Estimation and inference of error-prone covariate effect in the presence of confounding variables[*]

Jianxuan Liu

Department of Statistics, University of South Carolina
e-mail: jianxuan@email.sc.edu

Yanyuan Ma[†]

Department of Statistics, The Pennsylvania State University
e-mail: yzm63@psu.edu

Liping Zhu[‡]

Research Center for Applied Statistical Science and Institute of Statistics and Big Data,
Renmin University of China
e-mail: zhu.liping@ruc.edu.cn

and

Raymond J. Carroll[§]

Department of Statistics, Texas A&M University, and School of Mathematical and Physical
Sciences, University of Technology Sydney, Australia
e-mail: carroll@stat.tamu.edu

Abstract: We introduce a general single index semiparametric measurement error model for the case that the main covariate of interest is measured with error and modeled parametrically, and where there are many other variables also important to the modeling. We propose a semiparametric bias-correction approach to estimate the effect of the covariate of interest. The resultant estimators are shown to be root-n consistent, asymptotically normal and locally efficient. Comprehensive simulations and an analysis of an empirical data set are performed to demonstrate the finite sample performance and the bias reduction of the locally efficient estimators.

MSC 2010 subject classifications: Primary 62H12, 62H15; secondary 62F12.

1. Introduction

Estimating and testing the effect of a covariate of interest while accommodating many other covariates is an important problem in statistical practice. The t-test and the analysis of variance are widely used to evaluate the covariate effect when the covariate of interest is binary or categorical and no confounders are present. When the covariate of interest is not necessarily binary or categorical, evaluating the covariate effect has been studied extensively in the context of linear model, partially linear model (Heckman, 1986; Härdle et al., 2000; Ma et al., 2006) and partially linear single-index model (Carroll et al., 1997; Yu and Ruppert, 2002; Li et al., 2011; Ma and Zhu, 2013), as long as both the covariate of interest and the confounders are measured precisely. In this work, we intend to generalize the partially linear single-index model to a larger class where the link function is not restricted to be linear, and we further consider measurement error issues.

When the covariate of interest is measured with error, to evaluate its effect precisely we must reduce the bias caused by measurement error and adjust for the confounding effects simultaneously. This is an interesting yet very challenging problem. To partially address this problem, Carroll et al. (2006) assumed the confounding effects are linear, and Liang et al. (1999) and Ma and Carroll (2006) assumed the confounders are in fact univariate. These assumptions restrict the usefulness of their methods. To the best of our knowledge, how to assess the covariate effect subject to measurement error while taking into account possibly nonlinear confounding effects still remains an open and difficult problem in the literature.

Estimating and testing the effect of a covariate of interest in the presence of possibly nonlinear confounding effects has many applications in a variety of scientific fields such as econometrics, biology, policy making, etc. Consider the Framingham Heart Study (http://www.framinghamheartstudy.org/) as a typical example. It is common knowledge that high systolic blood pressure (SBP) is directly linked to the occurrence of coronary heart disease (Y). To quantify the effect is however not necessarily straightforward. One difficulty is that SBP can vary significantly from time to time, hence a clinically meaningful covariate is the long term average of SBP $(\widetilde{X})$, which is unfortunately impossible to measure precisely. A widely used practice is to use the average of several measured SBP values $(\widetilde{W})$ during a reasonably long time course as a substitute. Thus, long term average SBP is a variable measured with error. Another difficulty comes from the presence of possibly nonlinear confounding effects $(\mathbf{Z})$ for heart disease, such as smoking status, family history, ethnicity, BMI, lung capacity, age and other laboratory variables. Because these effects are not of medical interest while their connection to the heart disease occurrence might be complex, a suitable modeling strategy is to use an unspecified function to summarize

their possibly nonlinear effect. Difficulty with such modeling strategy naturally arises when the dimension of $\mathbf{Z}$ is more than one, since it is well known that nonparametrically estimating a function of multivariate confounding variables suffers from the curse of dimensionality. To tackle this issue, we follow the single index modeling strategy and assume that the combined effect of the covariates in $\mathbf{Z}$ is manifested through a linear combination $\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z}$, where $\widetilde{\boldsymbol{\gamma}}$ is a length p vector. For identifiability, we assume that $\mathbf{Z}$ contains at least one continuous variable, the first component of $\widetilde{\boldsymbol{\gamma}}$ is one, and we use $\boldsymbol{\gamma}$ to denote the vector of the last $p-1$ components. Let H be the logistic distribution function. In this Framingham data example, we assume that, given $\widetilde{X}$ and $\mathbf{Z}$, the probability of the occurrence of the coronary heart disease (Y) admits a model of the form

$$
\begin{aligned}
\mathrm{pr}(Y = 1 \mid \widetilde{X}, \mathbf{Z}) &= H\{\widetilde{X}\beta + \theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z})\}, \\
\log(\widetilde{W} - 50) &= \log(\widetilde{X} - 50) + U.
\end{aligned}
$$

Here we adopt the general assumption that after the transformation from the raw systolic blood pressure, the relation between $W \equiv \log(\widetilde{W} - 50)$ and $X \equiv \log(\widetilde{X} - 50)$ is additive with a normal measurement error, i.e. $U \sim N(0, \sigma_u^2)$, and we assume the error is nondifferential. This relation is verified by Carroll et al. (2006, chapter 6).

The above model can be viewed as a special case of the following general semiparametric measurement error model. To be specific, we write the general probability density/mass function of the response variable Y, for example disease status, conditional on the covariate set $(X, \mathbf{S}^{\mathrm{T}}, \mathbf{Z}^{\mathrm{T}})^{\mathrm{T}}$ as

$$
g\{y, x, \mathbf{s}, \theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}), \boldsymbol{\beta}\}, \tag{1.1}
$$

where X is an error-prone covariate whose effect on Y is of central research interest, $\mathbf{Z}, \mathbf{S}$ contain additional covariates that may be related to Y and may be confounded with X. We model part of these confounders $(\mathbf{S})$ parametrically, such as the categorical variables, and part of these confounders $(\mathbf{Z})$ nonparametrically through an unspecified smooth function θ. Both $\mathbf{S}$ and $\mathbf{Z}$ are measured precisely. In model (1.1), g is a known conditional probability density/mass function, θ is an unspecified smooth function, $\widetilde{\boldsymbol{\gamma}} = (1, \boldsymbol{\gamma}^{\mathrm{T}})^{\mathrm{T}}$, where $\boldsymbol{\gamma}$ is an unknown length $p-1$ vector, and $\boldsymbol{\beta}$ is an unknown parameter. In this notation, the example above can be written as $g\{y, x, \mathbf{s}, \theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}), \boldsymbol{\beta}\} = \exp[y\{\widetilde{x}\beta + \theta(\widetilde{\boldsymbol{\gamma}}\mathbf{z})\}]/[1 + \exp\{\widetilde{x}\beta + \theta(\widetilde{\boldsymbol{\gamma}}\mathbf{z})\}]$. In our context, we assume the covariate X is of our primary interest but is unobservable. Instead, we observe its erroneous version W, where the relation between W and X is specified, i.e. $f_{W|X}(w \mid x)$ is a known model. In practice, the specification of $f_{W|X}(w \mid x)$ is usually obtained through validation data, instruments or repeated measurements. We treat $\theta(\cdot)$ as an infinite dimensional nuisance parameter. We further make the surrogacy assumption that W and Y are independent given $X, \mathbf{S}, \mathbf{Z}$. The primary interest is in $\boldsymbol{\beta}$, which describes the effect of X on Y. In many applications, $\boldsymbol{\beta}$ enters the model as multiplication coefficient of a linear function of the covariates, such as through $\beta_1 X + \beta_2 S$.

Model (1.1) is an extension of the generalized single index model proposed by Cui et al. (2011) in which neither X nor $\mathbf{S}$ is present. In addition, Tsiatis and

Ma (2004) studied a simpler version of model (1.1) where $\mathbf{Z}$ does not appear, and Ma and Carroll (2006) considered a simpler version of model (1.1) where Z is univariate. The generalization to multivariate $\mathbf{Z}$ in model (1.1) is important in practice since it accomodates more realistic applications; see, for example, the Framingham Heart Study in Section 5. In particular, model (1.1) allows us to handle the possible nonlinearity of the confounding variables through the unspecified function θ, while the single index structure $\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}$ facilitates non-parametric modeling. Nevertherless, the extension also poses several challenging technical and computational problems. Indeed, when the index vector appears inside an unknown function, its estimation is more complex and interaction between the estimation of the indices and the function has to be taken into account. The variability in estimating these quantities further affects the estimation quality of the parameter of interest. Overall, the three sets of parameters, namely the parameter of interest, the index vector and the unknown smooth function link together intrinsically, which complicates the estimation procedure, the computational treatment and the theoretical development. Compared with the case when the index vector does not appear, such additional complexity can be viewed as a price paid to overcome the curse of dimensionality.

We design a general methodology for the semiparametric measurement error model (1.1), and introduce a bias-correction approach to construct a class of locally efficient estimators. This bias-correction approach is motivated by the projected score idea in semiparametrics (Tsiatis and Ma, 2004) and does not have to resort to a deconvolution method or to correctly specify a distributional model for the error-prone covariate of interest. We further generalize the bias-correction approach to estimating $\widetilde{\boldsymbol{\gamma}}$ in model (1.1), which is a component that does not appear in the models considered in Tsiatis and Ma (2004) or Ma and Carroll (2006). In their studies, $\mathbf{Z}$ is either absent or univariate, hence the issue of estimating $\widetilde{\boldsymbol{\gamma}}$ does not occur. In the presence of multivariate $\mathbf{Z}$, the conditional density of X given $\mathbf{S}$ and $\mathbf{Z}$, denoted $f_{X|\mathbf{S},\mathbf{Z}}(x,\mathbf{s},\mathbf{z})$, is required in implementing the bias-correction approach. However, with a multivariate $\mathbf{Z}$, regardless whether $\mathbf{S}$ is discrete or continuous, estimating $f_{X|\mathbf{S},\mathbf{Z}}(x,\mathbf{s},\mathbf{z})$ is a thorny issue even if X were observed due to the curse of dimensionality. To alleviate the difficulty in estimating $f_{X|\mathbf{S},\mathbf{Z}}(x,\mathbf{s},\mathbf{z})$, a working model is adopted. If this working model happens to be the underlying true one, the resultant estimator is semiparametrically efficient, whereas if this working model is unfortunately misspecified, then the resultant estimator is still root-n consistent and asymptotically normal. In other words, the resultant estimator is locally efficient. To put the bias-correction approach into practice, we suggest a profiling algorithm for estimating $\boldsymbol{\beta}$.

The article is organized as the following. In Section 2 we introduce the bias-correction approach for estimating $\boldsymbol{\beta}$ in the semiparametric measurement error model (1.1). The asymptotic properties of the resultant estimators are given in Section 3. We report several simulation studies in Section 4 and revisit the Framingham data in Section 5. This paper is concluded with a brief discussion in Section 6. All technical details are given in an Appendix.

2. Estimation

In this section we discuss estimation of the covariate effect at the sample level. Write the observation as $(y_i, w_i, \mathbf{s}_i, \mathbf{z}_i)$, $i = 1, \ldots, n$. We propose to estimate the effect of the covariate of interest as well as other nuisance parameters through solving the estimating equations derived from the semiparametric log-likelihood.

The surrogacy assumption and the model specification in Section 1 directly lead to the semiparametric log-likelihood, subject to an additive term that does not involve the parameters $\boldsymbol{\beta}, \boldsymbol{\gamma}, \theta$,

$$
\begin{aligned}
& l(\boldsymbol{\beta}, \boldsymbol{\gamma}, \theta, f_{X|\mathbf{S},\mathbf{z}}) \\
& = \sum_{i=1}^{n} \log \int g\{y_i, x, \mathbf{s}_i, \theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_i), \boldsymbol{\beta}\} f_{W|X}(w_i \mid x) f_{X|\mathbf{S},\mathbf{z}}(x \mid \mathbf{s}_i, \mathbf{z}_i) dx.
\end{aligned}
$$

Recall that $\boldsymbol{\gamma}$ is defined in Section 1 as a vector of the free parameters in $\widetilde{\boldsymbol{\gamma}}$. Here $f_{X|\mathbf{S},\mathbf{z}}$ and $f_{W|X}$ represent the probability density function of X conditional on $(\mathbf{S}, \mathbf{Z})$ and the probability density function of W conditional on X respectively. If both θ and $f_{X|\mathbf{S},\mathbf{z}}$ had been known, the simple maximum likelihood estimator (MLE) would have provided a most natural estimator for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. Let

$$
\begin{aligned}
& \mathbf{S}_{\beta}(w, \mathbf{s}, \mathbf{z}, y; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta, f_{X|\mathbf{S},\mathbf{z}}) \\
& \quad = \frac{\partial \log \int g\{y, x, \mathbf{s}, \theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}), \boldsymbol{\beta}\} f_{W|X}(w \mid x) f_{X|\mathbf{S},\mathbf{z}}(x \mid \mathbf{s}, \mathbf{z}) dx}{\partial \boldsymbol{\beta}}, \\
& \mathbf{S}_{\gamma}(w, \mathbf{s}, \mathbf{z}, y; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta, f_{X|\mathbf{S},\mathbf{z}}) \\
& \quad = \frac{\partial \log \int g\{y, x, \mathbf{s}, \theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}), \boldsymbol{\beta}\} f_{W|X}(w \mid x) f_{X|\mathbf{S},\mathbf{z}}(x \mid \mathbf{s}, \mathbf{z}) dx}{\partial \boldsymbol{\gamma}}
\end{aligned}
$$

be the score functions with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, then we could modify the MLE through localization to handle the issue caused by the unknown functional form of θ. Specifically, let us adopt a local parametric model $\theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}) = \nu(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}; \boldsymbol{\alpha})$. For example, the most widely used local polynomial model in Fan and Gijbels (1996) can be used as $\nu(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}; \boldsymbol{\alpha})$. Here $\boldsymbol{\alpha}$ depends on $\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}$, but we suppress the dependence of $\boldsymbol{\alpha}$ on $\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}$ for notational clarity. Then we could estimate θ together with $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, through iteratively solving

$$
\begin{aligned}
\mathbf{0} &= \sum_{i=1}^{n} \mathbf{S}_{\beta}\{w_i, \mathbf{s}_i, \mathbf{z}_i, y_i; \boldsymbol{\beta}, \boldsymbol{\gamma}, \widehat{\theta}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_i), f_{X|\mathbf{S},\mathbf{z}}(x_i \mid \mathbf{s}_i, \mathbf{z}_i)\} \\
\mathbf{0} &= \sum_{i=1}^{n} \mathbf{S}_{\gamma}\{w_i, \mathbf{s}_i, \mathbf{z}_i, y_i; \boldsymbol{\beta}, \boldsymbol{\gamma}, \widehat{\theta}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_i), f_{X|\mathbf{S},\mathbf{z}}(x_i \mid \mathbf{s}_i, \mathbf{z}_i)\}
\end{aligned}
$$

to obtain $\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}$, and

$$
\mathbf{0} = \sum_{i=1}^{n} K_h(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_i - \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_0) \mathbf{S}_{\alpha}(w_i, \mathbf{s}_i, \mathbf{z}_i, y_i; \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}, \boldsymbol{\alpha}, f_{X|\mathbf{S},\mathbf{z}})
$$

at $\mathbf{z}_0$ to obtain $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\theta}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0) = \nu(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0; \widehat{\boldsymbol{\alpha}})$ for $\mathbf{z}_0 = \mathbf{z}_1, \ldots, \mathbf{z}_n$. Here, $K_h(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z} - \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0) = h^{-1}K\{(\mathbf{z}^{\mathrm{T}}\boldsymbol{\gamma} - \mathbf{z}_0^{\mathrm{T}}\boldsymbol{\gamma})/h\}$, K is a kernel function and h is a bandwidth. In the above display, $\mathbf{S}_\alpha$ is defined analogously as $\mathbf{S}_\beta$ except that $\theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z})$ is replaced by $\nu(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}; \boldsymbol{\alpha})$ and the derivative is with respect to $\boldsymbol{\alpha}$, i.e.

$$\mathbf{S}_\alpha(w, \mathbf{s}, \mathbf{z}, y; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, f_{X|\mathbf{S},\mathbf{Z}})$$
$$= \frac{\partial \log \int g\{y, x, \mathbf{s}, \nu(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}; \boldsymbol{\alpha}), \boldsymbol{\beta}\} f_{W|X}(w \mid x) f_{X|\mathbf{S},\mathbf{Z}}(x \mid \mathbf{s}, \mathbf{z}) dx}{\partial \boldsymbol{\alpha}}.$$

The above idea would have worked if we knew how to actually calculate the score functions. However, without an explicit form of $f_{X|\mathbf{S},\mathbf{Z}}$, the calculation of the score vectors is not an easy task. A natural approach is to estimate $f_{X|\mathbf{S},\mathbf{Z}}$ and then use the estimated version to obtain the corresponding estimated score functions. This is not entirely out of the question, especially when $f_{W|X}(w, x)$ happens to describe an additive independent error model, i.e. when $f_{W|\mathbf{X}}(w, x) = f_U(w - x)$. In this case, from the relation $f_{W|\mathbf{S},\mathbf{Z}}(w, \mathbf{s}, \mathbf{z}) = \int f_U(w - x) f_{X|\mathbf{S},\mathbf{Z}}(x, \mathbf{s}, \mathbf{z}) dx$, we use the Fourier transform to obtain $\mathcal{F}_w(t, \mathbf{s}, \mathbf{z}) = \mathcal{F}_u(t)\mathcal{F}_x(t, \mathbf{s}, \mathbf{z})$, $\mathcal{F}_w(t, \mathbf{s}, \mathbf{z}) = \int f_{W|\mathbf{S},\mathbf{Z}}(w, \mathbf{s}, \mathbf{z}) e^{-2\pi itw} dw$, $\mathcal{F}_u(t) = \int f_U(u) e^{-2\pi itu} du$ and $\mathcal{F}_x(t, \mathbf{s}, \mathbf{z}) = \int f_{X|\mathbf{S},\mathbf{Z}}(x, \mathbf{s}, \mathbf{z}) e^{-2\pi itx} dx$. Thus, if we estimate $f_{W|\mathbf{S},\mathbf{Z}}(w, \mathbf{s}, \mathbf{z})$ nonparametrically, then we can obtain an estimated version of $\mathcal{F}_w(t, \mathbf{s}, z)$ and an estimated version of $\mathcal{F}_x(t, \mathbf{s}, \mathbf{z}) = \mathcal{F}_w(t, \mathbf{s}, \mathbf{z})/\mathcal{F}_u(t)$. Performing an inverse Fourier transform on $\mathcal{F}_x(t, \mathbf{s}, \mathbf{z})$ would then yield an estimate of $f_{X|\mathbf{S},\mathbf{Z}}(x, \mathbf{s}, \mathbf{z})$.

The above analysis reveals some hidden obstacles in estimating $f_{X|\mathbf{S},\mathbf{Z}}(x, \mathbf{s}, \mathbf{z})$. First of all, the deconvolution procedure is only applicable when the measurement error is additive and independent of X. When the measurement error model $f_{W|X}(w, x)$ goes beyond this structure, it is unclear how to recover $f_{X|\mathbf{S},\mathbf{Z}}(x, \mathbf{s}, \mathbf{z})$. Second, the procedure requires estimating $f_{W|\mathbf{S},\mathbf{Z}}(w, \mathbf{s}, \mathbf{z})$ nonparametrically. However, when the dimension of $(\mathbf{s}, \mathbf{z})$ is moderate or high, in other words, the confounding variables are multivariate, this is again a problem suffering from the curse of dimensionality and is not practically feasible in finite samples. Finally, even when the dimension of $(\mathbf{s}, \mathbf{z})$ is sufficiently low and the deconvolution procedure can be carried out in practice, the resulting estimate of $f_{X|\mathbf{S},\mathbf{Z}}(x, \mathbf{s}, \mathbf{z})$ has very slow convergence rate (Carroll and Hall, 1988; Fan, 1991), hence using the estimated $\widehat{f}_{X|\mathbf{S},\mathbf{Z}}(x, \mathbf{s}, \mathbf{z})$ may yield very different results from using the true $f_{X|\mathbf{S},\mathbf{Z}}(x, \mathbf{s}, \mathbf{z})$, which is required in the original score function calculation.

Due to these inherent difficulties involved with estimating $f_{X|\mathbf{S},\mathbf{Z}}(x, \mathbf{s}, \mathbf{z})$, we decide not to pursue this route. Instead, we take a somewhat counter-intuitive approach. Instead of striving to obtain an approximation of $f_{X|\mathbf{S},\mathbf{Z}}(x, \mathbf{s}, \mathbf{z})$, we propose to simply guess a model $f_{X|\mathbf{S},\mathbf{Z}}^*(x, \mathbf{s}, \mathbf{z})$, which may or may not reflect the true conditional density function, and calculate the score functions $\mathbf{S}_\beta, \mathbf{S}_\gamma, \mathbf{S}_\alpha$ under this guessed model. Of course, this simple replacement of the true score functions with the guessed version is not guaranteed to yield consistent estimation of $\boldsymbol{\beta}, \boldsymbol{\gamma}$ and θ. To correct the possible bias, we form

$$\mathbf{L}_\beta(w, \mathbf{s}, \mathbf{z}, y; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta, f^*_{X|\mathbf{S},\mathbf{Z}})$$
$$= \mathbf{S}_\beta(w, \mathbf{s}, \mathbf{z}, y; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta, f^*_{X|\mathbf{S},\mathbf{Z}}) - E^*\{\mathbf{a}_\beta(X, \mathbf{s}, \mathbf{z}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta) \mid w, \mathbf{s}, \mathbf{z}, y\},$$
$$\mathbf{L}_\gamma(w, \mathbf{s}, \mathbf{z}, y; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta, f^*_{X|\mathbf{S},\mathbf{Z}})$$
$$= \mathbf{S}_\gamma(w, \mathbf{s}, \mathbf{z}, y; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta, f^*_{X|\mathbf{S},\mathbf{Z}}) - E^*\{\mathbf{a}_\gamma(X, \mathbf{s}, \mathbf{z}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta) \mid w, \mathbf{s}, \mathbf{z}, y\}, \quad (2.1)$$
$$\mathbf{L}_\alpha(w, \mathbf{s}, \mathbf{z}, y; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, f^*_{X|\mathbf{S},\mathbf{Z}})$$
$$= \mathbf{S}_\alpha(w, \mathbf{s}, \mathbf{z}, y; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, f^*_{X|\mathbf{S},\mathbf{Z}}) - E^*\{\mathbf{a}_\alpha(X, \mathbf{s}, \mathbf{z}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}) \mid w, \mathbf{s}, \mathbf{z}, y\},$$

where $\mathbf{a}_\beta, \mathbf{a}_\gamma, \mathbf{a}_\alpha$ are functions of $(X, \mathbf{S}^\mathrm{T}, \mathbf{Z}^\mathrm{T})^\mathrm{T}$ that satisfy

$$E\{\mathbf{S}_\beta(W, \mathbf{s}, \mathbf{z}, Y; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta, f^*_{X|\mathbf{S},\mathbf{Z}}) \mid x, \mathbf{s}, \mathbf{z}\}$$
$$= E[E^*\{\mathbf{a}_\beta(X, \mathbf{s}, \mathbf{z}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta)|W, \mathbf{s}, \mathbf{z}, Y\} \mid x, \mathbf{s}, \mathbf{z}], \quad (2.2)$$
$$E\{\mathbf{S}_\gamma(W, \mathbf{s}, \mathbf{z}, Y; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta, f^*_{X|\mathbf{S},\mathbf{Z}}) \mid x, \mathbf{s}, \mathbf{z}\}$$
$$= E[E^*\{\mathbf{a}_\gamma(X, \mathbf{s}, \mathbf{z}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta)|W, \mathbf{s}, \mathbf{z}, Y\} \mid x, \mathbf{s}, \mathbf{z}],$$
$$E\{\mathbf{S}_\alpha(W, \mathbf{s}, \mathbf{z}, Y; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, f^*_{X|\mathbf{S},\mathbf{Z}}) \mid x, \mathbf{s}, \mathbf{z}\}$$
$$= E[E^*\{\mathbf{a}_\alpha(X, \mathbf{s}, \mathbf{z}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha})|W, \mathbf{s}, \mathbf{z}, Y\} \mid x, \mathbf{s}, \mathbf{z}],$$

and E^* represents expectation calculated using $f^*_{X|\mathbf{S},\mathbf{Z}}(x, \mathbf{s}, \mathbf{z})$. $E(\mathbf{a}_\beta \mid w, \mathbf{s}, \mathbf{z}, y)$, $E(\mathbf{a}_\gamma \mid w, \mathbf{s}, \mathbf{z}, y)$ and $E(\mathbf{a}_\alpha \mid w, \mathbf{s}, \mathbf{z}, y)$ are respectively the projections of the score vectors $\mathbf{S}_\beta$, $\mathbf{S}_\gamma$ and $\mathbf{S}_\alpha$ onto the tangent space Λ described in Appendix A.1, and has an no explicit form except in some special cases. We give one such special example at the end of this section. It is easy to see that the definition of $\mathbf{a}_\beta, \mathbf{a}_\gamma, \mathbf{a}_\alpha$ in (2.2) guarantees the consistency of $\mathbf{L}_\beta, \mathbf{L}_\gamma$, and $\mathbf{L}_\alpha$ automatically, whether or not $f^*_{X|\mathbf{S},\mathbf{Z}}$ reflects the truth. We then use $\mathbf{L}_\beta$, $\mathbf{L}_\gamma$ and $\mathbf{L}_\alpha$ to replace $\mathbf{S}_\beta, \mathbf{S}_\gamma, \mathbf{S}_\alpha$ in the iterative procedure described above to estimate $\boldsymbol{\beta}, \boldsymbol{\gamma}$ and θ. That is, we solve

$$\mathbf{0} = \sum_{i=1}^n \mathbf{L}_\beta(w_i, \mathbf{s}_i, \mathbf{z}_i, y_i; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta, f^*_{X|\mathbf{S},\mathbf{Z}}),$$
$$\mathbf{0} = \sum_{i=1}^n \mathbf{L}_\gamma(w_i, \mathbf{s}_i, \mathbf{z}_i, y_i; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta, f^*_{X|\mathbf{S},\mathbf{Z}}) \quad (2.3)$$

to estimate $\boldsymbol{\beta}, \boldsymbol{\gamma}$ and solve

$$\mathbf{0} = \sum_{i=1}^n K_h(\widetilde{\boldsymbol{\gamma}}^\mathrm{T}\mathbf{z}_i - \widetilde{\boldsymbol{\gamma}}^\mathrm{T}\mathbf{z}_0)\mathbf{L}_\alpha(w_i, \mathbf{s}_i, \mathbf{z}_i, y_i; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, f^*_{X|\mathbf{S},\mathbf{Z}}) \quad (2.4)$$

at $\mathbf{z}_0 = \mathbf{z}_1, \ldots, \mathbf{z}_n$ to obtain $\widehat{\theta}(\widetilde{\boldsymbol{\gamma}}^\mathrm{T}\mathbf{z}_0) = \nu(\widetilde{\boldsymbol{\gamma}}^\mathrm{T}\mathbf{z}_0; \widehat{\boldsymbol{\alpha}})$. Because different $\mathbf{z}_0$ yields different $\boldsymbol{\alpha}$, hence we could have used a more precise notation $\boldsymbol{\alpha}(\mathbf{z}_0)$ in (2.4). We suppressed the dependence of $\boldsymbol{\alpha}$ on $\mathbf{z}_0$ for notational brevity. The estimation procedure can be either iteratively solving (2.3) and (2.4) (backfitting), or using (2.4) to obtain $\widehat{\theta}$ as a function of $\boldsymbol{\beta}, \boldsymbol{\gamma}$, and then using (2.3) to solve for $\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}$ (profiling). In the following, we carry out all the procedures using the profiling approach.

The bias correction through forming $\mathbf{L}_\beta$ etc. is rooted in the projected score idea in semiparametrics (Bickel et al., 1993; Tsiatis and Ma, 2004; Tsiatis, 2006). Given any function, say $\mathbf{S}_\beta$, we can calculate its residual after projecting it onto the nuisance tangent space associated with the model. The projection of $(\mathbf{S}_\beta^{\mathrm{T}}, \mathbf{S}_\gamma^{\mathrm{T}}, \mathbf{S}_\alpha^{\mathrm{T}})^{\mathrm{T}}$ indeed would have been $(\mathbf{L}_\beta^{\mathrm{T}}, \mathbf{L}_\gamma^{\mathrm{T}}, \mathbf{L}_\alpha^{\mathrm{T}})^{\mathrm{T}}$, if we had used $f_{X|\mathbf{S},\mathbf{z}}$ throughout all the calculations. We defer the detail of this calculation in Appendix A.1. However, due to the lack of knowledge on $f_{X|\mathbf{S},\mathbf{z}}$, we are forced to perform all the calculations using a proposed $f_{X|\mathbf{S},\mathbf{z}}^*$. The fortunate fact is that even using the possibly misspecified conditional density, $(\mathbf{L}_\beta^{\mathrm{T}}, \mathbf{L}_\gamma^{\mathrm{T}}, \mathbf{L}_\alpha^{\mathrm{T}})^{\mathrm{T}}$ still has mean zero because this property is enforced by its very construction reflected on the definitions of $\mathbf{a}_\beta, \mathbf{a}_\gamma, \mathbf{a}_\alpha$ in (2.2). It is worth mentioning that if $f_{X|\mathbf{S},\mathbf{z}}$ happens to be the truth, then $\mathbf{S}_\beta, \mathbf{S}_\gamma, \mathbf{S}_\alpha$ are indeed the score functions. Thus, as the orthogonal projection of the score functions, $\mathbf{L}_\beta, \mathbf{L}_\gamma$ and $\mathbf{L}_\alpha$ are the efficient score functions. Hence the resulting estimator is not only consistent, but also efficient.

To further illustrate the estimator, we now investigate the partially linear single index model with normal measurement error. We will show that in this special case, many quantities simplify and a set of explicit estimating equations can be obtained.

Consider an alternative form of Model (1.1) in this case, where $Y = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta} + \theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z}) + \epsilon$, ϵ follows a normal distribution with mean zero, known constant variance σ^2 and is independent of $\mathbf{X}$. We adopt an additive normal measurement error $\mathbf{W} = \mathbf{X} + \mathbf{U}$, where $\mathbf{U}$ follows a normal distribution with mean zero and known constant covariance matrix $\boldsymbol{\Sigma}$ and is independent of $\mathbf{X}$. For estimating $\theta(\cdot)$, we adopt the familiar local linear form $\theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}) = \alpha_0 + \alpha_1\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}$.

Define $\boldsymbol{\Delta} = \mathbf{W} + Y\boldsymbol{\Sigma}\boldsymbol{\beta}/\sigma^2$. Following Stefanski and Carroll (1987), the forms of $\mathbf{L}_\beta$ is

$$\mathbf{L}_\beta(\mathbf{w}, \mathbf{z}, y; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta, f_{\mathbf{X}|\mathbf{Z}}^*) = \left\{ y - \frac{\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{\beta} + \theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z})}{1 + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\beta}/\sigma^2} \right\} E^*(\mathbf{X}|\boldsymbol{\delta}),$$

where E^* is computed under the model $f_{\mathbf{X}|\mathbf{Z}}^*(\mathbf{x}, \mathbf{z})$. Using similar derivation, we can further obtain

$$\mathbf{L}_\gamma(\mathbf{w}, \mathbf{z}, y; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta, f_{\mathbf{X}|\mathbf{Z}}^*) = \left\{ y - \frac{\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{\beta} + \theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z})}{1 + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\beta}/\sigma^2} \right\} \alpha_1\mathbf{z}_{-1},$$

$$\mathbf{L}_\alpha(\mathbf{w}, \mathbf{z}, y; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta, f_{\mathbf{X}|\mathbf{Z}}^*) = \left\{ y - \frac{\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{\beta} + \theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z})}{1 + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\beta}/\sigma^2} \right\} (1, \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z})^{\mathrm{T}}.$$

Then the estimation can be carried out through jointly solving

$$\mathbf{0} = \sum_{i=1}^{n} \left\{ y_i - \frac{\boldsymbol{\delta}_i^{\mathrm{T}}\boldsymbol{\beta} + \theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i)}{1 + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\beta}/\sigma^2} \right\} E^*(\mathbf{X}_i|\boldsymbol{\delta}_i),$$

$$\mathbf{0} = \sum_{i=1}^{n} \left\{ y_i - \frac{\boldsymbol{\delta}_i^{\mathrm{T}}\boldsymbol{\beta} + \theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i)}{1 + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\beta}/\sigma^2} \right\} \theta'(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i)\mathbf{z}_{-1,i}$$

to estimate $\boldsymbol{\beta}, \boldsymbol{\gamma}$ and

$$
\mathbf{0} = \sum_{i=1}^{n} K_h(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i - \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0)\left(y_i - \frac{\boldsymbol{\delta}_i^{\mathrm{T}}\boldsymbol{\beta} + \alpha_0 + \alpha_1\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i}{1 + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\beta}/\sigma^2}\right)(1, \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i)^{\mathrm{T}}
$$

at $\mathbf{z}_0 = \mathbf{z}_1, \ldots, \mathbf{z}_n$ to estimate $\widehat{\theta}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0) = \widehat{\alpha}_0 + \widehat{\alpha}_1\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0$.

   Similar calculations can also be made regarding the Poisson model $Y \sim$ Poisson$[\exp\{\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta} + \theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z})\}]$. In this case, $\mathbf{L}_\beta$ takes the form

$$
\mathbf{L}_\beta(\mathbf{w}, \mathbf{z}, y; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta, f_{\mathbf{X}|\mathbf{Z}}^*) = a(\mathbf{w}, \mathbf{z}, y; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta)E^*(\mathbf{X}|\boldsymbol{\delta}),
$$

where

$$
a(\mathbf{w}, \mathbf{z}, y; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta) = y - \frac{\sum_{y=0}^{\infty} y\exp[\{\boldsymbol{\delta}^{\mathrm{T}}\beta + \theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z})\}y - y^2\beta^{\mathrm{T}}\boldsymbol{\Sigma}\beta/2 - \log(y!)]}{\sum_{y=0}^{\infty}\exp[\{\boldsymbol{\delta}^{\mathrm{T}}\beta + \theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z})\}y - y^2\beta^{\mathrm{T}}\boldsymbol{\Sigma}\beta/2 - \log(y!)]},
$$

E^* is computed under the model $f_{\mathbf{X}|\mathbf{Z}}^*(\mathbf{x}, \mathbf{z})$. Using similar derivation, we can further obtain

$$
\begin{aligned}
\mathbf{L}_\gamma(\mathbf{w}, \mathbf{z}, y; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta, f_{\mathbf{X}|\mathbf{Z}}^*) &= a(\mathbf{w}, \mathbf{z}, y; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta)\alpha_1\mathbf{z}_{-1}, \\
\mathbf{L}_\alpha(\mathbf{w}, \mathbf{z}, y; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta, f_{\mathbf{X}|\mathbf{Z}}^*) &= a(\mathbf{w}, \mathbf{z}, y; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta)(1, \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z})^{\mathrm{T}}.
\end{aligned}
$$

Then the estimation can be carried out through jointly solving

$$
\begin{aligned}
\mathbf{0} &= \sum_{i=1}^{n} a(\mathbf{w}_i, \mathbf{z}_i, y_i; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta)E^*(\mathbf{X}_i|\boldsymbol{\delta}_i), &\quad (2.5) \\
\mathbf{0} &= \sum_{i=1}^{n} a(\mathbf{w}_i, \mathbf{z}_i, y_i; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta)\theta'(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i)\mathbf{z}_{-1,i}
\end{aligned}
$$

to estimate $\boldsymbol{\beta}, \boldsymbol{\gamma}$ and

$$
\mathbf{0} = \sum_{i=1}^{n} a(\mathbf{w}_i, \mathbf{z}_i, y_i; \boldsymbol{\beta}, \boldsymbol{\gamma}, \theta)(1, \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i)^{\mathrm{T}}
$$

at $\mathbf{z}_0 = \mathbf{z}_1, \ldots, \mathbf{z}_n$ to estimate $\widehat{\theta}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0) = \widehat{\alpha}_0 + \widehat{\alpha}_1\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0$.

3. Asymptotic properties and inference

In this section we show that the estimated covariate effect is asymptotically normal in Theorem 3.1 and locally efficient in Theorem 3.2. A by-product of the asymptotic normality property is that it facilitates testing if the estimated covariate effect is statistically significant.

   Viewing $\theta(\cdot)$ as a one dimensional parameter, we have $\mathbf{L}_\alpha = L_\theta\boldsymbol{\theta}_\alpha$, where L_θ is obtained the same way as $\mathbf{L}_\alpha$ by replacing $\boldsymbol{\alpha}$ with θ, and $\boldsymbol{\theta}_\alpha$ is the partial derivative of $\theta(\cdot, \boldsymbol{\alpha})$ with respect to $\boldsymbol{\alpha}$. Let $\boldsymbol{\theta}_{\alpha\alpha} = \partial\boldsymbol{\theta}_\alpha/\partial\boldsymbol{\alpha}^{\mathrm{T}}$. Let $\mathbf{L}_{\beta\beta}$, $\mathbf{L}_{\beta\gamma}$,

$\mathbf{L}_{\beta\alpha}$ and $\mathbf{L}_{\beta\theta}$ be the partial derivative of $\mathbf{L}_{\beta}$ with respect to $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ respectively. Similarly define $\mathbf{L}_{\gamma\beta}$, $\mathbf{L}_{\gamma\gamma}$, $\mathbf{L}_{\gamma\alpha}$, $\mathbf{L}_{\gamma\theta}$, $\mathbf{L}_{\alpha\beta}$, $\mathbf{L}_{\alpha\gamma}$, $\mathbf{L}_{\alpha\alpha}$ and $\mathbf{L}_{\alpha\theta}$. Let $\Omega(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z}) = E(L_{\theta\theta} \mid \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z})$, $\mathbf{U}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z}) = E\{(\mathbf{L}_{\beta\theta}^{\mathrm{T}} \quad \mathbf{L}_{\gamma\theta}^{\mathrm{T}})^{\mathrm{T}} \mid \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z}\}\Omega(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z})^{-1}$, and $\boldsymbol{\theta}_{\beta}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z}) = -\Omega(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z})^{-1}E(\mathbf{L}_{\theta\beta} \mid \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z})$, $\boldsymbol{\theta}_{\gamma}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z}) = -\Omega(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z})^{-1}E(\mathbf{L}_{\theta\gamma} \mid \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z})$. Define

$$
\begin{aligned}
\mathbf{A} &= E\left(\left[\begin{array}{cc} \mathbf{L}_{\beta\beta}\{Y,W,\mathbf{S},\mathbf{Z};\boldsymbol{\beta},\boldsymbol{\gamma},\theta(\cdot)\} & \mathbf{L}_{\beta\gamma}\{Y,W,\mathbf{S},\mathbf{Z};\boldsymbol{\beta},\boldsymbol{\gamma},\theta(\cdot)\} \\ \mathbf{L}_{\gamma\beta}\{Y,W,\mathbf{S},\mathbf{Z};\boldsymbol{\beta},\boldsymbol{\gamma},\theta(\cdot)\} & \mathbf{L}_{\gamma\gamma}\{Y,W,\mathbf{S},\mathbf{Z};\boldsymbol{\beta},\boldsymbol{\gamma},\theta(\cdot)\} \end{array}\right]\right) \\
&\quad + E\left(\left[\begin{array}{c} \mathbf{L}_{\beta\theta}\{Y,W,\mathbf{S},\mathbf{Z};\boldsymbol{\beta},\boldsymbol{\gamma},\theta(\cdot)\} \\ \mathbf{L}_{\gamma\theta}\{Y,W,\mathbf{S},\mathbf{Z};\boldsymbol{\beta},\boldsymbol{\gamma},\theta(\cdot)\} \end{array}\right]\{\boldsymbol{\theta}_{\beta}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z}) \quad \boldsymbol{\theta}_{\gamma}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z})\}\right). \quad (3.1)
\end{aligned}
$$

Theorem 3.1. *Under the regularity conditions listed in the Appendix, we have the expansion*

$$
\begin{aligned}
-\mathbf{A}n^{1/2}\left(\begin{array}{c} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} \end{array}\right) &= n^{-1/2}\sum_{i=1}^{n}\left(\left[\begin{array}{c} \mathbf{L}_{\beta}\{Y_i,W_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},\boldsymbol{\gamma},\theta(\cdot)\} \\ \mathbf{L}_{\gamma}\{Y_i,W_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},\boldsymbol{\gamma},\theta(\cdot)\} \end{array}\right]\right. \\
&\quad \left. -\mathbf{U}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z}_i)L_{\theta}\{Y_i,W_i,\mathbf{S}_i,\mathbf{Z}_i;\boldsymbol{\beta},\boldsymbol{\gamma},\theta(\cdot)\}\right) + o_p(1).
\end{aligned}
$$

Consequently, when $n \to \infty$,

$$
n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to N\{\mathbf{0}, (\mathbf{I}_{\beta} \quad \mathbf{0})\mathbf{A}^{-1}\mathbf{B}(\mathbf{A}^{\mathrm{T}})^{-1}(\mathbf{I}_{\beta} \quad \mathbf{0})^{\mathrm{T}}\}
$$

in distribution. Here, $\mathbf{I}_{\beta}$ *is the identity matrix with dimension being the length of* $\boldsymbol{\beta}$, *and* $\mathbf{B}$ *is equal to*

$$
\left(\left[\begin{array}{c} \mathbf{L}_{\beta}\{Y,W,\mathbf{S},\mathbf{Z};\boldsymbol{\beta},\boldsymbol{\gamma},\theta(\cdot)\} \\ \mathbf{L}_{\gamma}\{Y,W,\mathbf{S},\mathbf{Z};\boldsymbol{\beta},\boldsymbol{\gamma},\theta(\cdot)\} \end{array}\right] - \mathbf{U}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z})L_{\theta}\{Y,W,\mathbf{S},\mathbf{Z};\boldsymbol{\beta},\boldsymbol{\gamma},\theta(\cdot)\}\right). \quad (3.2)
$$

Theorem 3.2. *If the conjectured model* $f^*_{X|\mathbf{S},\mathbf{Z}}(x \mid \mathbf{s},\mathbf{z})$ *is correct, the subsequent estimator* $\widehat{\boldsymbol{\beta}}$ *has the additional property that it is semiparametric efficient.*

The proofs of Theorems 3.1 and 3.2 are given in the Appendix.

In practice, the matrices $\mathbf{A}$ and $\mathbf{B}$ can be estimated through their sample versions, while $\Omega, \mathbf{U}, \boldsymbol{\theta}_{\beta}$ and $\boldsymbol{\theta}_{\gamma}$ need to be estimated via their corresponding nonparametric regression.

Knowing the asymptotic properties of $\widehat{\boldsymbol{\beta}}$ allows us to perform various tests. Specifically, we can test the covariate effect described as $H_0 : \mathbf{M}\boldsymbol{\beta} = \mathbf{c}$, where $\mathbf{M}$ and $\mathbf{c}$ are the corresponding matrices or vectors used to describe the particular test of interest. As an example, we have the following Chi-square test result.

Theorem 3.3. *Under* H_0, *the test statistic*

$$
T = n(\mathbf{M}\widehat{\boldsymbol{\beta}} - \mathbf{c})^{\mathrm{T}}\{\mathbf{M}(\mathbf{I}_{\beta} \quad \mathbf{0})\widehat{\mathbf{A}}^{-1}\widehat{\mathbf{B}}(\widehat{\mathbf{A}}^{-1})^{\mathrm{T}}(\mathbf{I}_{\beta} \quad \mathbf{0})^{\mathrm{T}}\mathbf{M}^{\mathrm{T}}\}^{-1}(\mathbf{M}\widehat{\boldsymbol{\beta}} - \mathbf{c})
$$

follows a chi-square distribution with degrees of freedom d_M, *where* d_M *is the number of rows in* $\mathbf{M}$.

We provide the proof of Theorem 3.3 in Appendix A.5.

4. Simulation

We perform four simulation studies to examine the finite sample performance of the proposed method.

In the first set of simulation studies, the response variable Y is binary with $Y = 0$ or 1, with the true g function of the form

$$g\{y, x, \theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}), \boldsymbol{\beta}\} = \frac{\exp[y\{\beta_1 x + \beta_2 x^2 + \theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z})\}]}{1 + \exp\{\beta_1 x + \beta_2 x^2 + \theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z})\}}.$$

Thus, the parameter of interest $\boldsymbol{\beta} = (\beta_1, \beta_2)^{\mathrm{T}}$ consists of two components. The function $\theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}) = \cos(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z})/2 - 1$.

Our first simulation is a relatively simple one, where the covariate vector $\mathbf{Z}$ has dimension $p = 2$. This yields a total of three parameters in addition to the univariate nonparametric function θ and the unknown distribution of X. In simulations 2 and 3, we increase the dimension of the covariate vector $\mathbf{Z}$ to three and four respectively, which yield four and five parameters in addition to the two unknown functions. In all the simulations, the covariate X and the measurement errors are generated from normal distributions, and the covariate vector $\mathbf{Z}$ is generated from uniform distributions.

To compare the performance of various estimators, we implemented a naive estimator, two versions of the regression calibration estimators and two versions of the semiparametric estimators. In the naive estimator, the presence of measurement error is simply ignored and a profile likelihood estimation procedure is implemented to estimate the parameter $\boldsymbol{\beta}$. In the regression calibration procedures, we first calculate $X^* = E(X \mid W)$ and $X^{*2} = E(X^2 \mid W)$ then treat X^* and X^{*2} as X and X^2, and perform the profile likelihood estimation under the error-free model. In calculating $E(X \mid W)$ and $E(X^2 \mid W)$, we experimented with two situations, where we used two different working distributions of X, respectively normal and uniform. This corresponds to the true and misspecified distributional assumption on X. Finally, we also implemented the proposed semiparametric estimator, with the same working distributions of X. The estimation and inference results of all five estimators are given in Tables 1-3 respectively, corresponding to the three simulation studies. All the results are based on 1,000 simulated data sets with sample size 500. To see how the estimation procedure behaves with increasing dimension of $\mathbf{Z}$, we also experimented with $p > 4$. In our observation, with all other aspects of the simulation fixed, the procedure performs well until $p = 10$, when we started to see significant biases. Throughout the numerical analysis, we used the bandwidth $h = 3\mathrm{sd}(w)n^{-1/3}$, where $\mathrm{sd}(w)$ is the sample standard deviation of w. We also experimented with the bandwidth $h = 1.5\mathrm{sd}(w)n^{-1/3}$ and $h = 4.5\mathrm{sd}(w)n^{-1/3}$, the results appear insensitive to the bandwidth changes so are omitted.

The common observation across all simulations is that the naive estimator and the two regression calibration estimators tend to produce larger biases while the semiparametric estimators, whether performed under the true or misspecified working model of the distribution of X, have much smaller biases. The

TABLE 1
Results of Simulation 1 with $p = 2$. The true parameter values, the estimates ($\widehat{\mu}$), the sample standard errors ("sd"), the mean of the estimated standard errors ($\widehat{sd}$) and the 95% confidence interval ("%") of five different estimators are reported. The five estimators are the naive estimator ("Naive"), the regression calibration estimators with two working distributions of X ("RC-nor" and "RC-Uni") and the semiparametric estimators with two working distributions of X ("Semi-nor" and "Semi-Uni").

	Naive		RC-nor		RC-Uni		Semi-nor		Semi-Uni	
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
true	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700
mean	0.518	0.567	0.675	0.673	1.064	0.845	0.706	0.712	0.720	0.719
sd	0.180	0.092	0.212	0.110	0.270	0.134	0.255	0.138	0.243	0.133
$\widehat{sd}$	0.175	0.089	0.205	0.105	0.262	0.129	0.251	0.135	0.238	0.130
CI	79.5	62.2	93.4	93.3	71.2	79.8	94.7	94.4	94.5	95.3

TABLE 2
Results of Simulation 2 with $p = 3$. The true parameter values, the estimates ($\widehat{\mu}$), the sample standard errors ("sd"), the mean of the estimated standard errors ($\widehat{sd}$) and the 95% confidence interval ("%") of five different estimators are reported.

	Naive		RC-nor		RC-Uni		Semi-nor		Semi-Uni	
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
true	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700
mean	0.508	0.564	0.662	0.669	1.049	0.841	0.692	0.708	0.705	0.715
sd	0.177	0.090	0.207	0.107	0.269	0.132	0.243	0.132	0.233	0.128
$\widehat{sd}$	0.174	0.088	0.204	0.105	0.261	0.129	0.254	0.137	0.2414	0.1344
CI	78.7	62.8	94.3	93.4	72.4	81.1	95.8	95.6	96.0	96.0

TABLE 3
Results of Simulation 3 with $p = 4$. The true parameter values, the estimates ($\widehat{\mu}$), the sample standard errors ("sd"), the mean of the estimated standard errors ($\widehat{sd}$) and the 95% confidence interval ("%") of five different estimators are reported.

	Naive		RC-nor		RC-Uni		Semi-nor		Semi-Uni	
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
true	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700	0.700
mean	0.503	0.561	0.656	0.666	1.044	0.838	0.686	0.704	0.700	0.711
sd	0.183	0.094	0.215	0.112	0.275	0.136	0.257	0.142	0.248	0.138
$\widehat{sd}$	0.174	0.089	0.203	0.105	0.262	0.128	0.262	0.139	0.246	0.135
CI	77.4	60.9	92.4	91.5	73.6	80.6	94.6	94.0	94.9	94.7

relatively large biases of the naive and regression calibration estimators directly lead to invalid inference results, reflected in the terrible empirical coverage of the 95% confidence intervals. On the contrary, the semiparametric estimators not only yield very small biases, it also provides a close match between the sample standard deviations and their corresponding asymptotic versions. This leads to reasonable approximation of the empirical coverage of the 95% confidence intervals to the nominal level. It is worth pointing out that although we implemented an efficient estimator through adopting the working model for X as normal, and a non-efficient estimator through using uniform as the working model for X, the estimation variability of the two estimators are very close. In other words, the method appears to have certain robustness to the working model, in that in addition to retaining consistency as our theory has promised, it

also seems to remain efficient regardless of the working model. The latter property is not within our expectation and whether this is a universal phenomenon with theoretical explanation deserves further investigation.

To further illustrate the generality of the results derived in this paper, we perform a fourth set of simulation studies concerning a Poisson model. We generate the counting response variable Y with mean $\exp\{\beta x + \theta(\widetilde{\gamma}^{\mathrm{T}}\mathbf{z})\}$ and generate X from $N(0, 1.1^2)$. We set $\beta = 1.1$, $\theta(\widetilde{\gamma}^{\mathrm{T}}\mathbf{z}) = -0.4\cos(2.75\widetilde{\gamma}^{\mathrm{T}}\mathbf{z} - 1.0)$ and allow substantial measurement error $\sigma_u = 0.8$. Following (2.5), we directly posit $E^*(X \mid \delta) = \delta^2$ and $E^*(X \mid \delta) = \delta\sin(\delta)$ for $E(X \mid \delta)$. We experimented with various dimension of $\mathbf{z}$ from 2 to 11 where $\mathbf{z}$ contain both continuous and discrete. Simulations results are summarized in Table 4. The consistency of our estimator, regardless if the posited models are correct or not, as well as the superioty of our method in contrast with the comparison methods are clear from these results.

5. Framingham heart study

We use our new methodology to analyse data from the Framingham Heart Study described in Section 1. The data set contains 1,126 male subjects. We use the occurrence of coronary heart disease as the response variable (Y), and systolic blood pressure, after subtracting 50 and taking logarithm transformation, as the covariate measured with error (W), see Carroll et al. (2006) who used this transformation, so that $W = X + U$, where X is the transformed true systolic blood pressure. We included age, the logarithm of $1 +$ the number of cigarettes smoked per day as reported by the subject and metropolitan relative weight as confounders $\mathbf{Z}$, with age chosen to be the leading component in $\mathbf{Z}$. Metropolitan relative weight is defined as the percentage of desirable weight (the ratio of actual weight to desirable weight times 100). Desirable weight was derived from the 1959 Metropolitan Life Insurance Company tables (Metropolitan Life Insurance Company, 1959) by taking the midpoint of the weight range for the medium build at a specified height, see also Hubert et al. (1983).

We fit the model with systolic blood pressure in its original scale. With $H(\cdot)$ being the logistic distribution function, the final model is

$$
\begin{aligned}
\mathrm{pr}(Y = 1 \mid X, \mathbf{Z}) &= H\left[\{\exp(X) + 50\}\,\beta + \theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z})\right], \\
W &= X + U.
\end{aligned}
$$

Using the available repeated measurements of W, we obtained the measurement error standard deviation to be 0.0745, and the Kolmogorov-Smirnov test for the normality of U yields a p-value of 0.701. We also include the qq-plot of the errors in Figure 1, which exhibits a linear pattern. Thus, we assume U has the centered normal distribution with standard deviation 0.0745.

The semiparametric analysis of the Framingham data, as well as the results from naive estimator and regression calibration estimators are given in Table 5. Not unexpectedly given the context, all results confirm the significance of

Results of simulation 4 with $p = 2$ to 11. The true parameter is $\beta = 1.1$. The the estimates("est"), the sample standard errors ("sd"), the mean of the estimated standard errors ($\widehat{sd}$) and the 95% confidence interval ("%") of six different estimators are reported. The six estimators are the naive estimator ("Naive"), the regression calibration estimators with two working distributions of X ("RC-Nor" and "RC-Uni"), the oracle estimator ("Oracle"), and the local estimators with two posited forms of $E(X \mid \delta)$ ("Local 1" and "Local 2").

|		Naive	RC-Nor	RC-Uni	Oracle	Local 1	Local 2
$p = 2$	est	0.773	1.115	1.198	1.113	1.092	1.107
	sd	0.163	0.190	0.226	0.116	0.110	0.106
	$\widehat{sd}$	0.078	0.120	0.149	0.166	0.174	0.182
	CI	22.0	76.2	71.6	95.6	94.0	93.8
$p = 3$	est	0.809	1.162	1.248	1.104	1.109	1.102
	sd	0.120	0.111	0.151	0.160	0.138	0.093
	$\widehat{sd}$	0.085	0.123	0.158	0.170	0.168	0.132
	CI	30.8	92.8	81.6	95.2	95.6	95.8
$p = 4$	est	0.819	1.185	1.273	1.105	1.106	1.091
	sd	0.110	0.115	0.142	0.128	0.088	0.114
	$\widehat{sd}$	0.079	0.112	0.134	0.155	0.140	0.130
	CI	26.2	89.2	74.0	97.4	97.4	94.0
$p = 5$	est	0.729	1.149	1.221	1.103	1.102	1.104
	sd	0.141	0.170	0.202	0.159	0.076	0.098
	$\widehat{sd}$	0.082	0.142	0.160	0.190	0.143	0.159
	CI	17.6	84.6	77.8	95.8	96.0	96.4
$p = 6$	est	0.827	1.170	1.261	1.122	1.080	1.102
	sd	0.125	0.125	0.158	0.186	0.142	0.098
	$\widehat{sd}$	0.081	0.127	0.153	0.168	0.168	0.199
	CI	31.4	87.6	76.8	96.2	91.2	96.4
$p = 7$	est	0.808	1.159	1.243	1.128	1.098	1.100
	sd	0.1332	0.150	0.182	0.218	0.126	0.089
	$\widehat{sd}$	0.083	0.126	0.148	0.195	0.196	0.228
	CI	29.0	83.4	76.6	95.4	93.6	97.8
$p = 8$	est	0.810	1.166	1.253	1.126	1.098	1.099
	sd	0.122	0.136	0.174	0.206	0.139	0.089
	$\widehat{sd}$	0.089	0.129	0.145	0.178	0.193	0.191
	CI	29.2	84.0	74.4	95.6	94.2	97.8
$p = 9$	est	0.808	1.153	1.231	1.133	1.094	1.102
	sd	0.139	0.149	0.180	0.231	0.151	0.073
	$\widehat{sd}$	0.085	0.125	0.146	0.196	0.191	0.224
	CI	31.2	82.6	77.4	95.8	92.2	98.0
$p = 10$	est	0.801	1.157	1.243	1.122	1.104	1.099
	sd	0.129	0.151	0.178	0.188	0.154	0.081
	$\widehat{sd}$	0.085	0.126	0.148	0.198	0.207	0.218
	CI	28.6	85.6	78.0	96.2	94.2	98.2
$p = 11$	est	0.796	1.146	1.231	1.121	1.092	1.102
	sd	0.136	0.150	0.178	0.211	0.136	0.092
	$\widehat{sd}$	0.086	0.131	0.146	0.210	0.198	0.248
	CI	29.0	86.8	77.4	95.8	91.6	98.2
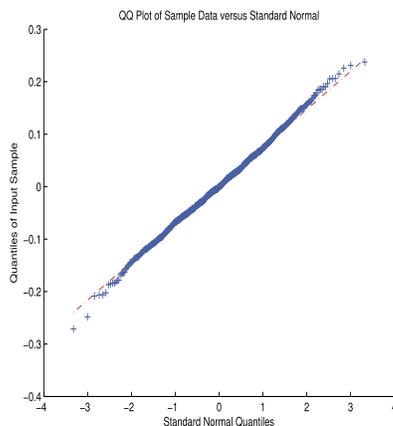
FIG 1. *QQ-plot of the measurement errors in Framingham data analysis.*

TABLE 5

Results of Framingham data analysis. The estimates $(\widehat{\mu})$ and the associated standard errors of five different estimators are reported. All values are multiplied by 100. In the table, $\widehat{\beta}_1$ is the regression coefficient for systolic blood pressure, $\widehat{\gamma}_1$ is the coefficient for transformed number of cigarettes smoked per day and $\widehat{\gamma}_2$ is the coefficient for metropolitan weight.

	$\widehat{\beta}_1$	sd($\widehat{\beta}_1$)	$\widehat{\gamma}_1$	sd($\widehat{\gamma}_1$)	$\widehat{\gamma}_2$	sd($\widehat{\gamma}_2$)
Naive	3.58	0.49	0.09	6.00	0.31	4.79
RC-Nor	4.22	0.60	0.11	5.99	0.25	4.77
RC-Uni	3.73	0.58	0.29	5.81	0.73	4.85
Semi-Nor	4.39	0.77	0.10	3.90	0.24	2.17
Semi-Uni	4.61	0.80	0.10	4.62	0.25	2.12

the systolic blood pressure as a risk factor for heart disease. In addition, the two estimates from the two semiparametric methods, conducted under a normal and a uniform working model for the distribution of X respectively, are very close. The naive estimator is attenuated towards zero by approximately 25%. Neither the effects from number of cigarettes smoked nor metropolitan weight is statistically significant. We also plot the estimated $\theta(\widehat{\gamma}^{\mathrm{T}}\mathbf{z})$ as a function of $\widehat{\gamma}^{\mathrm{T}}\mathbf{z}$, as well as the 95% pointwise confidence bands in Figure 2 from both semiparametric methods, and we can see a general trend of increasing risk with increasing age.

6. Discussion

We have developed both estimation and inference tools to analyse covariate effect when the covariate under study is measured with error and also subject to confounding effects. The method is completely general, reflected in the generality of the main regression model. Specifically, we allow arbitrary regression relation between the response variable and the covariate under study, and we do not require a specific parametric model strategy for the confounding effects.
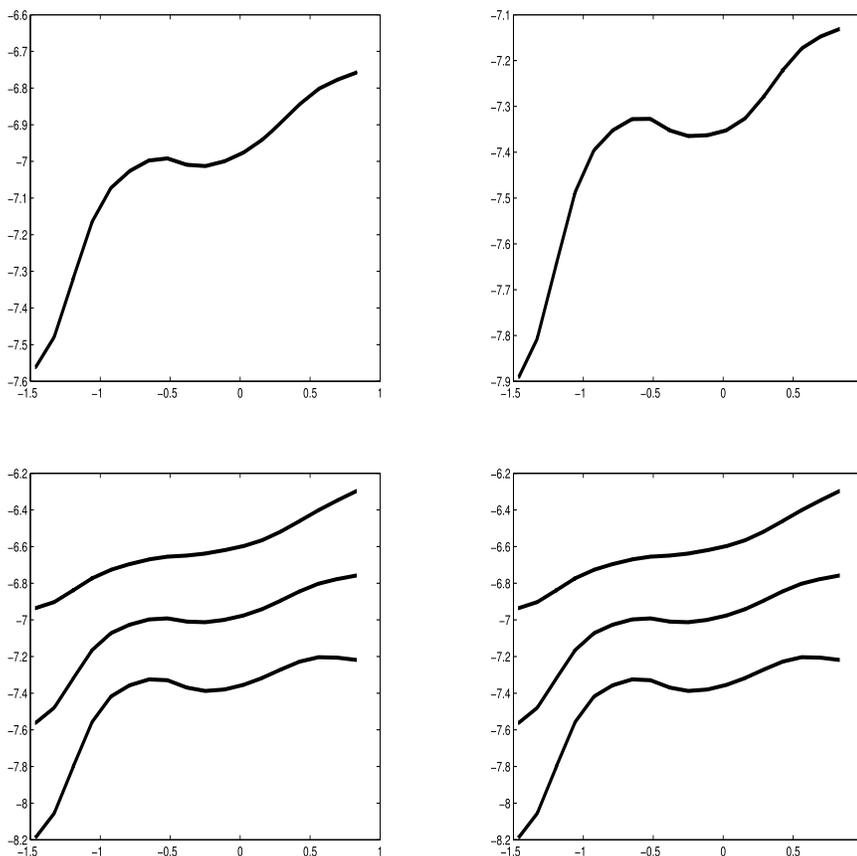
FIG 2. *The estimated $\theta(\widehat{\gamma}^{\mathrm{T}}\mathbf{z})$ as a function of $(\widehat{\gamma}^{\mathrm{T}}\mathbf{z})$ in Framingham data analysis. Vertical axis stands for $\theta(\widehat{\gamma}^{\mathrm{T}}\mathbf{z})$ and horizontal axis stands for $(\widehat{\gamma}^{\mathrm{T}}\mathbf{z})$. In the left panels, $\widehat{\gamma}$ is obtained with a normal working model on X and in the right panels $\widehat{\gamma}$ is obtained with uniform working model on X. The plots in the lower panels contain the 95% confidence bands.*

Our procedure does not require any model assumption on the unobservable covariate of interest, and the framework can allow arbitrary measurement error structure. Under the special situation, when the regression model has a generalized partially linear form, and the measurement error is normal additive, great simplification occurs (Ma and Tsiatis, 2006) and the estimation procedure degenerates to a backfitted or profiled version of the estimator given in Stefanski and Carroll (1987).

We would like to point out that to solve the estimating equations, one could choose to use backfitting or profiling procedures. In our construction of the estimator, these are only two ways of solving the estimating equations jointly. Upon convergence, the solutions from backfitting and profiling are identical. They are both roots of the estimating equations. This is very different from

using backfitting versus profiling before estimating equations are derived, where using profiling or backfitting could result in different sets of estimating equations and hence both the theoretical and empirical performance can be different. The latter issue is well studied in Van Keilegom and Carroll (2007). LIkewise, the nonparametric estimation of $\theta(\cdot)$ can also be carried out via splines, wavelets, etc., and research along these lines are certainly needed.

Appendix A: Appendix

A.1. Calculation of the projection of $(\mathbf{S}_\beta^\mathbf{T}, \mathbf{S}_\gamma^\mathbf{T}, \mathbf{S}_\alpha^\mathbf{T})^\mathbf{T}$

Replacing $\theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z})$ with $\nu(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}; \boldsymbol{\alpha})$, the conditional model of Y on $(\mathbf{X}, \mathbf{S}, \mathbf{Z})$ in (1.1) is a fully parametric model. Following Tsiatis and Ma (2004), we know that the nuisance tangent space Λ and its orthogonal complement $\Lambda^\perp$ are respectively

$$
\begin{aligned}
\Lambda &= [E\{\mathbf{f}(X, \mathbf{s}, \mathbf{z}) \mid w, \mathbf{s}, \mathbf{z}, y\} : E\{\mathbf{f}(X, \mathbf{s}, \mathbf{z}) \mid \mathbf{s}, \mathbf{z}\} = \mathbf{0}], \\
\Lambda^\perp &= [\mathbf{f}(w, \mathbf{s}, \mathbf{z}, y) : E\{\mathbf{f}(W, \mathbf{s}, \mathbf{z}, Y) \mid x, \mathbf{s}, \mathbf{z}\} = \mathbf{0}].
\end{aligned}
$$

We can then easily verify from the definition of $\mathbf{L}_\beta, \mathbf{L}_\gamma, \mathbf{L}_\alpha$ that $\{\mathbf{L}_\beta^{\mathrm{T}}(w, \mathbf{s}, \mathbf{z}, y; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, f_{X|\mathbf{S},\mathbf{Z}}), \mathbf{L}_\gamma^{\mathrm{T}}(w, \mathbf{s}, \mathbf{z}, y; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, f_{X|\mathbf{S},\mathbf{Z}}),$ $\mathbf{L}_\alpha^{\mathrm{T}}(w, \mathbf{s}, \mathbf{z}, y; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, f_{X|\mathbf{S},\mathbf{Z}})\}^{\mathrm{T}}$ is an element of $\Lambda^\perp$ and $[E\{\mathbf{a}_\beta^{\mathrm{T}}(X, \mathbf{s}, \mathbf{z}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}) \mid w, \mathbf{s}, \mathbf{z}, y\}, E\{\mathbf{a}_\gamma^{\mathrm{T}}(X, \mathbf{s}, \mathbf{z}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}) \mid w, \mathbf{s}, \mathbf{z}, y\}, E\{\mathbf{a}_\alpha^{\mathrm{T}}(X, \mathbf{s}, \mathbf{z}; \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}) \mid w, \mathbf{s}, \mathbf{z}, y\}]^{\mathrm{T}}$ is an element of Λ. Equivalently, the projection of $(\mathbf{S}_\beta^{\mathrm{T}}, \mathbf{S}_\gamma^{\mathrm{T}}, \mathbf{S}_\alpha^{\mathrm{T}})^{\mathrm{T}}$ is indeed $(\mathbf{L}_\beta^{\mathrm{T}}, \mathbf{L}_\gamma^{\mathrm{T}}, \mathbf{L}_\alpha^{\mathrm{T}})^{\mathrm{T}}$.

A.2. List of regularity conditions

1. The function $\theta(\cdot)$ is twice differentiable and its second derivative is Lipschitz-continuous.
2. The density function of $\mathbf{Z}$ has a compact support and is positive on the support.
3. The matrix $\mathbf{A}$ and $\mathbf{B}$ defined in (3.1) and (3.2) are non-singular and their elements are bounded away from infinity.
4. The kernel function $K(\cdot)$ has compact support, is bounded on its support, and satisfies $\int K(x)dx = 1$, $\int xK(x)dx = 0$ and $\int x^2 K(x)dx > 0$.
5. The bandwidth $h = O(n^{-r})$ for $1/8 < r < 1/2$.

Condition 1 is a standard smoothness requirement on $\theta(\cdot)$ required for general nonparametric smoothing methods. Condition 2 requires the distribution of $\mathbf{Z}$ to have some properties to avoid technical issues such as dividing by zero. This requirement can be slightly relaxed at the price of more tedious technical treatment. Condition 3 ensures that the estimators of the parameters do not degenerate. Condition 4 rquires the kernel function to be the usual compactly supported second order kenel. Condition 5 states the bandwidth reuirement and illustrates that the method does not require under smoothing.

A.3. Proof of Theorem 3.1

For notational simplicity, we define $\boldsymbol{\zeta} = (\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\gamma}^{\mathrm{T}})^{\mathrm{T}}$, $\mathbf{L}_{\zeta} = (\mathbf{L}_{\beta}^{\mathrm{T}}, \mathbf{L}_{\gamma}^{\mathrm{T}})^{\mathrm{T}}$, $\mathbf{L}_{\zeta\zeta} = \partial\mathbf{L}_{\zeta}/\partial\boldsymbol{\zeta}^{\mathrm{T}}$, $\mathbf{L}_{\zeta\alpha} = \partial\mathbf{L}_{\zeta}/\partial\boldsymbol{\alpha}^{\mathrm{T}}$. Let $\boldsymbol{\theta}_{\zeta}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z}) = \{\boldsymbol{\theta}_{\beta}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z}) \ \ \boldsymbol{\theta}_{\gamma}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z})\}$. When solving for $\boldsymbol{\alpha}$ in (2.4), we have

$$\mathbf{0} = n^{-1/2}\sum_{i=1}^{n}K_h(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i - \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0)\mathbf{L}_{\alpha}\{y_i, w_i, \mathbf{s}_i, \mathbf{z}_i; \boldsymbol{\zeta}, \widehat{\boldsymbol{\alpha}}(\boldsymbol{\zeta})\}$$

at any $\boldsymbol{\zeta}$, therefore

$$\mathbf{0} = n^{-1}\sum_{i=1}^{n}K_h(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i - \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0)\mathbf{L}_{\alpha\zeta}\{y_i, w_i, \mathbf{s}_i, \mathbf{z}_i; \boldsymbol{\zeta}, \widehat{\boldsymbol{\alpha}}(\boldsymbol{\zeta})\}$$

$$+n^{-1}\sum_{i=1}^{n}K_h(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i - \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0)\mathbf{L}_{\alpha\alpha}\{y_i, w_i, \mathbf{s}_i, \mathbf{z}_i; \boldsymbol{\zeta}, \widehat{\boldsymbol{\alpha}}(\boldsymbol{\zeta})\}\partial\widehat{\boldsymbol{\alpha}}(\boldsymbol{\zeta})/\partial\boldsymbol{\zeta}^{\mathrm{T}}$$

$$+n^{-1}\sum_{i=1}^{n}K_h(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i - \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0)\mathbf{L}_{\alpha}\{y_i, w_i, \mathbf{s}_i, \mathbf{z}_i; \boldsymbol{\zeta}, \widehat{\boldsymbol{\alpha}}(\boldsymbol{\zeta})\}$$

$$\left[\mathbf{0}_{\beta}^{\mathrm{T}} \quad \frac{K'\{\mathbf{e}_1^{\mathrm{T}}\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}(\mathbf{z}_i - \mathbf{z}_0)/h\}(\mathbf{z}_i - \mathbf{z}_0)^{\mathrm{T}}}{K\{\mathbf{e}_1^{\mathrm{T}}\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}(\mathbf{z}_i - \mathbf{z}_0)/h\}h} \quad \cdots \quad \frac{K'\{\mathbf{e}_d^{\mathrm{T}}\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}(\mathbf{z}_i - \mathbf{z}_0)/h\}(\mathbf{z}_i - \mathbf{z}_0)^{\mathrm{T}}}{K\{\mathbf{e}_d^{\mathrm{T}}\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}(\mathbf{z}_i - \mathbf{z}_0)/h\}h}\right],$$

where $\mathbf{0}_{\beta}$ is a zero vector with the same length as $\boldsymbol{\beta}$. Note also that

$$\mathbf{L}_{\alpha\alpha}\{y_i, w_i, \mathbf{s}_i, \mathbf{z}_i; \boldsymbol{\zeta}, \widehat{\boldsymbol{\alpha}}(\boldsymbol{\zeta})\}$$
$$= \mathbf{L}_{\theta\theta}\{y_i, w_i, \mathbf{s}_i, \mathbf{z}_i; \boldsymbol{\zeta}, \widehat{\boldsymbol{\alpha}}(\boldsymbol{\zeta})\}\boldsymbol{\theta}_{\alpha}\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i; \widehat{\boldsymbol{\alpha}}(\boldsymbol{\zeta})\}\boldsymbol{\theta}_{\alpha}^{\mathrm{T}}\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i; \widehat{\boldsymbol{\alpha}}(\boldsymbol{\zeta})\}$$
$$+\mathbf{L}_{\theta}\{y_i, w_i, \mathbf{s}_i, \mathbf{z}_i; \boldsymbol{\zeta}, \widehat{\boldsymbol{\alpha}}(\boldsymbol{\zeta})\}\boldsymbol{\theta}_{\alpha\alpha}\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i; \widehat{\boldsymbol{\alpha}}(\boldsymbol{\zeta})\}.$$

Taking into account that $E\{\mathbf{L}_{\theta}(Y, W, \mathbf{S}, \mathbf{Z}; \boldsymbol{\zeta}, \boldsymbol{\alpha}) \mid \mathbf{z}\} = \mathbf{0}$, this yields

$$\boldsymbol{\theta}_{\alpha}\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0, \widehat{\boldsymbol{\alpha}}(\boldsymbol{\zeta})\}^{\mathrm{T}}\frac{\partial\widehat{\boldsymbol{\alpha}}(\boldsymbol{\zeta})}{\partial\boldsymbol{\zeta}^{\mathrm{T}}}$$

$$= -\{\boldsymbol{\theta}_{\alpha}^{\mathrm{T}}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0, \boldsymbol{\alpha})\boldsymbol{\theta}_{\alpha}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0, \boldsymbol{\alpha})\}^{-1}\left[E\{\mathbf{L}_{\theta\theta}(Y_i, W_i, \mathbf{S}_i, \mathbf{Z}_i) \mid \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z}_i = \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0\}\right]^{-1}$$
$$\boldsymbol{\theta}_{\alpha}^{\mathrm{T}}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0, \boldsymbol{\alpha})E\{\mathbf{L}_{\alpha\zeta}(Y_i, W_i, \mathbf{S}_i, \mathbf{Z}_i) \mid \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z}_i = \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0\} + o_p(1)$$
$$= -\{\boldsymbol{\theta}_{\alpha}^{\mathrm{T}}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0, \boldsymbol{\alpha})\boldsymbol{\theta}_{\alpha}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0, \boldsymbol{\alpha})\}^{-1}\Omega(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0)^{-1}$$
$$\boldsymbol{\theta}_{\alpha}^{\mathrm{T}}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0, \boldsymbol{\alpha})\boldsymbol{\theta}_{\alpha}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0, \boldsymbol{\alpha})E\{\mathbf{L}_{\theta\zeta}(Y_i, W_i, \mathbf{S}_i, \mathbf{Z}_i) \mid \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0\} + o_p(1)$$
$$= \boldsymbol{\theta}_{\zeta}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0) + o_p(1).$$

Now we expand (2.3) and obtain

$$\mathbf{0} = n^{-1/2}\sum_{i=1}^{n}\mathbf{L}_{\zeta}\{y_i, w_i, \mathbf{s}_i, \mathbf{z}_i; \boldsymbol{\zeta}, \widehat{\boldsymbol{\alpha}}(\boldsymbol{\zeta})\} + n^{-1}\sum_{i=1}^{n}[\mathbf{L}_{\zeta\zeta}\{y_i, w_i, \mathbf{s}_i, \mathbf{z}_i; \boldsymbol{\zeta}, \widehat{\boldsymbol{\alpha}}(\boldsymbol{\zeta})\}$$

$$+\mathbf{L}_{\zeta\theta}\{y_i, w_i, \mathbf{s}_i, \mathbf{z}_i; \boldsymbol{\zeta}, \widehat{\boldsymbol{\alpha}}(\boldsymbol{\zeta})\}\boldsymbol{\theta}_\alpha^{\mathrm{T}}\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i; \widehat{\boldsymbol{\alpha}}(\boldsymbol{\zeta})\}\frac{\partial\widehat{\boldsymbol{\alpha}}(\boldsymbol{\zeta})}{\partial\boldsymbol{\zeta}^{\mathrm{T}}}\Bigg] n^{1/2}(\widehat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}) + o_p(1)$$

$$= n^{-1/2}\sum_{i=1}^{n}\mathbf{L}_\zeta\{y_i, w_i, \mathbf{s}_i, \mathbf{z}_i; \boldsymbol{\zeta}, \widehat{\boldsymbol{\alpha}}(\boldsymbol{\zeta})\} + \mathbf{A}n^{1/2}(\widehat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}) + o_p(1)$$

$$= \mathbf{A}n^{1/2}(\widehat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}) + n^{-1/2}\sum_{i=1}^{n}\mathbf{L}_\zeta\{y_i, w_i, \mathbf{s}_i, \mathbf{z}_i; \boldsymbol{\zeta}, \theta(\cdot)\}$$

$$+n^{-1/2}\sum_{i=1}^{n}[\mathbf{L}_\zeta\{y_i, w_i, \mathbf{s}_i, \mathbf{z}_i; \boldsymbol{\zeta}, \widehat{\boldsymbol{\alpha}}(\boldsymbol{\zeta})\} - \mathbf{L}_\zeta\{y_i, w_i, \mathbf{s}_i, \mathbf{z}_i; \boldsymbol{\zeta}, \theta(\cdot)\}] + o_p(1).$$

$$\text{(A.1)}$$

From (2.4), we also have

$$\mathbf{0} = n^{-1/2}\sum_{i=1}^{n}K_h(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i - \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0)\mathbf{L}_\alpha\{y_i, w_i, \mathbf{s}_i, \mathbf{z}_i; \boldsymbol{\zeta}, \boldsymbol{\alpha}(\boldsymbol{\zeta})\}$$

$$+n^{-1/2}\sum_{i=1}^{n}K_h(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i - \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0)\mathbf{L}_{\alpha\theta}[y_i, w_i, \mathbf{s}_i, \mathbf{z}_i; \boldsymbol{\zeta}, \theta\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i; \boldsymbol{\alpha}(\boldsymbol{\zeta})\}]$$

$$\boldsymbol{\theta}_\alpha^{\mathrm{T}}\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i; \boldsymbol{\alpha}(\boldsymbol{\zeta})\}\{\widehat{\boldsymbol{\alpha}}(\boldsymbol{\zeta}) - \boldsymbol{\alpha}(\boldsymbol{\zeta})\} + o_p(1)$$

$$= n^{-1/2}\sum_{i=1}^{n}K_h(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i - \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0) < \mathbf{L}_\alpha\{y_i, w_i, \mathbf{s}_i, \mathbf{z}_i; \boldsymbol{\zeta}, \boldsymbol{\alpha}(\boldsymbol{\zeta})\}$$

$$+n^{1/2}E\left(\mathbf{L}_{\theta\theta}[Y, W, \mathbf{S}, \mathbf{Z}; \boldsymbol{\zeta}, \theta\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z}; \boldsymbol{\alpha}(\boldsymbol{\zeta})\}] \mid \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0\right)$$

$$\times\boldsymbol{\theta}_\alpha\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0; \boldsymbol{\alpha}(\boldsymbol{\zeta})\}\boldsymbol{\theta}_\alpha^{\mathrm{T}}\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0; \boldsymbol{\alpha}(\boldsymbol{\zeta})\}f_{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z}}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_0)\{\widehat{\boldsymbol{\alpha}}(\boldsymbol{\zeta}) - \boldsymbol{\alpha}(\boldsymbol{\zeta})\} + o_p(1),$$

hence

$$\boldsymbol{\theta}_\alpha^{\mathrm{T}}\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i; \boldsymbol{\alpha}(\boldsymbol{\zeta})\}n^{1/2}\{\widehat{\boldsymbol{\alpha}}(\boldsymbol{\zeta}) - \boldsymbol{\alpha}(\boldsymbol{\zeta})\}$$

$$= -\left(E\left[\mathbf{L}_{\theta\theta}\{Y, W, \mathbf{S}, \mathbf{Z}; \boldsymbol{\zeta}, \theta\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z})\} \mid \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i\right]\right)^{-1}$$

$$[\boldsymbol{\theta}_\alpha^{\mathrm{T}}\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i; \boldsymbol{\alpha}(\boldsymbol{\zeta})\}\boldsymbol{\theta}_\alpha\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i; \boldsymbol{\alpha}(\boldsymbol{\zeta})\}]^{-1}$$

$$n^{-1/2}\sum_{j=1}^{n}f_{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z}}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i)^{-1}\boldsymbol{\theta}_\alpha^{\mathrm{T}}\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i; \boldsymbol{\alpha}(\boldsymbol{\zeta})\}K_h(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_j - \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i)$$

$$\mathbf{L}_\alpha\{y_j, w_j, \mathbf{s}_j, \mathbf{z}_j; \boldsymbol{\zeta}, \boldsymbol{\alpha}(\boldsymbol{\zeta})\} + o_p(1)$$

$$= -\Omega^{-1}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i)f_{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z}}^{-1}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i)[\boldsymbol{\theta}_\alpha^{\mathrm{T}}\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i; \boldsymbol{\alpha}(\boldsymbol{\zeta})\}\boldsymbol{\theta}_\alpha\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i; \boldsymbol{\alpha}(\boldsymbol{\zeta})\}]^{-1}\boldsymbol{\theta}_\alpha^{\mathrm{T}}\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i; \boldsymbol{\alpha}(\boldsymbol{\zeta})\}$$

$$n^{-1/2}\sum_{j=1}^{n}K_h(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_j - \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i)\mathbf{L}_\alpha\{y_j, w_j, \mathbf{s}_j, \mathbf{z}_j; \boldsymbol{\zeta}, \boldsymbol{\alpha}(\boldsymbol{\zeta})\} + o_p(1).$$

Incorporating the above, we have

$$n^{-1/2}\sum_{i=1}^{n}\left(\mathbf{L}_\zeta[y_i, w_i, \mathbf{s}_i, \mathbf{z}_i; \boldsymbol{\zeta}, \theta\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i; \widehat{\boldsymbol{\alpha}}(\boldsymbol{\zeta})\}] - \mathbf{L}_\zeta\{y_i, w_i, \mathbf{s}_i, \mathbf{z}_i; \boldsymbol{\zeta}, \theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{z}_i)\}\right)$$

$$= n^{-1/2} \sum_{i=1}^{n} \mathbf{L}_{\zeta\theta}\{y_i, w_i, \mathbf{s}_i, \mathbf{z}_i; \boldsymbol{\zeta}, \theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_i)\} \boldsymbol{\theta}_{\alpha}^{\mathrm{T}}\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_i; \boldsymbol{\alpha}(\boldsymbol{\zeta})\}\{\widehat{\boldsymbol{\alpha}}(\boldsymbol{\zeta}) - \boldsymbol{\alpha}(\boldsymbol{\zeta})\} + o_p(1)$$

$$= -n^{-1/2} \sum_{i=1}^{n} \mathbf{L}_{\zeta\theta}\{y_i, w_i, \mathbf{s}_i, \mathbf{z}_i; \boldsymbol{\zeta}, \theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_i)\} \Omega(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_i)^{-1}$$

$$[\boldsymbol{\theta}_{\alpha}^{\mathrm{T}}\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_i; \boldsymbol{\alpha}(\boldsymbol{\zeta})\} \boldsymbol{\theta}_{\alpha}\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_i; \boldsymbol{\alpha}(\boldsymbol{\zeta})\}]^{-1} f_{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{Z}}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_i)^{-1} \boldsymbol{\theta}_{\alpha}^{\mathrm{T}}\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_i; \boldsymbol{\alpha}(\boldsymbol{\zeta})\}$$

$$\left[n^{-1} \sum_{j=1}^{n} K_h(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_j - \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_i) \mathbf{L}_{\alpha}\{y_j, w_j, \mathbf{s}_j, \mathbf{z}_j; \boldsymbol{\zeta}, \boldsymbol{\alpha}(\boldsymbol{\zeta})\} \right] + o_p(1)$$

$$= -n^{-3/2} \sum_{j=1}^{n} \sum_{i=1}^{n} K_h(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_j - \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_i) \mathbf{L}_{\zeta\theta}\{y_i, w_i, \mathbf{s}_i, \mathbf{z}_i; \boldsymbol{\zeta}, \theta(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_i)\} \Omega(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_i)^{-1}$$

$$f_{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{Z}}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_i)^{-1} [\boldsymbol{\theta}_{\alpha}^{\mathrm{T}}\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_i; \boldsymbol{\alpha}(\boldsymbol{\zeta})\} \boldsymbol{\theta}_{\alpha}\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_i; \boldsymbol{\alpha}(\boldsymbol{\zeta})\}]^{-1} \boldsymbol{\theta}_{\alpha}^{\mathrm{T}}\{\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_i; \boldsymbol{\alpha}(\boldsymbol{\zeta})\}$$

$$\mathbf{L}_{\alpha}\{y_j, w_j, \mathbf{s}_j, \mathbf{z}_j; \boldsymbol{\zeta}, \boldsymbol{\alpha}(\boldsymbol{\zeta})\} + o_p(1)$$

$$= -n^{-1/2} \sum_{j=1}^{n} E(\mathbf{L}_{\zeta\theta} \mid \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_j) \Omega(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_j)^{-1} \{\boldsymbol{\theta}_{\alpha}^{\mathrm{T}}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_j) \boldsymbol{\theta}_{\alpha}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_j)\}^{-1} \boldsymbol{\theta}_{\alpha}^{\mathrm{T}}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_j; \boldsymbol{\alpha})$$

$$\mathbf{L}_{\alpha}\{y_j, w_j, \mathbf{s}_j, \mathbf{z}_j; \boldsymbol{\zeta}, \boldsymbol{\alpha}(\boldsymbol{\zeta})\} + o_p(1)$$

$$= -n^{-1/2} \sum_{i=1}^{n} \mathbf{U}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{z}_i) L_{\theta}\{y_i, w_i, \mathbf{s}_i, \mathbf{z}_i; \boldsymbol{\zeta}, \theta(\cdot)\} + o_p(1).$$

Plugging the above into (A.1), we obtain the expansion in Theorem 3.1. The subsequent results in Theorem 3.1 are easy to obtain. Therefore, their proofs are omitted. □

A.4. Proof of Theorem 3.2

The asymptotic expansion in Theorem 3.1 indicates that $\mathbf{A}^{-1}\mathbf{S}_{\mathrm{eff}}\{Y, W, \mathbf{S}, \mathbf{Z}; \boldsymbol{\zeta}, \theta(\cdot)\}$ is an influence function (Newey, 1990), where

$$\mathbf{S}_{\mathrm{eff}}\{Y, W, \mathbf{S}, \mathbf{Z}; \boldsymbol{\zeta}, \theta(\cdot)\}$$
$$\equiv \mathbf{L}_{\zeta}\{Y, W, \mathbf{S}, \mathbf{Z}; \boldsymbol{\zeta}, \theta(\cdot)\} - \mathbf{U}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{Z}) L_{\theta}\{Y, W, \mathbf{S}, \mathbf{Z}; \boldsymbol{\zeta}, \theta(\cdot)\}.$$

To show the efficiency, we need to show that when $f_{X|\mathbf{S},\mathbf{Z}}^*(x, \mathbf{s}, \mathbf{z}) = f_{X|\mathbf{S},\mathbf{Z}}(x, \mathbf{s}, \mathbf{z})$, $\mathbf{S}_{\mathrm{eff}}$ is the residual of the orthogonal projection of $\mathbf{S}_{\zeta}$ onto the nuisance tangent space, denoted Λ. Following Tsiatis and Ma (2004), the nuisance tangent space with respect to $f_{X|\mathbf{S},\mathbf{Z}}(x, \mathbf{s}, \mathbf{z})$ is

$$\Lambda_f = [E\{\mathbf{a}(X, \mathbf{S}, \mathbf{Z}) \mid Y, W, \mathbf{S}, \mathbf{Z}\} : E(\mathbf{a}) = \mathbf{0}],$$

and $\mathbf{L}_{\zeta}$ is the orthogonal projection of $\mathbf{S}_{\zeta}$ onto $\Lambda_f^{\perp}$, the orthogonal complement of Λ_f. Taking derivative of $l^*(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, y, w, \mathbf{s}, \mathbf{z})$ with respect to $\boldsymbol{\alpha}$ and considering all possible $\boldsymbol{\alpha}$, we obtain the nuisance tangent space with respect to $\theta(\cdot)$ as

$$\Lambda_{\theta} = \{S_{\theta}(Y, W, \mathbf{S}, \mathbf{Z}) \mathbf{a}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}} \mathbf{Z})\}.$$

Thus, the nuisance tangent space is $\Lambda = \Lambda_f + \Lambda_\theta$. Defining

$$\begin{aligned}
\widetilde{\Lambda}_\theta &= \{L_\theta(Y, W, \mathbf{S}, \mathbf{Z})\mathbf{a}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z})\} \\
&= \left([S_\theta(Y, W, \mathbf{S}, \mathbf{Z}) - E\{a_\theta(X, \mathbf{S}, \mathbf{Z}) \mid Y, W, \mathbf{S}, \mathbf{Z}\}]\mathbf{a}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z})\right),
\end{aligned}$$

where a_θ satisfies

$$E\{S_\theta(Y, W, \mathbf{S}, \mathbf{Z}) \mid X, \mathbf{S}, \mathbf{Z}\} = E[E\{a_\theta(X, \mathbf{S}, \mathbf{Z}) \mid Y, W, \mathbf{S}, \mathbf{Z}\} \mid X, \mathbf{S}, \mathbf{Z}],$$

and

$$L_\theta(Y, W, \mathbf{S}, \mathbf{Z}) = S_\theta(Y, W, \mathbf{S}, \mathbf{Z}) - E\{a_\theta(X, \mathbf{S}, \mathbf{Z}) \mid Y, W, \mathbf{S}, \mathbf{Z}\},$$

then $\Lambda = \Lambda_f \oplus \widetilde{\Lambda}_\theta$. Subsequently, the orthogonal complement of Λ is

$$\Lambda^\perp = \{\mathbf{b}(Y, W, \mathbf{S}, \mathbf{Z}) : E(\mathbf{b} \mid \mathbf{X}, \mathbf{S}, \mathbf{Z}) = E(\mathbf{b}S_\theta \mid \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z}) = \mathbf{0}\}.$$

It is easy to see that $\mathbf{U}(\widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z})L_\theta\{Y, W, \mathbf{S}, \mathbf{Z}; \boldsymbol{\zeta}, \theta(\cdot)\} \in \widetilde{\Lambda}_\theta \cap \Lambda_f^\perp$. On the other hand, we already have $\mathbf{S}_{\mathrm{eff}} \in \Lambda^\perp = \widetilde{\Lambda}_\theta^\perp \cap \Lambda_f^\perp$. Thus, $\mathbf{S}_{\mathrm{eff}}$ is the orthogonal projection of $\mathbf{L}_\zeta$ on $\Lambda^\perp$, hence equivalently, the orthogonal projection of $\mathbf{S}_\zeta$ on $\Lambda^\perp$. This proves the efficiency result. $\qquad\square$

A.5. *Proof of Theorem 3.3*

Following the results in Theorem 3.1, under H_0, $n^{1/2}(\mathbf{M}\widehat{\boldsymbol{\beta}} - \mathbf{c})$ follows a normal distribution with mean zero and covariance matrix $\mathbf{M}(\mathbf{I}_\beta, \mathbf{0})\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1\,\mathrm{T}}(\mathbf{I}_\beta, \mathbf{0})^{\mathrm{T}}\mathbf{M}^{\mathrm{T}}$ asymptotically. Consequently, T given in Theorem 3.3 has an asymptotic Chi-square distribution with d_M degrees of freedom. $\qquad\square$

References

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore, MD: Johns Hopkins University Press. MR1245941

CARROLL, R. J., FAN, J., GIJBELS, I. and WAND, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, **92**, 477–489. MR1467842

CARROLL, R. J. and HALL, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, **83**, 1184–1186. MR0997599

CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A. and CRAINICEANU, C. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, Second edition. Boca Raton, CRC Press. MR2243417

CUI, X., HÄRDLE, W. and ZHU, L. X. (2011). The EFM approach for single-index models'. *Annals of Statistics*, **39**, 1658–1688. MR2850216

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Annals of Statistics*, **19**, 1257–1272. MR1126324

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman and Hall, London.

Härdle, W., Liang, H. & Gao, J. (2000). *Partially Linear Models.* Heidelberg, Physica-Verlag. MR1787637

Heckman, N. E. (1986). Spline smoothing in a partly linear model. *Journal of the Royal Statistical Society, Series B*, **48**, 244–8.

Hubert, H. B., Feinleib, M., McNamara, P. M. and Castelli, W. P. (1983). Obesity as an independent risk factor for cardiovascular disease: a 26-year follow-up of participants in the Framingham Heart Study. *Circulation*, 67, 968–977.

Li, L., Zhu, L. P. and Zhu, L. X. (2011). Inference on the primary parameter of interest with the aid of dimension reduction estimation. *Journal of the Royal Statistical Society, Series B*, **73**, 59–80.

Liang, H., Härdle, W., and Carroll, R. J. (1999). Estimation in a semiparametric partially linear errors-in-variables model. *Annals of Statistics*, **27**, 1519–1535.

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.

Ma, Y., Chiou, J.-M. and Wang, N. (2006). Efficient semiparametric estimator for heteroscedastic partially-linear models. *Biometrika*, **93**, 75–84.

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.

Metropolitan Life Insurance Company (1959). New weight standards for men and women. *Statistical Bulletin of the Metropolitan Life Insurance Company*, 40, 1.

Newey, W. K. (1990). Semiparametric efficiency bounds. *Journal of Applied Econometrics*, **5**, 99–135.

Stefanski, L. A. and Carroll, R. J. (1987). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika*, **74**, 703–716. MR0919838

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

Tsiatis, A. A., and Ma, Y. (2004). Locally efficient semiparametric estimators for functional measurement error model. *Biometrika*, **91**, 835–848. MR2126036

Van Keilegom, I. and Carroll, R. J. (2007) Backfitting versus profiling in general criterion functions. *Statistica Sinica*, **17**, 797–816.

Yu, Y. and Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association*, **97**, 1042–1054.