Locally efficient estimation in generalized partially linear model with measurement error in nonlinear function

Qianqian Wang[1] · Yanyuan Ma[2] · Guangren Yang[3]

Abstract

We investigate the errors in covariates issues in a generalized partially linear model. Different from the usual literature (Ma and Carroll in J Am Stat Assoc 101:1465–1474, 2006), we consider the case where the measurement error occurs to the covariate that enters the model nonparametrically, while the covariates precisely observed enter the model parametrically. To avoid the deconvolution type operations, which can suffer from very low convergence rate, we use the B-splines representation to approximate the nonparametric function and convert the problem into a parametric form for operational purpose. We then use a parametric working model to replace the distribution of the unobservable variable, and devise an estimating equation to estimate both the model parameters and the functional dependence of the response on the latent variable. The estimation procedure is devised under the functional model framework without assuming any distribution structure of the latent variable. We further derive theories on the large sample properties of our estimator. Numerical simulation studies are carried out to evaluate the finite sample performance, and the practical performance of the method is illustrated through a data example.

✉ Guangren Yang
   tygr@jnu.edu.cn

1   Department of Statistics, University of South Carolina, Columbia, USA

2   Department of Statistics, Pennsylvania State University, University Park, USA

3   Department of Statistics, School of Economics, Jinan University, Guangzhou, China

1 Introduction

Generalized partially linear models have been widely used in statistics. Such models enrich the more classic generalized linear models by allowing a covariate to enter the link function through a nonparametric form. This is useful when the dependence of the response to some covariates, even after transformation through a suitable link function, is still not linear and difficult to specify. At the same time, the model also allows the more classic generalized linear dependence on some other covariates. Many works exist in the literature for estimation and inference for generalized partially linear models, see, for example, Carroll et al. (1995), Liang et al. (2009), Apanasovich et al. (2009), and Yu and Ruppert (2012).

When one of the covariates involved in the generalized partially linear model cannot be measured precisely, the problem becomes much more difficult. In fact, most of the works in handling measurement error issues in the generalized partially linear model considered only the case that measurement error occurs to a covariate involved in the linear component (Ma and Carroll 2006; Liu et al. 2017; Liang and Ren 2005; Liu 2007; Liang and Thurston 2008). When the model degenerates to simply the generalized linear model, even more literatures exist to handle the measurement error issues (Carroll et al. 2006; Stefanski and Carroll 1985, 1987; Huang and Wang 2001; Ma and Tsiatis 2006; Buonaccorsi 2010; Xu and Ma 2015). When handled properly, it can be shown that the parameters can be estimated at the root-n convergence rate despite of the presence of the measurement error and the possible presence of the nonparametric function in the model. However, it is a different story when the covariate inside the nonparametric function itself is measured with error. We conjecture that this is because as soon as the covariate inside an unknown function is subject to error, the problem falls into the general framework of nonparametric measurement error models and the standard practice for estimation and inference is through deconvolution. Deconvolution method is widely used in handling latent components and has been used to show that nonparametric regression with errors in covariates can have very slow convergence rate. Possibly due to these inherent difficulties, generalized partially linear models with errors in the covariate inside the nonparametric function have not been studied systematically.

We tackle this difficult problem where the error occurs to the covariate inside the nonparametric component of the generalized partially linear model through a novel approach that avoids the deconvolution treatment completely. Two key ideas lead to our success in this endeavor. The first is the idea of using B-splines expansion to approximate the nonparametric function of the latent covariate. The B-spline nature allows us to write out the approximation form without having to perform the estimation simultaneously. This is different from nonparametric estimation via kernel method, where the approximation and estimation is integrated and inseparable. The second idea is the recognition that after the B-spline approximation, the error-free model is effectively a parametric model, or at least a parametric model in terms of operation, hence the only nonparametric component in the measurement error model is the distribution of the latent covariate. This implies that the semiparametric approach in Tsiatis and Ma (2004) can be adopted here to help establishing the estimation procedure. The encouraging discovery is that we not only can bypass the difficulties caused by

nonparametric function of a covariate measured with error in terms of estimation, we also prove that the procedure can retain the root-n convergence rate of the parameter estimation in the original model.

The structure of this paper is as follows. We describe the model and the estimation methodology in Sect. 2, following with establishing the large sample properties of the parameter estimation in Sect. 3. Two simulation studies are conducted in Sect. 4, and we analyze the AIDS Clinical Trials Group (ACTG) study in Sect. 5. We finish the paper with some discussions in Sect. 6. All the technical details and proofs are provided in "Appendix".

2 Main results

2.1 The model

We work in a measurement error model framework, sometimes also referred to as errors-in-covariate model. It is different from the standard regression model; in that, at least one of the covariates is not directly observable. Instead, a measurement of this covariate that contains error is observed only. Generally speaking, in a standard regression problem, we would observe independent and identically distributed (i.i.d.) observations $(X_i, \mathbf{Z}_i, Y_i)$, $i = 1, \ldots, n$, where $(X_i, \mathbf{Z}_i)$ is the covariate and Y_i is the response. Then, with a specific model of Y_i given X_i and $\mathbf{Z}_i$, we can then proceed to estimate the unknown components in this regression relation. However, in a measurement error model, X_i is no longer available, instead, only an errored version of X_i, say W_i is observed. Thus, the goal is to still perform the estimation of the parameters in the model of Y_i given $(X_i, \mathbf{Z}_i)$, but using $(W_i, \mathbf{Z}_i, Y_i)$'s, instead of (X_i, Z_i, Y_i)'s.

In this paper, we study the generalized partially linear model

$$f_{Y|X,\mathbf{Z}}(y, x, \mathbf{z}, \boldsymbol{\alpha}, \boldsymbol{\beta}, g) = f\{y, \mathbf{z}^{\mathrm{T}}\boldsymbol{\beta} + g(x), \boldsymbol{\alpha}\}. \tag{1}$$

Here, Y is the univariate response variable and X and $\mathbf{Z}$ are covariates. We assume the univariate variable X to be compactly supported. Without loss of generality, let the support be [0, 1]. We assume $\mathbf{Z} \in \mathscr{R}^{p_z}$, $p_z \geq 1$. The unknown components in (1) are $\boldsymbol{\beta} \in \mathscr{R}^{p_z}$ and $\boldsymbol{\alpha}$, whose estimation and inference are of the main interest to us, and the nuisance function $g(\cdot)$, which contributes to the name "partially linear." The link function $f(\cdot)$ is assumed to be known. Here, the parameter $\boldsymbol{\beta}$ describes the linear effect of the covariate in $\mathbf{Z}$, $g(\cdot)$ describes the unspecified effect of X and $\boldsymbol{\alpha}$ arises according to the link function f. For example, $f(\cdot)$ can be the inverse logit link function $f(\cdot) = 1 - 1/\{\exp(\cdot) + 1\}$ or the normal link function $f(\cdot) = \exp\{-(\cdot)^2/(2\alpha^2)\}/\{(2\pi\alpha^2)^{1/2}\}$. Note that in the logistic example, the parameter $\boldsymbol{\alpha}$ does not appear, while in the normal example, α captures the standard error of Y. Now, although Y and $\mathbf{Z}$'s are observable, $\mathbf{X}$ is not. Instead, X is a random variable measured with error. Thus, lieu of observing X, we observe W, where

$$W = X + U, \tag{2}$$

and U is a normal random error independent of X and $\mathbf{Z}$ with mean zero and variance σ_U^2. For ease of the presentation of the main methodology, we assume σ_U^2 is known. When σ_U^2 is unknown, a common approach is to use repeated measurements to estimate σ_U^2 first and then plug in. The observed data are $(W_i, \mathbf{Z}_i, Y_i)$, $i = 1, \ldots, n$, which are iid. Our goal is to estimate $\boldsymbol{\alpha}, \boldsymbol{\beta}$, together with $g(\cdot)$ hence to understand the dependence of Y on the covariates $(X, \mathbf{Z})$.

2.2 Efficient score derivation

For preparation, we first approximate $g(x)$ with a B-spline representation, i.e., $g(x) \approx \mathbf{B}(x)^{\mathrm{T}}\boldsymbol{\gamma}$. Under this approximation, model (1) becomes

$$f_{Y|X,\mathbf{Z}}(y, x, \mathbf{z}, \boldsymbol{\alpha}, \boldsymbol{\beta}, g) \approx f_{Y|X,\mathbf{Z}}(y, x, \mathbf{z}, \boldsymbol{\theta}) \equiv f\{y, \mathbf{z}^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{B}(x)^{\mathrm{T}}\boldsymbol{\gamma}, \boldsymbol{\alpha}\}, \quad (3)$$

which is a complete parametric model with unknown parameters $\boldsymbol{\theta} \equiv (\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\gamma}^{\mathrm{T}})^{\mathrm{T}}$. This model falls in the general framework of Tsiatis and Ma (2004), hence their estimation procedure can be adopted here. Specifically, the joint distribution of the observed variables conditional on $\mathbf{Z}$ is

$$f_{W,Y|\mathbf{Z}}(y, w, \mathbf{z}, \boldsymbol{\theta}) = \int f\{y, \mathbf{z}^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{B}(x)^{\mathrm{T}}\boldsymbol{\gamma}, \boldsymbol{\alpha}\} f_{W|X}(w, x) f_{X|\mathbf{Z}}(x, \mathbf{z}) d\mu(x),$$

with the condition distribution function $f_{X|\mathbf{Z}}(x, \mathbf{z})$ being a nuisance parameter. The nuisance tangent space Λ and its orthogonal complement $\Lambda^{\perp}$ can be written as

$$\Lambda = [E\{\mathbf{a}(X, \mathbf{Z})|Y, W, \mathbf{Z}\} : E\{\mathbf{a}(X, \mathbf{Z}) \mid \mathbf{Z}\} = 0],$$
$$\Lambda^{\perp} = [\mathbf{h}(Y, W, \mathbf{Z}) : E\{\mathbf{h}(Y, W, \mathbf{Z}) \mid X, \mathbf{Z}\} = \mathbf{0} \text{ almost everywhere}].$$

The efficient score for $\boldsymbol{\theta}$ is the residual of its score vector $\mathbf{S}_{\boldsymbol{\theta}}(y, w, \mathbf{z})$ after projecting it on to the nuisance tangent space Λ, denoted by

$$\mathbf{S}_{\mathrm{res}}(y, w, \mathbf{z}, \boldsymbol{\theta}) \equiv \mathbf{S}_{\boldsymbol{\theta}}(y, w, \mathbf{z}, \boldsymbol{\theta}) - \Pi\{\mathbf{S}_{\boldsymbol{\theta}}(Y, W, \mathbf{Z}, \boldsymbol{\theta})|\Lambda\},$$

where $\mathbf{S}_{\boldsymbol{\theta}}(y, w, \mathbf{z}, \boldsymbol{\theta}) \equiv \partial \log f_{W,Y|\mathbf{Z}}(y, w, \mathbf{z}, \boldsymbol{\theta})/\partial\boldsymbol{\theta}$. Here, "res" stands for residual. The detailed form of $\mathbf{S}_{\mathrm{res}}(y, w, \mathbf{z}, \boldsymbol{\theta})$ is given as

$$\mathbf{S}_{\mathrm{res}}(Y, W, \mathbf{Z}, \boldsymbol{\theta}) = \mathbf{S}_{\boldsymbol{\theta}}(Y, W, \mathbf{Z}, \boldsymbol{\theta}) - E\{\mathbf{a}(X, \mathbf{Z}, \boldsymbol{\theta})|Y, W, \mathbf{Z}\}, \quad (4)$$

where $\mathbf{a}(X, \mathbf{Z}, \boldsymbol{\theta})$ satisfies

$$E\{\mathbf{S}_{\boldsymbol{\theta}}(Y, W, \mathbf{Z}, \boldsymbol{\theta}) \mid X, \mathbf{Z}\} = E[E\{\mathbf{a}(X, \mathbf{Z}, \boldsymbol{\theta})|Y, W, \mathbf{Z}\} \mid X, \mathbf{Z}]. \quad (5)$$

Now, noting that the above derivation is obtained from the approximate model (3), we hence perform some further analysis. Separating the components corresponding to

α, β and γ in θ, we can write $\mathbf{S}_\theta(y, w, z, \theta) = \{\mathbf{S}_{\alpha,\beta}(y, w, z, \theta)^{\mathrm{T}}, \mathbf{S}_\gamma(y, w, z, \theta)^{\mathrm{T}}\}^{\mathrm{T}}$, which leads to the corresponding relation as follows:

$$\mathbf{S}_{\mathrm{res}}(y, w, \mathbf{z}, \theta) = \{\mathbf{S}_{\mathrm{res}1}(y, w, \mathbf{z}, \theta)^{\mathrm{T}}, \mathbf{S}_{\mathrm{res}2}(y, w, \mathbf{z}, \theta)^{\mathrm{T}}\}^{\mathrm{T}}.$$

The estimating equation of the approximate model can be written as

$$\sum_{i=1}^{n} \mathbf{S}_{\mathrm{res}}(Y_i, W_i, \mathbf{Z}_i, \theta) = \sum_{i=1}^{n}\{\mathbf{S}_{\mathrm{res}1}(Y_i, W_i, \mathbf{Z}_i, \theta)^{\mathrm{T}}, \mathbf{S}_{\mathrm{res}2}(Y_i, W_i, \mathbf{Z}_i, \theta)^{\mathrm{T}}\}^{\mathrm{T}} = \mathbf{0}.$$

$$(6)$$

Remember that our original model contains an unknown function $g(z)$. Thus, for the estimation of α, β, it is beneficial to treat g as a nuisance parameter as well first, and estimate α, β using profiling. We then plug in the estimated values of α and β and estimate g via the B-spline approximation. Of course, in addition to g, the distribution of the unobservable covariate conditional on the observable covariate $\mathbf{Z}$ is also a nuisance component and still has to be taken into account.

Let $\delta \equiv (\alpha^{\mathrm{T}}, \beta^{\mathrm{T}})^{\mathrm{T}}$ be a p-dimensional parameter. We propose to solve for γ from $\sum_{i=1}^{n} \mathbf{S}_{\mathrm{res}2}(Y_i, W_i, \mathbf{Z}_i, \theta) = \mathbf{0}$ to obtain $\widehat{\gamma}(\delta)$ first. Now from

$$f_{W,Y|\mathbf{Z}}(w, \mathbf{z}, y, \delta, g, f_X) = \int f\{y, \mathbf{z}^{\mathrm{T}}\beta + g(x), \alpha\} f_{W|X}(w, x) f_{X|\mathbf{Z}}(x, \mathbf{z}) d\mu(x),$$

we can construct the nuisance tangent space as $\Lambda = \Lambda_{f_X} + \Lambda_g$, where

$$\Lambda_{f_X} = [E\{\mathbf{a}(X, \mathbf{Z})|Y, W, \mathbf{Z}\} : E\{\mathbf{a}(X, \mathbf{Z}) \mid \mathbf{Z}\} = \mathbf{0}]$$
$$\Lambda_g = \left(E\left[s\{Y, \mathbf{Z}^{\mathrm{T}}\beta + g(X), \alpha\}\mathbf{b}(X)|Y, W, \mathbf{Z}\right] : \forall \mathbf{b}(X)\right),$$

where $s(y, t, \alpha) \equiv \partial \log f(y, t, \alpha)/\partial t$. Note that Λ_{f_X} and Λ_g are not orthogonal to each other. We can further verify that

$$\Lambda_{f_X}^{\perp} = [\mathbf{h}(Y, W, \mathbf{Z}) : E\{\mathbf{h}(Y, W, \mathbf{Z}) \mid X, \mathbf{Z}\} = \mathbf{0} \text{ a.e.}],$$
$$\Lambda_g^{\perp} = \left(\mathbf{h}(Y, W, \mathbf{Z}) : E\left[\mathbf{h}(Y, W, \mathbf{Z})s\{Y, \mathbf{Z}^{\mathrm{T}}\beta + g(X), \alpha\} \mid X, \mathbf{Z}\right] = \mathbf{0} \text{ almost everywhere}\right).$$

The efficient score for δ is now the residual of the score vector $\mathbf{S}_\delta$ after projecting it on to the nuisance tangent space Λ, denoted by

$$\mathbf{S}_{\mathrm{eff}}(Y, W, \mathbf{Z}, \delta, g) \equiv \mathbf{S}_\delta(Y, W, \mathbf{Z}, \delta, g) - \Pi\{\mathbf{S}_\delta(Y, W, \mathbf{Z}, \delta, g) \mid \Lambda\}. \quad (7)$$

Its explicit form is given as

$$\mathbf{S}_{\mathrm{eff}}(Y, W, \mathbf{Z}, \delta, g) = \mathbf{S}_\delta(Y, W, \mathbf{Z}, \delta, g) - E\{\mathbf{a}(X, \mathbf{Z})|Y, W, \mathbf{Z}\}$$
$$- E\left[s\{Y, \mathbf{Z}^{\mathrm{T}}\beta + g(X), \alpha\}\mathbf{b}(X)|Y, W, \mathbf{Z}\right],$$

where $\mathbf{a}(X, \mathbf{Z})$ and $\mathbf{b}(X)$ satisfy

$$E\{\mathbf{S}_\delta(Y, W, \mathbf{Z}, \delta, g) \mid X, \mathbf{Z}\}$$
$$= E[E\{\mathbf{a}(X, \mathbf{Z})|Y, W, \mathbf{Z}\} \mid X, \mathbf{Z}]]$$
$$+ E(E[s\{Y, \mathbf{Z}^\mathrm{T}\boldsymbol{\beta} + g(X), \boldsymbol{\alpha}\}\mathbf{b}(X)|Y, W, \mathbf{Z}] \mid X, \mathbf{Z})$$

and

$$E[\mathbf{S}_\delta(Y, W, \mathbf{Z}, \delta, g)s\{Y, \mathbf{Z}^\mathrm{T}\boldsymbol{\beta} + g(X), \boldsymbol{\alpha}\} \mid X, \mathbf{Z}]$$
$$= E[E\{\mathbf{a}(X, \mathbf{Z})|Y, W, \mathbf{Z}\}s\{Y, \mathbf{Z}^\mathrm{T}\boldsymbol{\beta} + g(X), \boldsymbol{\alpha}\} \mid X, \mathbf{Z}]$$
$$+ E(E[s\{Y, \mathbf{Z}^\mathrm{T}\boldsymbol{\beta} + g(X), \boldsymbol{\alpha}\}\mathbf{b}(X)|Y, W, \mathbf{Z}]s\{Y, \mathbf{Z}^\mathrm{T}\boldsymbol{\beta} + g(X), \boldsymbol{\alpha}\} \mid X, \mathbf{Z}).$$
(8)

We can then form the estimating equation $\sum_{i=1}^n \mathbf{S}_\mathrm{eff}\{Y_i, W_i, \mathbf{Z}_i, \delta, \widehat{\boldsymbol{\gamma}}(\delta)\} = \mathbf{0}$ to solve for $\widehat{\delta}$ as the estimator, where $\mathbf{a}(X, \mathbf{Z})$, $\mathbf{b}(X)$ are the solutions to the integral equations in (8).

2.3 Estimation under working model

The above derivations are based on efficient score calculation and hence will yield the efficient estimator. However, a close look at the procedure reveals that the procedure is not practical because the implementation relies on the unknown function $f_{X|\mathbf{Z}}(x, \mathbf{z})$. Thus, our estimator needs to be calculated under a posited working model of $\mathbf{f}^*_{X|\mathbf{Z}}(x, \mathbf{z})$. The procedure is described below, where we use * to denote a quantity whose calculation is carried out using $\mathbf{f}^*_{X|\mathbf{Z}}(x, \mathbf{z})$ instead of $\mathbf{f}_{X|\mathbf{Z}}(x, \mathbf{z})$.

1. Posit a working model $\mathbf{f}^*_{X|\mathbf{Z}}(x, \mathbf{z})$.
2. Solve for $\boldsymbol{\gamma}$ from $\sum_{i=1}^n \mathbf{S}_{\mathrm{res}2}^*(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\theta}) = \mathbf{0}$ to obtain $\widehat{\boldsymbol{\gamma}}(\delta)$.
3. Calculate the score function $\mathbf{S}_\delta^*(Y, W, \mathbf{Z}, \delta, g)$ under the working model $\mathbf{f}^*_{X|\mathbf{Z}}(x, \mathbf{z})$.
4. Solve the integral equation (8) to get $\mathbf{a}(X, \mathbf{Z})$ and $\mathbf{b}(X)$.
5. Calculate the approximate efficient score function $\mathbf{S}_\mathrm{eff}^*(Y, W, \mathbf{Z}, \delta, \widehat{g})$ following (7), where $\widehat{g}(\cdot) = \mathbf{B}(\cdot)^\mathrm{T}\widehat{\boldsymbol{\gamma}}(\delta)$.
6. Solve the estimating equation $\sum_{i=1}^n \mathbf{S}_\mathrm{eff}^*(Y_i, W_i, \mathbf{Z}_i, \delta, \widehat{g}) = \mathbf{0}$ to obtain $\widehat{\delta}$.

When we calculate $\mathbf{a}(X, \mathbf{Z})$ at each observed $\mathbf{z}$ value and calculate $\mathbf{b}(X)$, we discretize the distribution of X on m equally spaced points on the support of $f_{X|\mathbf{Z}}(x, \mathbf{z})$ and calculate the probability mass function $\pi_j(\mathbf{Z})$ at each of the m points. We of course normalize the $\pi_j(\mathbf{Z})$ in order to ensure $\sum_{j=1}^m \pi_j(\mathbf{Z}) = 1$. Note that using the discretization,

$$f^*_{X,Y,W|\mathbf{Z}}(x_j, y, w, \mathbf{z}) \approx f\{y, \mathbf{z}^\mathrm{T}\boldsymbol{\beta} + g(x_j), \boldsymbol{\alpha}\}f_{W|X=x_j}(w, x_j)\pi_j(\mathbf{Z}).$$

Further, $\mathbf{S}_\delta^*(Y, W, \mathbf{Z}, \delta, g)$, $E^*\{\mathbf{a}(X, \mathbf{Z})|Y, W, \mathbf{Z}\}$ and $E^*[s\{Y, \mathbf{Z}^\mathrm{T}\boldsymbol{\beta} + g(X, \delta), \boldsymbol{\alpha}\}\mathbf{b}(X)|Y, W, \mathbf{Z}]$ can be approximated by

$$\mathbf{S}_{\boldsymbol{\delta}}^{*}(Y, W, \mathbf{Z}, \boldsymbol{\delta}, g)$$
$$\approx \frac{\partial \log[\sum_{i=1}^{m} f\{y, \mathbf{z}^{\mathrm{T}}\boldsymbol{\beta} + g(x_i), \boldsymbol{\alpha}\} f_{W|X}(w, x_i)\pi_i(\mathbf{Z})]}{\partial \boldsymbol{\delta}},$$
$$E^{*}\{\mathbf{a}(X, \mathbf{Z})|Y, W, \mathbf{Z}\}$$
$$\approx \frac{\sum_{i=1}^{m} \mathbf{a}(x_i, \mathbf{Z}) f_{X,Y,W|\mathbf{Z}}^{*}(x_i, Y, W, \mathbf{Z})}{\sum_{i=1}^{m} f_{X,Y,W|\mathbf{Z}}^{*}(x_i, Y, W, \mathbf{Z})},$$
$$E^{*}[s\{Y, \mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta} + g(X), \boldsymbol{\alpha}\}\mathbf{b}(X)|Y, W, \mathbf{Z}]$$
$$\approx \frac{\sum_{i=1}^{m} s\{Y, \mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta} + g(x_i), \boldsymbol{\alpha}\}\mathbf{b}(x_i) f_{X,Y,W|\mathbf{Z}}^{*}(x_i, Y, W, \mathbf{Z})}{\sum_{i=1}^{m} f_{X,Y,W|\mathbf{Z}}^{*}(x_i, Y, W, \mathbf{Z})}.$$

Let $\mathbf{A}(X, \mathbf{Z}) \equiv \{\mathbf{a}(x_1, \mathbf{Z}), \ldots, \mathbf{a}(x_m, \mathbf{Z})\}^{\mathrm{T}}$ and $\mathbf{B}(X) \equiv \{\mathbf{b}(x_1), \ldots, \mathbf{b}(x_m)\}^{\mathrm{T}}$. Let $\mathbf{M}_1(X, \mathbf{Z}) \equiv \{\mathbf{m}_1(x_1, \mathbf{Z}), \ldots, \mathbf{m}_1(x_m, \mathbf{Z})\}^{\mathrm{T}}$ be a $m \times p_{\boldsymbol{\delta}}$ matrix, where $p_{\boldsymbol{\delta}}$ is the length of $\boldsymbol{\delta}$ and $\mathbf{m}_1(x_i, \mathbf{Z}) \equiv E\{\mathbf{S}_{\boldsymbol{\delta}}^{*}(Y, W, \mathbf{Z}, \boldsymbol{\delta}, g) \mid x_i, \mathbf{Z}\}$. Further, let $\mathbf{M}_2(X, \mathbf{Z}) \equiv \{\mathbf{m}_2(x_1, \mathbf{Z}), \ldots, \mathbf{m}_2(x_m, \mathbf{Z})\}^{\mathrm{T}}$ be a $m \times p_{\boldsymbol{\delta}}$ matrix, where we define $\mathbf{m}_2(x_i, \mathbf{Z}) \equiv E[\mathbf{S}_{\boldsymbol{\delta}}^{*}(Y, W, \mathbf{Z}, \boldsymbol{\delta}, g)s\{Y, \mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta} + g(x_i)\} \mid x_i, \mathbf{Z}]$. Finally, let $\mathbf{C}(X, \mathbf{Z})$ be a $m \times m$ matrix with the (i, j) block equal to

$$E\left\{\frac{f_{X,Y,W|\mathbf{Z}}^{*}(x_j, Y, W, \mathbf{Z})}{\sum_{i=1}^{m} f_{X,Y,W|\mathbf{Z}}^{*}(x_i, Y, W, \mathbf{Z})} \mid x_i, \mathbf{Z}\right\},$$

let $\mathbf{D}(X, \mathbf{Z})$ be an $m \times m$ matrix with the (i, j) block equal to

$$E\left[\frac{s\{Y, \mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta} + g(x_j), \boldsymbol{\alpha}\}f_{X,Y,W|\mathbf{Z}}^{*}(x_j, Y, W, \mathbf{Z})}{\sum_{i=1}^{m} f_{X,Y,W|\mathbf{Z}}^{*}(x_i, Y, W, \mathbf{Z})} \mid x_i, \mathbf{Z}\right],$$

let $\mathbf{F}(X, \mathbf{Z})$ be an $m \times m$ matrix with the (i, j) block equal to

$$E\left[\frac{f_{X,Y,W|\mathbf{Z}}^{*}(x_j, Y, W, \mathbf{Z})s\{Y, \mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta} + g(x_i)\}}{\sum_{i=1}^{m} f_{X,Y,W|\mathbf{Z}}^{*}(x_i, Y, W, \mathbf{Z})} \mid x_i, \mathbf{Z}\right],$$

and let $\mathbf{G}(X, \mathbf{Z})$ be an $m \times m$ matrix with the (i, j) block

$$E\left[\frac{s\{Y, \mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta} + g(x_j), \boldsymbol{\alpha}\}f_{X,Y,W|\mathbf{Z}}^{*}(x_j, Y, W, \mathbf{Z})s\{Y, \mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta} + g(x_i)\}}{\sum_{i=1}^{m} f_{X,Y,W|\mathbf{Z}}^{*}(x_i, Y, W, \mathbf{Z})} \mid x_i, \mathbf{Z}\right].$$

We can then get $\mathbf{a}(x_i, \mathbf{Z})$ and $\mathbf{b}(x_i)$ by solving

$$\begin{bmatrix} \mathbf{C}(X, \mathbf{Z}) & \mathbf{D}(X, \mathbf{Z}) \\ \mathbf{F}(X, \mathbf{Z}) & \mathbf{G}(X, \mathbf{Z}) \end{bmatrix} \begin{bmatrix} \mathbf{A}(X, \mathbf{Z}) \\ \mathbf{B}(X) \end{bmatrix} = \begin{bmatrix} \mathbf{M}_1(X, \mathbf{Z}) \\ \mathbf{M}_2(X, \mathbf{Z}) \end{bmatrix}.$$

3 Asymptotic properties

Let $\mathbf{S}_{\text{res2}}(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, g)$ be $\mathbf{S}_{\text{res2}}(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ with all the appearance of $\mathbf{B}(X)^{\mathrm{T}}\boldsymbol{\gamma}$ in it replaced by $g(X)$. We first list the set of regularity conditions required for establishing the large sample properties of our estimator.

(C1) The true density $f_X(x)$ is bounded with compact support. Without loss of generality, we assume the support of $f_X(x)$ is $[0, 1]$.

(C2) The function $g(x) \in C^q([0, 1])$, $q > 1$, is bounded.

(C3) The spline order $r \geq q$.

(C4) Define the knots $t_{-r+1} = \cdots = t_0 = 0 < t_1 < \cdots < t_N < 1 = t_{N+1} = \cdots = t_{N+r}$, where N is the number of interior knots that satisfies $N \to \infty$, $N^{-1}n(\log n)^{-1} \to \infty$ and $Nn^{-1/(2q)} \to \infty$ as $n \to \infty$. Denote the number of spline bases $d_{\boldsymbol{\gamma}}$, i.e., $d_{\boldsymbol{\gamma}} = N + r$.

(C5) Let h_j be the distance between the jth and $(j-1)$th interior knots. Let $h_b = \max_{1 \leq j \leq N} h_j$ and $h_s = \min_{1 \leq j \leq N} h_j$. There exists a constant $c_h \in (0, \infty)$ such that $h_b/h_s < c_h$. Hence, $h_b = O_p(N^{-1})$ and $h_s = O_p(N^{-1})$.

(C6) $\boldsymbol{\gamma}_0$ is a $d_{\boldsymbol{\gamma}}$-dimensional spline coefficient vector such that $\sup_{x \in [0,1]} |\mathbf{B}(x)^{\mathrm{T}}\boldsymbol{\gamma}_0 - g(x)| = O_p(h_b^q)$.

(C7) The equation set

$$E\{\mathbf{S}_{\text{eff}}^*(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}, \boldsymbol{\gamma})\} = \mathbf{0},$$
$$E\{\mathbf{S}_{\text{res2}}^*(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}, \boldsymbol{\gamma})\} = \mathbf{0}$$

has unique root for $\boldsymbol{\theta}$ in the neighborhood of $\boldsymbol{\theta}_0$. Recall that $\boldsymbol{\theta} = (\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\gamma}^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{\delta} = (\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}$. The derivatives with respect to $\boldsymbol{\theta}$ of the left-hand side are smooth functions of $\boldsymbol{\theta}$, with its singular values bounded and bounded away from $\mathbf{0}$. Let the unique root be $\boldsymbol{\theta}^*$. Note that $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}^*$ are functions of N, that is, for any sufficiently large N, there is a unique root $\boldsymbol{\theta}^*$ in the neighborhood of $\boldsymbol{\theta}_0$.

(C8) The maximum absolute row sum of the matrix $\partial \mathbf{S}_{\text{eff}}^*(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}_0, \boldsymbol{\gamma}_0)/\partial \boldsymbol{\gamma}_0^{\mathrm{T}}$, i.e., $\|\partial \mathbf{S}_{\text{eff}}^*(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}_0, \boldsymbol{\gamma}_0)/\partial \boldsymbol{\gamma}_0^{\mathrm{T}}\|_\infty$, is integrable.

The conditions listed above are all standard bounded, smoothness conditions on functions and some classical conditions imposed on the spline order and number of knots. These are commonly used conditions in spline approximation and semiparametric regression literature. We now establish the consistency of $\widehat{\boldsymbol{\delta}}_n$ and $\widehat{\boldsymbol{\gamma}}_n$ as well as the asymptotic distribution property of $\widehat{\boldsymbol{\delta}}_n$.

Theorem 1 *Assume Conditions* (C1)$-$(C7) *to hold. Let* $\widehat{\boldsymbol{\theta}}_n$ *satisfy*

$$\frac{1}{n}\sum_{i=1}^n \mathbf{S}_{\text{eff}}^*(Y_i, W_i, \mathbf{Z}_i, \widehat{\boldsymbol{\delta}}_n, \widehat{\boldsymbol{\gamma}}_n) = \mathbf{0}$$

$$\frac{1}{n}\sum_{i=1}^n \mathbf{S}_{\text{res2}}^*(Y_i, W_i, \mathbf{Z}_i, \widehat{\boldsymbol{\delta}}_n, \widehat{\boldsymbol{\gamma}}_n) = \mathbf{0}.$$

Then, $\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = o_p(1)$ *element-wise.*

The result in Theorem 1 is used to further establish the asymptotic properties of the estimator of the parameters of interest $\widehat{\boldsymbol{\delta}}_n$ and the estimator of the function of interest $\mathbf{B}(\cdot)^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_n$.

Theorem 2 *Assume Conditions* (C1) − (C8) *to hold and let*

$$\mathbf{Q} \equiv E\left\{\left.\frac{\partial \mathbf{S}^*_{\mathrm{eff}}(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}_0, \boldsymbol{\gamma})}{\partial \boldsymbol{\delta}_0^{\mathrm{T}}}\right|_{\mathbf{B}(\cdot)^{\mathrm{T}}\boldsymbol{\gamma}=g(\cdot)}\right\}.$$

Then,

$$\sqrt{n}(\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_0) = -\mathbf{Q}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{S}^*_{\mathrm{eff}}(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}_0, g) + o_p(1).$$

Consequently, $\sqrt{n}(\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_0) \to N(\mathbf{0}, \mathbf{V})$ in distribution when $n \to \infty$, where

$$\mathbf{V} = \mathbf{Q}^{-1}var\{\mathbf{S}^*_{\mathrm{eff}}(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}_0, g)\}(\mathbf{Q}^{-1})^{\mathrm{T}}.$$

Theorem 2 indicates that $\boldsymbol{\delta}$ is estimated at the root-n rate. The proofs of Theorems 1 and 2 are given in "Appendix." Because the B-spline estimation of $g(\cdot)$ is at a slower rate than root-n, the estimation of $\boldsymbol{\delta}$ does not have any impact on the first-order asymptotic properties of $\widehat{g}$. Thus, for the analysis of the asymptotic properties of $\widehat{g}$, we can treat $\boldsymbol{\delta}$ as known. Then, the proof of Theorem 2 in Jiang and Ma (2018) can be directly used. We skip the details of the proof and provide the specific convergence property of the estimation of g in Theorem 3.

Theorem 3 *Assume Conditions* (C1) − (C8) *to hold and let*

$$\mathbf{P} \equiv E\left\{\left.\frac{\partial \mathbf{S}^*_{\mathrm{res2}}(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}_0, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^{\mathrm{T}}}\right|_{\mathbf{B}(\cdot)^{\mathrm{T}}\boldsymbol{\gamma}=g(\cdot)}\right\}.$$

Then, $\|\widehat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_0\|_2 = O_p\{(nh_b)^{-1/2}\}$. Further,

$$\widehat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_0 = -\mathbf{P}^{-1}n^{-1}\sum_{i=1}^n \mathbf{S}^*_{\mathrm{res2}}(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}_0, \boldsymbol{\gamma})\{1 + o_p(1)\}.$$

This leads to that $\widehat{g}(x)$, which equals $\mathbf{B}(x)^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_n$, satisfies $\sup_{x\in[0,1]} |\widehat{g}(x) - g(x)| = O_p\{(nh_b)^{-1/2}\}$. Specifically, $bias\{\widehat{g}(x)\} = E\{\widehat{g}(x) - g(x)\} = O(h_b^{q-1/2})$ and

$$\sqrt{nh_b}[\widehat{g}(x) - g(x) - bias\{\widehat{g}(x)\}]$$
$$= \sqrt{nh_b}\mathbf{B}(x)^{\mathrm{T}}\left\{-\mathbf{P}^{-1}n^{-1}\sum_{i=1}^n \mathbf{S}^*_{\mathrm{res2}}(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}_0, g)\right\} + o_p(1).$$

4 Numerical study

In our first simulation, we generated the observations $(W_i, \mathbf{Z}_i, Y_i)$ from the model

$$\mathrm{pr}(Y_i = 1 | X_i = x_i, Z_i = z_i) = H\{g(x_i) + \beta_1 z_{1i} + \beta_2 z_{2i} + \beta_3 z_{3i} + \beta_4 z_{4i}\}, \quad (9)$$

where $W = X + U$ and $U = \mathrm{normal}(0, 0.03)$. The true function is: $g(x) = -5 \exp\{-0.8(x - 2.5)^2\}$ and $H(t)$ is the inverse logit link function. We set $\beta_1 = 1$, $\beta_2 = 0.5$, $\beta_3 = 1$ and $\beta_4 = -0.3$. The sample size is 1000, and we run 1000 simulations. X_i is generated from a truncated normal distribution with mean 0.5 and variance $1/36$ on [0,1] independently of $\mathbf{Z}_i$. We implemented our method using a normal working model, corresponding to a correct working model case. In order to investigate the performance of our method under a misspecified working model, we also performed another study, in which we have X_i generated from a truncated student t distribution with degrees of freedom 5. Covariates Z_{1i}, Z_{2i} and Z_{4i} are generated from the standard normal distribution. The covariate Z_{3i} is generated from a uniform distribution on $[-1, 1]$. In both studies, we estimated both the parameters $\beta_1, \beta_2, \beta_3, \beta_4$ and the function $g(x)$.

In the second simulation, we set the true g function to be $g(x) = -5 \exp(-0.2x^2) + 5$, while all other settings remain the same. Similarly to the first simulation, we compared the performance of a correct working model and a misspecified working model in terms of estimating both $\beta_1, \beta_2, \beta_3, \beta_4$ and $g(x)$.

In the third simulation, we increase the sample size to 2000 to see the performance of our method while having a g function that has more nonlinear feature. Specifically, we set $g(x) = \sin(2\pi x)$, while keep all other settings unchanged.

In simulations 1, 2 and 3, we discretized the distribution of X on [0, 1] to $m = 15$ equal segments and we use the truncated normal distribution discussed earlier as our working model. We used quadratic splines with seven equally spaced knots on [0, 1] to estimate $g(x)$. The number of knots is chosen to be larger than $n^{1/4}$ to reflect condition (C4). When we further increase the number of knots, the results do not change much. The simulation results are shown in Tables 1, 2 and Figs. 1, 2 and 3.

The results in Tables 1 and 2 show little bias for the $\boldsymbol{\beta}$ estimation, regardless a correct working model or a misspecified working model is used. Figures 1, 2 and 3 show that the estimators of $g(x)$ have somewhat large bias on the boundary in both methods, which are within our expectation when factoring in the boundary effect. The performance of $g(x)$ estimation is satisfactory in the interior of the function domain. The simulation results show no big difference between the performance of the correct working model of $f_X(x)$ and a misspecified one, confirming our theory on consistency in both cases.

5 Data analysis

The data set we analyzed is from an AIDS Clinical Trials Group (ACTG) study. The goal of this study is to compare four different treatments, "ZDV," "ZDV+ddI," "ZDV+ddC" and "ddC," on HIV-infected adults whose CD4 cell counts were from

Table 1 Simulation results under a correct working model

Truth	β_1 1.0	β_2 0.5	β_3 1.0	β_4 -0.3
Simulation 1				
Mean	1.0183	0.5096	1.0106	-0.3103
Median	1.0125	0.5075	1.010	-0.3086
SE	0.0955	0.0817	0.1231	0.0792
MSE	0.0095	0.0068	0.0153	0.0064
Simulation 2				
Mean	1.0177	0.5052	1.0074	-0.3051
Median	1.0136	0.5043	1.0068	-0.3029
SE	0.0858	0.0467	0.0917	0.0425
MSE	0.0077	0.0022	0.0085	0.0018
Simulation 3				
Mean	1.0231	0.5127	1.0210	-0.3105
Median	1.0199	0.5122	1.0144	-0.3088
SE	0.0846	0.0652	0.0979	0.0666
MSE	0.0077	0.0044	0.0100	0.0045

Table 2 Simulation results under a misspecified working model

Truth	β_1 1.0	β_2 0.5	β_3 1.0	β_4 -0.3
Simulation 1				
Mean	1.0145	0.5122	1.0147	-0.3100
Median	1.0113	0.5117	1.0172	-0.3087
SE	0.0941	0.0823	0.1240	0.0822
MSE	0.0091	0.0069	0.0156	0.0069
Simulation 2				
Mean	1.0149	0.5051	1.0063	-0.3035
Median	1.0083	0.5038	1.0059	-0.3017
SE	0.0762	0.0463	0.0882	0.0320
MSE	0.0060	0.0022	0.0078	0.0010
Simulation 3				
Mean	1.0279	0.5115	1.0214	-0.3093
Median	1.0220	0.5105	1.0237	-0.3080
SE	0.0880	0.0673	0.0992	0.0667
MSE	0.0085	0.0047	0.0103	0.0045

200 to 500 per cubic millimeter. We labeled those treatments as treatment 1, treatment 2, treatment 3 and treatment 4. We used treatment 1 as the base treatment because it is a standard treatment. There were 1036 patients enrolled in the study and they had no antiretroviral therapy at enrollment. The criterion that we used to compare the four treatments is whether a patient has his or her CD4 count drop below 50%, which is an
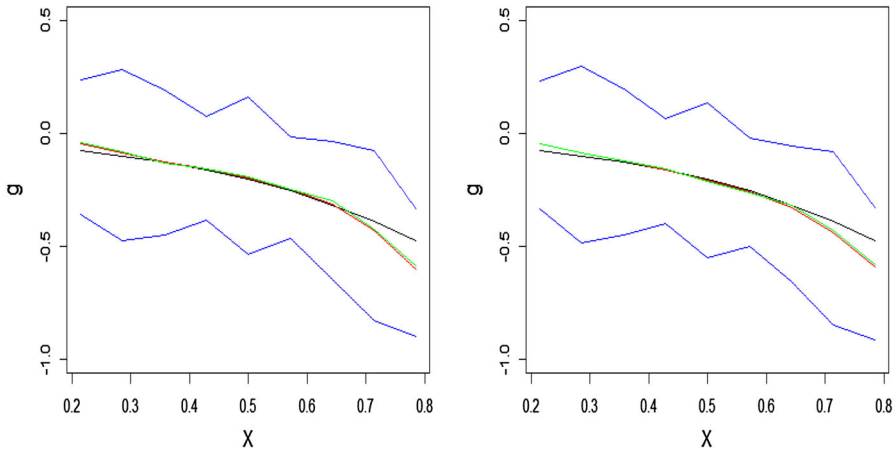
Fig. 1 True function (black line), median estimation (green line), mean estimation (red line) and 90% confidence band (blue line) of $g(x)$ in simulation 1. Correct working model on the left and misspecified working model on the right (color figure online)
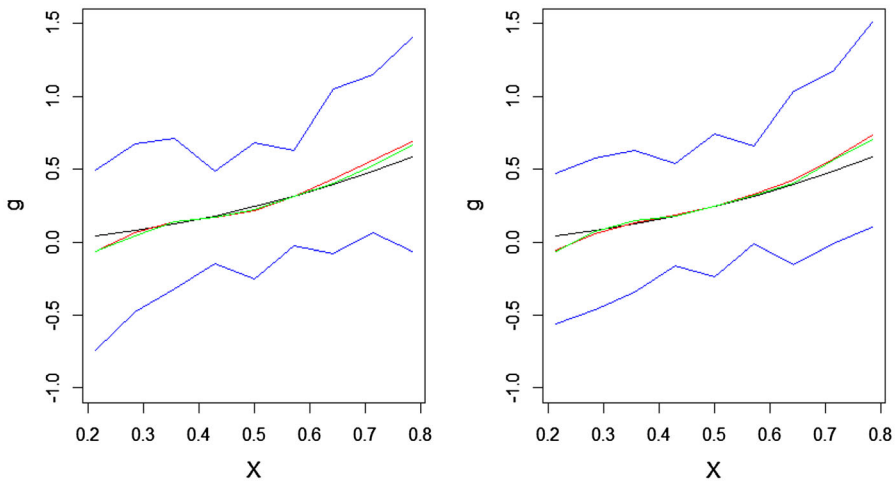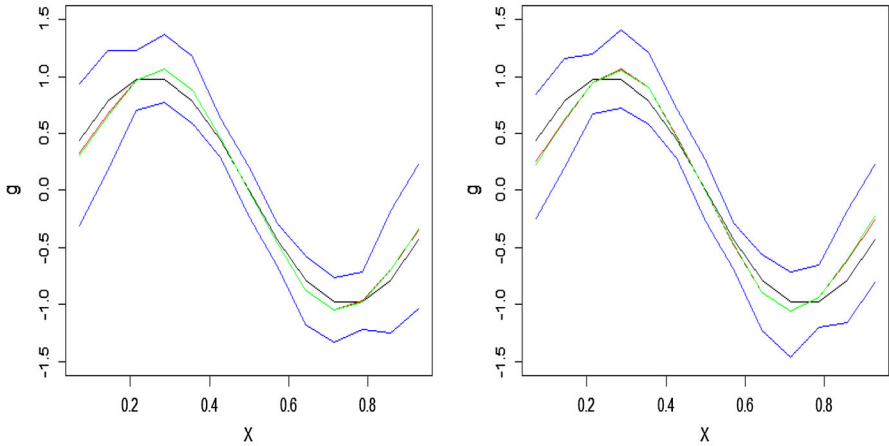

Fig. 2 True function (black line), median estimation (green line), mean estimation (red line) and 90% confidence band (blue line) of $g(x)$ in simulation 2. Correct working model on the left and misspecified working model on the right (color figure online)

important indicator for HIV-infected patients to develop AIDS or die. We have $Y = 1$ if a patient has his or her CD4 count drop below 50%, and $Y = 0$ otherwise.

Our model has the form:

$$\text{pr}(Y_i = 1 | X_i = x_i, Z_i = z_i) = H\{g(x_i) + \beta_1 z_{1i} + \beta_2 z_{2i} + \beta_3 z_{3i}\}, \qquad (10)$$

where $W = X + U$ and $U = \text{normal}(0, \sigma_U^2)$. The covariates Z_1, Z_2 and Z_3 are dichotomous variables. $Z_{1i} = Z_{2i} = Z_{3i} = 0$ indicates that the ith individual receives

Fig. 3 True function (black line), median estimation (green line), mean estimation (red line) and 90% confidence band (blue line) of $g(x)$ in simulation 3. Correct working model on the left and misspecified working model on the right (color figure online)

Table 3 Real data analysis results	Estimates	β_1	β_2	β_3
		-0.8076	-1.0970	-0.5150
	Bootstrap mean	-0.7876	-1.1012	-0.4896
	Bootstrap median	-0.7770	-1.0953	-0.4827
	Bootstrap SE	0.3575	0.3133	0.2947
	P-value	0.0239	<0.0001	0.0805

treatment 1, the base treatment; $Z_{1i} = 1$ and $Z_{2i} = Z_{3i} = 0$ indicates that the ith individual receives treatment 2; $Z_{1i} = 0$, $Z_{2i} = 1$ and $Z_{3i} = 0$ indicates that the ith individual receives treatment 3; $Z_{1i} = Z_{2i} = 0$ and $Z_{3i} = 1$ indicates that the ith individual receives treatment 4. The covariate X is the baseline log(CD4 count) prior to the start of treatment. Because CD4 count cannot be measured precisely, X is considered as our unobservable covariate. We use the average of two available measurements of log(CD4 count) as W.

First, we estimated the variance of U using the two repeated measurements and we got $\widehat{\sigma}_U^2 = 0.3$. Then, we constructed our working model of unobservable variance X. We assume that X follows a truncated normal distribution and estimated its variance by $\widehat{\sigma}_X^2 = \widehat{\sigma}_W^2 - \widehat{\sigma}_U^2$.

Table 3 shows that treatment 2, treatment 3 and treatment 4 are more efficient than the baseline treatment, i.e., treatment 1, at 90% confidence level according to the P-values of β_1, β_2 and β_3. The estimated index function $g(x)$ is in Fig. 4. We generated 1000 bootstrapped samples and calculated the bootstrapped mean, median and 90% confidence band for $g(x)$. It shows that $g(x)$ is an decreasing function, indicating that a large baseline CD4 count leads to a smaller risk of developing AIDS or having his/her CD4 counts drop below 50%. Thus, our analysis indicates that in general, the alternative treatments and a higher baseline CD4 count are beneficial to a patient.
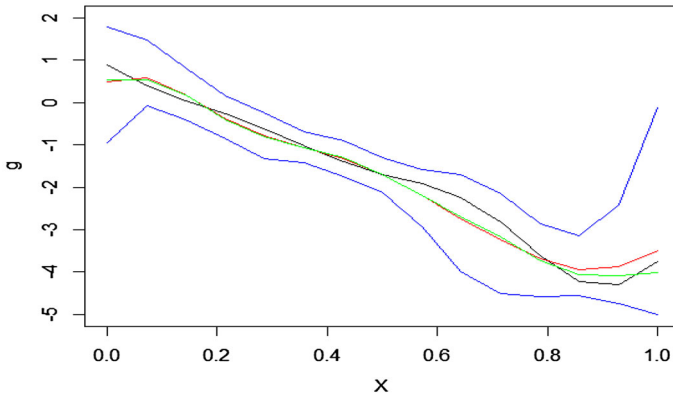
Fig. 4 Estimated $g(x)$ for real data (black line), median estimation (green line), mean estimation (red line) and 90% confidence band (blue line) of $g(x)$ from 1000 bootstrapped samples (color figure online)

6 Discussion

We devised a consistent and locally efficient estimation procedure to estimate both parameters and functions in a generalized partially linear model where the covariate inside the nonparametric function is subject to measurement error. The method does not make any assumption on the distribution of the covariate measured with error other than its finite support, which is easily satisfied in practice. The method is efficient in terms of estimating the model parameters if a correct working model is used, and retains its consistency even if this working model is misspecified. The estimation procedure breaks free from the deconvolution approach, which is the norm of practice in handling nonparametric problems with measurement errors. Instead, a novel usage of B-spline approach in combination with semiparametric method is exploited to push through the analysis.

Many possible extensions can be explored further. Possibilities include handling multivariate covariates measured with error, via multivariate B-splines, or incorporating index modeling approach or additive structures. Although our method is developed conceptually for generalized linear models, we did not really make use of the linear structure, hence any model of the form $f(Y, g(X), \mathbf{Z}, \boldsymbol{\beta})$ can be treated in a similar way. To this end, the continuous Y case typically involves normal error and has been widely studied, while the binary response case is studied in the main text of this work. When Y is count data, many computational issues arise, and is worth careful investigation further.

We have assumed the measurement error U to either have a known distribution, or to have its model parameters estimable from multiple observations. Of course, any other available information to identify the measurement error distributional model parameter also works and the plug-in procedure is largely "blind" to how the parameter is estimated. Of course, the estimated distributional model parameter will alter the estimation variability of $\boldsymbol{\delta}$, which can be take into account in a standard way (Yi et al. 2015).

Appendix

A.1 Proof of Theorem 1

From the definitions of $\mathbf{S}^*_{\text{eff}}(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}, g)$ and $\mathbf{S}_{\text{res}2}^*(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}, \boldsymbol{\gamma})$, we have

$$E\{\mathbf{S}^*_{\text{eff}}(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}_0, g)|X_i, \mathbf{Z}_i\} = \mathbf{0},$$
$$E_a\{\mathbf{S}_{\text{res}2}^*(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}_0, \boldsymbol{\gamma}_0)|X_i, \mathbf{Z}_i\} = \mathbf{0},$$

where $_a$ here and throughout the text stands for "approximate," and E_a indicates the expectation calculated with $g(\cdot)$ replaced by the approximate model $\mathbf{B}(\cdot)^{\mathrm{T}}\boldsymbol{\gamma}_0$. Taking another expectation, we get

$$E\{\mathbf{S}^*_{\text{eff}}(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}_0, g)\} = \mathbf{0},$$
$$E_a\{\mathbf{S}_{\text{res}2}^*(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}_0, \boldsymbol{\gamma}_0)\} = \mathbf{0}.$$

Using Condition (C6), we further get

$$E\{\mathbf{S}^*_{\text{eff}}(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}_0, \boldsymbol{\gamma}_0)\} = o(1),$$
$$E\{\mathbf{S}_{\text{res}2}^*\{Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}_0, \boldsymbol{\gamma}_0)\} = o(1),$$

component-wise. Condition (C7) ensures that $[E\{\mathbf{S}^*_{\text{eff}}(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}, \boldsymbol{\gamma})\}^{\mathrm{T}}, E\{\mathbf{S}_{\text{res}2}^*(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}, \boldsymbol{\gamma})\}^{\mathrm{T}}]^{\mathrm{T}}$ is invertible near its zero $\boldsymbol{\theta}^*$ as a vector function of $\boldsymbol{\theta}$, and the first derivative of the inverse function is bounded in the neighborhood of $\boldsymbol{\theta}^*$. Therefore, $\|\boldsymbol{\theta}^* - \boldsymbol{\theta}_0\|_2 = o_p(1)$. On the other hand, since

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{S}^*_{\text{eff}}(Y_i, W_i, \mathbf{Z}_i, \widehat{\boldsymbol{\delta}}_n, \widehat{\boldsymbol{\gamma}}_n) = \mathbf{0},$$

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{S}_{\text{res}2}^*(Y_i, W_i, \mathbf{Z}_i, \widehat{\boldsymbol{\delta}}_n, \widehat{\boldsymbol{\gamma}}_n) = \mathbf{0},$$

we have

$$E\{\mathbf{S}^*_{\text{eff}}(Y_i, W_i, \mathbf{Z}_i, \widehat{\boldsymbol{\delta}}_n, \widehat{\boldsymbol{\gamma}}_n)\} = o(1),$$
$$E\{\mathbf{S}_{\text{res}2}^*(Y_i, W_i, \mathbf{Z}_i, \widehat{\boldsymbol{\delta}}_n, \widehat{\boldsymbol{\gamma}}_n)\} = o(1)$$

element-wise. Using exactly the same argument as above, we can also obtain $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*\|_2 = o_p(1)$. Hence, combining the two results, we get $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|_2 = o_p(1)$. $\qquad\square$

A.2 Proof of Theorem 2

We first write

$$
\mathbf{0} = n^{-1/2} \sum_{i=1}^{n} \mathbf{S}_{\text{eff}}^{*}\{Y_i, W_i, \mathbf{Z}_i, \widehat{\boldsymbol{\delta}}_n, \widehat{\boldsymbol{\gamma}}_n(\widehat{\boldsymbol{\delta}}_n)\}
$$
$$
= \mathbf{T}_1 + \mathbf{T}_2(\widetilde{\boldsymbol{\delta}}_n)\sqrt{n}(\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_0),
$$

where

$$
\mathbf{T}_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{S}_{\text{eff}}^{*}\{Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}_0, \widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\delta}_0)\},
$$
$$
\mathbf{T}_2(\boldsymbol{\delta}) = \mathbf{T}_{21}(\boldsymbol{\delta}) + \mathbf{T}_{22}(\boldsymbol{\delta})\frac{\partial \widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}^{\text{T}}},
$$

where

$$
\mathbf{T}_{21}(\boldsymbol{\delta}) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \mathbf{S}_{\text{eff}}^{*}(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}, \widehat{\boldsymbol{\gamma}}_n)}{\partial \boldsymbol{\delta}^{\text{T}}},
$$
$$
\mathbf{T}_{22}(\boldsymbol{\delta}) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \mathbf{S}_{\text{eff}}^{*}\{Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}, \widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\delta})\}}{\partial \widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\delta})^{\text{T}}},
$$

and $\widetilde{\boldsymbol{\delta}}_n$ is on the line connecting $\boldsymbol{\delta}_0$ and $\widehat{\boldsymbol{\delta}}_n$.

We further expand $\mathbf{T}_1$ as a function of $\widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\delta}_0)$ about $\boldsymbol{\gamma}_0(\boldsymbol{\delta}_0)$ to obtain

$$
\mathbf{T}_1 = \mathbf{T}_{11} + \mathbf{T}_{12}\{\widetilde{\boldsymbol{\gamma}}_n(\boldsymbol{\delta}_0)\}\sqrt{n}\{\widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\delta}_0) - \boldsymbol{\gamma}_0(\boldsymbol{\delta}_0)\},
$$

where

$$
\mathbf{T}_{11} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{S}_{\text{eff}}^{*}\{Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}_0, \boldsymbol{\gamma}_0(\boldsymbol{\delta}_0)\},
$$
$$
\mathbf{T}_{12}\{\boldsymbol{\gamma}(\boldsymbol{\delta}_0)\} = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \mathbf{S}_{\text{eff}}^{*}\{Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}_0, \boldsymbol{\gamma}(\boldsymbol{\delta}_0)\}}{\partial \boldsymbol{\gamma}(\boldsymbol{\delta}_0)^{\text{T}}},
$$

and $\widetilde{\boldsymbol{\gamma}}_n(\boldsymbol{\delta}_0)$ is on the line connects $\widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\delta}_0)$ and $\boldsymbol{\gamma}_0(\boldsymbol{\delta}_0)$.

Because of the consistency of $\mathbf{B}(x)^{\text{T}}\widetilde{\boldsymbol{\gamma}}_n$ to $g(x)$ derived from Condition (C6) and Theorem 1, and the weak law of large numbers, for arbitrary $d_{\boldsymbol{\gamma}} \times p$ matrix $\mathbf{G}$ with $\|\mathbf{G}\|_2 = 1$, we have

$$
\mathbf{T}_{12}\{\widetilde{\boldsymbol{\gamma}}_n(\boldsymbol{\delta}_0)\}\mathbf{G} = E\left\{ \frac{\partial \mathbf{S}_{\text{eff}}^{*}(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}_0, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^{\text{T}}}\mathbf{G}\bigg|_{\mathbf{B}(\cdot)^{\text{T}}\boldsymbol{\gamma}=g(\cdot)} \right\}\{1 + o_p(1)\},
$$

where

$$
E\left\{\frac{\partial \mathbf{S}^*_{\text{eff}}(Y_i, W_i, \mathbf{Z}_i, \delta_0, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^{\text{T}}}\mathbf{G}\bigg|_{\mathbf{B}(\cdot)^{\text{T}}\boldsymbol{\gamma}=g(\cdot)}\right\}
$$

$$
= \int \left\{\frac{\partial \mathbf{S}^*_{\text{eff}}(y_i, w_i, \mathbf{z}_i, \delta_0, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^{\text{T}}}\mathbf{G}\bigg|_{\mathbf{B}(\cdot)^{\text{T}}\boldsymbol{\gamma}=g(\cdot)}\right\} f(y_i, w_i, \mathbf{z}_i, \delta_0, g, f_X) dy_i dw_i d\mathbf{z}_i
$$

$$
= \int \left\{\frac{\partial \mathbf{S}^*_{\text{eff}}(y_i, w_i, \mathbf{z}_i, \delta_0, \boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}_0^{\text{T}}}\mathbf{G} + O_p(h_b^q)\right\}\{f(y_i, w_i, \mathbf{z}_i, \delta_0, \boldsymbol{\gamma}_0, f_X)
$$

$$
+ O_p(h_b^q)\}dy_i dw_i d\mathbf{z}_i
$$

$$
= \int \frac{\partial \mathbf{S}^*_{\text{eff}}(y_i, w_i, \mathbf{z}_i, \delta_0, \boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}_0^{\text{T}}}\mathbf{G} f(y_i, w_i, \mathbf{z}_i, \delta_0, \boldsymbol{\gamma}_0, f_X) dy_i dw_i d\mathbf{z}_i + O_p(h_b^q)
$$

$$
= \frac{\partial}{\partial \boldsymbol{\gamma}_0^{\text{T}}}\int \{\mathbf{S}^*_{\text{eff}}(y_i, w_i, \mathbf{z}_i, \delta_0, g) + O_P(h_b^q)\}\mathbf{G}\{f(y_i, w_i, \mathbf{z}_i, \beta_0, g, f_X)
$$

$$
+ O_p(h_b^q)\}dy_i dw_i d\mathbf{z}_i
$$

$$
- \int \{\mathbf{S}^*_{\text{eff}}(y_i, w_i, \mathbf{z}_i, \delta_0, g) + O_P(h_b^q)\}\mathbf{G}\frac{\partial f(y_i, w_i, \mathbf{z}_i, \delta_0, \boldsymbol{\gamma}_0, f_X)}{\partial \boldsymbol{\gamma}_0^{\text{T}}}dy_i dw_i d\mathbf{z}_i
$$

$$
+ O_P(h_b^q)
$$

$$
= - \int \mathbf{S}^*_{\text{eff}}(y_i, w_i, \mathbf{z}_i, \delta_0, g)\left\{\mathbf{G}^{\text{T}}\mathbf{S}_{a,\boldsymbol{\gamma}}(y_i, w_i, \mathbf{z}_i, \delta_0, \boldsymbol{\gamma}_0)\right\}^{\text{T}}
$$

$$
f(y_i, w_i, \mathbf{z}_i, \delta_0, g, f_X) dy_i dw_i d\mathbf{z}_i + O_p(h_b^p)
$$

$$
= O_p(h_b^q). \tag{A.1}
$$

Here, like before, $f(y_i, w_i, \mathbf{z}_i, \delta_0, \boldsymbol{\gamma}, f_X)$ stands for $f(y_i, w_i, \mathbf{z}_i, \delta_0, g, f_X)$ with $g(\cdot)$ replaced by $\mathbf{B}(\cdot)^{\text{T}}\boldsymbol{\gamma}$, and $\mathbf{S}_{a,\boldsymbol{\gamma}}(y_i, w_i, \mathbf{z}_i, \delta_0, \boldsymbol{\gamma}_0) \equiv \partial \log f(y_i, w_i, \mathbf{z}_i, \delta_0, \boldsymbol{\gamma}, f_X)/\partial \boldsymbol{\gamma}$. The second equality holds by condition (C6).

The third equality holds because $\|\partial \mathbf{S}^*_{\text{eff}}(y_i, w_i, \mathbf{z}_i, \delta_0, \boldsymbol{\gamma}_0)/\partial \boldsymbol{\gamma}_0^{\text{T}}\|_\infty$ is integrable by condition (C8) and $f(y_i, w_i, \mathbf{z}_i, \delta_0, \boldsymbol{\gamma}_0, f_X)$ is absolutely integrable. The fourth equality holds also by condition (C6). The fifth equality holds because $E\{\mathbf{S}^*_{\text{eff}}(y_i, w_i, \mathbf{z}_i, \delta, g)\} = \mathbf{0}$. For the last equality, we note that

$$
\mathbf{G}^{\text{T}}\mathbf{S}_{a,\boldsymbol{\gamma}}(y_i, w_i, \mathbf{z}_i, \delta_0, \boldsymbol{\gamma}_0) = E[s\{y_i, \mathbf{z}_i^{\text{T}}\boldsymbol{\beta}_0 + \mathbf{B}(X)^{\text{T}}\boldsymbol{\gamma}_0, \boldsymbol{\alpha}_0\}\mathbf{G}^{\text{T}}\mathbf{B}(X) \mid y_i, w_i, \mathbf{z}_i].
$$

By Condition (C6) and definitions of Λ_g and $\Lambda_{a,\boldsymbol{\gamma}}$, for any $d_{\boldsymbol{\gamma}} \times p$ matrix $\mathbf{G}$, there exists a function $\mathbf{h}(y_i, w_i, \mathbf{z}_i, \delta_0, g) \equiv E[s\{y_i, \mathbf{z}_i^{\text{T}}\boldsymbol{\beta}_0 + g(X), \boldsymbol{\alpha}_0\}\mathbf{G}^{\text{T}}\mathbf{B}(X) \mid y_i, w_i, \mathbf{z}_i] \in \Lambda_g$ such that

$$
\sup |\mathbf{G}^{\text{T}}\mathbf{S}_{a,\boldsymbol{\gamma}}(y_i, w_i, \mathbf{z}_i, \delta_0, \boldsymbol{\gamma}_0) - \mathbf{h}(y_i, w_i, \mathbf{z}_i, \delta_0, g)| = O_P(h_b^q).
$$

Further, $\mathbf{S}^*_{\text{eff}}(y_i, w_i, \mathbf{z}_i, \delta_0, g)$ is orthogonal to any function in Λ_g, thus the last equality holds. Hence, we obtain $\|\mathbf{T}_{12}\{\widetilde{\boldsymbol{\gamma}}(\delta_0)\}\|_2 = O_p(h_b^q)$.

Based on the asymptotic results of Proposition 4 in Jiang and Ma (2018), we have
$\|\widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\delta}_0) - \boldsymbol{\gamma}_0(\boldsymbol{\delta}_0)\|_2 = O_p\{(nh_b)^{-1/2}\}$. Then, we have

$$\|\mathbf{T}_{12}\{\widetilde{\boldsymbol{\gamma}}_n(\boldsymbol{\delta}_0)\}\sqrt{n}\{\widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\delta}_0) - \boldsymbol{\gamma}_0(\boldsymbol{\delta}_0)\}\|_2 = O_p(h_b^{q-1/2}).$$

Further, by (C6) we have $\mathbf{T}_{11} = n^{-1/2}\sum_{i=1}^{n}\mathbf{S}_{\text{eff}}^*(Y_i, W_i, Z_i, \boldsymbol{\delta}_0, g) + O_p(n^{1/2}h_b^q)$.
Since $h_b^{q-1/2} = o_p(n^{1/2}h_b^q)$, and $n^{1/2}h_b^q = o_p(1)$ by conditions (C4) and (C5), then

$$\mathbf{T}_1 = n^{-1/2}\sum_{i=1}^{n}\mathbf{S}_{\text{eff}}^*(Y_i, W_i, Z_i, \boldsymbol{\delta}_0, g) + o_p(1). \qquad (A.2)$$

We next consider each term in $\mathbf{T}_2(\widetilde{\boldsymbol{\delta}}_n)$. Since $\widehat{\boldsymbol{\gamma}}_n(\cdot)$ satisfies

$$n^{-1}\sum_{i=1}^{n}\mathbf{S}_{\text{res}2}^*\{Y_i, W_i, Z_i, \boldsymbol{\delta}, \widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\delta})\} = \mathbf{0}$$

for any $\boldsymbol{\delta}$,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\mathbf{S}_{\text{res}2}^*(Y_i, W_i, Z_i, \boldsymbol{\delta}, \widehat{\boldsymbol{\gamma}}_n)}{\partial\boldsymbol{\delta}^{\mathrm{T}}} + \frac{1}{n}\sum_{i=1}^{n}\frac{\partial\mathbf{S}_{\text{res}2}^*\{Y_i, W_i, Z_i, , \boldsymbol{\delta}, \widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\delta})\}}{\partial\widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\delta})^{\mathrm{T}}}\frac{\partial\widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\delta})}{\partial\boldsymbol{\delta}^{\mathrm{T}}} = \mathbf{0}.$$

Then,

$$\frac{\partial\widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\delta})}{\partial\boldsymbol{\delta}^{\mathrm{T}}} = -\{\mathbf{T}_{23}(\boldsymbol{\delta})\}^{-1}\mathbf{T}_{24}(\boldsymbol{\delta}),$$

where

$$\mathbf{T}_{23}(\boldsymbol{\delta}) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial\mathbf{S}_{\text{res}2}^*\{Y_i, W_i, Z_i, , \boldsymbol{\delta}, \widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\delta})\}}{\partial\widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\delta})^{\mathrm{T}}},$$

$$\mathbf{T}_{24}(\boldsymbol{\delta}) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial\mathbf{S}_{\text{res}2}^*(Y_i, W_i, Z_i, \boldsymbol{\delta}, \widehat{\boldsymbol{\gamma}}_n)}{\partial\boldsymbol{\delta}^{\mathrm{T}}}.$$

Hence,

$$\mathbf{T}_2(\widetilde{\boldsymbol{\delta}}_n) = \mathbf{T}_{21}(\widetilde{\boldsymbol{\delta}}_n) - \mathbf{T}_{22}(\widetilde{\boldsymbol{\delta}}_n)\{\mathbf{T}_{23}(\widetilde{\boldsymbol{\delta}}_n)\}^{-1}\mathbf{T}_{24}(\widetilde{\boldsymbol{\delta}}_n).$$

By the consistency of $\widetilde{\boldsymbol{\delta}}_n$ to $\boldsymbol{\delta}_0$ and $\mathbf{B}(x)^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_n$ to $g(x)$, we have

$$\mathbf{T}_{21}(\widetilde{\boldsymbol{\delta}}_n) = E\left\{\frac{\partial\mathbf{S}_{\text{eff}}^*(Y_i, W_i, Z_i, \boldsymbol{\delta}_0, \boldsymbol{\gamma})}{\partial\boldsymbol{\delta}_0^{\mathrm{T}}}\bigg|_{\mathbf{B}(\cdot)^{\mathrm{T}}\boldsymbol{\gamma}=g(\cdot)}\right\}\{1 + o_p(1)\},$$

and

$$\mathbf{T}_{24}(\widetilde{\boldsymbol{\delta}}_n) = E\left\{\left.\frac{\partial \mathbf{S}_{\mathrm{res}_2}^*(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}_0, \boldsymbol{\gamma})}{\partial \boldsymbol{\delta}_0^{\mathrm{T}}}\right|_{\mathbf{B}(\cdot)^{\mathrm{T}}\boldsymbol{\gamma}=g(\cdot)}\right\}\{1 + o_p(1)\}.$$

From (A.1), we also have

$$\mathbf{T}_{22}(\widetilde{\boldsymbol{\delta}}_n) = E\left\{\left.\frac{\partial \mathbf{S}_{\mathrm{eff}}^*(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}_0, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^{\mathrm{T}}}\right|_{\mathbf{B}(\cdot)^{\mathrm{T}}\boldsymbol{\gamma}=g(\cdot)}\right\}\{1 + o_p(1)\} = O_p(h_b^q).$$

Based on the proof of Proposition 4 in Jiang and Ma (2018), we have $\|\mathbf{T}_{23}(\widetilde{\boldsymbol{\delta}}_n)^{-1}\|_2 = O_p(h_b^{-1})$. Therefore, we have $\mathbf{T}_{22}(\widetilde{\boldsymbol{\delta}}_n)\{\mathbf{T}_{23}(\widetilde{\boldsymbol{\delta}}_n)\}^{-1}\mathbf{T}_{24}(\widetilde{\boldsymbol{\delta}}_n) = O_p(h_b^{q-1})$, where $q > 1$ by condition (C2). Thus,

$$\mathbf{T}_2(\widetilde{\boldsymbol{\delta}}_n) = E\left\{\left.\frac{\partial \mathbf{S}_{\mathrm{eff}}^*(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}_0, \boldsymbol{\gamma})}{\partial \boldsymbol{\delta}_0^{\mathrm{T}}}\right|_{\mathbf{B}(\cdot)^{\mathrm{T}}\boldsymbol{\gamma}=g(\cdot)}\right\}\{1 + o_p(1)\} + O(h_b^{q-1}).$$

Therefore,

$$
\begin{aligned}
&\sqrt{n}(\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_0)\\
&= -\left[E\left\{\left.\frac{\partial \mathbf{S}_{\mathrm{eff}}^*(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}_0, \boldsymbol{\gamma})}{\partial \boldsymbol{\delta}_0^{\mathrm{T}}}\right|_{\mathbf{B}(\cdot)^{\mathrm{T}}\boldsymbol{\gamma}=g(\cdot)}\right\}\right]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{S}_{\mathrm{eff}}^*(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}_0, g)\\
&\quad + o_p(1).
\end{aligned}
$$

Since $n^{-1/2}\sum_{i=1}^{n}\mathbf{S}_{\mathrm{eff}}^*(Y_i, W_i, \mathbf{Z}_i, \boldsymbol{\delta}_0, g)$ is the sum of zero-mean random vectors, this will converge in distribution to a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{V}$ given in Theorem 2. $\qquad\square$

References

Apanasovich TV, Carroll RJ, Maity A (2009) SIMEX and standard error estimation in semiparametric measurement error models. Electron J Stat 3:318–348

Buonaccorsi JP (2010) Measurement error: models, methods, and applications. Chapman & Hall/CRC, New York

Carroll RJ, Fan J, Gijbels I, Wand MP (1995) Generalized partially linear single-index models. J Am Stat Assoc 92:477–489

Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu C (2006) Measurement error in nonlinear models: a modern perspective, 2nd edn. CRC Press, London

Huang Y, Wang CY (2001) Consistent functional methods for logistic regression with errors in covariates. J Am Stat Assoc 96:1469–1482

Jiang F, Ma Y (2018) A spline-assisted semiparametric approach to nonparametric measurement error models. arXiv:1804.00793

Liang H, Ren H (2005) Generalized partially linear measurement error models. J Comput Graph Stat 14:237–250

Liang H, Thurston SW (2008) Additive partial linear models with measurement errors. Biometrika 95:667–678

Liang H, Qin Y, Zhang X, Ruppert D (2009) Empirical likelihood-based inferences for generalized partially linear models. Scand Stat Theory Appl 36:433–443

Liu L (2007) Estimation of generalized partially linear models with measurement error using sufficiency scores. Stat Probab Lett 77:1580–1588

Liu J, Ma Y, Zhu L, Carroll RJ (2017) Estimation and inference of error-prone covariate effect in the presence of confounding variables. Electron J Stat 11:480–501

Ma Y, Carroll RJ (2006) Locally efficient estimators for semiparametric models with measurement error. J Am Stat Assoc 101:1465–1474

Ma Y, Tsiatis AA (2006) Closed form semiparametric estimators for measurement error models. Stat Sin 16:183–193

Stefanski LA, Carroll RJ (1985) Covariate measurement error in logistic regression. Ann Stat 13:1335–1351

Stefanski LA, Carroll RJ (1987) Conditional scores and optimal scores in generalized linear measurement error models. Biometrika 74:703–716

Tsiatis AA, Ma Y (2004) Locally efficient semiparametric estimators for functional measurement error models. Biometrika 91:835–848

Xu K, Ma Y (2015) Instrument assisted regression for errors in variables models with binary response. Scand J Stat 42:104–117

Yi G, Ma Y, Spiegelman D, Carroll RJ (2015) Functional and structural methods with mixed measurement error and misclassification in covariates. J Am Stat Assoc 110:681–696

Yu Y, Ruppert D (2012) Penalized spline estimation for partially linear single-index models. J Am Stat Assoc 97:1042–1054