index models for repeated measurements. *J. Multivariate Anal.* **130** 354–375. [MR3229543](#)
- [22] 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](#)
- [23] McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, 2nd ed. *Monographs on Statistics and Applied Probability*. London: Chapman & Hall. [MR3223057](#)
- [24] Powell, J.L., Stock, J.H. and Stoker, T.M. (1989). Semiparametric estimation of index coefficients. *Econometrica* **57** 1403–1430. [MR1035117](#)

- [25] Sheng, W. and Yin, X. (2013). Direction estimation in single-index models via distance covariance. *J. Multivariate Anal.* **122** 148–161. [MR3189314](#)
- [26] Wang, J.-L., Xue, L., Zhu, L. and Chong, Y.S. (2010). Estimation for a partial-linear single-index model. *Ann. Statist.* **38** 246–274. [MR2589322](#)
- [27] Wang, L. and Cao, G. (2016). Supplement to “Efficient estimation for generalized partially linear single-index models.” DOI:10.3150/16-BEJ873SUPP.
- [28] Wang, L., Liu, X., Liang, H. and Carroll, R.J. (2011). Estimation and variable selection for generalized additive partial linear models. *Ann. Statist.* **39** 1827–1851. [MR2893854](#)
- [29] Wang, L. and Yang, L. (2007). Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *Ann. Statist.* **35** 2474–2503. [MR2382655](#)
- [30] Wang, L. and Yang, L. (2009). Spline estimation of single-index models. *Statist. Sinica* **19** 765–783. [MR2514187](#)
- [31] Wood, S.N. (2006). *Generalized Additive Models: An Introduction with R. Texts in Statistical Science Series*. Boca Raton, FL: Chapman & Hall/CRC. [MR2206355](#)
- [32] Xia, Y. and Härdle, W. (2004). Semi-parametric estimation of generalized partially linear single-index models. Technical report, Humboldt Universities, Berlin.
- [33] Xia, Y., Li, W.K., Tong, H. and Zhang, D. (2004). A goodness-of-fit test for single-index models. *Statist. Sinica* **14** 1–39. [MR2036761](#)
- [34] Xia, Y., Tong, H., Li, W.K. and Zhu, L.-X. (2002). An adaptive estimation of dimension reduction space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 363–410. [MR1924297](#)
- [35] Xue, L. and Yang, L. (2006). Additive coefficient modeling via polynomial spline. *Statist. Sinica* **16** 1423–1446. [MR2327498](#)
- [36] Yu, Y. and Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *J. Amer. Statist. Assoc.* **97** 1042–1054. [MR1951258](#)

Received July 2015 and revised February 2016