Brazilian Journal of Probability and Statistics
2020, Vol. 34, No. 4, 770–794
https://doi.org/10.1214/19-BJPS463

Generalized partially linear single index model with measurement error, instruments and binary response

Guangren Yang[a], **Qianqian Wang**[b], **Xia Cui**[c], **and Yanyuan Ma**[d]

[a] *Jinan University*
[b] *University of South Carolina*
[c] *Guangzhou University*
[d] *Pennsylvania State University*

Abstract. Partially linear generalized single index models are widely used and have attracted much attention in the literature. However, when the covariates are subject to measurement error, the problem is much less studied. On the other hand, instrumental variables are important elements in studying many errors-in-variables problems. We use the relation between the unobservable variables and the instruments to devise consistent estimators for partially linear generalized single index models with binary response. We establish the consistency, asymptotic normality of the estimator and illustrate the numerical performance of the method through simulation studies and a data example. Despite the connection to (*Scand. J. Statist.* **42** (2015) 104–117) in its general layout, the mathematical derivations are much more challenging in the context studied here.

1 Introduction

Generalized linear models are familiar tools that are widely used in statistical applications. The model becomes complicated when the dependence of the response to some covariates, even after the transformation with a suitable link function, is not linear. A feasible and flexible approach to this is through introducing a partially linear single index structure, so that some covariates are modeled linearly, while some other covariates are summarized into an index, and the relation of the index to the response is modeled nonparametrically. This leads to the generalized partially linear single index model. A further complexity is when some of the covariates are measured with errors. Ignoring the measurement errors can generally lead to biased results, while taking the measurement error into account is also hard without specifying the measurement error variability exactly. Specifically, we denote the binary response variable Y, and let the $q \times 1$ covariate vector observed without error be $\mathbf{Z}$. We further let $\mathbf{X}$ be a $p \times 1$ latent variable. The model we study then is explicitly written as

$$\mathrm{pr}(Y = 1|\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}) = H\{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + g(\mathbf{z}^{\mathrm{T}}\boldsymbol{\gamma})\}, \tag{1.1}$$

where $\boldsymbol{\beta} \in R^p$ and $\boldsymbol{\gamma} \in R^q$ are unknown parameters of interest, $H(\cdot)$ is a known inverse link function, for example, the inverse logit link function $H(\cdot) = 1 - 1/\{\exp(\cdot) + 1\}$ or the inverse probit link function $H(\cdot) = \Phi(\cdot)$, and $g(\cdot)$ is an unknown function. Because $\boldsymbol{\gamma}$ is not identifiable when incorporated with an unspecified g, the constraint $\|\boldsymbol{\gamma}\| = 1$ or the first component of $\boldsymbol{\gamma}$ is positive is often imposed. Here, we use the latter choice, which fixes the first component of $\boldsymbol{\gamma}$ to be 1 and leave the remaining components arbitrary. We denote the vector formed by the second to last components of $\boldsymbol{\gamma}$ as $\boldsymbol{\gamma}_{-1}$.

When $\mathbf{X}$ is latent or observed with error, the parameters in model (1.1) are generally hard to identify in practice. However, the existence of instruments is often very helpful and can

Key words and phrases. Errors in variables, generalized linear models, instrumental variables, measurement errors, partially linear models, single index models.
Received August 2018; accepted October 2019.

save the situation. Instead of observing $\mathbf{X}$, we observe an erroneous version of $\mathbf{X}$, written as $\mathbf{W}$ and an instrumental variable $\mathbf{S}$. Here we do not impose any restriction on the dimension of $\mathbf{S}$. The variables $\mathbf{W}$ and $\mathbf{S}$ are linked to $\mathbf{X}$ through

$$\mathbf{W} = \mathbf{X} + \mathbf{U} \quad \text{and} \quad \mathbf{X} = \mathbf{m}(\mathbf{S}, \mathbf{Z}; \boldsymbol{\alpha}) + \boldsymbol{\varepsilon}, \tag{1.2}$$

where $\mathbf{m}(\cdot)$ is a known function up to an unknown parameter $\boldsymbol{\alpha}$. Here, we assume the conditional mean of $\boldsymbol{\varepsilon}$ and the marginal mean of $\mathbf{U}$ to be zero, that is, $E(\boldsymbol{\varepsilon}|\mathbf{S}, \mathbf{Z}) = \mathbf{0}$, $E(\mathbf{U}) = \mathbf{0}$. Further assume that $(\mathbf{X}, \mathbf{S}, \mathbf{Z})$ is independent of $\mathbf{U}$, $\mathbf{U}$ is independent of $\boldsymbol{\varepsilon}$, $\mathbf{W}$ is independent of $(\mathbf{S}, \mathbf{Z})$ given $\mathbf{X}$, and Y is independent of $(\mathbf{W}, \mathbf{S})$ given $(\mathbf{X}, \mathbf{Z})$. The independence between $\mathbf{U}$ and $(\mathbf{X}, \mathbf{S}, \mathbf{Z})$ is a standard assumption in the measurement error literature and whether or not it is valid is based on empirical knowledge. In most cases when the measurement error $\mathbf{U}$ occurs due to reasons external to the covariates, this is a reasonable assumption. The independence between $\mathbf{U}$ and ϵ is often a reasonable assumption as well, because they are errors resulting from two different procedures. The independence of $\mathbf{W}$ and $(\mathbf{S}, \mathbf{Z})$ given $\mathbf{X}$ indicates that the only link between $\mathbf{W}$ and $(\mathbf{S}, \mathbf{Z})$ is through $\mathbf{X}$, and it is indeed the case when $\mathbf{W}$ is $\mathbf{X}$ plus a pure error that is independent of everything else. The model in (1.1), in combination with the instrumental variable condition studied here, has much resemblance with the problem setting in Xu, Ma and Wang (2015). However, the critical difference lies in the presence of the unknown function g as well as the unknown index vector $\boldsymbol{\gamma}$. This seemingly small change actually brings much more complexity in all aspects of the analysis, including the method development, the theoretical proofs and the numerical implementation. To appreciate this fact, one can link to the additional hurdles encountered and overcome in the literature when moving from linear regression to single index models. Indeed, single index models are much more complex than linear regression models, due to the additional unspecified function involved, and subsequently all the different treatments involved in nonparametric function estimation.

As a field of much practical importance, measurement error models in general have been extensively studied. However, as far as we are aware, no work exists in studying measurement error models when the experiment model is of the generalized partially linear single index type with binary response, while an instrumental variable exists to provide additional information. In fact, the only works in handling binary response models with measurement errors that we are aware are Stefanski and Carroll (1985, 1987), Buzas and Stefanski (1996), Huang and Wang (2001), Ma and Tsiatis (2006), in addition to Xu, Ma and Wang (2015) mentioned above. However, none of these works contains a partially linear single index component, and most of these works do not consider instruments.

In this paper, we demonstrate that by employing a prediction relation for the unobserved covariates using available instruments, we can construct consistent estimators for all the parameters in the generalized linear single index model. In addition, we also provide a nonparametric estimator for the unspecified function of the estimated index. The method we devise incorporates instrumental variables in a different way from most traditional methods in handling instruments. In fact, our work is the first in using instruments in handling the generalized linear single index regression models with measurement error and binary response.

The rest of paper is organized as follows. We describe our main methodology and the asymptotic properties of our estimator in Section 2. Simulation studies are given in Section 3 to provide finite sample performance of our method. We analyze an AIDs study data in Section 4 and conclude the paper in Section 5.

2 Estimation procedure via profiling and the asymptotic properties

2.1 Methodology development

Denote the ith observed data $\mathbf{O}_i = (Y_i, \mathbf{W}_i, \mathbf{S}_i, \mathbf{Z}_i)$, for $i = 1, \dots, n$. These observations are independent and identically distributed (i.i.d.) according to the model described in (1.1) and (1.2). Our main interest is in estimating $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\gamma}_{-1}^{\mathrm{T}})^{\mathrm{T}}$. However, $g(\cdot)$ is an unknown nuisance function.

First of all, we have

$$\mathbf{W} = \mathbf{m}(\mathbf{S}, \mathbf{Z}; \boldsymbol{\alpha}) + \mathbf{U} + \boldsymbol{\varepsilon},$$

where $E(\mathbf{U} + \boldsymbol{\varepsilon} | \mathbf{S}, \mathbf{Z}) = \mathbf{0}$. We can use the least squares method to estimate $\hat{\boldsymbol{\alpha}}$, for example (2.1) is the estimating equation to obtain $\hat{\boldsymbol{\alpha}}$,

$$\sum_{i=1}^{n} \mathcal{S}_{\boldsymbol{\alpha}}(\mathbf{S}_i, \mathbf{Z}_i; \boldsymbol{\alpha}) = \sum_{i=1}^{n} \frac{\partial \mathbf{m}^{\mathrm{T}}(\mathbf{S}_i, \mathbf{Z}_i; \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \boldsymbol{\Omega}(\mathbf{S}_i, \mathbf{Z}_i)\{\mathbf{W}_i - \mathbf{m}(\mathbf{S}_i, \mathbf{Z}_i; \boldsymbol{\alpha})\} = \mathbf{0}, \tag{2.1}$$

where $\boldsymbol{\Omega}(\mathbf{S}, \mathbf{Z})$ is any weight matrix. We can choose to use ordinary least squares (OLS) or weighted least squares (WLS) method by using different weight matrices. Specifically, we can use identity matrix as weight matrix to obtain the OLS estimator and use the inverse of the error variance-covariance matrix conditional on $(\mathbf{S}, \mathbf{Z})$ as weight matrix to obtain the WLS estimator.

After we have an estimate $\hat{\boldsymbol{\alpha}}$, we can write $\mathbf{X}$ in the form of $\hat{\boldsymbol{\alpha}}$ and $(\mathbf{S}, \mathbf{Z})$, and insert into model (1.1) to obtain the joint distribution of $(Y, \mathbf{S}, \mathbf{Z})$ as

$$\begin{aligned} &\mathrm{pr}(Y = y, \mathbf{S} = \mathbf{s}, \mathbf{Z} = \mathbf{z}) \\ &\quad = f_{\mathbf{S},\mathbf{Z}}(\mathbf{s}, \mathbf{z}) \int [1 - y + (2y - 1) H\{\mathbf{m}(\mathbf{S}, \mathbf{Z}, \hat{\boldsymbol{\alpha}})^{\mathrm{T}} \boldsymbol{\beta} + \boldsymbol{\varepsilon}^{\mathrm{T}} \boldsymbol{\beta} + g(\mathbf{Z}^{\mathrm{T}} \boldsymbol{\gamma})\}] \\ &\qquad \times f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon} | \mathbf{s}, \mathbf{z})\, d\mu(\boldsymbol{\varepsilon}), \end{aligned} \tag{2.2}$$

where $f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon} | \mathbf{s}, \mathbf{z})$ is a conditional probability density function that satisfies

$$\int \boldsymbol{\varepsilon} f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon} | \mathbf{s}, \mathbf{z})\, d\mu(\boldsymbol{\varepsilon}) = \mathbf{0}$$

and $f_{\mathbf{S},\mathbf{Z}}(\mathbf{s}, \mathbf{z})$ is the joint p.d.f. of $(\mathbf{S}, \mathbf{Z})$.

Now we move to construct the estimation procedure for $\boldsymbol{\theta}$ and $g(\cdot)$. Borrowing the ideas in Ma and Carroll (2006) and Xu, Ma and Wang (2015), we will construct two sets of estimating equations in order to estimate $\boldsymbol{\theta}$ and $g(\cdot)$.

Treating (2.2) as a semiparametric model, the nuisance tangent space is

$$\begin{aligned} \Lambda &= \Lambda_1 \oplus \Lambda_2 \\ &= \{\mathbf{f}(\mathbf{S}, \mathbf{Z}) : E(\mathbf{f}) = 0, E(\mathbf{f}^{\mathrm{T}} \mathbf{f}) < \infty, \forall \mathbf{f} \in \mathcal{R}^{p+q-1}\} \\ &\quad \oplus \{E\{\mathbf{f}(\boldsymbol{\varepsilon}, \mathbf{S}, \mathbf{Z}) | Y, \mathbf{S}, \mathbf{Z}\} : E(\mathbf{f} | \mathbf{S}, \mathbf{Z}) = \mathbf{0}, E(\boldsymbol{\varepsilon} \mathbf{f}^{\mathrm{T}} | \mathbf{S}, \mathbf{Z}) = \mathbf{0}, \\ &\quad E(\mathbf{f}^{\mathrm{T}} \mathbf{f}) < \infty, \forall \mathbf{f} \in \mathcal{R}^{p+q-1}\}. \end{aligned}$$

Notation $\oplus$ is used to emphasize that an arbitrary function $\mathbf{f}_1(\mathbf{S}, \mathbf{Z})$ in Λ_1 and an arbitrary function $\mathbf{f}_2(\boldsymbol{\varepsilon}, \mathbf{S}, \mathbf{Z})$ in Λ_2 satisfy $E\{\mathbf{f}_1(\mathbf{S}, \mathbf{Z}) \mathbf{f}_2^{\mathrm{T}}(\boldsymbol{\varepsilon}, \mathbf{S}, \mathbf{Z})\} = \mathbf{0}$, that is, $\Lambda_1 \perp \Lambda_2$. Here Λ_1 results from the nuisance component $f_{\mathbf{S},\mathbf{Z}}(\mathbf{s}, \mathbf{z})$ in (2.2) and Λ_2 from $f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon} | \mathbf{s}, \mathbf{z})$ in (2.2). Specifically, for a parametric submodel of $f_{\mathbf{S},\mathbf{Z}}(\mathbf{s}, \mathbf{z})$, $f_{\mathbf{S},\mathbf{Z}}(\mathbf{s}, \mathbf{z}, \boldsymbol{\xi})$, because the only requirement we have is $f_{\mathbf{S},\mathbf{Z}}(\mathbf{s}, \mathbf{z}, \boldsymbol{\xi}) \geq 0$ and $\int f_{\mathbf{S},\mathbf{Z}}(\mathbf{s}, \mathbf{z}, \boldsymbol{\xi}) = 1$, hence any mean zero function can be a resulting nuisance score with respect to $\boldsymbol{\xi}$. This leads to the result of Λ_1. The same can be seen regarding $f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}|\mathbf{s},\mathbf{z})$, except that all the operations are conditional on $\mathbf{S},\mathbf{Z}$. Specifically, any function $f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}|\mathbf{s},\mathbf{z},\boldsymbol{\xi})$ can be a parametric submodel of $f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}|\mathbf{s},\mathbf{z})$, as long as $f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}|\mathbf{s},\mathbf{z},\boldsymbol{\xi})\geq 0$ and $\int f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}|\mathbf{s},\mathbf{z},\boldsymbol{\xi})\,d\boldsymbol{\varepsilon}=1$. Thus, its score function with respect to $\boldsymbol{\xi}$ can lead to any function in Λ_2. To understand the concept of parametric submodel and the meaning of nuisance tangent space, readers are referred to Chapter 4 of Tsiatis (2006). The orthogonal complement of Λ is

$$\Lambda^{\perp}=\bigl\{\mathbf{f}(Y,\mathbf{S},\mathbf{Z}):E(\mathbf{f}|\boldsymbol{\varepsilon},\mathbf{S},\mathbf{Z})=\mathbf{a}(\mathbf{S},\mathbf{Z})\boldsymbol{\varepsilon},\ \|E\mathbf{a}^{\mathrm{T}}\mathbf{a}\|_{\infty}<\infty,\ \forall\mathbf{f}\in\mathcal{R}^{p+q-1},\forall\mathbf{a}\in\mathcal{R}^{(p+q-1)\times p}\bigr\}.$$

Note that unlike the derivation of Λ, which is highly technical and requires understanding the semiparametric theory, the derivation of $\Lambda^{\perp}$ is purely mathematical and involves the verification of the orthogonality of Λ and $\Lambda^{\perp}$, and that $\Lambda+\Lambda^{\perp}$ forms the Hilbert space of all the mean zero finite variance functions.

Let $\mathcal{S}_{\boldsymbol{\theta}}\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\theta},g(\cdot)\}$, and $\mathcal{S}_g\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\theta},g(\cdot)\}$ be the score functions for $\boldsymbol{\theta}$ and $g(\cdot)$ respectively obtained by taking derivative with respect to $\boldsymbol{\theta}$ and g of the log likelihood of one observation. Specifically,

$$\mathcal{S}_{\boldsymbol{\theta}}\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\theta},g(\cdot)\}=(2Y-1)\frac{\int\binom{\mathbf{m}(\mathbf{S},\mathbf{Z},\hat{\boldsymbol{\alpha}})+\boldsymbol{\varepsilon}}{g'(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\gamma})\mathbf{Z}_{-1}}H'\{\Delta\}f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}|\mathbf{s},\mathbf{z})\,d\mu(\boldsymbol{\varepsilon})}{\int[1-Y+(2Y-1)H\{\Delta\}]f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}|\mathbf{s},\mathbf{z})\,d\mu(\boldsymbol{\varepsilon})},$$
$$S_g\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\theta},g(\cdot)\}=(2Y-1)\frac{\int H'\{\Delta\}f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}|\mathbf{s},\mathbf{z})\,d\mu(\boldsymbol{\varepsilon})}{\int[1-Y+(2Y-1)H\{\Delta\}]f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}|\mathbf{s},\mathbf{z})\,d\mu(\boldsymbol{\varepsilon})},$$

where $\Delta\equiv\mathbf{m}(\mathbf{S},\mathbf{Z},\hat{\boldsymbol{\alpha}})^{\mathrm{T}}\boldsymbol{\beta}+\boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{\beta}+g(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\gamma})$.

We get the efficient score by projecting $\mathcal{S}_{\boldsymbol{\theta}}$ and S_g to $\Lambda^{\perp}$ as detailed below. Let

$$\mathcal{L}\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\theta},g(\cdot)\}=\mathcal{S}_{\boldsymbol{\theta}}\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\theta},g(\cdot)\}-E[\mathbf{b}_{\boldsymbol{\theta}}\{\boldsymbol{\varepsilon},\mathbf{S},\mathbf{Z};\boldsymbol{\theta},g(\cdot)\}|Y,\mathbf{S},\mathbf{Z}],$$
$$\Phi\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\theta},g(\cdot)\}=S_g\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\theta},g(\cdot)\}-E[b_g\{\boldsymbol{\varepsilon},\mathbf{S},\mathbf{Z};\boldsymbol{\theta},g(\cdot)\}|Y,\mathbf{S},\mathbf{Z}],$$

where $\mathbf{b}_{\boldsymbol{\theta}}\{\boldsymbol{\varepsilon},\mathbf{S},\mathbf{Z};\boldsymbol{\theta},g(\cdot)\}\in\mathcal{R}^{p+q-1}$ and $b_g\{\boldsymbol{\varepsilon},\mathbf{S},\mathbf{Z};\boldsymbol{\theta},g(\cdot)\}\in\mathcal{R}$ satisfy

$$E\{\mathcal{S}_{\boldsymbol{\theta}}\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\theta},g(\cdot)\}-E[\mathbf{b}_{\boldsymbol{\theta}}\{\boldsymbol{\varepsilon},\mathbf{S},\mathbf{Z};\boldsymbol{\theta},g(\cdot)\}|Y,\mathbf{S},\mathbf{Z}]|\boldsymbol{\varepsilon},\mathbf{S},\mathbf{Z}\}=\mathbf{a}_{\boldsymbol{\theta}}(\mathbf{S},\mathbf{Z})\boldsymbol{\varepsilon},$$
$$E\{S_g\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\theta},g(\cdot)\}-E[b_g\{\boldsymbol{\varepsilon},\mathbf{S},\mathbf{Z};\boldsymbol{\theta},g(\cdot)\}|Y,\mathbf{S},\mathbf{Z}]|\boldsymbol{\varepsilon},\mathbf{S},\mathbf{Z}\}=\mathbf{a}_g(\mathbf{S},\mathbf{Z})\boldsymbol{\varepsilon},$$

where $\mathbf{a}_{\boldsymbol{\theta}}(\mathbf{S},\mathbf{Z})\in\mathcal{R}^{(p+q-1)\times p}$ and $\mathbf{a}_g(\mathbf{S},\mathbf{Z})\in\mathcal{R}^{1\times p}$. Here we have to specify the following terms $\mathbf{b}_{\boldsymbol{\theta}},\mathbf{a}_{\boldsymbol{\theta}},b_g$ and $\mathbf{a}_g$. By multiplying $\boldsymbol{\varepsilon}$ on both sides of the above formulas and taking expectation conditional on $(\mathbf{S},\mathbf{Z})$, we obtain

$$\begin{aligned}&\mathbf{a}_{\boldsymbol{\theta}}(\mathbf{S},\mathbf{Z})E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}|\mathbf{S},\mathbf{Z})\\&\quad=E\{\mathcal{S}_{\boldsymbol{\theta}}\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\theta},g(\cdot)\}\boldsymbol{\varepsilon}^{\mathrm{T}}-E[\mathbf{b}_{\boldsymbol{\theta}}\{\boldsymbol{\varepsilon},\mathbf{S},\mathbf{Z};\boldsymbol{\theta},g(\cdot)\}|Y,\mathbf{S},\mathbf{Z}]\boldsymbol{\varepsilon}^{\mathrm{T}}|\mathbf{S},\mathbf{Z}\},\\&\mathbf{a}_g(\mathbf{S},\mathbf{Z})E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}|\mathbf{S},\mathbf{Z})\\&\quad=E\{S_g\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\theta},g(\cdot)\}\boldsymbol{\varepsilon}^{\mathrm{T}}-E[b_g\{\boldsymbol{\varepsilon},\mathbf{S},\mathbf{Z};\boldsymbol{\theta},g(\cdot)\}|Y,\mathbf{S},\mathbf{Z}]\boldsymbol{\varepsilon}^{\mathrm{T}}|\mathbf{S},\mathbf{Z}\}.\end{aligned}$$

Then, we have

$$\begin{aligned}\mathbf{a}_{\boldsymbol{\theta}}(\mathbf{S},\mathbf{Z})&=E\{\mathcal{S}_{\boldsymbol{\theta}}\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\theta},g(\cdot)\}\boldsymbol{\varepsilon}^{\mathrm{T}}\\&\qquad-E[\mathbf{b}_{\boldsymbol{\theta}}\{\boldsymbol{\varepsilon},\mathbf{S},\mathbf{Z};\boldsymbol{\theta},g(\cdot)\}|Y,\mathbf{S},\mathbf{Z}]\boldsymbol{\varepsilon}^{\mathrm{T}}|\mathbf{S},\mathbf{Z}\}\{E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}|\mathbf{S},\mathbf{Z})\}^{-1},\\\mathbf{a}_g(\mathbf{S},\mathbf{Z})&=E\{S_g\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\theta},g(\cdot)\}\boldsymbol{\varepsilon}^{\mathrm{T}}\\&\qquad-E[b_g\{\boldsymbol{\varepsilon},\mathbf{S},\mathbf{Z};\boldsymbol{\theta},g(\cdot)\}|Y,\mathbf{S},\mathbf{Z}]\boldsymbol{\varepsilon}^{\mathrm{T}}|\mathbf{S},\mathbf{Z}\}\{E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}|\mathbf{S},\mathbf{Z})\}^{-1}.\end{aligned}$$

Inserting the form of $\mathbf{a}_{\boldsymbol\theta}(\mathbf{S},\mathbf{Z})$ and $\mathbf{a}_g(\mathbf{S},\mathbf{Z})$ respectively, we obtain the following equations

$$\begin{aligned}
&E\{\mathcal{S}_{\boldsymbol\theta}\{Y,\mathbf{S},\mathbf{Z};\boldsymbol\theta,g(\cdot)\}-E[\mathbf{b}_{\boldsymbol\theta}\{\boldsymbol\varepsilon,\mathbf{S},\mathbf{Z};\boldsymbol\theta,g(\cdot)\}|Y,\mathbf{S},\mathbf{Z}]|\boldsymbol\varepsilon,\mathbf{S},\mathbf{Z}\}\\
&\quad=E\{\mathcal{S}_{\boldsymbol\theta}\{Y,\mathbf{S},\mathbf{Z};\boldsymbol\theta,g(\cdot)\}\boldsymbol\varepsilon^{\mathrm T}\\
&\qquad-E[\mathbf{b}_{\boldsymbol\theta}\{\boldsymbol\varepsilon,\mathbf{S},\mathbf{Z};\boldsymbol\theta,g(\cdot)\}|Y,\mathbf{S},\mathbf{Z}]\boldsymbol\varepsilon^{\mathrm T}|\mathbf{S},\mathbf{Z}\}\{E(\boldsymbol\varepsilon\boldsymbol\varepsilon^{\mathrm T}|\mathbf{S},\mathbf{Z})\}^{-1}\boldsymbol\varepsilon, \qquad (2.3)\\
&E\{S_g\{Y,\mathbf{S},\mathbf{Z};\boldsymbol\theta,g(\cdot)\}-E[b_g\{\boldsymbol\varepsilon,\mathbf{S},\mathbf{Z};\boldsymbol\theta,g(\cdot)\}|Y,\mathbf{S},\mathbf{Z}]|\boldsymbol\varepsilon,\mathbf{S},\mathbf{Z}\}\\
&\quad=E\{S_g\{Y,\mathbf{S},\mathbf{Z};\boldsymbol\theta,g(\cdot)\}\boldsymbol\varepsilon^{\mathrm T}\\
&\qquad-E[b_g\{\boldsymbol\varepsilon,\mathbf{S},\mathbf{Z};\boldsymbol\theta,g(\cdot)\}|Y,\mathbf{S},\mathbf{Z}]\boldsymbol\varepsilon^{\mathrm T}|\mathbf{S},\mathbf{Z}\}\{E(\boldsymbol\varepsilon\boldsymbol\varepsilon^{\mathrm T}|\mathbf{S},\mathbf{Z})\}^{-1})\boldsymbol\varepsilon.
\end{aligned}$$

Then we can obtain the terms $\mathbf{b}_{\boldsymbol\theta}$ and b_g by solving the equations in (2.3). Unfortunately, the integral equations in (2.3) do not have a closed form solution hence numerical methods are required to obtain approximate solutions. In fact, these are first type Fredholm integral equations and require regularization to obtain stable solutions (Kress, 1991). Nevertheless, such integral equations are well studied in numerical analysis and many methods exist. Here, our final goal is to obtain $E[\mathbf{b}_{\boldsymbol\theta}\{\boldsymbol\varepsilon,\mathbf{S},\mathbf{Z};\boldsymbol\theta,g(\cdot)\}|Y,\mathbf{S},\mathbf{Z}]$ and $E[b_g\{\boldsymbol\varepsilon,\mathbf{S},\mathbf{Z};\boldsymbol\theta,g(\cdot)\}|Y,\mathbf{S},\mathbf{Z}]$ instead of $\mathbf{b}_{\boldsymbol\theta}$ and b_g, hence the numerical problem is an easier one to handle than the typical Type I Fredholm integral equations. For details on how to solve the integral equations in (2.3), we refer to Kress (1991).

Obviously, it follows that

$$\mathbf{0}=E[\mathcal{L}\{Y,\mathbf{S},\mathbf{Z};\boldsymbol\theta,g(\cdot)\}|\mathbf{S},\mathbf{Z}], \qquad (2.4)$$

$$0=E[\Phi\{Y,\mathbf{S},\mathbf{Z};\boldsymbol\theta,g(\cdot)\}|\mathbf{S},\mathbf{Z}]. \qquad (2.5)$$

Equations (2.4) and (2.5) form the backbone of our method that allows for a general unknown function $g(\cdot)$. Because $g(\cdot)$ is modeled nonparametrically, we use the local linear method to estimate $\widehat{g}(\cdot)$. Let $K(z)$ be a smooth symmetric density function, let h be a bandwidth, and define $K_h(z)=h^{-1}K(z/h)$. Let $U=\mathbf{Z}^{\mathrm T}\boldsymbol\gamma$ and $u_0=\mathbf{z}_0^{\mathrm T}\boldsymbol\gamma$. We approximate $g(\cdot)$ locally by a linear function

$$g(U,\boldsymbol\beta)\approx g(u_0,\boldsymbol\beta)+g'(u_0,\boldsymbol\beta)(U-u_0).$$

The nonparametric function estimator $\widehat{g}(\cdot)$ is then defined as the solution to $g(u_0,\boldsymbol\beta)$ of the local linear estimating equation

$$\begin{aligned}
\mathbf{0}=\sum_{i=1}^{n}&\left\{\begin{matrix}1\\(U_i-u_0)/h\end{matrix}\right\}K_h(U_i-u_0)\\
&\times\Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol\theta,g(u_0,\boldsymbol\beta)+g'(u_0,\boldsymbol\theta)(U_i-u_0)\}. \qquad (2.6)
\end{aligned}$$

Note that although $g(\cdot)$ depends on $\boldsymbol\beta$, its estimator depends on $\boldsymbol\theta$, hence we write it as $\widehat{g}(\cdot,\boldsymbol\theta)$. The estimate $\widehat{\boldsymbol\theta}$ is subsequently obtained as the solution to

$$\mathbf{0}=\sum_{i=1}^{n}\mathcal{L}\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol\theta,\widehat{g}(U_i;\boldsymbol\beta)\}. \qquad (2.7)$$

In the above description, we have developed the whole methodology as if the computation can be carried through. However, a closer look at the expressions involving conditional expectation given $\mathbf{S},\mathbf{Z}$ reveals that these quantities are not computable without knowing the distribution of $\boldsymbol\varepsilon$ given $\mathbf{S},\mathbf{Z}$. Instead of estimating the error distribution $f_{\boldsymbol\varepsilon|\mathbf{S},\mathbf{Z}}(\boldsymbol\varepsilon,\mathbf{S},\mathbf{Z})$, which is difficult, we propose to use a working model and carry out all the calculations under this working model. Thus, in summary, we first estimate the function g through the local linear method, by treating $\boldsymbol\theta$ as parameters that are held fixed. The set of estimating equations are exactly (2.6) at $\mathbf{z}_0=\mathbf{z}_1,\dots,\mathbf{z}_n$, and the solutions are $\widehat{g}(\boldsymbol\gamma^{\mathrm T}\mathbf{z}_1,\boldsymbol\theta),\dots,\widehat{g}(\boldsymbol\gamma^{\mathrm T}\mathbf{z}_n,\boldsymbol\theta)$. We then estimate $\boldsymbol\theta$ through solving (2.7). Obviously, this is a type of profiling estimation procedure.

2.2 Algorithm

We provide the detailed algorithm below.

1. Posit a working model $f_\epsilon^*(\epsilon \mid \mathbf{S}, \mathbf{Z})$ that has mean $\mathbf{0}$.
2. Calculate the score function $\mathbf{S}_{\boldsymbol{\theta}}^*(Y, \mathbf{S}, \mathbf{Z}, \boldsymbol{\theta}, g)$ and $\mathbf{S}_g^*(Y, \mathbf{S}, \mathbf{Z}, \boldsymbol{\theta}, g)$ under the working model $f_\epsilon^*(\epsilon \mid \mathbf{S}, \mathbf{Z})$.
3. Solve the integral equation (2.3) to obtain $\mathbf{b}_{\boldsymbol{\theta}}$ and b_g.
4. Solve the estimating equation (2.6) at $\mathbf{z}_0 = \mathbf{z}_1, \dots, \mathbf{z}_n$, and get solutions $\widehat{g}(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{z}_1, \boldsymbol{\theta}), \dots, \widehat{g}(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{z}_n, \boldsymbol{\theta})$.
5. Estimate $\boldsymbol{\theta}$ through solving (2.7).

Any working model that satisfies the mean zero property can be used in step 1. For convenience, we suggest to simply use a symmetric distribution that is independent of $\mathbf{S}$, $\mathbf{Z}$. The above steps are straightforward except step 3. To carry out the calculation in step 3, we discretize the distribution of ϵ on m equally spaced points on the support of the distribution and calculate the probability mass $\pi_i(\mathbf{S}, \mathbf{Z})$ at each point. Then we approximate $f^*_{\epsilon,Y|\mathbf{S},\mathbf{Z}}(\epsilon_i, Y \mid \mathbf{S}, \mathbf{Z})$ and $E^*\{b_\theta(\epsilon, \mathbf{S}, \mathbf{Z})|Y, \mathbf{S}, \mathbf{Z}\}$ by

$$f^*_{\epsilon,Y|\mathbf{S},\mathbf{Z}}(\epsilon_i, Y \mid \mathbf{S}, \mathbf{Z}) \approx \big[1 - y + (2y-1)H\big\{\mathbf{m}(\mathbf{S}, \mathbf{Z}, \hat{\boldsymbol{\alpha}})^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i^{\mathrm{T}}\boldsymbol{\beta} + g\big(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\gamma}\big)\big\}\big]\pi_i(\mathbf{S}, \mathbf{Z})$$

and

$$E^*\big\{b_\theta(\epsilon, \mathbf{S}, \mathbf{Z})|Y, \mathbf{S}, \mathbf{Z}\big\} \approx \frac{\sum_{i=1}^m b_\theta(\epsilon_i, \mathbf{S}, \mathbf{Z}) f^*_{\epsilon_i,Y|\mathbf{S},\mathbf{Z}}(\epsilon_i, Y \mid \mathbf{S}, \mathbf{Z})}{\sum_{i=1}^m f^*_{\epsilon_i,Y|\mathbf{S},\mathbf{Z}}(\epsilon_i, Y \mid \mathbf{S}, \mathbf{Z})}.$$

Note that here $f^*_{\epsilon,Y|\mathbf{S},\mathbf{Z}}(\epsilon, Y \mid \mathbf{S}, \mathbf{Z})$ denotes the conditional p.d.f. of (ϵ, Y) given $(\mathbf{S}, \mathbf{Z})$, calculated under the working model $f_\epsilon^*(\epsilon \mid \mathbf{S}, \mathbf{Z})$.

Let $\mathbf{B}(\mathbf{S}, \mathbf{Z}) \equiv \{\mathbf{b}_\theta(\epsilon_\mathbf{1}, \mathbf{S}, \mathbf{Z}), \dots, \mathbf{b}_\theta(\epsilon_\mathbf{m}, \mathbf{S}, \mathbf{Z})\}^{\mathrm{T}}$ and $\mathbf{C}(\mathbf{S}, \mathbf{Z}) \equiv \{\mathbf{c}(\epsilon_\mathbf{1}, \mathbf{S}, \mathbf{Z}), \dots, \mathbf{c}(\epsilon_\mathbf{m}, \mathbf{S}, \mathbf{Z})\}^{\mathrm{T}}$, where

$$c(\epsilon_i, \mathbf{S}, \mathbf{Z}) = E\big\{\mathcal{S}_{\boldsymbol{\theta}}^*\{Y, \mathbf{S}, \mathbf{Z}; \boldsymbol{\theta}, g(\cdot)\}|\boldsymbol{\varepsilon}_i, \mathbf{S}, \mathbf{Z}\big\} - E\big\{\mathcal{S}_{\boldsymbol{\theta}}^*\{Y, \mathbf{S}, \mathbf{Z}; \boldsymbol{\theta}, g(\cdot)\}\boldsymbol{\varepsilon}^{\mathrm{T}}|\mathbf{S}, \mathbf{Z}\big\}\big\{E^*\big(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}|\mathbf{S}, \mathbf{Z}\big)\big\}^{-1}\boldsymbol{\varepsilon}_i.$$

Let $\mathbf{A}(\mathbf{S}, \mathbf{Z})$ be an $m \times m$ matrix with the (i, j) block equal to

$$E\left\{\frac{f^*_{\epsilon,Y,\mathbf{S},\mathbf{Z}}(\epsilon_j, Y, \mathbf{S}, \mathbf{Z})}{\sum_{i=1}^m f^*_{\epsilon,Y,\mathbf{S},\mathbf{Z}}(\epsilon_i, Y, \mathbf{S}, \mathbf{Z})}\middle|\epsilon_i, \mathbf{S}, \mathbf{Z}\right\} - E^*\left\{\frac{f^*_{\epsilon,Y,\mathbf{S},\mathbf{Z}}(\epsilon_j, Y, \mathbf{S}, \mathbf{Z})}{\sum_{i=1}^m f^*_{\epsilon,Y,\mathbf{S},\mathbf{Z}}(\epsilon_i, Y, \mathbf{S}, \mathbf{Z})}\epsilon^{\mathrm{T}}\middle|\mathbf{S}, \mathbf{Z}\right\}\big\{E^*\big(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}|\mathbf{S}, \mathbf{Z}\big)\big\}^{-1}\boldsymbol{\varepsilon}_i.$$

We then get $\mathbf{b}_{\boldsymbol{\theta}}(\epsilon_i, \mathbf{S}, \mathbf{Z})$ by solving $\mathbf{A}(\mathbf{S}, \mathbf{Z})\mathbf{B}(\mathbf{S}, \mathbf{Z}) = \mathbf{C}(\mathbf{S}, \mathbf{Z})$. $b_g(\epsilon_i, \mathbf{S}, \mathbf{Z})$ is calculated similarly.

2.3 Asymptotic properties

We now study the asymptotic properties of the proposed estimator, which is computed under the working model of $f_{\boldsymbol{\varepsilon}|\mathbf{S},\mathbf{Z}}(\boldsymbol{\varepsilon}, \mathbf{S}, \mathbf{Z})$. We first list the regularity conditions required.

(C1) The kernel function $K(\cdot)$ is non-negative, has compact support, and satisfies $\int K(s)\,ds = 1$, $\int sK(s)\,ds = 0$, $0 < \mu_2 = \int s^2K(s)\,ds < \infty$ and $\int sK^2(s)\,ds < \infty$.

(C2) The bandwidth h in the kernel smoothing satisfies $nh^2 \to \infty$ and $nh^4 \to 0$ when $n \to \infty$.

(C3) The link function $H(\cdot)$ is differentiable.
(C4) The nonparametric function $g(\cdot)$ has continuous first order derivative.
(C5) The random variable $U = \mathbf{Z}^{\mathrm{T}}\boldsymbol{\gamma}$ has compact support and its marginal density function $f_U(\cdot)$ is bounded away from zero on the support.

Let $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^{\mathrm{T}}$ for any matrix or vector $\mathbf{a}$ throughout the text. Then we have the following result.

Theorem 1. *Under the regularity conditions* (C1)–(C5), $\widehat{\boldsymbol{\theta}}$ *satisfy*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \to N\big(\mathbf{0}, \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1^{\mathrm{T}}}\big), \tag{2.8}$$

in distribution when $n \to \infty$, where

$$\boldsymbol{A} = \left\{ \begin{aligned} & E\left(\frac{\partial \mathcal{L}\{Y_i, \boldsymbol{S}_i, \boldsymbol{Z}_i; \boldsymbol{\beta}, g(U_i;\boldsymbol{\beta})\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} - \frac{\partial \mathcal{L}\{Y_i, \boldsymbol{S}_i, \boldsymbol{Z}_i; \boldsymbol{\beta}, g(U_i;\boldsymbol{\beta})\}}{\partial g(U_i;\boldsymbol{\beta})} \frac{E[\partial \Phi\{Y_j, \boldsymbol{S}_j, \boldsymbol{Z}_j; \boldsymbol{\beta}, g(U_i;\boldsymbol{\beta})\}/\partial \boldsymbol{\beta}^{\mathrm{T}} | U_i]}{E[\partial \Phi\{Y_j, \boldsymbol{S}_j, \boldsymbol{Z}_j; \boldsymbol{\beta}, g(U_i;\boldsymbol{\beta})\}/\partial g(U_i;\boldsymbol{\beta}) | U_i]} \right) \\ & - E\left(\frac{\partial \mathcal{L}\{Y_j, \boldsymbol{S}_j, \boldsymbol{Z}_j; \boldsymbol{\beta}, g(U_i;\boldsymbol{\beta})\}}{\partial g(U_i;\boldsymbol{\beta})} \frac{E[(\boldsymbol{Z}_j - \boldsymbol{Z}_i)_{-1}^{\mathrm{T}} \partial \Phi\{Y_j, \boldsymbol{S}_j, \boldsymbol{Z}_j; \boldsymbol{\beta}, g(U_i,\boldsymbol{\beta})\}/\partial g(U_i,\boldsymbol{\beta}) g'(U_i,\boldsymbol{\beta}) | U_i]}{E[\partial \Phi\{Y_j, \boldsymbol{S}_j, \boldsymbol{Z}_j; \boldsymbol{\beta}, g(U_i,\boldsymbol{\beta})\}/\partial g(U_i,\boldsymbol{\beta}) | U_i]} \right) \end{aligned} \right\}$$

$$\begin{aligned} \boldsymbol{B} = E\big\{ & \big(\mathcal{L}\{Y_i, \boldsymbol{S}_i, \boldsymbol{Z}_i; \boldsymbol{\theta}, g(U_i;\boldsymbol{\theta})\} - E\big[\partial \mathcal{L}\{Y, \boldsymbol{S}, \boldsymbol{Z}; \boldsymbol{\theta}, g(U;\boldsymbol{\theta})\}/\partial g(U;\boldsymbol{\theta}) | U = U_i\big] \\ & \times \big[E\{\partial \Phi\{Y, \boldsymbol{S}, \boldsymbol{Z}; \boldsymbol{\theta}, g(U;\boldsymbol{\theta})\}/\partial g(U;\boldsymbol{\theta}) | U = U_i\}\big]^{-1} \Phi\{Y_i, \boldsymbol{S}_i, \boldsymbol{Z}_i; \boldsymbol{\theta}, g(U_i;\boldsymbol{\theta})\}\big)^{\otimes 2}\big\}. \end{aligned}$$

3 Numerical study

We performed three sets of simulation studies to evaluate the finite sample performance of the proposed estimator. In all the simulations, we set the sample size $n = 1000$ and we repeated the experiments 1000 times. In the first simulation, we generated the observations (Y_i, W_i, S_i, Z_i) from the model

$$\begin{aligned} \mathrm{pr}(Y_i = 1 | X_i = x_i, Z_i = z_i) &= H\{\beta x_i + g(\gamma_1 z_{1i} + \gamma_2 z_{2i} \\ &\qquad + \gamma_3 z_{3i} + \gamma_4 z_{4i})\}, \\ W_i &= X_i + U_i, \\ X_i &= \alpha_1 + \alpha_2 S_i + \epsilon_i, \end{aligned} \tag{3.1}$$

where $H(t)$ is the inverse logit link function and $\alpha_1 = 1$, $\alpha_2 = 1$, $\beta = 0.3$, $\gamma_1 = 1$, $\gamma_2 = 0.5$, $\gamma_3 = 1$, $\gamma_4 = -0.3$. Also, we experimented with different parameters values using inverse probit link function and $\alpha_1 = 1$, $\alpha_2 = 1$, $\beta = 1$, $\gamma_1 = 1$, $\gamma_2 = 0.5$, $\gamma_3 = 1$, $\gamma_4 = 0.3$. The true function $g(t) = t$. Thus, we experiment with a simple linear function with slope 1 and intercept 0 for g. The observable covariates Z_{1i}, Z_{2i}, Z_{4i} and the instrument variable S_i are generated from the standard normal distribution. The observable covariate Z_{3i} is generated from the uniform distribution on domain $[-1, 1]$. U_i is generated from a normal distribution with mean zero and variance 0.6. ϵ_i is generated from a standard normal distribution with mean 0 and variance ${S_i}^2/2$ and a t_5 distribution multiplied by $|S_i|/\sqrt{2}$, respectively. Since our working model for ϵ_i is set to be normal, it corresponds to a correct working model in the simulation where ϵ_i's are normally distributed, and corresponds to a misspecified working model in the simulation where ϵ_i's are non-normally distributed.

Our second simulation resembles the first simulation, but with a nonlinear g function. Our true g function in the second simulation is $g(t) = \exp(-0.3t^2)$.

Further, in the third simulation, we set $\alpha_1 = 1$, $\alpha_2 = 1$, $\beta = 0.3$, $\gamma_1 = 1.0$, $\gamma_2 = 0.5$, $\gamma_3 = 1.0$, $\gamma_4 = -0.3$ using both inverse logit link function and inverse probit link function. We generated the data the same as in the first simulation, except that the true function form of

Table 1 *Simulation* 1 *results* (*link function*: *logit*)

		α_1	α_2	β	γ_2	γ_3	γ_4
	Truth	1.0	1.0	0.3	0.5	1.0	−0.3
				ϵ: Normal distribution			
α_0	mean			0.3218	0.5052	1.0103	−0.3033
	median			0.3022	0.5008	1.0009	−0.2997
	se			0.1437	0.0690	0.1226	0.0668
OLS	mean	1.0267	1.0254	0.3331	0.5171	1.0399	−0.3128
	median	0.9992	0.9969	0.2995	0.5007	1.0008	−0.2998
	se	0.0325	0.0438	0.1585	0.0686	0.1239	0.0622
WLS	mean	0.9989	0.9996	0.3234	0.5064	1.0097	−0.3037
	median	0.9986	0.9987	0.3007	0.5007	1.0008	−0.2997
	se	0.0302	0.0414	0.1545	0.0663	0.1188	0.0650
				ϵ: Student t distribution t_5			
α_0	mean			0.3180	0.5095	1.0116	−0.3030
	median			0.3007	0.5010	1.0005	−0.2998
	se			0.1437	0.0691	0.1166	0.0642
OLS	mean	0.9988	0.9979	0.3246	0.5029	1.0106	−0.3022
	median	0.9990	0.9965	0.3013	0.5005	1.0009	−0.2995
	se	0.0384	0.0536	0.1533	0.0738	0.1260	0.0657
WLS	mean	0.9989	1.0004	0.3178	0.5085	1.0125	−0.3029
	median	0.9987	1.0015	0.3002	0.5008	1.0005	−0.2997
	se	0.0321	0.0474	0.1473	0.0687	0.1231	0.0648

g is now $g(t) = 1.5\sin(t)$ for normally distributed ϵ_i and $g(t) = \sin(t)$ for t distributed ϵ_i. Thus we also experiment with a nonlinear function form for g.

We used respectively, OLS and WLS to estimate α_1 and α_2, and compared the subsequent performance with the estimation result under the known α for all three simulation studies described above. The results of the three simulations are summarized in Tables 1–6 and Figure 1, where the confidence bands are obtained via bootstrap procedures. Histograms of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are plotted in Figures 2, 3, and 4. From these results, it is quite clear that the proposed estimators indeed yield consistent estimation, regardless a correct or a misspecified working model is used, in that the biases are quite small and the mean and median estimated curves track the true curves very well. Even though WLS produces more efficient estimators for α_1 and α_2, the efficiency in the parameter estimation of $\boldsymbol{\alpha}$ does not really translate to the efficiency difference in estimating the main parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. In fact, even when $\boldsymbol{\alpha}$ is completely known, we do not see a significant advantage in estimating $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. This is quite encouraging since this confirms our theoretical discovery. This result also provides practical significance since $\boldsymbol{\alpha}$ is typically not known in reality.

We further experimented using the asymptotic results to assess the estimation variability of our estimator. The results were not satisfactory. This is not surprising given that partially linear single index model without measurement error already requires very large sample size to demonstrate the asymptotic properties (Ma and Zhu, 2013), and measurement error typically have much larger sample size requirement to enable precise variability assessment. Thus, in practice, if variability needs to be assessed, we recommend using standard bootstrap methods, which is often observed to perform better than asymptotic results in finite samples.

Table 2 *Simulation* 2 *results* (*link function*: *logit*)

	Truth	α_1 1.0	α_2 1.0	β 0.3	γ_2 0.5	γ_3 1.0	γ_4 −0.3
				ϵ: Normal distribution			
α_0	mean			0.3467	0.5643	1.1072	−0.3472
	median			0.3377	0.5131	1.0217	−0.3044
	se			0.1078	0.3156	0.3930	0.2820
OLS	mean	0.9982	0.9978	0.3502	0.5560	1.0932	−0.3291
	median	0.9985	0.9964	0.3391	0.5102	1.0221	−0.3027
	se	0.03279	0.0453	0.1043	0.3078	0.4014	0.2638
WLS	mean	0.9993	0.9973	0.3469	0.5656	1.0987	−0.3490
	median	0.9990	0.9971	0.3365	0.5153	1.0216	−0.3048
	se	0.0299	0.0397	0.1025	0.3008	0.3961	0.2798
				ϵ: Student t distribution t_5			
α_0	mean			0.3451	0.5460	1.0936	−0.3285
	median			0.3371	0.5059	1.0130	−0.3026
	se			0.1011	0.3083	0.3922	0.2642
OLS	mean	1.0002	0.9994	0.3491	0.5583	1.0969	−0.3269
	median	0.9992	0.9978	0.3398	0.5123	1.0243	−0.3016
	se	0.0384	0.0542	0.0994	0.3147	0.3882	0.2662
WLS	mean	1.0008	0.9990	0.3422	0.5457	1.0889	−0.3323
	median	0.9992	1.0005	0.3353	0.5080	1.0121	−0.3012
	se	0.0332	0.0465	0.0849	0.3020	0.3842	0.2658

Table 3 *Simulation* 3 *results* (*link function*: *logit*)

	Truth	α_1 1.0	α_2 1.0	β 0.3	γ_2 0.5	γ_3 1.0	γ_4 −0.3
				ϵ: Normal distribution			
α_0	mean			0.2595	0.5184	1.0355	−0.3102
	median			0.2606	0.5028	1.0113	−0.3013
	se			0.0756	0.0788	0.1419	0.0669
OLS	mean	0.9989	0.9990	0.2602	0.5195	1.0413	−0.3123
	median	0.9984	0.9972	0.2629	0.5063	1.0117	−0.3021
	se	0.0350	0.0452	0.0865	0.0761	0.1424	0.0707
WLS	mean	0.9997	0.9999	0.2553	0.5198	1.0462	−0.3110
	median	0.9995	0.9995	0.2590	0.5049	1.0153	−0.3015
	se	0.0293	0.0407	0.0789	0.0810	0.1470	0.0692
				ϵ: Student t distribution t_5			
α_0	mean			0.2586	0.5234	1.0414	−0.3122
	median			0.2574	0.5073	1.0131	−0.3020
	se			0.0824	0.0810	0.1467	0.0697
OLS	mean	0.9994	0.9996	0.2597	0.5220	1.0480	−0.3152
	median	0.9990	0.9998	0.2593	0.5072	1.0175	−0.3040
	se	0.0396	0.0555	0.0821	0.0811	0.1427	0.0770
WLS	mean	0.9984	1.0003	0.2546	0.5168	1.0417	−0.3119
	median	0.9987	1.0009	0.2546	0.5038	1.0171	−0.3027
	se	0.0330	0.0471	0.0844	0.0743	0.1368	0.0645

Table 4 *Simulation* 1 *results* (*link function*: *probit*)

		α_1	α_2	β	γ_2	γ_3	γ_4
	Truth	1.0	1.0	1.0	0.5	1.0	0.3
			ϵ: Normal distribution				
α_0	mean			1.2308	0.5041	1.0067	0.3039
	median			1.0643	0.5007	1.0006	0.3006
	se			0.4098	0.0741	0.1279	0.0644
OLS	mean	0.9988	0.9980	1.2354	0.5029	1.0085	0.3038
	median	0.9985	0.9971	1.0758	0.5004	1.0006	0.3008
	se	0.0341	0.0451	0.4083	0.0748	0.1307	0.0592
WLS	mean	1.0001	0.9986	1.0823	0.5037	1.00082	0.3042
	median	0.9992	0.9993	1.0349	0.5007	1.0007	0.3005
	se	0.0314	0.0407	0.2220	0.0701	0.1177	0.0605
			ϵ: Student t distribution t_5				
α_0	mean			1.2417	0.5019	1.0086	0.3022
	median			1.0650	0.50041	1.0012	0.3004
	se			0.4269	0.0739	0.1276	0.0643
OLS	mean	1.0049	1.0049	1.2302	0.5039	1.0190	0.3061
	median	0.9991	0.9979	1.0650	0.5002	1.0011	0.3007
	se	0.0378	0.0557	0.4166	0.0726	0.1268	0.0653
WLS	mean	0.9992	1.0012	1.0926	0.5012	1.0042	0.3035
	median	0.9991	1.0012	1.0378	0.5006	1.0005	0.3006
	se	0.0340	0.0480	0.2364	0.0703	0.1171	0.0615

Table 5 *Simulation* 2 *results* (*link function*: *probit*)

		α_1	α_2	β	γ_2	γ_3	γ_4
	Truth	1.0	1.0	1.0	0.5	1.0	0.3
			ϵ: Normal distribution				
α_0	mean			1.1590	0.5519	1.0491	0.3432
	median			1.1189	0.5052	01.0061	0.3019
	se			0.2134	0.2988	0.4150	0.2788
OLS	mean	0.9982	0.9970	1.1678	0.5582	1.0589	0.3301
	median	0.9985	0.9963	1.1136	0.5057	1.0077	0.3017
	se	0.0321	0.0447	0.2206	0.3137	0.4118	0.2853
WLS	mean	0.9980	0.9985	1.1306	0.5214	1.0193	0.3199
	median	0.9978	0.9988	1.0979	0.5027	1.0020	0.3008
	se	0.0291	0.0396	0.1629	0.1869	0.2490	0.1827
			ϵ: Student t distribution t_5				
α_0	mean			1.1581	0.5461	1.0682	0.3097
	median			1.1064	0.5041	1.0111	0.3010
	se			0.2199	0.3112	0.3971	0.2625
OLS	mean	1.0004	1.0011	1.1590	0.5466	1.0678	0.3098
	median	0.9993	0.9999	1.1027	0.5072	1.0128	0.3016
	se	0.0379	0.0569	0.2234	0.3099	0.4076	0.2767
WLS	mean	0.9998	1.0001	1.1378	0.5216	1.0283	0.3208
	median	0.9984	1.0007	1.1069	0.5024	1.0018	0.3010
	se	0.0330	0.0481	0.1725	0.1938	0.2365	0.1809

Table 6 *Simulation* 3 *results* (*link function*: *probit*)

		α_1	α_2	β	γ_2	γ_3	γ_4
	Truth	1.0	1.0	0.3	0.5	1.0	−0.3
			ϵ: Normal distribution				
α_0	mean			0.2648	0.5186	1.0386	−0.3111
	median			0.2631	0.5059	1.0172	−0.3036
	se			0.0575	0.0716	0.1276	0.0641
OLS	mean	1.0009	0.9973	0.2646	0.5164	1.0415	−0.3104
	median	1.0010	0.9965	0.2630	0.5057	1.0197	−0.3018
	se	0.0336	0.0461	0.0650	0.0720	0.1345	0.0646
WLS	mean	0.9985	0.9987	0.2646	0.5186	1.0368	−0.3123
	median	0.9988	0.9996	0.2651	0.5077	1.0166	−0.3048
	se	0.0306	0.0395	0.0583	0.0704	0.1338	0.0636
			ϵ: Student t distribution t_5				
α_0	mean			0.2605	0.5169	1.0458	−0.3157
	median			0.2573	0.5052	1.0277	−0.3064
	se			0.0663	0.0751	0.1253	0.0693
OLS	mean	1.0015	0.9985	0.2611	0.5178	1.0414	−0.3117
	median	1.0004	0.9980	0.2578	0.5055	1.0135	−0.3017
	se	0.0376	0.0557	0.0581	0.0716	0.1304	0.0637
WLS	mean	1.0007	0.9987	0.2603	0.5131	1.0387	−0.3130
	median	1.0018	0.9991	0.2587	0.5031	1.0176	−0.3043
	se	0.0323	0.0451	0.0581	0.0722	0.1298	0.0644

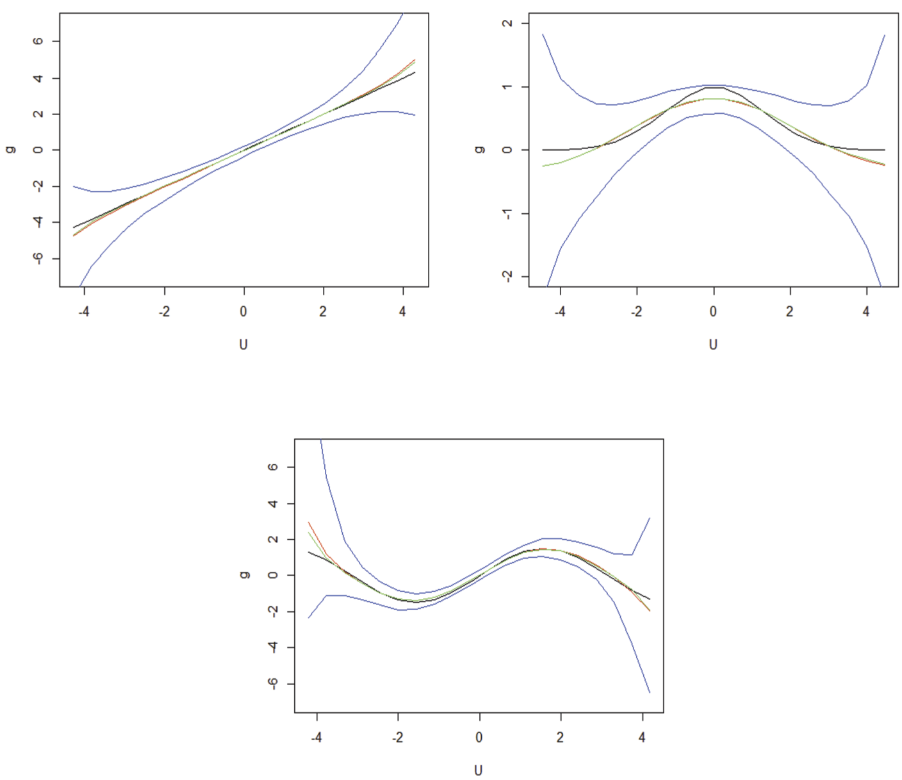

Figure 1 *True function* (*black line*), *median estimation* (*green line*), *mean estimation* (*red line*) *and* 95% *confidence band* (*blue line*) *of* $g(u)$ *in simulations* 1 (*upper-left*), 2 (*upper-right*) *and* 3 (*lower*) *when link function is inverse logit*, ϵ *is normal distributed and with OLS method applied*.

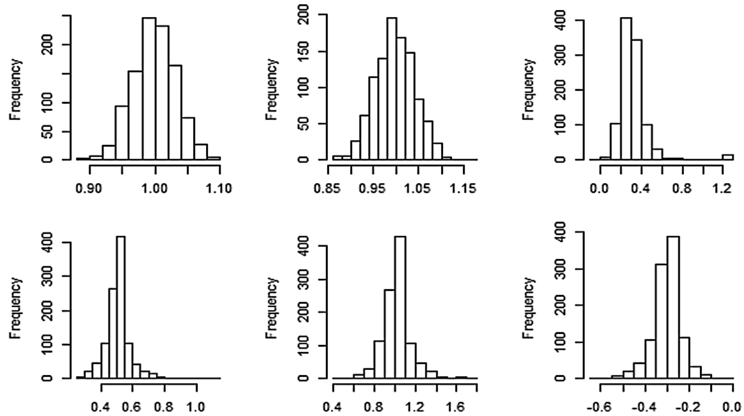

Figure 2 *Histograms of* α_1 (*upper-left*), α_2 (*upper-middle*), β_1 (*upper-right*), β_2 (*lower-left*), β_3 (*lower-middle*), *and* β_4 (*lower-right*) *in simulation* 1 *when link function is inverse logit*, ϵ *is normal distributed and with OLS method applied.*

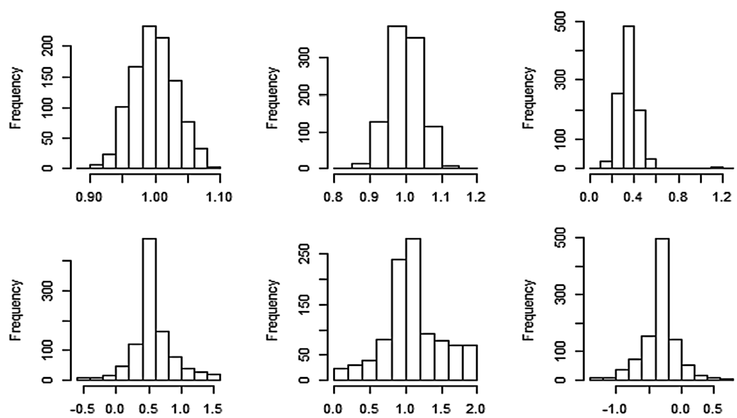

Figure 3 *Histograms of* α_1 (*upper-left*), α_2 (*upper-middle*), β_1 (*upper-right*), β_2 (*lower-left*), β_3 (*lower-middle*), *and* β_4 (*lower-right*) *in simulation* 2 *when link function is inverse logit*, ϵ *is normal distributed and with OLS method applied.*

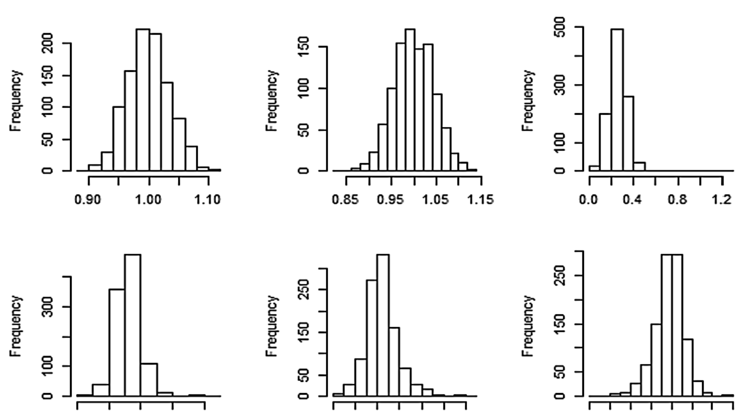

Figure 4 *Histograms of* α_1 (*upper-left*), α_2 (*upper-middle*), β_1 (*upper-right*), β_2 (*lower-left*), β_3 (*lower-middle*), *and* β_4 (*lower-right*) *in simulation* 3 *when link function is inverse logit*, ϵ *is normal distributed and with OLS method applied.*

4 Real data analysis

The data set we analyze here is from the AIDS Clinical Trials Group (ACTG) study. In this study four different treatments 'ZDV', 'ZDV+ddI', 'ZDV+ddC', and 'ddC' were used on HIV infected adults whose CD4 cell counts were between 200 and 500 per cubic millimetre. 'ZDV' is a standard treatment, and is considered as the reference treatment. For convenience, we name 'ZDV' treatment 1, 'ZDV+ddI' treatment 2, 'ZDV+ddC' treatment 3 and 'ddC' treatment 4. 'Age' was included as an explanatory variable. There were 1036 patients in our sample who had no antiretroviral therapy prior to the study. The purpose of this study is to see whether there is any difference among the four treatments in terms of preventing a patient's CD4 count from dropping below 50%. CD4 count is an important indicator for HIV positive patients to develop AIDS or to die from HIV caused disease. This is considered an endpoint event and when it occurs, our response variable Y is set to 1. We use Z_1, Z_2, Z_3 as three treatment indicators besides the reference treatment.

Let X be the baseline log(CD4 count) before the start of the treatment. Here, we treat X as a latent variable, since it cannot be measured precisely. Instead of observing X, we observe W, which is the average of two available measurements of X. We thus assume W is X plus a random noise. We also have an instrumental variable S, which is the screening log(CD4 count). Figure 5 suggests that there is a linear relationship between W and S, thus we further assume a linear regression model to link X and S. Finally, to model the relation between the occurrence of AIDS or death with the covariates and treatments, we used the familiar logistic model. The complete form of the model that is used to describe the ACTG data is

$$\begin{aligned} \mathrm{pr}(Y_i = 1 | X_i = x_i, Z_i = z_i) &= H\{\beta x_i + g(\gamma_1 z_{1i} + \gamma_2 z_{2i} \\ &\qquad + \gamma_3 z_{3i} + \gamma_{\mathrm{age}} z_{\mathrm{age}})\}, \\ W_i &= X_i + U_i, \\ X_i &= \alpha_1 + \alpha_2 S_i + \epsilon_i. \end{aligned} \tag{4.1}$$

We used the methodology developed earlier in the paper to analyze the data. Using OLS method, we got the estimates for α_1 and α_2 to be $(0.001, 0.674)$. The estimates for the main parameters β and γ are shown in Table 7. These results indicate that there is significant difference between the four different treatments.

We further fixed the variable 'Age' at 41 years old, which is the mid point on the range of all observed ages to compare the four treatments. The estimated g function of treatments 1,

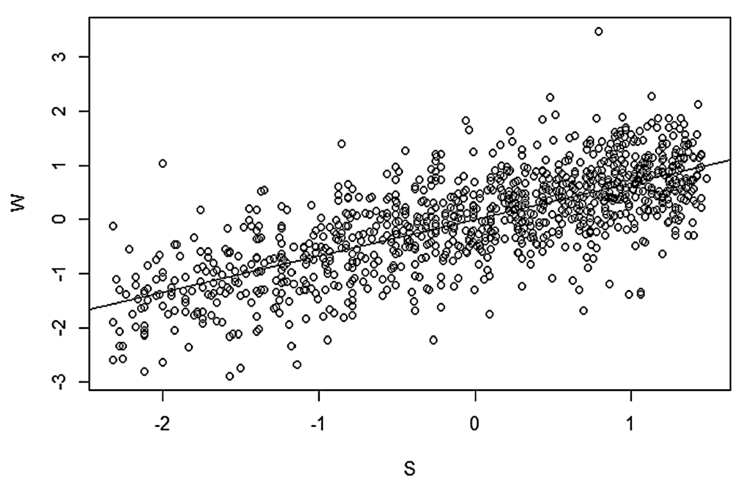

Figure 5 *Plot of averaged baseline CD*4 *count versus screening CD*4 *count.*

Table 7 *Realdata analysis results*

	β	γ_1	γ_2	γ_3	γ_{age}
Realdata estimates	−0.72808	1.0	1.511	2.4915	−2.3035
Bootstrapped mean	−0.7219		1.7897	2.6411	−2.3819
Bootstrapped median	−0.7082		1.6383	2.5455	−2.3029
Bootstrapped se	0.1274		0.3737	0.2919	0.41061
95% CI	(−0.9779, −0.4783)		(0.8186, 2.2836)	(1.9194, 3.0636)	(−3.0996, −1.5075)

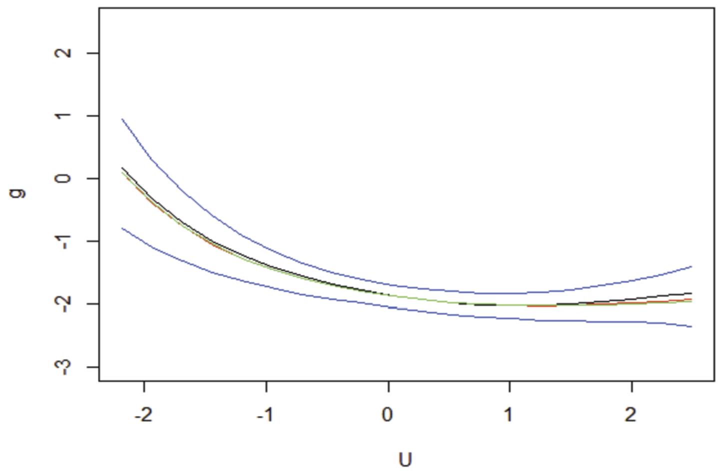

Figure 6 *Estimated* $g(u)$ *for real data* (*black line*), *median estimation* (*green line*), *mean estimation* (*red line*) *and* 90% *confidence band* (*blue line*) *of* $g(u)$ *using* 1000 *bootstrapped samples*.

2, 3 and 4 are −1.23, −1.78, −1.94 and −2.20 respectively. Their corresponding 90% confidence intervals are (−1.65, −0.96), (−2.01, −1.63), (−2.17, −1.78) and (−2.26, −1.80). It indicates that treatments 2, 3, 4 are not as efficient as treatment 1 for 41 year old patients in general.

We also plot the estimated g as a function of the estimated index in Figure 6, together with its 90% confidence band. We can see that g is decreasing and nonlinear and has a general decreasing trend, indicating a protective effect of the index in terms of risk of CD4 counts decreasing or death.

5 Discussion

Measurement error issue is a widely encountered problem in statistical applications. When the magnitude of the error is known or estimable, either from multiple measurements or from validation data, many methods are available to proceed with the subsequent analysis that adjust for the known measurement error issue. However, when the measurement error magnitude is unknown and un-estimable, which is often the case in practice, instruments are often indispensable. In this paper, we demonstrate that instrumental variable can be used in estimation in the generalized linear single index model context with binary response, which is unsolved in the literature before. In addition, the estimation of the model parameters is conducted without making any parametric assumption for the distribution of the unobserved variables in the model, that is, we have worked in the functional model framework.

The simulation studies show satisfactory performance of the proposed estimator in finite sample situation. Further, despite the fact that we present our main estimator in the context of logistic and probit models, the method is not restricted to these models only. In fact, any generalized partially linear regression model of Y conditional on X and $\mathbf{Z}$ can be handled by our method via a suitable link function H, thus Y is not restricted to binary variables and the method can be further extended to arbitrary generalized semiparametric single index models in terms of methodology. However, we foresee computational challenges in the more general cases.

Appendix: Proof of Theorem 1

Let $\widehat{\boldsymbol\beta}$, $\widehat{\boldsymbol\gamma}$ and $\widehat{g}(\mathbf{Z}_i^{\mathrm T}\widehat{\boldsymbol\gamma}, \widehat{\boldsymbol\beta})$, $i = 1, \ldots, n$ solve

$$\mathbf{0} = n^{-1/2}\sum_{i=1}^{n} \mathcal{L}\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \widehat{\boldsymbol\beta}, \widehat{g}(\mathbf{Z}_i^{\mathrm T}\widehat{\boldsymbol\gamma}, \widehat{\boldsymbol\beta})\}.$$

Then we have

$$\begin{aligned}
\mathbf{0} &= n^{-1/2}\sum_{i=1}^{n} \mathcal{L}\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \widehat{\boldsymbol\beta}, \widehat{g}(\mathbf{Z}_i^{\mathrm T}\widehat{\boldsymbol\gamma}, \widehat{\boldsymbol\beta})\} \\
&= n^{-1/2}\sum_{i=1}^{n} \mathcal{L}\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol\beta, \widehat{g}(\mathbf{Z}_i^{\mathrm T}\boldsymbol\gamma, \boldsymbol\beta)\} \\
&\quad + \Bigg[n^{-1}\sum_{i=1}^{n} \frac{\partial \mathcal{L}\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol\beta, \widehat{g}(\mathbf{Z}_i^{\mathrm T}\boldsymbol\gamma, \boldsymbol\beta)\}}{\partial \boldsymbol\beta^{\mathrm T}} \\
&\quad + n^{-1}\sum_{i=1}^{n} \frac{\partial \mathcal{L}\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol\beta, \widehat{g}(\mathbf{Z}_i^{\mathrm T}\boldsymbol\gamma; \boldsymbol\beta)\}}{\partial \widehat{g}(\mathbf{Z}_i^{\mathrm T}\boldsymbol\gamma; \boldsymbol\beta)} \frac{\partial \widehat{g}(\mathbf{Z}_i^{\mathrm T}\boldsymbol\gamma; \boldsymbol\beta)}{\partial \boldsymbol\beta^{\mathrm T}}\Bigg]\Bigg|_{\boldsymbol\theta=\boldsymbol\theta^*} \sqrt{n}(\widehat{\boldsymbol\beta} - \boldsymbol\beta) \\
&\quad + n^{-1}\sum_{i=1}^{n} \frac{\partial \mathcal{L}\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol\beta, \widehat{g}(\mathbf{Z}_i^{\mathrm T}\boldsymbol\gamma; \boldsymbol\beta)\}}{\partial \widehat{g}(\mathbf{Z}_i^{\mathrm T}\boldsymbol\gamma; \boldsymbol\beta)} \frac{\partial \widehat{g}(\mathbf{Z}_i^{\mathrm T}\boldsymbol\gamma; \boldsymbol\beta)}{\partial \boldsymbol\gamma^{\mathrm T}}\Bigg|_{\boldsymbol\theta=\boldsymbol\theta^*} \sqrt{n}(\widehat{\boldsymbol\gamma} - \boldsymbol\gamma) \\
&= n^{-1/2}\sum_{i=1}^{n} \mathcal{L}\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol\beta, \widehat{g}(\mathbf{Z}_i^{\mathrm T}\boldsymbol\gamma; \boldsymbol\beta)\} \\
&\quad + \Bigg\{E\Bigg[\frac{\partial \mathcal{L}\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol\beta, g(\mathbf{Z}_i^{\mathrm T}\boldsymbol\gamma; \boldsymbol\beta)\}}{\partial \boldsymbol\beta^{\mathrm T}} \\
&\quad + \frac{\partial \mathcal{L}\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol\beta, g(\mathbf{Z}_i^{\mathrm T}\boldsymbol\gamma; \boldsymbol\beta)\}}{\partial g(\mathbf{Z}_i^{\mathrm T}\boldsymbol\gamma; \boldsymbol\beta)} \frac{\partial \widehat{g}(\mathbf{Z}_i^{\mathrm T}\boldsymbol\gamma; \boldsymbol\beta)}{\partial \boldsymbol\beta^{\mathrm T}}\Bigg] + o_p(1)\Bigg\}\sqrt{n}(\widehat{\boldsymbol\beta} - \boldsymbol\beta) \\
&\quad + \Bigg\{E\Bigg[\frac{\partial \mathcal{L}\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol\beta, g(\mathbf{Z}_i^{\mathrm T}\boldsymbol\gamma; \boldsymbol\beta)\}}{\partial g(\mathbf{Z}_i^{\mathrm T}\boldsymbol\gamma; \boldsymbol\beta)} \frac{\partial \widehat{g}(\mathbf{Z}_i^{\mathrm T}\boldsymbol\gamma; \boldsymbol\beta)}{\partial \boldsymbol\gamma^{\mathrm T}}\Bigg] + o_p(1)\Bigg\}\sqrt{n}(\widehat{\boldsymbol\gamma} - \boldsymbol\gamma) \\
&= n^{-1/2}\sum_{i=1}^{n} \mathcal{L}\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol\beta, \widehat{g}(U_i; \boldsymbol\beta)\} + \Bigg\{E\Bigg[\frac{\partial \mathcal{L}\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol\beta, g(U_i; \boldsymbol\beta)\}}{\partial \boldsymbol\beta^{\mathrm T}} \\
&\quad + \frac{\partial \mathcal{L}\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol\beta, g(U_i; \boldsymbol\beta)\}}{\partial g(U_i; \boldsymbol\beta)} \frac{\partial \widehat{g}(U_i; \boldsymbol\beta)}{\partial \boldsymbol\beta^{\mathrm T}}\Bigg] + o_p(1)\Bigg\}\sqrt{n}(\widehat{\boldsymbol\beta} - \boldsymbol\beta) \\
&\quad + \Bigg\{E\Bigg[\frac{\partial \mathcal{L}\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol\beta, g(U_i; \boldsymbol\beta)\}}{\partial g(U_i; \boldsymbol\beta)} \frac{\partial \widehat{g}(U_i; \boldsymbol\beta)}{\partial \boldsymbol\gamma^{\mathrm T}}\Bigg] + o_p(1)\Bigg\}\sqrt{n}(\widehat{\boldsymbol\gamma} - \boldsymbol\gamma), \qquad \text{(A.1)}
\end{aligned}$$

where $U_i = \mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\gamma}$. Because of the estimation process, we have

$$\mathbf{0} = \sum_{i=1}^{n} \begin{Bmatrix} 1 \\ (U_i - u_0)/h \end{Bmatrix} K_h(U_i - u_0) \times \Phi\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol{\beta}, \widehat{g}(u_0, \boldsymbol{\beta}) + \widehat{g}'(u_0, \boldsymbol{\beta})(U_i - u_0)\}, \tag{A.2}$$

at all $u_0 = \mathbf{z}_0^{\mathrm{T}}\boldsymbol{\gamma}$ and all parameter values of $\boldsymbol{\beta}$, say $\boldsymbol{\beta}^*$. Thus, we have

$$\begin{aligned}
\mathbf{0} = n^{-1} \sum_{i=1}^{n} & \begin{Bmatrix} 1 \\ (U_i - u_0)/h \end{Bmatrix} K_h(U_i - u_0) \\
&\times \Bigg[\frac{\partial \Phi\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol{\beta}^*, \widehat{g}(u_0, \boldsymbol{\beta}^*) + \widehat{g}'(u_0, \boldsymbol{\beta}^*)(U_i - u_0)\}}{\partial \boldsymbol{\beta}^{*\mathrm{T}}} \\
&+ \frac{\partial \Phi\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol{\beta}^*, \widehat{g}(u_0; \boldsymbol{\beta}^*) + \widehat{g}'(u_0, \boldsymbol{\beta}^*)(U_i - u_0)\}}{\partial\{\widehat{g}(u_0, \boldsymbol{\beta}) + \widehat{g}'(u_0, \boldsymbol{\beta})(U_i - u_0)\}} \frac{\partial \widehat{g}(u_0; \boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}^{*\mathrm{T}}} \\
&+ \frac{\partial \Phi\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol{\beta}^*, \widehat{g}(u_0, \boldsymbol{\beta}^*) + \widehat{g}'(u_0, \boldsymbol{\beta}^*)(U_i - u_0)\}}{\partial\{\widehat{g}(u_0, \boldsymbol{\beta}) + \widehat{g}'(u_0, \boldsymbol{\beta})(U_i - u_0)\}} \frac{\partial \widehat{g}'(u_0; \boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}^{*\mathrm{T}}} (U_i - u_0) \Bigg]
\end{aligned}$$

at all $\boldsymbol{\beta}^*$. Let $f_U(\cdot)$ be the probability density function of U. Then

$$\frac{\partial \widehat{g}(u_0; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -\frac{E[\frac{\partial \Phi\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol{\beta}, g(u_0; \boldsymbol{\beta})\}}{\partial \boldsymbol{\beta}} \mid U_i = u_0]}{E[\frac{\partial \Phi\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol{\beta}, g(u_0; \boldsymbol{\beta})\}}{\partial g(u_0; \boldsymbol{\beta})} \mid U_i = u_0]} + o_p(1). \tag{A.3}$$

On the other hand, we have

$$\begin{aligned}
\mathbf{0} = n^{-1} \sum_{i=1}^{n} & \begin{Bmatrix} 1 \\ (U_i - u_0)/h \end{Bmatrix} \Bigg[\frac{\partial K_h(U_i - u_0)}{\partial (U_i - u_0)} (\mathbf{Z}_i - \mathbf{z}_0) \\
&\times \Phi\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol{\beta}, \widehat{g}(u_0, \boldsymbol{\beta}) + \widehat{g}'(u_0, \boldsymbol{\beta})(U_i - u_0)\} \\
&+ K_h(U_i - u_0) \frac{\partial \Phi\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol{\beta}, \widehat{g}(u_0, \boldsymbol{\beta}) + \widehat{g}'(u_0, \boldsymbol{\beta})(U_i - u_0)\}}{\partial\{\widehat{g}(u_0, \boldsymbol{\beta}) + \widehat{g}'(u_0, \boldsymbol{\beta})(U_i - u_0)\}} \frac{\partial \widehat{g}(u_0, \boldsymbol{\beta})}{\partial \boldsymbol{\gamma}} \\
&+ K_h(U_i - u_0) \frac{\partial \Phi\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol{\beta}, \widehat{g}(u_0, \boldsymbol{\beta}) + \widehat{g}'(u_0, \boldsymbol{\beta})(U_i - u_0)\}}{\partial\{\widehat{g}(u_0, \boldsymbol{\beta}) + \widehat{g}'(u_0, \boldsymbol{\beta})(U_i - u_0)\}} \frac{\partial \widehat{g}'(u_0, \boldsymbol{\beta})}{\partial \boldsymbol{\gamma}} (U_i - u_0) \\
&+ K_h(U_i - u_0) \frac{\partial \Phi\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol{\beta}, \widehat{g}(u_0, \boldsymbol{\beta}) + \widehat{g}'(u_0, \boldsymbol{\beta})(U_i - u_0)\}}{\partial\{\widehat{g}(u_0, \boldsymbol{\beta}) + \widehat{g}'(u_0, \boldsymbol{\beta})(U_i - u_0)\}} \\
&\times \widehat{g}'(u_0, \boldsymbol{\beta})(\mathbf{Z}_i - \mathbf{z}_0) \Bigg]_{-1}^{\mathrm{T}} \\
&+ n^{-1} \sum_{i=1}^{n} \begin{Bmatrix} \mathbf{0}_{1\times(q-1)} \\ (\mathbf{Z}_i - \mathbf{z}_0)_{-1}^{\mathrm{T}} \end{Bmatrix} K_h(U_i - u_0) \\
&\times \Phi\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol{\beta}, \widehat{g}(u_0, \boldsymbol{\beta}) + \widehat{g}'(u_0, \boldsymbol{\beta})(U_i - u_0)\} \\
= n^{-1} \sum_{i=1}^{n} & \begin{Bmatrix} 1 \\ (U_i - u_0)/h \end{Bmatrix} \frac{\partial K_h(U_i - u_0)}{\partial (U_i - u_0)} (\mathbf{Z}_i - \mathbf{z}_0)_{-1}^{\mathrm{T}} \\
&\times \Phi\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol{\beta}, \widehat{g}(u_0, \boldsymbol{\beta}) + \widehat{g}'(u_0, \boldsymbol{\beta})(U_i - u_0)\} \\
&+ \left\{ E \begin{pmatrix} 1 \\ 0 \end{pmatrix} \left[\frac{\partial \Phi\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol{\beta}, g(u_0, \boldsymbol{\beta})\}}{\partial g(u_0, \boldsymbol{\beta})} \Big| U_i = u_0 \right] \right\} \frac{\partial \widehat{g}(u_0, \boldsymbol{\beta})}{\partial \boldsymbol{\gamma}_{-1}^{\mathrm{T}}} f_U(u_0)
\end{aligned}$$

$$
\begin{aligned}
&+\left\{E\begin{pmatrix}1\\0\end{pmatrix}\left[\frac{\partial\Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(u_0,\boldsymbol{\beta})\}}{\partial g(u_0,\boldsymbol{\beta})}g'(u_0,\boldsymbol{\beta})(\mathbf{Z}_i-\mathbf{z}_0)_{-1}^{\mathrm{T}}\Big|U_i=u_0\right]\right\}\times f_U(u_0)\\
&+E\left[\begin{Bmatrix}\mathbf{0}_{1\times(q-1)}\\(\mathbf{Z}_i-\mathbf{z}_0)_{-1}^{\mathrm{T}}/h\end{Bmatrix}\Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(u_0,\boldsymbol{\beta})\}\Big|U_i=u_0\right]f_U(u_0)\\
&+o_p(1).
\end{aligned}
$$

Now, to analyze the first term above, we introduce $f_{U,\mathbf{Z}_{-1}}(u,\mathbf{z}_{-1})$ as the joint p.d.f. of $U,\mathbf{Z}_{-1}$. Note that $f_{U,\mathbf{Z}_{-1}}(u,\mathbf{z}_{-1})=f_{\mathbf{Z}}(\mathbf{z})$ for $u=\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{z}$ where $\boldsymbol{\gamma}$ is any parameter with the first component 1. Then

$$
\begin{aligned}
&n^{-1}\sum_{i=1}^{n}\frac{\partial K_h(U_i-u_0)}{\partial(U_i-u_0)}(\mathbf{Z}_i-\mathbf{z}_0)_{-1}^{\mathrm{T}}\\
&\qquad\times\Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},\widehat{g}(u_0,\boldsymbol{\beta})+\widehat{g}'(u_0,\boldsymbol{\beta})(U_i-u_0)\}\\
&\quad=n^{-1}\sum_{i=1}^{n}h^{-2}K'\{(U_i-u_0)/h\}(\mathbf{Z}_i-\mathbf{z}_0)_{-1}^{\mathrm{T}}\\
&\qquad\times\Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},\widehat{g}(u_0,\boldsymbol{\beta})+\widehat{g}'(u_0,\boldsymbol{\beta})(U_i-u_0)\}\\
&\quad=\int h^{-2}K'\{(u-u_0)/h\}(\mathbf{z}-\mathbf{z}_0)_{-1}^{\mathrm{T}}\\
&\qquad\times\Phi\{y,\mathbf{s},u,\mathbf{z}_{-1};\boldsymbol{\beta},g(u_0,\boldsymbol{\beta})+g'(u_0,\boldsymbol{\beta})(u-u_0)\}\\
&\qquad\times f_{U,\mathbf{Z}_{-1},\mathbf{S},Y}(u,\mathbf{z}_{-1},\mathbf{s},y)\,du\,d\mathbf{z}_{-1}\,d\mathbf{s}\,dy\{1+o_p(1)\}\\
&\quad=\int h^{-1}K'(t)(\mathbf{z}-\mathbf{z}_0)_{-1}^{\mathrm{T}}\Phi\{y,\ \mathbf{s},u_0+ht,\mathbf{z}_{-1};\boldsymbol{\beta},g(u_0,\boldsymbol{\beta})+g'(u_0,\boldsymbol{\beta})ht\}\\
&\qquad\times f_{U,\mathbf{Z}_{-1},\mathbf{S},Y}(u_0+ht,\mathbf{z}_{-1},\mathbf{s},y)\,dt\,d\mathbf{z}_{-1}\,d\mathbf{s}\,dy\{1+o_p(1)\}\\
&\quad=\Big(\int h^{-1}K'(t)(\mathbf{z}-\mathbf{z}_0)_{-1}^{\mathrm{T}}\Phi\{y,\ \mathbf{s},u_0,\mathbf{z}_{-1};\boldsymbol{\beta},g(u_0,\boldsymbol{\beta})\}\\
&\qquad\times f_{U,\mathbf{Z}_{-1},\mathbf{S},Y}(u_0,\mathbf{z}_{-1},\mathbf{s},y)\,dt\,d\mathbf{z}_{-1}\,d\mathbf{s}\,dy\\
&\qquad+\int K'(t)(\mathbf{z}-\mathbf{z}_0)_{-1}^{\mathrm{T}}\Big[\frac{\partial\Phi\{y,\mathbf{s},u_0,\mathbf{z}_{-1};\boldsymbol{\beta},g(u_0,\boldsymbol{\beta})\}}{\partial u_0}f_{U,\mathbf{Z}_{-1},\mathbf{S},Y}(u_0,\mathbf{z}_{-1},\mathbf{s},y)\\
&\qquad+\frac{\partial\Phi\{y,\mathbf{s},u_0,\mathbf{z}_{-1};\boldsymbol{\beta},g(u_0,\boldsymbol{\beta})\}}{\partial g(u_0,\boldsymbol{\beta})}g'(u_0,\boldsymbol{\beta})f_{U,\mathbf{Z}_{-1},\mathbf{S},Y}(u_0,\mathbf{z}_{-1},\mathbf{s},y)\\
&\qquad+\Phi\{y,\mathbf{s},u_0,\mathbf{z}_{-1};\boldsymbol{\beta},g(u_0,\boldsymbol{\beta})\}\frac{\partial f_{U,\mathbf{Z}_{-1},\mathbf{S},Y}(u_0,\mathbf{z}_{-1},\mathbf{s},y)}{\partial u_0}\Big]t\,dt\,d\mathbf{z}_{-1}\,d\mathbf{s}\,dy+O_p(h)\Big)\\
&\qquad\times\{1+o_p(1)\}\\
&\quad=-\Big(\int(\mathbf{z}-\mathbf{z}_0)_{-1}^{\mathrm{T}}\Big[\frac{\partial\Phi\{y,\mathbf{s},u_0,\mathbf{z}_{-1};\boldsymbol{\beta},g(u_0,\boldsymbol{\beta})\}}{\partial u_0}f_{U,\mathbf{Z}_{-1},\mathbf{S},Y}(u_0,\mathbf{z}_{-1},\mathbf{s},y)\\
&\qquad+\frac{\partial\Phi\{y,\mathbf{s},u_0,\mathbf{z}_{-1};\boldsymbol{\beta},g(u_0,\boldsymbol{\beta})\}}{\partial g(u_0,\boldsymbol{\beta})}g'(u_0,\boldsymbol{\beta})f_{U,\mathbf{Z}_{-1},\mathbf{S},Y}(u_0,\mathbf{z}_{-1},\mathbf{s},y)\\
&\qquad+\Phi\{y,\mathbf{s},u_0,\mathbf{z}_{-1};\boldsymbol{\beta},g(u_0,\boldsymbol{\beta})\}\frac{\partial f_{U,\mathbf{Z}_{-1},\mathbf{S},Y}(u_0,\mathbf{z}_{-1},\mathbf{s},y)}{\partial u_0}\Big]d\mathbf{z}_{-1}\,d\mathbf{s}\,dy+O_p(h)\Big)\\
&\qquad\times\{1+o_p(1)\}
\end{aligned}
$$

$$
\begin{aligned}
&= -\Big(\int (\mathbf{z}-\mathbf{z}_0)_{-1}^{\mathrm{T}} \\
&\quad \times \frac{d[\Phi\{y,\mathbf{s},u_0,\mathbf{z}_{-1};\boldsymbol{\beta},g(u_0,\boldsymbol{\beta})\}f_{U,\mathbf{Z}_{-1},\mathbf{S},Y}(u_0,\mathbf{z}_{-1},\mathbf{s},y)]}{du_0}\,d\mathbf{z}_{-1}\,d\mathbf{s}\,dy + O_p(h)\Big) \\
&\quad \times \{1+o_p(1)\} \\
&= -\frac{d}{du_0}\Big(\int (\mathbf{z}-\mathbf{z}_0)_{-1}^{\mathrm{T}} \\
&\quad \times [\Phi\{y,\mathbf{s},u_0,\mathbf{z}_{-1};\boldsymbol{\beta},g(u_0,\boldsymbol{\beta})\}f_{U,\mathbf{Z}_{-1},\mathbf{S},Y}(u_0,\mathbf{z}_{-1},\mathbf{s},y)]\,d\mathbf{z}_{-1}\,d\mathbf{s}\,dy\Big)\{1+o_p(1)\} \\
&\quad + O_p(h) \\
&= -\frac{d}{du_0}E\big((\mathbf{Z}-\mathbf{z}_0)_{-1}^{\mathrm{T}}E[\Phi\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\beta},g(U,\boldsymbol{\beta})\}|\mathbf{Z},\mathbf{S}]|U=u_0\big)\{1+o_p(1)\}+O_p(h) \\
&= o_p(1),
\end{aligned}
$$

where the last equality is because Φ has expectation zero conditional on $\mathbf{S}, \mathbf{Z}$. Further

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}(U_i-u_0)/h\,\frac{\partial K_h(U_i-u_0)}{\partial(U_i-u_0)}(\mathbf{Z}_i-\mathbf{z}_0)_{-1}^{\mathrm{T}} \\
&\qquad \times \Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},\widehat{g}(u_0,\boldsymbol{\beta})+\widehat{g}'(u_0,\boldsymbol{\beta})(U_i-u_0)\} \\
&\quad = n^{-1}\sum_{i=1}^{n}h^{-1}(U_i-u_0)K'\{(U_i-u_0)/h\}(\mathbf{Z}_i-\mathbf{z}_0)_{-1}^{\mathrm{T}} \\
&\qquad \times \Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},\widehat{g}(u_0,\boldsymbol{\beta})+\widehat{g}'(u_0,\boldsymbol{\beta})(U_i-u_0)\} \\
&\quad = \int h^{-1}(U_i-u_0)K'\{(u-u_0)/h\}(\mathbf{z}-\mathbf{z}_0)_{-1}^{\mathrm{T}} \\
&\qquad \times \Phi\{y,\mathbf{s},u,\mathbf{z}_{-1};\boldsymbol{\beta},g(u_0,\boldsymbol{\beta})+g'(u_0,\boldsymbol{\beta})(u-u_0)\} \\
&\qquad \times f_{U,\mathbf{Z}_{-1},\mathbf{S},Y}(u,\mathbf{z}_{-1},\mathbf{s},y)\,du\,d\mathbf{z}_{-1}\,d\mathbf{s}\,dy\{1+o_p(1)\} \\
&\quad = \int tK'(t)(\mathbf{z}-\mathbf{z}_0)_{-1}^{\mathrm{T}}\Phi\{y,\mathbf{s},u_0+ht,\mathbf{z}_{-1};\boldsymbol{\beta},g(u_0,\boldsymbol{\beta})+g'(u_0,\boldsymbol{\beta})ht\} \\
&\qquad \times f_{U,\mathbf{Z}_{-1},\mathbf{S},Y}(u_0+ht,\mathbf{z}_{-1},\mathbf{s},y)\,dt\,d\mathbf{z}_{-1}\,d\mathbf{s}\,dy\{1+o_p(1)\} \\
&\quad = \Big(\int tK'(t)(\mathbf{z}-\mathbf{z}_0)_{-1}^{\mathrm{T}}\Phi\{y,\mathbf{s},u_0,\mathbf{z}_{-1};\boldsymbol{\beta},g(u_0,\boldsymbol{\beta})\} \\
&\qquad \times f_{U,\mathbf{Z}_{-1},\mathbf{S},Y}(u_0,\mathbf{z}_{-1},\mathbf{s},y)\,dt\,d\mathbf{z}_{-1}\,d\mathbf{s}\,dy\Big)\{1+o_p(1)\}+O_p(h) \\
&\quad = -\Big(\int (\mathbf{z}-\mathbf{z}_0)_{-1}^{\mathrm{T}}\Phi\{y,\mathbf{s},u_0,\mathbf{z}_{-1};\boldsymbol{\beta},g(u_0,\boldsymbol{\beta})\} \\
&\qquad \times f_{U,\mathbf{Z}_{-1},\mathbf{S},Y}(u_0,\mathbf{z}_{-1},\mathbf{s},y)\,d\mathbf{z}_{-1}\,d\mathbf{s}\,dy\Big)\{1+o_p(1)\}+O_p(h) \\
&\quad = -E\big((\mathbf{Z}-\mathbf{z}_0)_{-1}^{\mathrm{T}}E[\Phi\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\beta},g(U,\boldsymbol{\beta})\}|\mathbf{Z},\mathbf{S}]|U=u_0\big)\{1+o_p(1)\} \\
&\qquad + O_p(h) \\
&\quad = o_p(1),
\end{aligned}
$$

where the last equality is because Φ has expectation zero conditional on $\mathbf{S}$, $\mathbf{Z}$. Similarly

$$E\left[\left\{\begin{matrix}\mathbf{0}_{1\times(q-1)}\\(\mathbf{Z}_i-\mathbf{z}_0)^{\mathrm{T}}_{-1}/h\end{matrix}\right\}\Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(u_0,\boldsymbol{\beta})\}\Big|U_i=u_0\right]$$
$$=E\left(\left\{\begin{matrix}\mathbf{0}_{1\times(q-1)}\\(\mathbf{Z}_i-\mathbf{z}_0)^{\mathrm{T}}_{-1}/h\end{matrix}\right\}E\big[\Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(U_i,\boldsymbol{\beta})\}|\mathbf{S}_i,\mathbf{Z}_i\big]\Big|U_i=u_0\right)=\mathbf{0}.$$

We thus obtain

$$\mathbf{0}=\left\{E\begin{pmatrix}1\\0\end{pmatrix}\left[\frac{\partial\Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(u_0,\boldsymbol{\beta})\}}{\partial g(u_0,\boldsymbol{\beta})}\Big|U_i=u_0\right]\right\}\frac{\partial\widehat{g}(u_0,\boldsymbol{\beta})}{\partial\boldsymbol{\gamma}^{\mathrm{T}}_{-1}}f_U(u_0)$$
$$+\left\{E\begin{pmatrix}1\\0\end{pmatrix}\left[\frac{\partial\Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(u_0,\boldsymbol{\beta})\}}{\partial g(u_0,\boldsymbol{\beta})}g'(u_0,\boldsymbol{\beta})(\mathbf{Z}_i-\mathbf{z}_0)^{\mathrm{T}}_{-1}\Big|U_i=u_0\right]\right\}$$
$$\times f_U(u_0)+o_p(1).$$

Hence

$$\frac{\partial\widehat{g}(u_0,\boldsymbol{\beta})}{\partial\boldsymbol{\gamma}_{-1}}$$
$$=-\frac{E[(\mathbf{Z}_i-\mathbf{z}_0)_{-1}\partial\Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(u_0,\boldsymbol{\beta})\}/\partial g(u_0,\boldsymbol{\beta})g'(u_0,\boldsymbol{\beta})|U_i=u_0]}{E[\partial\Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(u_0,\boldsymbol{\beta})\}/\partial g(u_0,\boldsymbol{\beta})|U_i=u_0]}$$
$$+o_p(1) \tag{A.4}$$

Continue from (A.1) by inserting (A.3) and (A.4), this leads to

$$\mathbf{0}=n^{-1/2}\sum_{i=1}^{n}\mathcal{L}\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})\}$$
$$+n^{-1/2}\sum_{i=1}^{n}\frac{\partial\mathcal{L}\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})\}}{\partial g(U_i;\boldsymbol{\beta})}\{\widehat{g}(U_i;\boldsymbol{\beta})-g(U_i;\boldsymbol{\beta})\}$$
$$\times\big(1+O_p(|\widehat{g}(U_i;\boldsymbol{\beta})-g(U_i;\boldsymbol{\beta})|)\big)$$
$$+\left\{E\left(\frac{\partial\mathcal{L}\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}-\frac{\partial\mathcal{L}\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})\}}{\partial g(U_i;\boldsymbol{\beta})}\right.\right.$$
$$\left.\left.\times\frac{E[\partial\Phi\{Y_j,\mathbf{S}_j,\mathbf{Z}_j;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})\}/\partial\boldsymbol{\beta}^{\mathrm{T}}|U_i]}{E[\partial\Phi\{Y_j,\mathbf{S}_j,\mathbf{Z}_j;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})\}/\partial g(U_i;\boldsymbol{\beta})|U_j=U_i]}\right)+o_p(1)\right\}\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})$$
$$-\left\{E\left(\frac{\partial\mathcal{L}\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})\}}{\partial g(U_i;\boldsymbol{\beta})}\right.\right.$$
$$\left.\left.\times\frac{E[(\mathbf{Z}_j-\mathbf{Z}_i)^{\mathrm{T}}_{-1}\partial\Phi\{Y_j,\mathbf{S}_j,\mathbf{Z}_j;\boldsymbol{\beta},g(U_i,\boldsymbol{\beta})\}/\partial g(U_i,\boldsymbol{\beta})g'(U_i,\boldsymbol{\beta})|U_i]}{E[\partial\Phi\{Y_j,\mathbf{S}_j,\mathbf{Z}_j;\boldsymbol{\beta},g(U_i,\boldsymbol{\beta})\}/\partial g(U_i,\boldsymbol{\beta})|U_i]}\right)+o_p(1)\right\}$$
$$\times\sqrt{n}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}), \tag{A.5}$$

where $g^*(u;\boldsymbol{\beta})$ lies on the line connecting $g(u;\boldsymbol{\beta})$ and $\widehat{g}(u;\boldsymbol{\beta})$. Let $g'^*(u_0,\boldsymbol{\beta})$ be on the line connecting $\widehat{g}'(u_0,\boldsymbol{\beta})$ and $g'(u_0,\boldsymbol{\beta})$. We now rewrite (A.2) as

$$\mathbf{0}=n^{-1}\sum_{i=1}^{n}\left\{\begin{matrix}1\\(U_i-u_0)/h\end{matrix}\right\}K_h(U_i-u_0)$$
$$\times\Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},\widehat{g}(u_0,\boldsymbol{\beta})+\widehat{g}'(u_0,\boldsymbol{\beta})(U_i-u_0)\}$$

$$
\begin{aligned}
&= n^{-1}\sum_{i=1}^{n}\begin{Bmatrix}1\\(U_i-u_0)/h\end{Bmatrix}K_h(U_i-u_0)\\
&\quad\times\Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(u_0;\boldsymbol{\beta})+g'(u_0,\boldsymbol{\beta})(U_i-u_0)\}\\
&\quad+\Bigg[n^{-1}\sum_{i=1}^{n}\begin{Bmatrix}1\\(U_i-u_0)/h\end{Bmatrix}K_h(U_i-u_0)\\
&\quad\times\frac{\partial\Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(u_0;\boldsymbol{\beta})+g'(u_0;\boldsymbol{\beta})(U_i-u_0)\}}{\partial\{g(u_0;\boldsymbol{\beta})+g'(u_0;\boldsymbol{\beta})(U_i-u_0)\}}\Bigg]\\
&\quad\times\{\widehat{g}(u_0;\boldsymbol{\beta})-g(u_0;\boldsymbol{\beta})\}\big(1+O_p\big(|\widehat{g}(u_0;\boldsymbol{\beta})-g(u_0;\boldsymbol{\beta})|\\
&\quad+|\widehat{g}'(u_0;\boldsymbol{\beta})-g'(u_0;\boldsymbol{\beta})|\big)\big)\\
&\quad+\Bigg[n^{-1}\sum_{i=1}^{n}\begin{Bmatrix}1\\(U_i-u_0)/h\end{Bmatrix}K_h(U_i-u_0)(U_i-u_0)\\
&\quad\times\frac{\partial\Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(u_0;\boldsymbol{\beta})+g'(u_0;\boldsymbol{\beta})(U_i-u_0)\}}{\partial\{g(u_0;\boldsymbol{\beta})+g'(u_0;\boldsymbol{\beta})(U_i-u_0)\}}\Bigg]\\
&\quad\times\{\widehat{g}'(u_0;\boldsymbol{\beta})-g'(u_0;\boldsymbol{\beta})\}\big(1+O_p\big(|\widehat{g}(u_0;\boldsymbol{\beta})-g(u_0;\boldsymbol{\beta})|\\
&\quad+|\widehat{g}'(u_0;\boldsymbol{\beta})-g'(u_0;\boldsymbol{\beta})|\big)\big)\\
&= n^{-1}\sum_{i=1}^{n}\begin{Bmatrix}1\\(U_i-u_0)/h\end{Bmatrix}K_h(U_i-u_0)\\
&\quad\times\Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(u_0;\boldsymbol{\beta})+g'(u_0,\boldsymbol{\beta})(U_i-u_0)\}\\
&\quad+\Bigg[n^{-1}\sum_{i=1}^{n}\begin{Bmatrix}1 & (U_i-u_0)/h\\(U_i-u_0)/h & (U_i-u_0)^2/h^2\end{Bmatrix}K_h(U_i-u_0)\\
&\quad\times\frac{\partial\Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(u_0;\boldsymbol{\beta})+g'(u_0;\boldsymbol{\beta})(U_i-u_0)\}}{\partial\{g(u_0;\boldsymbol{\beta})+g'(u_0;\boldsymbol{\beta})(U_i-u_0)\}}\Bigg]\\
&\quad\times\begin{bmatrix}\widehat{g}(u_0;\boldsymbol{\beta})-g(u_0;\boldsymbol{\beta})\\h\{\widehat{g}'(u_0;\boldsymbol{\beta})-g'(u_0;\boldsymbol{\beta})\}\end{bmatrix}\big(1+O_p\big(|\widehat{g}(u_0;\boldsymbol{\beta})-g(u_0;\boldsymbol{\beta})|\\
&\quad+|\widehat{g}'(u_0;\boldsymbol{\beta})-g'(u_0;\boldsymbol{\beta})|\big)\big)\\
&= n^{-1}\sum_{i=1}^{n}\begin{Bmatrix}1\\(U_i-u_0)/h\end{Bmatrix}K_h(U_i-u_0)\\
&\quad\times\Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(u_0;\boldsymbol{\beta})+g'(u_0,\boldsymbol{\beta})(U_i-u_0)\}\\
&\quad+\Bigg[\int h^{-1}\begin{Bmatrix}1 & (u-u_0)/h\\(u-u_0)/h & (u-u_0)^2/h^2\end{Bmatrix}K\left(\frac{u-u_0}{h}\right)\\
&\quad\times\frac{\partial\Phi\{y,\mathbf{s},u,\mathbf{z}_{-1};\boldsymbol{\beta},g(u_0;\boldsymbol{\beta})+g'(u_0,\boldsymbol{\beta})(u-u_0)\}}{\partial\{g(u_0;\boldsymbol{\beta})+g'(u_0;\boldsymbol{\beta})(u-u_0)\}}\\
&\quad\times f_{U,\mathbf{Z}_{-1},\mathbf{S},Y}(u,\mathbf{z}_{-1},\mathbf{s},y)\,du\,d\mathbf{z}_{-1}\,d\mathbf{s}\,dy+o_p(1)\Bigg]\begin{bmatrix}\widehat{g}(u_0;\boldsymbol{\beta})-g(u_0;\boldsymbol{\beta})\\h\{\widehat{g}'(u_0;\boldsymbol{\beta})-g'(u_0;\boldsymbol{\beta})\}\end{bmatrix}\\
&\quad\times\big(1+O_p\big(|\widehat{g}(u_0;\boldsymbol{\beta})-g(u_0;\boldsymbol{\beta})|+|\widehat{g}'(u_0;\boldsymbol{\beta})-g'(u_0;\boldsymbol{\beta})|\big)\big)
\end{aligned}
$$

$$
\begin{aligned}
&= n^{-1} \sum_{i=1}^{n} \begin{Bmatrix} 1 \\ (U_i - u_0)/h \end{Bmatrix} K_h(U_i - u_0) \\
&\quad \times \Phi\{Y_i, \mathbf{S}_i, \mathbf{Z}_i; \boldsymbol{\beta}, g(u_0; \boldsymbol{\beta}) + g'(u_0, \boldsymbol{\beta})(U_i - u_0)\} \\
&\quad + \bigg[\int \begin{pmatrix} 1 & t \\ t & t^2 \end{pmatrix} K(t) \frac{\partial \Phi\{y, \mathbf{s}, u_0 + th, \mathbf{z}_{-1}; \boldsymbol{\beta}, g(u_0; \boldsymbol{\beta}) + g'(u_0, \boldsymbol{\beta})th\}}{\partial \{g(u_0; \boldsymbol{\beta}) + g'(u_0; \boldsymbol{\beta})th\}} \\
&\quad \times f_{U,\mathbf{Z}_{-1},\mathbf{S},Y}(u_0 + th, \mathbf{z}_{-1}, \mathbf{s}, y)\, dt\, d\mathbf{z}_{-1}\, d\mathbf{s}\, dy + o_p(1)\bigg] \\
&\quad \times \begin{bmatrix} \widehat{g}(u_0; \boldsymbol{\beta}) - g(u_0; \boldsymbol{\beta}) \\ h\{\widehat{g}'(u_0; \boldsymbol{\beta}) - g'(u_0; \boldsymbol{\beta})\} \end{bmatrix} \big(1 + O_p\big(|\widehat{g}(u_0; \boldsymbol{\beta}) - g(u_0; \boldsymbol{\beta})| \\
&\quad + |\widehat{g}'(u_0; \boldsymbol{\beta}) - g'(u_0; \boldsymbol{\beta})|\big)\big)
\end{aligned}
$$

where

$$
\begin{aligned}
&\int K(t) \frac{\partial \Phi\{y, \mathbf{s}, u_0 + th, \mathbf{z}_{-1}; \boldsymbol{\beta}, g(u_0; \boldsymbol{\beta}) + g'(u_0, \boldsymbol{\beta})th\}}{\partial \{g(u_0; \boldsymbol{\beta}) + g'(u_0; \boldsymbol{\beta})th\}} \\
&\qquad \times f_{U,\mathbf{Z}_{-1},\mathbf{S},Y}(u_0 + th, \mathbf{z}_{-1}, \mathbf{s}, y)\, dt\, d\mathbf{z}_{-1}\, d\mathbf{s}\, dy \\
&\quad = \int K(t) \frac{\partial \Phi\{y, \mathbf{s}, u_0, \mathbf{z}_{-1}; \boldsymbol{\beta}, g(u_0; \boldsymbol{\beta})\}}{\partial g(u_0; \boldsymbol{\beta})} \\
&\qquad \times f_{U,\mathbf{Z}_{-1},\mathbf{S},Y}(u_0, \mathbf{z}_{-1}, \mathbf{s}, y)\, dt\, d\mathbf{z}_{-1}\, d\mathbf{s}\, dy + O(h^2) \\
&\quad = \int \frac{\partial \Phi\{y, \mathbf{s}, u_0, \mathbf{z}_{-1}; \boldsymbol{\beta}, g(u_0; \boldsymbol{\beta})\}}{\partial g(u_0; \boldsymbol{\beta})} \\
&\qquad \times f_{U,\mathbf{Z}_{-1},\mathbf{S},Y}(u_0, \mathbf{z}_{-1}, \mathbf{s}, y)\, d\mathbf{z}_{-1}\, d\mathbf{s}\, dy + O(h^2) \\
&\quad = E\bigg[\frac{\partial \Phi\{Y, \mathbf{S}, U, \mathbf{Z}_{-1}; \boldsymbol{\beta}, g(U; \boldsymbol{\beta})\}}{\partial g(U; \boldsymbol{\beta})} \bigg| U = u_0\bigg] f_U(u_0) + O(h^2).
\end{aligned}
$$

Similarly

$$
\begin{aligned}
&\int t K(t) \frac{\partial \Phi\{y, \mathbf{s}, u_0 + th, \mathbf{z}_{-1}; \boldsymbol{\beta}, g(u_0; \boldsymbol{\beta}) + g'(u_0, \boldsymbol{\beta})th\}}{\partial \{g(u_0; \boldsymbol{\beta}) + g'(u_0, \boldsymbol{\beta})th\}} \\
&\qquad \times f_{U,\mathbf{Z}_{-1},\mathbf{S},Y}(u_0 + th, \mathbf{z}_{-1}, \mathbf{s}, y)\, dt\, d\mathbf{z}_{-1}\, d\mathbf{s}\, dy = O(h^2),
\end{aligned}
$$

and

$$
\begin{aligned}
&\int t^2 K(t) \frac{\partial \Phi\{y, \mathbf{s}, u_0 + th, \mathbf{z}_{-1}; \boldsymbol{\beta}, g(u_0; \boldsymbol{\beta}) + g'(u_0, \boldsymbol{\beta})th\}}{\partial \{g(u_0; \boldsymbol{\beta}) + g'(u_0, \boldsymbol{\beta})th\}} \\
&\qquad \times f_{U,\mathbf{Z}_{-1},\mathbf{S},Y}(u_0 + th, \mathbf{z}_{-1}, \mathbf{s}, y)\, dt\, d\mathbf{z}_{-1}\, d\mathbf{s}\, dy \\
&\quad = \int t^2 K(t) \frac{\partial \Phi\{y, \mathbf{s}, u_0, \mathbf{z}_{-1}; \boldsymbol{\beta}, g(u_0; \boldsymbol{\beta})\}}{\partial g(u_0; \boldsymbol{\beta})} \\
&\qquad \times f_{U,\mathbf{Z}_{-1},\mathbf{S},Y}(u_0, \mathbf{z}_{-1}, \mathbf{s}, y)\, dt\, d\mathbf{z}_{-1}\, d\mathbf{s}\, dy + O(h^4) \\
&\quad = \int \mu_2 \frac{\partial \Phi\{y, \mathbf{s}, u_0, \mathbf{z}_{-1}; \boldsymbol{\beta}, g(u_0; \boldsymbol{\beta})\}}{\partial g(u_0; \boldsymbol{\beta})} \\
&\qquad \times f_{U,\mathbf{Z}_{-1},\mathbf{S},Y}(u_0, \mathbf{z}_{-1}, \mathbf{s}, y)\, d\mathbf{z}_{-1}\, d\mathbf{s}\, dy\{1 + O(h^2)\} \\
&\quad = E\bigg[\frac{\partial \Phi\{Y, \mathbf{S}, U, \mathbf{Z}_{-1}; \boldsymbol{\beta}, g(U; \boldsymbol{\beta})\}}{\partial g(U; \boldsymbol{\beta})} \mu_2 \bigg| U = u_0\bigg] f_U(u_0)\{1 + O_p(h^2)\}.
\end{aligned}
$$

So

$$
\begin{aligned}
\mathbf{0} &= n^{-1}\sum_{i=1}^{n}\left\{\begin{matrix}1\\(U_i-u_0)/h\end{matrix}\right\}K_h(U_i-u_0)\\
&\quad\times\Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(u_0;\boldsymbol{\beta})+g'(u_0,\boldsymbol{\beta})(U_i-u_0)\}\\
&\quad+\left(E\left[\frac{\partial\Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(u_0;\boldsymbol{\beta})\}}{\partial g(u_0;\boldsymbol{\beta})}\Big|U_i=u_0\right]\begin{pmatrix}1&0\\0&\mu_2\end{pmatrix}f_U(u_0)+o_p(1)\right)\\
&\quad\times\left[\begin{matrix}\widehat{g}(u_0;\boldsymbol{\beta})-g(u_0;\boldsymbol{\beta})\\h\{\widehat{g}'(u_0;\boldsymbol{\beta})-g'(u_0;\boldsymbol{\beta})\}\end{matrix}\right](1+O_p(|\widehat{g}(u_0;\boldsymbol{\beta})-g(u_0;\boldsymbol{\beta})|\\
&\quad+|\widehat{g}'(u_0;\boldsymbol{\beta})-g'(u_0;\boldsymbol{\beta})|)).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\widehat{g}(u_0;\boldsymbol{\beta})-g(u_0;\boldsymbol{\beta})\\
&\quad=-\left(E\left[\frac{\partial\Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(u_0;\boldsymbol{\beta})\}}{\partial g(u_0;\boldsymbol{\beta})}\Big|U_i=u_0\right]f_U(u_0)+o_p(1)\right)^{-1}\\
&\qquad\times\left[n^{-1}\sum_{i=1}^{n}K_h(U_i-u_0)\Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(u_0;\boldsymbol{\beta})+g'(u_0,\boldsymbol{\beta})(U_i-u_0)\}\right].
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&n^{-1/2}\sum_{i=1}^{n}\frac{\partial\mathcal{L}\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})\}}{\partial g(U_i;\boldsymbol{\beta})}\{\widehat{g}(U_i;\boldsymbol{\beta})-g(U_i;\boldsymbol{\beta})\}\\
&\quad=-n^{-3/2}\sum_{i,j=1}^{n}K_h(U_j-U_i)\frac{\partial\mathcal{L}\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})\}}{\partial g(U_i;\boldsymbol{\beta})}\\
&\qquad\times\left[E\left\{\frac{\partial\Phi\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\beta},g(U,\boldsymbol{\beta})\}}{\partial g(U;\boldsymbol{\beta})}\Big|U=U_i\right\}f_U(U_i)+o_p(1)\right]^{-1}\\
&\qquad\times\Phi\{Y_j,\mathbf{S}_j,\mathbf{Z}_j;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})+g'(U_i,\boldsymbol{\beta})(U_j-U_i)\}.
\end{aligned}\tag{A.6}
$$

Note that, under the conditions (C1) and (C2),

$$
\begin{aligned}
&n^{-1/2}\sum_{j=1}^{n}E\Big\{K_h(U_j-U_i)\frac{\partial\mathcal{L}\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})\}}{\partial g(U_i;\boldsymbol{\beta})}\\
&\qquad\times\left[E\left\{\frac{\partial\Phi\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\beta},g(U;\boldsymbol{\beta})\}}{\partial g(U;\boldsymbol{\beta})}\Big|U=U_i\right\}f_U(U_i)+o_p(1)\right]^{-1}\\
&\qquad\times\Phi\{Y_j,\mathbf{S}_j,\mathbf{Z}_j;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})+g'(U_i,\boldsymbol{\beta})(U_j-U_i)\}\Big|\mathbf{S}_j,\mathbf{Z}_j,Y_j\Big\}\\
&\quad=n^{-1/2}\sum_{j=1}^{n}\Big\{E\left[\frac{\partial\mathcal{L}\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\beta},g(U;\boldsymbol{\beta})\}}{\partial g(U;\boldsymbol{\beta})}\Big|U=U_j\right]\\
&\qquad\times\left[E\left\{\frac{\partial\Phi\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\beta},g(U;\boldsymbol{\beta})\}}{\partial g(U;\boldsymbol{\beta})}\Big|U=U_j\right\}\right]^{-1}+o_p(1)\Big\}\\
&\qquad\times\Phi\{Y_j,\mathbf{S}_j,\mathbf{Z}_j;\boldsymbol{\beta},g(U_j;\boldsymbol{\beta})\},
\end{aligned}
$$

and

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n} E\bigg\{K_h(U_j-U_i)\frac{\partial\mathcal{L}\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})\}}{\partial g(U_i;\boldsymbol{\beta})} \\
&\quad\times\bigg[E\bigg\{\frac{\partial\Phi\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\beta},g(U;\boldsymbol{\beta})\}}{\partial g(U;\boldsymbol{\beta})}\Big|U=U_i\bigg\}f_U(U_i)+o_p(1)\bigg]^{-1} \\
&\quad\times\Phi\{Y_j,\mathbf{S}_j,\mathbf{Z}_j;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})+g'(U_i,\boldsymbol{\beta})(U_j-U_i)\}|\mathbf{S}_i,\mathbf{Z}_i,Y_i\bigg\} \\
&= n^{-1/2}\sum_{i=1}^{n}\frac{\partial\mathcal{L}\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})\}}{\partial g(U_i;\boldsymbol{\beta})} \\
&\quad\times\bigg[E\bigg\{\frac{\partial\Phi\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\beta},g(U;\boldsymbol{\beta})\}}{\partial g(U;\boldsymbol{\beta})}\Big|U=U_i\bigg\}f_U(U_i)+o_p(1)\bigg]^{-1} \\
&\quad\times E\big[K_h(U_j-U_i)\Phi\{Y_j,\mathbf{S}_j,\mathbf{Z}_j;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})+g'(U_i,\boldsymbol{\beta})(U_j-U_i)\}|\mathbf{S}_i,\mathbf{Z}_i,Y_i\big] \\
&= n^{-1/2}\sum_{i=1}^{n}\frac{\partial\mathcal{L}\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})\}}{\partial g(U_i;\boldsymbol{\beta})} \\
&\quad\times\bigg[E\bigg\{\frac{\partial\Phi\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\beta},g(U;\boldsymbol{\beta})\}}{\partial g(U;\boldsymbol{\beta})}\Big|U=U_i\bigg\}f_U(U_i)+o_p(1)\bigg]^{-1} \\
&\quad\times\big(E\big[\Phi\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})\}|U=U_i\big]f_U(U_i)+O\big(h^2\big)\big) \\
&= O_p\big(n^{1/2}h^2\big)=o_p(1).
\end{aligned}
$$

Inserting these results to (A.6), in combination with U-statistic properties, we have

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\frac{\partial\mathcal{L}\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})\}}{\partial g(U_i;\boldsymbol{\beta})}\{\widehat{g}(U_i;\boldsymbol{\beta})-g(U_i;\boldsymbol{\beta})\} \\
&= -n^{-1/2}\sum_{i=1}^{n}\bigg(E\bigg[\frac{\partial\mathcal{L}\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\beta},g(U;\boldsymbol{\beta})\}}{\partial g(U;\boldsymbol{\beta})}\Big|U=U_i\bigg] \\
&\qquad\times\bigg[E\bigg\{\frac{\partial\Phi\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\beta},g(U;\boldsymbol{\beta})\}}{\partial g(U;\boldsymbol{\beta})}\Big|U=U_i\bigg\}\bigg]^{-1}\Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})\}\bigg) \\
&\qquad+o_p(1) \\
&= -n^{-1/2}\sum_{i=1}^{n}E\bigg[\frac{\partial\mathcal{L}\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\beta},g(U;\boldsymbol{\beta})\}}{\partial g(U;\boldsymbol{\beta})}\Big|U=U_i\bigg] \\
&\qquad\times\bigg[E\bigg\{\frac{\partial\Phi\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\beta},g(U;\boldsymbol{\beta})\}}{\partial g(U;\boldsymbol{\beta})}\Big|U=U_i\bigg\}\bigg]^{-1}\Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})\} \\
&\qquad+o_p(1).
\end{aligned}
$$

Continuing from (A.5), using the property that $nh^4\to 0$, we obtain

$$
\begin{aligned}
\mathbf{0} &= n^{-1/2}\sum_{i=1}^{n}\bigg(\mathcal{L}\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})\} \\
&\quad - E\bigg[\frac{\partial\mathcal{L}\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\beta},g(U;\boldsymbol{\beta})\}}{\partial g(U;\boldsymbol{\beta})}\Big|U=U_i\bigg]
\end{aligned}
$$

$$\times \left[E\left\{\frac{\partial\Phi\{Y,\mathbf{S},\mathbf{Z};\boldsymbol{\beta},g(U;\boldsymbol{\beta})\}}{\partial g(U;\boldsymbol{\beta})}\Big|U=U_i\right\}\right]^{-1}\Phi\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})\}\Bigg)$$

$$+\Bigg\{E\Bigg[\frac{\partial\mathcal{L}\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})\}}{\partial\boldsymbol{\beta}^{\mathrm{T}}}-\frac{\partial\mathcal{L}\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})\}}{\partial g(U_i;\boldsymbol{\beta})}$$

$$\times\frac{E[\partial\Phi\{Y_j,\mathbf{S}_j,\mathbf{Z}_j;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})\}/\partial\boldsymbol{\beta}^{\mathrm{T}}|U_i]}{E[\partial\Phi\{Y_j,\mathbf{S}_j,\mathbf{Z}_j;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})\}/\partial g(U_i;\boldsymbol{\beta})|U_i]}\Bigg]+o_p(1)\Bigg\}\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})$$

$$-\Bigg\{E\Bigg[\frac{\partial\mathcal{L}\{Y_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},g(U_i;\boldsymbol{\beta})\}}{\partial g(U_i;\boldsymbol{\beta})}$$

$$\times\frac{E[(\mathbf{Z}_j-\mathbf{Z}_i)_{-1}^{\mathrm{T}}\partial\Phi\{Y_j,\mathbf{S}_j,\mathbf{Z}_j;\boldsymbol{\beta},g(U_i,\boldsymbol{\beta})\}/\partial g(U_i,\boldsymbol{\beta})g'(U_i,\boldsymbol{\beta})|U_i]}{E[\partial\Phi\{Y_j,\mathbf{S}_j,\mathbf{Z}_j;\boldsymbol{\beta},g(U_i,\boldsymbol{\beta})\}/\partial\{g(U_i,\boldsymbol{\beta})\}|U_i]}\Bigg]+o_p(1)\Bigg\}$$

$$\times\sqrt{n}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})+o_p(1).$$

This leads to the result stated in the theorem. □

Acknowledgments

Yang's research was supported by the National Social Science Foundation of China grant 16BTJ032, the Guangdong Province Nature Science Foundation of China grants 2019A1515010721 and the Fundamental Research Funds of Central University 19JNYH08. Cui is the corresponding author. Cui's Research was supported by grants from National Natural Science Foundation of China grant 11871173. Ma's research was partially supported by the national science foundation and the national institute of neurological disorder and stroke.

References

Buzas, J. S. and Stefanski, L. A. (1996). Instrumental variable estimation in generalized linear measurement error models. *Journal of the American Statistical Association* **91**, 999–1006. MR1340507

Huang, Y. and Wang, C. Y. (2001). Consistent functional methods for logistic regression with errors in covariates. *Journal of the American Statistical Association* **96**, 1469–1482. MR1946591 https://doi.org/10.1198/016214501753382372

Kress, R. (1991). *Linear Integral Equations*. New York: Springer. MR3184286 https://doi.org/10.1007/978-1-4614-9593-2

Ma, Y. and Carroll, R. J. (2006). Locally efficient estimators for semiparametric models with measurement error. *Journal of the American Statistical Association* **101**, 1465–1474. MR2279472 https://doi.org/10.1198/016214506000000519

Ma, Y. and Tsiatis, A. A. (2006). Closed form semiparametric estimators for measurement error models. *Statistica Sinica* **16**, 183–193. MR2256086

Ma, Y. and Zhu, L. (2013). Doubly robust and efficient estimators for heteroscedastic partially linear single-index model allowing high dimensional covariates. *Journal of the Royal Statistical Society, Series B* **75**, 305–322. MR3021389 https://doi.org/10.1111/j.1467-9868.2012.01040.x

Stefanski, L. A. and Carroll, R. J. (1985). Covariate measurement error in logistic regression. *The Annals of Statistics* **13**, 1335–1351. MR0811496 https://doi.org/10.1214/aos/1176349741

Stefanski, L. A. and Carroll, R. J. (1987). Conditional scores and optimal scores in generalized linear measurement error models. *Biometrika* **74**, 703–716. MR0919838 https://doi.org/10.1093/biomet/74.4.703

Tsiatis, A. A. (2006). *Semiparametric Theory and Missing Data*. New York: Springer. MR2233926

Xu, K., Ma, Y. and Wang, L. (2015). Instrument assisted regression for errors in variables models with binary response. *Scandinavian Journal of Statistics* **42**, 104–117. MR3318027 https://doi.org/10.1111/sjos.12097

G. Yang
Department of Statistics
School of Economics
Jinan University
Guangzhou
China
E-mail: tygr@jnu.edu.cn

Q. Wang
Department of Statistics
University of South Carolina
1523 Greene St
Columbia, South Carolina
USA
E-mail: wangqq07@gmail.com

X. Cui
School of Economics and Statistics
Guangzhou University
Guangzhou
China
E-mail: cuixia@gzhu.edu.cn

Y. Ma
Department of Statistics
Pennsylvania State University
University Park, Pennsylvania
USA
E-mail: yanyuanma@gmail.com