Journal of Multivariate Analysis 171 (2019) 320–338

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Semiparametric regression for measurement error model with heteroscedastic error

Mengyan Li, Yanyuan Ma, Runze Li *

Department of Statistics, Pennsylvania State University, University Park, PA 16802-2111, USA

ARTICLE INFO

Article history:
Received 5 April 2018
Available online 8 January 2019

Keywords:
B-splines
Efficient score
Heteroscedasticity
Measurement error
Semiparametrics

AMS 2010 subject classifications:
62G05
62G08

ABSTRACT

Covariate measurement error is a common problem. Improper treatment of measurement errors may affect the quality of estimation and the accuracy of inference. Extensive literature exists on homoscedastic measurement error models, but little research exists on heteroscedastic measurement. In this paper, we consider a general parametric regression model allowing for a covariate measured with heteroscedastic error. We allow both the variance function of the measurement errors and the conditional density function of the error-prone covariate given the error-free covariates to be completely unspecified. We treat the variance function using B-spline approximation and propose a semiparametric estimator based on efficient score functions to deal with the heteroscedasticity of the measurement error. The resulting estimator is consistent and enjoys good inference properties. Its finite-sample performance is demonstrated through simulation studies and a real data example.

1. Introduction

Covariate measurement errors arise frequently in areas such as chemistry, biological sciences, medicine and epidemiological studies. Instead of a precise measurement, we have only error-prone surrogates of the unobservable covariate. Measurement errors in covariates have received extensive attention, e.g., in linear models [7] and in nonlinear models [3]; see also [16] for a wider range of applications, such as survival analysis and case-control studies. Existing methods often make restrictive and unrealistic assumptions about the measurement error distribution such as normality and homoscedasticity. However, in practice, the error density may violate these assumptions, which leads to erroneous estimation and inference [2].

Some research has been completed on the difficult problem of measurement error heteroscedasticity. Staudenmayer et al. [14] found bias issues in density estimation in the presence of incorrect assumptions of homoscedasticity of the measurement errors. Similarly, in regression problems, ignoring the heteroscedasticity of the measurement errors can affect both the accuracy of the estimation and the quality of the inference. In an attempt to properly treat the heteroscedastic measurement error, Devanarayan and Stefanski [6] proposed the empirical SIMEX method to accommodate heteroscedastic measurement error. However, they assumed normality of the measurement error and only provided an approximate solution. Cheng and Riu [4] studied linear relationships in which both the response variable and the covariates are subject to heteroscedastic errors using the maximum likelihood method, method-of-moments, and generalized least squares method, under a critical but restrictive assumption that the variances of the normally distributed measurement and regression errors for each observation are known. Guo and Little [8] extended the regression calibration and multiple imputation methods to allow heteroscedastic measurement error, while assuming the normality of the conditional density of the measurement

* Corresponding author.
E-mail addresses: mql5616@psu.edu (M. Li), yzm63@psu.edu (Y. Ma), rzli@psu.edu (R. Li).

https://doi.org/10.1016/j.jmva.2018.12.012
0047-259X/

errors given the unobservable covariate and assuming that the variance function is a power function of the unobservable covariate. Sarkar et al. [13] studied the regression model with heteroscedastic errors in covariates in a Bayesian hierarchical framework and avoided assumptions about normality and homoscedasticity of the measurement and regression errors. However, due to the complexity of using both B-splines and Dirichlet processes, the theoretical properties of the estimator were not established.

In this article, we consider a general parametric regression model with one covariate measured with heteroscedastic error when both the conditional density of the unobservable covariate given the error-free covariates and the variance function of the measurement errors are unknown. We allow the distributions of the regression error and the measurement error to have any form; in particular, they are not limited to the normal distribution family. This problem requires that we estimate two nuisance functions. If the unknown variance function were parametric, this problem would be simpler. Hence, we approximate the variance function with B-splines to convert the model to a simpler setting operationally. Using a semiparametric approach, we also avoid performing deconvolution. A B-spline approximation has been used in the measurement error model in the Bayesian framework [1,13], although the theoretical impact of the B-spline approximation is unclear.

Recently, in the nonparametric model framework, Jiang and Ma [12] proposed a spline-assisted semiparametric approach to measurement error models, where the asymptotic properties of the nonparametric estimator were established in the homoscedastic measurement error case. Although the B-spline approximation makes the estimation of the variance function feasible, we still face the challenge of handling the conditional density function of the unobservable covariate. We avoid having to estimate it by proposing an estimator that allows for misspecification of the conditional density function. The final estimator is obtained by solving estimating equations based on estimating functions that are approximated without a closed form. Hence it is very challenging to establish the asymptotic properties of the estimator. The method of analysis is also very different from that of typical analyses of splines.

Our method enjoys good asymptotic properties in terms of convergence rate. Generally, estimating nuisance parameters will alter (often inflate) the estimation variance for the parameter of interest. However, our method does not. Further, if the density function of the unobservable covariate is correctly specified, the estimator of the parameter of interest is efficient. We specify the model for data with mismeasured covariate and heteroscedastic measurement error in Section 2 and develop the estimation procedure in Section 3. Regularity conditions and asymptotic properties of the estimator are described in Section 4. We explain how to implement our method in Section 5. To assess the performance of our method, we conduct simulation studies and perform an empirical data analysis in Section 6.

2. Notation and model setup

2.1. Model specification

Throughout this paper, we use bold fonts for vectors and matrices and regular fonts for scalars. Let Y be the response variable and $\mathbf{Z}$ be the vector of observed error-free covariates. Let X be an unobservable latent covariate that is measured in an error-prone way. Let W be the observed surrogate of X. We assume the support of X to be finite; without loss of generality, let the support be $[0, 1]$. We are interested in the relationship between the response variable Y and the true covariates $(X, \mathbf{Z})$. In particular, we link the response to the covariates using a parametric model

$$f_{Y|X,\mathbf{Z}}(y, x, \mathbf{z}, \boldsymbol{\beta}), \tag{1}$$

where $f_{Y|X,\mathbf{Z}}$ is the specified conditional probability density function or probability mass function of Y given X and $\mathbf{Z}$, and $\boldsymbol{\beta}$ is a p-dimensional vector of unknown parameters.

We further assume that W, the observed surrogate of X, is linked to X through a heteroscedastic measurement error model

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

where U is a random variable with a known density function $f_U(u)$, which is not restricted to the normal family. Without loss of generality, we assume that U has mean zero and variance 1. We also assume that U is independent of X and $\mathbf{Z}$. Let $\sigma(X)$ be an unknown positive nuisance function that describes the heteroscedasticity of the measurement error. Our goal is to estimate the regression parameter $\boldsymbol{\beta}$ based on a sample $(Y_1, W_1, \mathbf{Z}_1), \ldots, (Y_n, W_n, \mathbf{Z}_n)$.

2.2. Identifiability considerations

To ensure the identifiability of the model, we assume that two repeated measurements of the error-prone covariate X are available. Let W_1 and W_2 be two measurements of X, which are independent with each other conditional on X. To prove identifiability, we aim to show that if the density or mass function of the observed data conditional on $\mathbf{Z}$ satisfies $f_{Y,W_1,W_2|\mathbf{Z}}(y, w_1, w_2, \mathbf{z}, \boldsymbol{\beta}, \sigma, f_{X|\mathbf{Z}}) = f_{Y,W_1,W_2|\mathbf{Z}}(y, w_1, w_2, \mathbf{z}, \widetilde{\boldsymbol{\beta}}, \widetilde{\sigma}, \widetilde{f}_{X|\mathbf{Z}})$, then $(\boldsymbol{\beta}, \sigma, \mathbf{f}_{X|\mathbf{Z}}) = (\widetilde{\boldsymbol{\beta}}, \widetilde{\sigma}, \widetilde{f}_{X|\mathbf{Z}})$. Here $f_{X|\mathbf{Z}}(x|\mathbf{z})$ is the conditional density of X given $\mathbf{Z}$. Note that

$$f_{Y,W_1,W_2|\mathbf{Z}}(y, w_1, w_2, \mathbf{z}, \boldsymbol{\beta}, \sigma, f_{X|\mathbf{Z}}) = \int f_{Y|X,\mathbf{Z}}(y|x, \mathbf{z}, \boldsymbol{\beta}) f_U\left\{\frac{w_1 - x}{\sigma(x)}\right\} f_U\left\{\frac{w_2 - x}{\sigma(x)}\right\} f_{X|\mathbf{Z}}(x|\mathbf{z}) \frac{1}{\sigma^2(x)}\, dx$$

and, likewise,

$$f_{Y,W_1,W_2|\mathbf{Z}}\,\{y, w_1, w_2, \mathbf{z}, \widetilde{\boldsymbol{\beta}}, \widetilde{\sigma}, \widetilde{f}_{X|\mathbf{Z}}\} = \int f_{Y|X,\mathbf{Z}}(y|x, \mathbf{z}, \widetilde{\boldsymbol{\beta}}) f_U\left\{\frac{w_1 - x}{\widetilde{\sigma}(x)}\right\} f_U\left\{\frac{w_2 - x}{\widetilde{\sigma}(x)}\right\} \widetilde{f}_{X|\mathbf{Z}}(x|\mathbf{z})\,\frac{1}{\widetilde{\sigma}^2(x)}\,dx.$$

When we leave the model of Y given X and $\mathbf{Z}$ as an arbitrary known parametric model, it is very difficult to prove identifiability without adding many conditions that are difficult to check. Thus, we believe it is a better strategy to establish identifiability in a case-by-case fashion. Here, as an example, we consider a specific situation where the main model is linear with heteroscedastic normal or Laplace measurement errors, i.e., $Y = \beta_0 + X\beta_1 + \mathbf{Z}^\top \boldsymbol{\beta}_2 + \epsilon$ and $W_j = X + \sigma(X)U_j$ for $j \in \{1, 2\}$, where ϵ has a mean zero normal distribution with variance σ_ϵ^2 and U_1, U_2 are standard normal random variables or Laplace random variables with mean 0 and variance 1. The identifiability of this specific model can be established by computing the first and the second moments of Y, W_1 and W_2 given $\mathbf{Z}$ and using Fourier transform. The detailed proof is given in Appendix A.1.

3. Methodology

3.1. Estimator of the original model

The score function of the parametric model given in (1) is $\partial \ln\{f_{Y|X,\mathbf{Z}}(y|x, \mathbf{z}, \boldsymbol{\beta})\}/\partial \boldsymbol{\beta}$. If the covariate X were observed precisely, a consistent estimate of the parameter $\boldsymbol{\beta}$ could be obtained by solving the sample version of

$$\mathrm{E}[\partial \ln\{f_{Y|X,\mathbf{Z}}(Y|X, \mathbf{Z}, \boldsymbol{\beta})\}/\partial \boldsymbol{\beta}] = \mathbf{0}.$$

However, since X is unobservable and only W_1 and W_2 are available, we have to rely on the conditional density of (Y, W_1, W_2) given $\mathbf{Z}$,

$$f_{Y,W_1,W_2|\mathbf{Z}}(y, w_1, w_2, \mathbf{z}, \boldsymbol{\beta}, \sigma, f_{X|\mathbf{Z}}) = \int \frac{1}{\sigma^2(x)} f_{Y|X,\mathbf{Z}}(y|x, \mathbf{z}, \boldsymbol{\beta}) f_U(u_1) f_U(u_2) f_{X|\mathbf{Z}}(x|\mathbf{z})dx,$$

where $u_j = (w_j - x)/\sigma(x)$, for $j \in \{1, 2\}$. Here the p-dimensional parameter $\boldsymbol{\beta}$ is of interest, and infinite-dimensional parameters $f_{X|\mathbf{Z}}$ and σ are nuisance. The "observed" score function with respect to $\boldsymbol{\beta}$ is given as

$$\mathbf{S}_{\beta}(y, w_1, w_2, \mathbf{z}, \boldsymbol{\beta}, \sigma, f_{X|\mathbf{Z}}) = \frac{\int \{\partial f_{Y|X,\mathbf{Z}}(y|x, \mathbf{z}, \boldsymbol{\beta})/\partial \boldsymbol{\beta}\} f_U(u_1) f_U(u_2) f_{X|\mathbf{Z}}(x|\mathbf{z})/\sigma^2(x)dx}{\int f_{Y|X,\mathbf{Z}}(y|x, \mathbf{z}, \boldsymbol{\beta}) f_U(u_1) f_U(u_2) f_{X|\mathbf{Z}}(x|\mathbf{z})/\sigma^2(x)dx},$$

which is the partial derivative of the logarithm of the likelihood $f_{Y,W_1,W_2|\mathbf{Z}}$ with respect to the parameter $\boldsymbol{\beta}$. Although $\mathrm{E}\{\mathbf{S}_{\beta}(Y, W_1, W_2, \mathbf{z}, \boldsymbol{\beta}, \sigma, f_{X|\mathbf{Z}})\} = \mathbf{0}$ at the true parameter values, it is impossible to estimate $\boldsymbol{\beta}$ by solving $\sum_{i=1}^n \mathbf{S}_{\beta}(y_i, w_{i1}, w_{i2}, \mathbf{z}, \boldsymbol{\beta}, \sigma, f_{X|\mathbf{Z}}) = \mathbf{0}$ directly due to the presence of the nuisance parameters.

We thus take a different approach and try to construct estimators $\widehat{\boldsymbol{\beta}}_n$ by directly identifying the influence functions. A regular asymptotic linear estimator $\widehat{\boldsymbol{\beta}}_n$ can be written as

$$\sqrt{n}\,(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = n^{-1/2} \sum_{i=1}^n \phi(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}) + o_p(1),$$

where $\phi(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta})$ is a p-dimensional zero-mean random vector referred to as the ith influence function of the estimator $\widehat{\boldsymbol{\beta}}_n$. From a geometric point of view, in the original model described by (1) and (2), an influence function of a single observation lies in the Hilbert space $\mathcal{H}$ of all p-dimensional zero-mean measurable functions of the observed data with finite variance, equipped with the inner product $\langle h_1, h_2\rangle = \mathrm{E}\{h_1{}^\top(Y, W_1, W_2, \mathbf{Z})h_2(Y, W_1, W_2, \mathbf{Z})|\mathbf{Z}\}$, where $h_1, h_2 \in \mathcal{H}$. Further, influence functions belong to the linear space orthogonal to the nuisance tangent space, which is defined as the mean squared closure of the nuisance tangent spaces of parametric sub-models spanned by the nuisance score vectors.

In the original models from (1) and (2), the nuisance space is given as $\Lambda = \Lambda_{f_{X|\mathbf{Z}}} + \Lambda_\sigma$, where

$$\Lambda_{f_{X|\mathbf{Z}}} = \{\mathrm{E}\{\mathbf{a}(X, \mathbf{Z})|Y, W_1, W_2, \mathbf{Z}\} : \mathrm{E}\{\mathbf{a}(X, \mathbf{Z})|\mathbf{Z}\} = \mathbf{0}\},$$
$$\Lambda_\sigma = \{\mathrm{E}\,\{V(U_1, U_2)\mathbf{b}(X)|Y, W_1, W_2, \mathbf{Z}\} : \forall \mathbf{b}(X)\}.$$

Here $V(U_1, U_2) = f_U'(U_1)U_1/f_U(U_1) + f_U'(U_2)U_2/f_U(U_2) + 2$, where $U_j = (W_j - X)/\sigma(X)$, for $j \in \{1, 2\}$. The detailed proof of the result concerning Λ is in Appendix A.2. Note that the two subspaces $\Lambda_{f_{X|\mathbf{Z}}}$ and Λ_σ are not orthogonal to each other, hence we write the orthogonal complement of Λ as $\Lambda^\perp = \Lambda_{f_{X|\mathbf{Z}}}^\perp \cap \Lambda_\sigma^\perp$. Here, $\Lambda_{f_{X|\mathbf{Z}}}^\perp$ is the orthogonal complement of $\Lambda_{f_{X|\mathbf{Z}}}$ and has the form

$$\Lambda_{f_{X|\mathbf{Z}}}^\perp = \{\mathbf{h}(Y, W_1, W_2, \mathbf{Z}) : \mathrm{E}\{\mathbf{h}(Y, W_1, W_2, \mathbf{Z})|X, \mathbf{Z}\} = \mathbf{0} \text{ almost everywhere}\},$$

while $\Lambda_\sigma^\perp$ is the orthogonal complement of Λ_σ and is given by

$$\Lambda_\sigma^\perp = \{\mathbf{h}(Y, W_1, W_2, \mathbf{Z}) : \mathrm{E}\,\{\mathbf{h}(Y, W_1, W_2, \mathbf{Z})V(U_1, U_2)|X, \mathbf{Z}\} = \mathbf{0} \text{ almost everywhere}\}.$$

Even without an explicit form of $\Lambda^{\perp}$, we can still derive the orthogonal projection of $\mathbf{S}_{\boldsymbol\beta}(y, w_1, w_2, \mathbf{z}, \boldsymbol\beta, \sigma, f_{X|\mathbf{Z}})$ onto $\Lambda^{\perp}$. We write the orthogonal projection as $\mathbf{S}_{\text{eff}}(y, w_1, w_2, \mathbf{z}, \boldsymbol\beta, \sigma, f_{X|\mathbf{Z}})$ and call it the efficient score. It is obvious that the asymptotic variance of a regular asymptotically linear estimator equals the variance of its influence function. Consequently, the optimal estimator among a class of regular asymptotically linear estimators is the one whose influence function has the smallest variance, which we call the efficient influence function. The efficient score directly leads to the efficient influence function through the form

$$\boldsymbol\phi_{\text{eff}}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol\beta) = [\mathrm{E}\{\mathbf{S}_{\text{eff}}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol\beta, \sigma, f_{X|\mathbf{Z}})^{\otimes 2}\}]^{-1}\mathbf{S}_{\text{eff}}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol\beta, \sigma, f_{X|\mathbf{Z}}),$$

where $\mathbf{a}^{\otimes 2} \equiv \mathbf{a}\mathbf{a}^{\top}$ for any vector or matrix $\mathbf{a}$, and this convention is used throughout this article. Thus, for the purpose of constructing estimating equations, we only need to identify $\mathbf{S}_{\text{eff}}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol\beta, \sigma, f_{X|\mathbf{Z}})$. It is easy to verify that

$$\begin{aligned}&\mathbf{S}_{\text{eff}}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol\beta, \sigma, f_{X|\mathbf{Z}})\\ &\quad = \mathbf{S}_{\boldsymbol\beta}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol\beta, \sigma, f_{X|\mathbf{Z}}) - \mathrm{E}\{\mathbf{a}(X, \mathbf{Z})|Y, W_1, W_2, \mathbf{Z}\} - \mathrm{E}\left\{V(U_1, U_2)\mathbf{b}(X)|Y, W_1, W_2, \mathbf{Z}\right\},\end{aligned}$$

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

$$\begin{aligned}&\mathrm{E}\{\mathbf{S}_{\boldsymbol\beta}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol\beta, \sigma, f_{X|\mathbf{Z}})|X, \mathbf{Z}\}\\ &\quad = \mathrm{E}[\mathrm{E}\{\mathbf{a}(X, \mathbf{Z})|Y, W_1, W_2, \mathbf{Z}\}|X, \mathbf{Z}] + \mathrm{E}[\mathrm{E}\{V(U_1, U_2)\mathbf{b}(X)|Y, W_1, W_2, \mathbf{Z}\}|X, \mathbf{Z}],\end{aligned}$$

and

$$\begin{aligned}&\mathrm{E}\{\mathbf{S}_{\boldsymbol\beta}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol\beta, \sigma, f_{X|\mathbf{Z}})V(U_1, U_2)|X, \mathbf{Z}\}\\ &\quad = \mathrm{E}[\mathrm{E}\{\mathbf{a}(X, \mathbf{Z})|Y, W_1, W_2, \mathbf{Z}\}V(U_1, U_2)|X, \mathbf{Z}] + \mathrm{E}[\mathrm{E}\{V(U_1, U_2)\mathbf{b}(X)|Y, W_1, W_2, \mathbf{Z}\}V(U_1, U_2)|X, \mathbf{Z}].\end{aligned}$$

By the definition of $\mathbf{a}(X, \mathbf{Z})$ and $\mathbf{b}(X)$, it is obvious that $\mathrm{E}\{\mathbf{S}_{\text{eff}}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol\beta, \sigma, f_{X|\mathbf{Z}})|X, \mathbf{Z}\} = \mathbf{0}$. Hence, an efficient estimator of $\boldsymbol\beta$ can be obtained by solving the estimating equation $\sum_{i=1}^{n} \mathbf{S}_{\text{eff}}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol\beta, \sigma, f_{X|\mathbf{Z}}) = \mathbf{0}$ if we know the true σ and $f_{X|\mathbf{Z}}$.

Remark 1. One way to ensure the identifiability of a problem is to ensure that the efficient score is not always equal to zero. This is true as long as $\Lambda^{\perp}$ is not a zero space. In Appendix A.3, we show this to be the case for the logistic regression model with normal measurement errors. In other models, when a rigorous proof for identifiability is difficult to obtain, numerical calculation of $\mathbf{S}^{*}_{\text{eff}}$ can provide some insights to the identifiability issue. For example, in all the simulation studies and real data examples conducted in this work, $\mathbf{S}^{*}_{\text{eff}}$ is not close to zero at any arbitrary parameter values, indicating the identifiability of the corresponding models.

An interesting discovery here is that even without knowing the true σ and $f_{X|\mathbf{Z}}$, we can still construct the estimating equation in the same fashion after adopting a working model $f^{*}_{X|\mathbf{Z}}$. A similar observation is made in [15]. Specifically, we find that

$$\mathrm{E}\{\mathbf{S}^{*}_{\text{eff}}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol\beta, \sigma, f^{*}_{X|\mathbf{Z}})|X, \mathbf{Z}\} = \mathrm{E}^{*}\{\mathbf{S}^{*}_{\text{eff}}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol\beta, \sigma, f^{*}_{X|\mathbf{Z}})|X, \mathbf{Z}\} = \mathbf{0},$$

where the superscript * denotes the corresponding quantities, such as the efficient score $\mathbf{S}_{\text{eff}}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol\beta, \sigma, f_{X|\mathbf{Z}})$ and the expectations, calculated with the unknown $f_{X|\mathbf{Z}}$ replaced by the possibly misspecified working model $f^{*}_{X|\mathbf{Z}}$ everywhere it appears in the construction. Thus,

$$\sum_{i=1}^{n} \mathbf{S}^{*}_{\text{eff}}(Y_i, W_{1i}, W_{2i}, \mathbf{Z}_i, \boldsymbol\beta, \sigma, f^{*}_{X|\mathbf{Z}}) = \mathbf{0}$$

is a consistent estimating equation set.

3.2. Estimator of the approximate model

We have seen that although we have obtained $\mathbf{S}_{\text{eff}}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol\beta, \sigma, f_{X|\mathbf{Z}})$, it is not realistic to use it. One reason is that it relies on the unknown conditional density function $f_{X|\mathbf{Z}}$. We have circumvented this difficulty by adopting a working model $f^{*}_{X|\mathbf{Z}}$ as in Section 3.1. The other obstacle we encounter in implementing $\mathbf{S}_{\text{eff}}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol\beta, \sigma, f_{X|\mathbf{Z}})$ lies in σ, and it also holds in implementing $\mathbf{S}^{*}_{\text{eff}}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol\beta, \sigma, f^{*}_{X|\mathbf{Z}})$. To overcome this obstacle, we propose to estimate $\sigma(X)$ using spline approximation $\mathbf{B}(X)^{\top}\boldsymbol\gamma$. If a suitable estimator $\widehat{\boldsymbol\gamma}$ can be obtained, then we can use $\widehat{\sigma}(X) = \mathbf{B}(X)^{\top}\widehat{\boldsymbol\gamma}$ in place of $\sigma(X)$ to facilitate the construction of the estimating equations using $\mathbf{S}^{*}_{\text{eff}}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol\beta, \widehat{\sigma}, f^{*}_{X|\mathbf{Z}})$.

To estimate $\boldsymbol\gamma$, we consider the approximate model

$$f_{Y|X,\mathbf{Z}}(y|x, \mathbf{z}, \boldsymbol\beta) \quad \text{and} \quad W = X + \mathbf{B}(X)^{\top}\boldsymbol\gamma U, \tag{3}$$

which allows us to estimate $\boldsymbol\gamma$ at any $\boldsymbol\beta$. In approximate model (3), the density of the observed data conditional on $\mathbf{Z}$ is given by

$$f_{a,Y,W_1,W_2|\mathbf{Z}}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol\beta, \boldsymbol\gamma, f_{X|\mathbf{Z}}) = \int f_{Y|X,\mathbf{Z}}(y|x, \mathbf{z}, \boldsymbol\beta) f_U\left\{\frac{w_1 - x}{\mathbf{B}(x)^{\top}\boldsymbol\gamma}\right\} f_U\left\{\frac{w_2 - x}{\mathbf{B}(x)^{\top}\boldsymbol\gamma}\right\} f_{X|\mathbf{Z}}(x|\mathbf{z}) \frac{1}{\{\mathbf{B}(x)^{\top}\boldsymbol\gamma\}^2}\, dx,$$

where the p-dimensional parameter $\boldsymbol{\beta}$ is of interest, while the d_{γ}-dimensional parameter $\boldsymbol{\gamma}$ and function $f_{X|\mathbf{Z}}$ are nuisance parameters. Here and throughout the text, we use the subscript $_a$ to denote quantities pertaining to approximate model (3). In this model, the influence functions of a single observation for regular asymptotically linear estimators of $\boldsymbol{\beta}$ lie in the Hilbert space $\mathcal{H}_a$ of all p-dimensional zero-mean measurable functions of observed data with finite variance, equipped with the inner product $\langle h_1, h_2\rangle = \mathrm{E}_a\{h_1{}^{\top}(Y, W_1, W_2, \mathbf{Z})h_2(Y, W_1, W_2, \mathbf{Z})|\mathbf{Z}\}$, where $h_1, h_2 \in \mathcal{H}_a$.

As in Section 3.1, the nuisance tangent space is $\Lambda_a = \Lambda_{a,f_{X|\mathbf{Z}}} + \Lambda_{a,\gamma}$, where

$$\Lambda_{a,f_{X|\mathbf{Z}}} = \{\mathrm{E}_a\{\mathbf{c}(X, \mathbf{Z})|Y, W_1, W_2, \mathbf{Z}\} : \mathrm{E}_a\{\mathbf{c}(X, \mathbf{Z})|\mathbf{Z}\} = \mathbf{0}\},$$
$$\Lambda_{a,\gamma} = \{\mathrm{E}_a\{V(U_{a,1}, U_{a,2})\mathbf{KB}(X)|Y, W_1, W_2, \mathbf{Z}\} : \mathbf{K} \text{ is a } p \times d_{\gamma} \text{ constant matrix}\},$$

and $U_{a,j} = (W_j - X)/\{\mathbf{B}(X)^{\top}\boldsymbol{\gamma}\}$, for $j \in \{1, 2\}$. If we treat $\boldsymbol{\gamma}$ as part of the parameters of interest, the efficient score for $\boldsymbol{\gamma}$ is the residual of the score vector for $\boldsymbol{\gamma}$ after projecting it on to $\Lambda_{a,f_{X|\mathbf{Z}}}$. Following a derivation similar to that for $\mathbf{S}_{\mathrm{eff}}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol{\beta}, \sigma, f_{X|\mathbf{Z}})$, the efficient score for $\boldsymbol{\gamma}$ is given as

$$\mathbf{S}_{a,\mathrm{eff},\gamma}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\gamma}, f_{X|\mathbf{Z}}) = \mathbf{S}_{a,\gamma}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\gamma}, f_{X|\mathbf{Z}}) - \mathrm{E}_a\{\mathbf{c}(X, \mathbf{Z})|Y, W_1, W_2, \mathbf{Z}\}, \tag{4}$$

where $\mathbf{S}_{a,\gamma}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\gamma}, f_{X|\mathbf{Z}}) \equiv \partial \ln\{f_{a,Y,W_1,W_2|\mathbf{Z}}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\gamma}, f_{X|\mathbf{Z}})\}/\partial\boldsymbol{\gamma}$ is the score vector for $\boldsymbol{\gamma}$, and $\mathbf{c}(X, \mathbf{Z})$ satisfies

$$\mathrm{E}_a\{\mathbf{S}_{a,\gamma}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\gamma}, f_{X|\mathbf{Z}})|X, \mathbf{Z}\} = \mathrm{E}_a[\mathrm{E}_a\{\mathbf{c}(X, \mathbf{Z})|Y, W_1, W_2, \mathbf{Z}\}|X, \mathbf{Z}]. \tag{5}$$

The detailed derivation is in Appendix A.4. We can estimate $\boldsymbol{\gamma}$ as a function of $\boldsymbol{\beta}$ through solving the estimating equations $\sum_{i=1}^n \mathbf{S}^*_{a,\mathrm{eff},\gamma}(Y_i, W_{1i}, W_{2i}, \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, f^*_{X|\mathbf{Z}}) = \mathbf{0}$, where $f^*_{X|\mathbf{Z}}$ is the working model.

Summarizing the above methods, we describe the detailed estimation procedure for $\boldsymbol{\beta}$ in approximate model (3) in the following algorithm.

Algorithm

1: Adopt a working model for $f_{X|\mathbf{Z}}(x|\mathbf{z})$ and denote it as $f^*_{X|\mathbf{Z}}(x|\mathbf{z})$.

2: Select a B-spline representation $\mathbf{B}(x)^{\top}\boldsymbol{\gamma}$ for $\sigma(x)$ with spline order r. Define the knots $t_{-r+1} = \cdots = t_0 = 0 < t_1 < \cdots < t_N < 1 = t_{N+1} = \cdots = t_{N+r}$, where N is the number of interior knots.

3: Solve the estimating equation for $\boldsymbol{\gamma}$

$$\begin{aligned}&\sum_{i=1}^n \mathbf{S}^*_{a,\mathrm{eff},\gamma}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, f^*_{X|\mathbf{Z}})\\ &\quad= \sum_{i=1}^n [\mathbf{S}^*_{a,\gamma}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, f^*_{X|\mathbf{Z}}) - \mathrm{E}^*_a\{\mathbf{c}(X, \mathbf{Z}_i)|Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i\}] = \mathbf{0}\end{aligned}$$

to obtain $\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta})$, where $\mathbf{c}(X, \mathbf{Z})$ satisfies

$$\mathrm{E}_a\{\mathbf{S}^*_{a,\gamma}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\gamma}, f^*_{X|\mathbf{Z}})|X, \mathbf{Z}\} = \mathrm{E}_a[\mathrm{E}^*_a\{\mathbf{c}(X, \mathbf{Z})|Y, W_1, W_2, \mathbf{Z}\}|X, \mathbf{Z}]. \tag{6}$$

4: Solve the estimating equation for $\boldsymbol{\beta}$

$$\begin{aligned}\sum_{i=1}^n \mathbf{S}^*_{\mathrm{eff}}\{Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}, \widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}), f^*_{X|\mathbf{Z}}\} &= \sum_{i=1}^n [\mathbf{S}^*_{\beta}\{Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}, \widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}), f^*_{X|\mathbf{Z}}\}\\ &\quad - \mathrm{E}^*_a\{\mathbf{a}(X, \mathbf{Z}_i)|Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i\} - \mathrm{E}^*_a\{V(U_{a,i1}, U_{a,i2})\mathbf{b}(X)|Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i\}] = \mathbf{0}\end{aligned} \tag{7}$$

to obtain $\widehat{\boldsymbol{\beta}}$, where $\mathbf{a}(X, \mathbf{Z})$ and matrix $\mathbf{b}(X)$ satisfy

$$\begin{aligned}&\mathrm{E}_a[\mathbf{S}^*_{\beta}\{Y, W_1, W_2, \mathbf{Z}, \boldsymbol{\beta}, \widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}), f^*_{X|\mathbf{Z}}\}|X, \mathbf{Z}]\\ &\quad= \mathrm{E}_a[\mathrm{E}^*_a\{\mathbf{a}(X, \mathbf{Z})|Y, W_1, W_2, \mathbf{Z}\}|X, \mathbf{Z}] + \mathrm{E}_a[\mathrm{E}^*_a\{V(U_{a,1}, U_{a,2})\mathbf{b}(X)|Y, W_1, W_2, \mathbf{Z}\}|X, \mathbf{Z}],\end{aligned} \tag{8}$$

and

$$\begin{aligned}&\mathrm{E}_a[\mathbf{S}^*_{\beta}\{Y, W_1, W_2, \mathbf{Z}, \boldsymbol{\beta}, \widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}), f^*_{X|\mathbf{Z}}\}V(U_{a,1}, U_{a,2})|X, \mathbf{Z}]\\ &\quad= \mathrm{E}_a[\mathrm{E}^*_a\{\mathbf{a}(X, \mathbf{Z})|Y, W_1, W_2, \mathbf{Z}\}V(U_{a,1}, U_{a,2})|X, \mathbf{Z}] + \mathrm{E}_a[\mathrm{E}^*_a\{V(U_{a,1}, U_{a,2})\mathbf{b}(X)|Y, W_1, W_2, \mathbf{Z}\}V(U_{a,1}, U_{a,2})|X, \mathbf{Z}].\end{aligned}$$

$\mathbf{S}^*_{\mathrm{eff}}\{Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}, \widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}), f^*_{X|\mathbf{Z}}\}$ is $\mathbf{S}^*_{\mathrm{eff}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}, \sigma, f^*_{X|\mathbf{Z}})$ with $\sigma(X_i)$ replaced by $\mathbf{B}(X_i)^{\top}\widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta})$.

Note that all the calculations with * should be conducted under the posited model $f^*_{X|\mathbf{Z}}$. We note that since the functions $\mathbf{a}(X, \mathbf{Z})$ and $\mathbf{c}(X, \mathbf{Z})$ satisfy (8) and (6), they automatically satisfy $\mathrm{E}^*_a\{\mathbf{a}(X, \mathbf{Z})|\mathbf{Z}\} = \mathbf{0}$, and $\mathrm{E}^*_a\{\mathbf{c}(X, \mathbf{Z})|\mathbf{Z}\} = \mathbf{0}$. Hence, $\mathrm{E}^*_a\{\mathbf{c}(X, \mathbf{Z})|Y, W_1, W_2, \mathbf{Z}\}$ and $\mathrm{E}^*_a\{\mathbf{a}(X, \mathbf{Z})|Y, W_1, W_2, \mathbf{Z}\}$ are indeed in $\Lambda^*_{a,f_{X|\mathbf{Z}}}$.

Remark 2. So far, we have assumed that there are at least two observations, W_1 and W_2, for each subject. Our method can also be applied to the cases with partial replication. For the subjects with replication, we construct estimating equations following the method given in Section 3 exactly. For the subjects that have only one W available, we simply modify our model using one W, then construct corresponding estimating equations. A consistent estimator for θ can be obtained by solving the two sets of estimating equations together. More details about the model with one W are given in Appendix A.7.

4. Asymptotic properties

To facilitate the proof of the asymptotic results, we first provide a list of regularity conditions.

(C1) The true density $f_{X|\mathbf{Z}}(x|\mathbf{z})$ at any $\mathbf{z}$ is a bounded function of x with compact support.
(C2) The function $\sigma(x) \in C^q([0,1])$, $q > 1$, is bounded with compact support.
(C3) The spline order $r \geqslant q$.
(C4) In B-splines approximation, let the number of interior knots N satisfy $N \to \infty$, $N^{-1}n\{\ln(n)\}^{-1} \to \infty$ and $Nn^{-1/(2q)} \to \infty$ as $n \to \infty$. Let $d_{\boldsymbol{\gamma}}$ denote the number of spline bases and $d_{\boldsymbol{\gamma}} = N + r$.
(C5) Let h_j be the distance between the jth and $(j-1)$th interior knots. Let $h_b = \max_{1 \leqslant j \leqslant N} h_j$ and $h_s = \min_{1 \leqslant j \leqslant N} h_j$. There exists a constant $c_h \in (0,\infty)$ such that $h_b/h_s < c_h$. Hence, $h_b = O(N^{-1})$ and $h_s = O(N^{-1})$.
(C6) The equation set

$$\mathrm{E}\{\mathbf{S}^*_{\mathrm{eff}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, f^*_{X|\mathbf{Z}})\} = \mathbf{0}, \quad \mathrm{E}\{\mathbf{S}^*_{a,\mathrm{eff},\boldsymbol{\gamma}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, f^*_{X|\mathbf{Z}})\} = \mathbf{0}$$

has a unique root for $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$ in the neighborhood of the true parameters $\boldsymbol{\theta}_0 = ({\boldsymbol{\beta}_0}^\top, {\boldsymbol{\gamma}_0}^\top)^\top$. The derivatives with respect to $\boldsymbol{\theta}$ on the left-hand side are smooth functions of $\boldsymbol{\theta}$, with singular values bounded above and bounded away from $\mathbf{0}$ in this neighborhood. Let the unique root be $\boldsymbol{\theta}^*$. Note that $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}^*$ are functions of N, that is, for any sufficiently large N; there is a unique root $\boldsymbol{\theta}^*$ in the neighborhood of $\boldsymbol{\theta}_0$.
(C7) For a matrix $\mathbf{A} = (a_{ij})$, denote $\|\mathbf{A}\|_\infty = \max_i \sum_j |a_{ij}|$ and $\|A\|_2 = \lambda_{\max}(\mathbf{A})$, where $\lambda_{\max}(\mathbf{A})$ represents the largest singular value of matrix $\mathbf{A}$. The following terms are integrable:

$$\|\partial \mathbf{S}^*_{\mathrm{eff}}(y_i, w_{i1}, w_{i2}, \mathbf{z}_i, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, f^*_{X|\mathbf{Z}})/\partial {\boldsymbol{\gamma}_0}^\top\|_\infty, \quad \|\mathbf{S}^*_{\mathrm{eff}}(y_i, w_{i1}, w_{i2}, \mathbf{z}_i, \boldsymbol{\beta}_0, \sigma, f^*_{X|\mathbf{Z}})\|_\infty,$$
$$\|\mathbf{S}^*_{\mathrm{eff}}(y_i, w_{i1}, w_{i2}, \mathbf{z}_i, \boldsymbol{\beta}_0, \sigma, f^*_{X|\mathbf{Z}})\,\{\mathbf{G}^\top \mathbf{S}_{a,\boldsymbol{\gamma}}(y_i, w_{i1}, w_{i2}, \mathbf{z}_i, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, f^*_{X|\mathbf{Z}})\}^\top\|_\infty$$

and

$$\|\{\mathbf{G}^\top \mathbf{S}_{a,\boldsymbol{\gamma}}(y_i, w_{i1}, w_{i2}, \mathbf{z}_i, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, f_{X|\mathbf{Z}})\}^\top f_{Y,W_1,W_2|\mathbf{Z}}(y_i, w_{i1}, w_{i2}, \mathbf{z}_i, \boldsymbol{\beta}_0, \sigma, f_{X|\mathbf{Z}}) f_{\mathbf{Z}}(\mathbf{z}_i)\|_\infty,$$

where $\mathbf{S}^*_{a,\mathrm{eff},\boldsymbol{\gamma}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}, \sigma, f^*_{X|\mathbf{Z}})$ is defined as $\mathbf{S}^*_{a,\mathrm{eff},\boldsymbol{\gamma}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, f^*_{X|\mathbf{Z}})$ with $\mathbf{B}(X_i)^\top\boldsymbol{\gamma}$ replaced by $\sigma(X_i)$, and $\mathbf{G}$ is arbitrary $d_{\boldsymbol{\gamma}} \times p$ matrix with $\|\mathbf{G}\|_2 = 1$.

Remark 3. From (C2) and (C5), there exists a $d_{\boldsymbol{\gamma}}$-dimensional spline coefficient vector $\boldsymbol{\gamma}_0$ such that $\sup_{x\in[0,1]} |\mathbf{B}(x)^\top \boldsymbol{\gamma}_0 - \sigma(x)| = O(h_b^q)$ [5]. Note that the dimension of $\boldsymbol{\gamma}_0$ goes to infinity, as $n \to \infty$.

We now establish the consistency of $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{\boldsymbol{\gamma}}_n$, as well as the asymptotic distribution property of $\widehat{\boldsymbol{\beta}}_n$. The proofs of the following results are in Appendices A.5 and A.6.

Theorem 1. *Assume that Conditions* (C1)–(C6) *hold. Let* $\widehat{\boldsymbol{\theta}}_n$ *satisfy*

$$\frac{1}{n}\sum_{i=1}^n \mathbf{S}^*_{\mathrm{eff}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\gamma}}_n, f^*_{X|\mathbf{Z}}) = \mathbf{0}, \quad \frac{1}{n}\sum_{i=1}^n \mathbf{S}^*_{a,\mathrm{eff},\boldsymbol{\gamma}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\gamma}}_n, f^*_{X|\mathbf{Z}}) = \mathbf{0}.$$

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

The result in Theorem 1 is used to further establish the asymptotic properties of the estimator of the parameters of interest $\widehat{\boldsymbol{\beta}}_n$.

Theorem 2. *Assume that Conditions* (C1)–(C7) *hold and let*

$$\mathbf{Q} \equiv \mathrm{E}\left[\left.\frac{\partial \mathbf{S}^*_{\mathrm{eff}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}_0, \boldsymbol{\gamma}, f^*_{X|\mathbf{Z}})}{\partial {\boldsymbol{\beta}_0}^\top}\right|_{\mathbf{B}(X)^\top\boldsymbol{\gamma}=\sigma(X)}\right].$$

Here the subscript $\mathbf{B}(X)^\top\boldsymbol{\gamma} = \sigma(X)$ *means replacing* $\mathbf{B}(X)^\top\boldsymbol{\gamma}$ *with the true function* $\sigma(X)$. *Then*

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = -\mathbf{Q}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{S}^*_{\mathrm{eff}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}_0, \sigma, f^*_{X|\mathbf{Z}}) + o_p(1).$$

Consequently, $\sqrt{n}\,(\widehat{\boldsymbol\beta}_n - \boldsymbol\beta_0) \rightsquigarrow \mathcal{N}(\mathbf{0}, \mathbf{V})$ *in distribution when* $n \to \infty$*, where*

$$\mathbf{V} = \mathbf{Q}^{-1}\mathrm{var}\{\mathbf{S}^*_{\mathrm{eff}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol\beta_0, \sigma, f^*_{X|\mathbf{Z}})\}(\mathbf{Q}^{-1})^\top.$$

In addition, if the working model $f^*_{X|\mathbf{Z}}$ *is correctly specified, i.e., if* $f^*_{X|\mathbf{Z}}(x|\mathbf{z}) = f_{X|\mathbf{Z}}(x|\mathbf{z})$*, then the estimator* $\widehat{\boldsymbol\beta}$ *is semiparametric efficient, where* $\widehat{\boldsymbol\beta}$ *achieves the optimal estimation variance bound* $[\mathrm{E}\{\mathbf{S}_{\mathrm{eff}}(Y, W_1, W_2, \mathbf{Z}, \sigma, f_{X|\mathbf{Z}})^{\otimes 2}\}]^{-1}$.

Remark 4. Typically, estimating the nuisance parameters will alter, often inflate, the estimation variance for the parameter of interest. However, in our construction, if we knew the true function $\sigma(x)$ and used it in the estimating equation derived in Section 3.1, the variance of the $\boldsymbol\beta$ estimation would not change, as shown in Theorem 2. In other words, the estimation of $\sigma(x)$ does not inflate the variance of $\widehat{\boldsymbol\beta}_n$ asymptotically.

Remark 5. The semiparametric efficiency of our estimator $\widehat{\boldsymbol\beta}$ relies on that the working model $f^*_{X|\mathbf{Z}}$ is correctly specified. In practice, the conditional density function $f_{X|\mathbf{Z}}$ can be estimated consistently using B-splines approximation $\mathbf{B}(X|\mathbf{Z})^\top\boldsymbol\zeta$. The density function of W_1, W_2 given $\mathbf{Z}$ is

$$f_{W_1,W_2|\mathbf{Z}}(w_1, w_2|\mathbf{z}) = \int f_{W_1|X,\mathbf{Z}}(w_1|x, \mathbf{z}) f_{W_2|X,\mathbf{Z}}(w_2|x, \mathbf{z}) f_{X|\mathbf{Z}}(x|\mathbf{z})dx.$$

Since W_1, W_2 and $\mathbf{Z}$ are observable, a consistent estimator of $\boldsymbol\zeta$ can be obtained by maximizing the log-likelihood of W_1 and W_2 given $\mathbf{Z}$ with respect to $\boldsymbol\zeta$. Specifically, the likelihood is given as

$$\prod_{i=1}^n \int f_U\left\{\frac{w_{i1} - x}{\widehat\sigma(x)}\right\} f_U\left\{\frac{w_{i2} - x}{\widehat\sigma(x)}\right\} \widehat\sigma(x)^{-2}\mathbf{B}(x|\mathbf{z}_i)^\top\boldsymbol\zeta dx,$$

where the estimated function $\widehat\sigma(x) = \mathbf{B}(x)^\top\widehat{\boldsymbol\gamma}$. One can then iteratively update the $\boldsymbol\gamma$ and $\boldsymbol\zeta$ estimates. However, it is not clear yet whether consistent estimation of $f_{X|\mathbf{Z}}$ is sufficient to achieve efficiency in estimating $\boldsymbol\beta$ or whether a certain convergence rate is needed. Therefore, we did not pursue this approach further.

The asymptotic covariance matrix $\mathbf{V}$ can be estimated using the sample version of the matrix $\mathbf{Q}$ and the variance of $\mathbf{S}^*_{\mathrm{eff}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol\beta_0, \sigma, f^*_{X|\mathbf{Z}})$. The detailed formulation is given in Section 6.1.

5. Implementation

To simplify the implementation, instead of profiling $\boldsymbol\gamma$ as a function of $\boldsymbol\beta$, we estimate $\boldsymbol\beta$ and $\boldsymbol\gamma$ together. Let $\boldsymbol\theta = (\boldsymbol\beta^\top, \boldsymbol\gamma^\top)^\top$. To estimate $\boldsymbol\theta$, we need to compute the efficient score for $\boldsymbol\theta$, denoted by $\mathbf{S}^*_{a,\mathrm{eff},\boldsymbol\theta}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol\theta, f^*_{X|\mathbf{Z}})$. This entails solving the integral equations

$$\mathrm{E}_a\{\mathbf{S}^*_{a,\boldsymbol\theta}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol\theta, f^*_{X|\mathbf{Z}})|X, \mathbf{Z}\} = \mathrm{E}_a[\mathrm{E}^*_a\{\mathbf{a}(X, \mathbf{Z})|Y, W_1, W_2, \mathbf{Z}\}|X, \mathbf{Z}]$$

to obtain $\mathbf{a}(X, \mathbf{Z})$, where

$$\mathbf{S}^*_{a,\boldsymbol\theta}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol\theta, f^*_{X|\mathbf{Z}}) \equiv \partial \ln\{f_{a,Y,W_1,W_2|\mathbf{Z}}(Y, W_1, W_2, \mathbf{Z}, \boldsymbol\theta, f^*_{X|\mathbf{Z}})\}/\partial\boldsymbol\theta$$

is the score vector for $\boldsymbol\theta$. A simple approach to solving integral equations is discretization, which is mathematically equivalent to approximating $f_{X|\mathbf{Z}}(x|\mathbf{z})$ with a discrete distribution with mass at L points $0 < x_1 < \cdots < x_L < 1$ with the corresponding weights $d_1, \ldots, d_L$. We write

$$f^*_{X|\mathbf{Z}}(x|\mathbf{z}) = \sum_{j=1}^L d_j \mathbf{1}(x = x_j),$$

where $d_j \geqslant 0$ and $d_1 + \cdots + d_L = 1$.

The joint density of Y, W_1, W_2 conditional on $X = x_j, \mathbf{Z} = \mathbf{z}$ is

$$f_{a,Y,W_1,W_2|X=x_j,\mathbf{Z}}(y, w_1, w_2, x_j, \mathbf{z}, \boldsymbol\theta) = f_{Y|X,\mathbf{Z}}(y|x_j, \mathbf{z}, \boldsymbol\beta) f_U\left\{\frac{w_1 - x_j}{\mathbf{B}(x_j)^\top\boldsymbol\gamma}\right\} f_U\left\{\frac{w_2 - x_j}{\mathbf{B}(x_j)^\top\boldsymbol\gamma}\right\}\frac{1}{\{\mathbf{B}(x_j)^\top\boldsymbol\gamma\}^2}.$$

Thus we obtain

$$\mathbf{S}^*_{a,\boldsymbol\beta}(y, w_1, w_2, \mathbf{z}, \boldsymbol\theta, f^*_{X|\mathbf{Z}}) = \frac{\sum_{j=1}^L \left[\{\partial f_{Y|X,\mathbf{Z}}(y|x_j, \mathbf{z}, \boldsymbol\beta)/\partial\boldsymbol\beta\}/f_{Y|X,\mathbf{Z}}(y|x_j, \mathbf{z}, \boldsymbol\beta)\right] f_{a,Y,W_1,W_2|X=x_j,\mathbf{Z}}(y, w_1, w_2, x_j, \mathbf{z}, \boldsymbol\theta)d_j}{\sum_{j=1}^L f_{a,Y,W_1,W_2|X=x_j,\mathbf{Z}}(y, w_1, w_2, x_j, \mathbf{z}, \boldsymbol\theta)d_j},$$

$$\mathbf{S}^*_{a,\boldsymbol\gamma}(y, w_1, w_2, \mathbf{z}, \boldsymbol\theta, f^*_{X|\mathbf{Z}}) = \frac{\sum_{j=1}^L \left[-V(u_{j1}, u_{j2})\mathbf{B}(x_j)/\{\mathbf{B}(x_j)^\top\boldsymbol\gamma\}\right] f_{a,Y,W_1,W_2|X=x_j,\mathbf{Z}}(y, w_1, w_2, x_j, \mathbf{z}, \boldsymbol\theta)d_j}{\sum_{j=1}^L f_{a,Y,W_1,W_2|X=x_j,\mathbf{Z}}(y, w_1, w_2, x_j, \mathbf{z}, \boldsymbol\theta)d_j},$$

and

$$\mathrm{E}_a^*\{\mathbf{a}(X,\mathbf{Z})|Y,W_1,W_2,\mathbf{Z},\boldsymbol{\theta}\}=\frac{\sum_{j=1}^{L}\mathbf{a}(x_j,\mathbf{z})f_{a,Y,W_1,W_2|X=x_j,\mathbf{z}}(y,w_1,w_2,x_j,\mathbf{z},\boldsymbol{\theta})d_j}{\sum_{j=1}^{L}f_{a,Y,W_1,W_2|X=x_j,\mathbf{z}}(y,w_1,w_2,x_j,\mathbf{z},\boldsymbol{\theta})d_j}, \tag{9}$$

where $u_{jk}=(w_k-x_j)/\{\mathbf{B}(x_j)^\top\boldsymbol{\gamma}\}$, for $j\in\{1,\ldots,L\}$ and $k\in\{1,2\}$. Note that

$$\mathbf{S}_{a,\boldsymbol{\theta}}^*(y,w_1,w_2,\mathbf{z},\boldsymbol{\theta},f_{X|\mathbf{Z}}^*)=\{\mathbf{S}_{a,\boldsymbol{\beta}}^*(y,w_1,w_2,\mathbf{z},\boldsymbol{\theta},f_{X|\mathbf{Z}}^*)^\top,\mathbf{S}_{a,\boldsymbol{\gamma}}^*(y,w_1,w_2,\mathbf{z},\boldsymbol{\theta},f_{X|\mathbf{Z}}^*)^\top\}^\top.$$

At any $\mathbf{Z}$, let $\mathbf{A}^{(\mathbf{Z})}(\boldsymbol{\theta})$ be a $L\times L$ matrix with its (i,j) entry

$$A_{ij}^{(\mathbf{Z})}(\boldsymbol{\theta})=\int\frac{f_{a,Y,W_1,W_2|X=x_j,\mathbf{Z}}(y,w_1,w_2,x_j,\mathbf{z},\boldsymbol{\theta})d_j}{\sum_{j=1}^{L}f_{a,Y,W_1,W_2|X=x_j,\mathbf{Z}}(y,w_1,w_2,x_j,\mathbf{z},\boldsymbol{\theta})d_j}f_{a,Y,W_1,W_2|X=x_i,\mathbf{Z}}(y,w_1,w_2,x_i,\mathbf{z},\boldsymbol{\theta})dydw_1dw_2.$$

Define $\mathbf{H}^{(\mathbf{Z})}(\boldsymbol{\theta})$ as a $(p+d_{\boldsymbol{\gamma}})\times L$ matrix whose ith column is given by

$$\mathbf{H}_i^{(\mathbf{Z})}(\boldsymbol{\theta})=\int\mathbf{S}_{a,\boldsymbol{\theta}}^*(y,w_1,w_2,\mathbf{z},\boldsymbol{\theta},f_{X|\mathbf{Z}}^*)f_{Y,W_1,W_2|X=x_i,\mathbf{Z}}(y,w_1,w_2,x_i,\mathbf{z},\boldsymbol{\theta})dydw_1dw_2.$$

Let $\mathbf{a}^{(\mathbf{Z})}=(\mathbf{a}(x_1,\mathbf{z}),\ldots,\mathbf{a}(x_L,\mathbf{z}))$. Then we obtain $\mathbf{H}^{(\mathbf{Z})}(\boldsymbol{\theta})=\mathbf{a}^{(\mathbf{Z})}\ \{\mathbf{A}^{(\mathbf{Z})}(\boldsymbol{\theta})\}^\top$ and $\mathbf{a}^{(\mathbf{Z})}=\mathbf{H}^{(\mathbf{Z})}(\boldsymbol{\theta})[\{\mathbf{A}^{(\mathbf{Z})}(\boldsymbol{\theta})\}^\top]^{-1}$, as long as $\mathbf{A}^{(\mathbf{Z})}(\boldsymbol{\theta})$ is nonsingular. To emphasize the dependence of the resulting $\mathbf{a}(X,\mathbf{Z})$ on $\boldsymbol{\theta}$, we write the solution as $\mathbf{a}(X,\mathbf{Z},\boldsymbol{\theta})$. This allows us to form $\mathrm{E}_a^*\{\mathbf{a}(X,\mathbf{Z},\boldsymbol{\theta})|Y,W_1,W_2,\mathbf{Z},\boldsymbol{\theta}\}$ using (9). The resulting estimating equation for $\boldsymbol{\theta}$ is thus

$$\frac{1}{n}\sum_{i=1}^{n}\left[\mathbf{S}_{a,\boldsymbol{\theta}}^*(y_i,w_{i1},w_{12},\mathbf{z}_i,\boldsymbol{\theta},f_{X|\mathbf{Z}}^*)-\mathrm{E}_a^*\{\mathbf{a}(X,\mathbf{Z}_i,\boldsymbol{\theta})|Y_i,W_{i1},W_{i2},\mathbf{Z}_i,\boldsymbol{\theta}\}\right]=\mathbf{0}. \tag{10}$$

By solving (10), we can obtain semiparametric estimator $\widehat{\boldsymbol{\theta}}_n=(\widehat{\boldsymbol{\beta}}_n{}^\top,\widehat{\boldsymbol{\gamma}}_n{}^\top)^\top$.

6. Empirical studies

6.1. Simulation studies

We conducted two simulation studies to investigate the finite-sample performance of the proposed method. We also compared the results of our method with those of the method that ignores the measurement error and the method that assumes the variance of the measurement error is constant. We set the sample size $n=1000$ and generated 1000 samples in each simulation study. The error-prone covariate X_i is generated from the uniform distribution from -2.7 to 0.7, and the error-free covariate Z_i is a Bernoulli random variable independent of X_i with success probability 0.5. Moreover, U_{i1} and U_{i2} are generated from independent standard normal distributions. We then formed $W_{ik}=X_i+\sigma(X_i)U_{ik}$, for each $k\in\{1,2\}$, and $i\in\{1,\ldots,n\}$.

We set discretization points of X to be $x_j=3.4j/L-2.7$ for $j\in\{1,\ldots,L\}$ and set the working model

$$f_{X|\mathbf{Z}}^*(x|\mathbf{z})=\sum_{j=1}^{L}d_j\mathbf{1}(x=x_j).$$

Two different working models are considered. In the first model, $d_j=1/L$, corresponding to a uniform working model. In the second model,

$$d_j=\frac{\phi\{(x_j+1)/3.4\}}{\sum_{j=1}^{L}\phi\{(x_j+1)/3.4\}},$$

where ϕ is the density of standard normal distribution, corresponding to a normal working model.

In the first simulation study, the response Y_i is generated from a logistic regression model

$$\mathrm{logit}\{\Pr(Y_i=1|X_i,Z_i)\}=\beta_0+\beta_1X_i+\beta_2Z_i,$$

with true values $\boldsymbol{\beta}=(\beta_0,\beta_1,\beta_2)^\top=(0.5,0.2,-0.2)^\top$ and the true $\sigma(X_i)=(X_i^2+3)/13.5$. We used quadratic splines with six interior knots to approximate $\sigma(x)$ and set L to be 20. We computed the sample mean and standard deviation of the estimates $\widehat{\boldsymbol{\beta}}_n$ over 1000 data sets and estimated the asymptotic covariance matrix using sandwich formula $\widehat{\mathbf{V}}=\widehat{\mathbf{Q}}^{-1}\widehat{\Sigma}(\widehat{\mathbf{Q}}^{-1})^\top$, where

$$\widehat{\mathbf{Q}}=\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\widehat{\boldsymbol{\beta}}_n{}^\top}\mathbf{S}_{\mathrm{eff}}^*(Y_i,W_{i1},W_{i2},Z_i,\widehat{\boldsymbol{\beta}}_n,\widehat{\boldsymbol{\gamma}}_n,f_{X|\mathbf{Z}}^*),$$

Table 1
Results of the first simulation study.

Truth	β_0 0.5		β_1 0.2		β_2 −0.2	
	No measurement error					
Mean	0.4858		0.1869		−0.1970	
Emp sd	0.1193		0.0629		0.1341	
Est sd	0.1115		0.0630		0.1281	
Emp cov	93.2%		94.0%		94.7%	
Truth	β_0 0.5	β_1 0.2	β_2 −0.2	β_0 0.5	β_1 0.2	β_2 −0.2
	Homoscedastic measurement error					
	Uniform working model			Normal working model		
Mean	0.5049	0.2058	−0.1970	0.5049	0.2058	−0.1970
Emp sd	0.1220	0.0678	0.1355	0.1220	0.0678	0.1355
Est sd	0.6240	0.1844	0.8492	0.6484	0.1914	0.8875
Emp cov	96.2%	90.3%	96.6%	96.4%	90.6%	96.6%
	Heteroscedastic measurement error					
	Uniform working model			Normal working model		
Mean	0.5024	0.2038	−0.1948	0.5032	0.2043	−0.1954
Emp sd	0.1183	0.0665	0.1317	0.1180	0.0668	0.1318
Est sd	0.1149	0.0680	0.1283	0.1149	0.0680	0.1283
Emp cov	95.3%	95.3%	94.8%	94.7%	95.0%	94.9%

In Table 1, "emp sd" denotes the empirical standard deviation of the estimates, "est sd" denotes the estimated asymptotic standard deviation and "emp cov" denotes the empirical coverage of the estimated 95% confidence intervals.

and

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{S}^*_{\rm eff}(Y_i, W_{i1}, W_{i2}, Z_i, \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\gamma}}_n, f^*_{X|\mathbf{Z}})\mathbf{S}^*_{\rm eff}(Y_i, W_{i1}, W_{i2}, Z_i, \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\gamma}}_n, f^*_{X|\mathbf{Z}})^\top.$$

Further, 95% confidence intervals for β_0, β_1, β_2 are constructed in each simulated data set based on the asymptotic normal distribution of $\widehat{\boldsymbol{\beta}}_n$ to compute the empirical percentage covering the true values. We compared the results of our method with those of the naive logistic regression method, which ignores the measurement error and existing estimating equation method, which incorrectly assumes a constant error variance, i.e., $W = X + \sigma U$, $\sigma > 0$. The estimating equations are constructed based on efficient scores that are derived by treating σ as a nuisance parameter and adopting working model $f^*_{X|\mathbf{Z}}(x|\mathbf{z})$. The results of the first simulation study are summarized in Table 1, where median of the estimated standard deviation was reported as "est sd".

In the second method assuming homoscedastic measurement error, the estimated standard deviation of measurement error is $\widehat{\sigma} = 0.4066$ under both uniform and normal working models. Fig. 1 shows the performance of the B-spline approximation of the nuisance function $\sigma(x)$ of our method under different working models of $f_{X|\mathbf{Z}}(x|\mathbf{z})$. The solid line represents the true function $\sigma(x)$, while the three dashed lines represent the 1/4, 1/2 and 3/4 sample quantiles of the estimated function $\mathbf{B}(x)^\top\widehat{\boldsymbol{\gamma}}_n$. Note that the median curve is almost overlapping with the true $\sigma(x)$.

Following the comment of a referee, we also experimented with heteroscedastic Laplace measurement errors. The results are given in Table 2 and Fig. 2. The performance of our method is still satisfactory.

In the second simulation study, we consider a relatively more complex model, where $\sigma(X_i) = 0.4\exp\{-0.15(X_i+1)^2\}$. We generate the response Y_i from a quadratic logistic regression model, viz.

$$\text{logit}\{\Pr(Y_i = 1|X_i, Z_i)\} = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 Z_i,$$

with true values for $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^\top = (-0.8, 0.5, 0.2, 2.0)^\top$. We used similar working distribution models $f^*_{X|Z}(x|\mathbf{z})$ with $L = 20$. A quadratic spline with seven knots was used to approximate the nuisance function $\sigma(x)$. The results of the three different methods in the second simulation study are summarized in Table 3.

In the second method, the estimated standard deviation of the measurement error is $\widehat{\sigma} = 0.3514$ under both uniform and normal working models, while Fig. 3 shows the performance of the B-splines approximation of the standard deviation function $\sigma(x)$ of the measurement error under different working models in our method.

In the simulation studies, our method performs well in both the simple and the complex model settings. The estimates have very small biases, the median of estimated standard deviations closely approximate the empirical standard deviations, and the empirical coverages of the estimated 95% confidence interval are close to the nominal level. Further, the results of using different working models are similar in both simulations, which suggests the insensitivity of our method to the misspecification of $f_{X|\mathbf{Z}}(x|\mathbf{z})$. We also see that ignoring the measurement error or incorrectly assuming homoscedastic measurement error can cause bias issues in estimation, especially the estimation of the coefficients associated with the unobservable covariate, and affect the quality of influence.

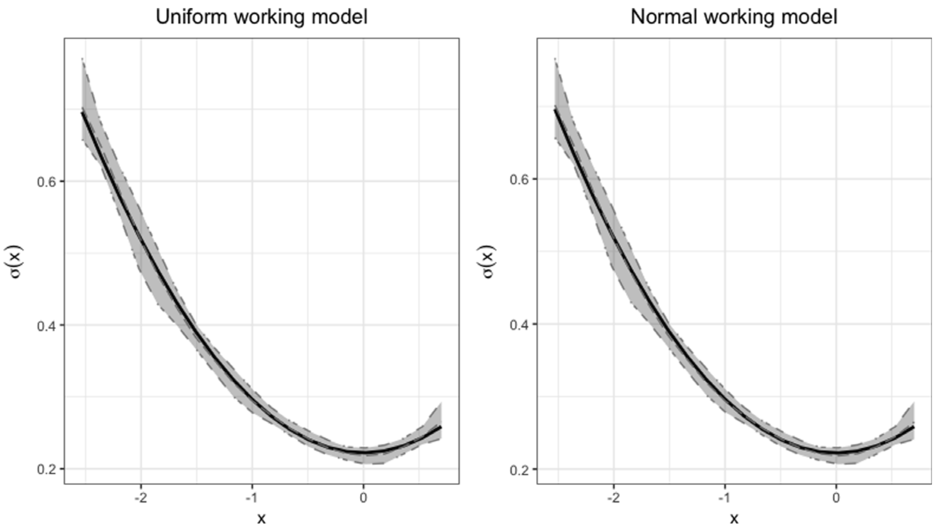

Fig. 1. Performance of the B-splines approximation of $\sigma(x)$ under two working models of $f_{X|\mathbf{Z}}$ in the first simulation study.

Table 2
Results of the first simulation study with heteroscedastic Laplace measurement errors.

	β_0		β_1		β_2	
Truth	0.5		0.2		−0.2	
	No measurement error					
Mean	0.4906		0.1852		−0.2046	
Emp sd	0.1130		0.0619		0.1252	
Est sd	0.1117		0.0631		0.1281	
Emp cov	95.2%		94.6%		95.4%	
	β_0	β_1	β_2	β_0	β_1	β_2
Truth	0.5	0.2	−0.2	0.5	0.2	−0.2
	Homoscedastic measurement errors					
	Uniform working model			Normal working model		
Mean	0.5054	0.2088	−0.2056	0.5053	0.2088	−0.2054
Emp sd	0.1168	0.0703	0.1298	0.1172	0.0703	0.1302
Est sd	0.1521	0.0553	0.1675	0.1518	0.0557	0.1705
Emp cov	93.8%	80.6%	93.9%	93.3%	80.1%	93.9%
	Heteroscedastic measurement error					
	Uniform working model			Normal working model		
Mean	0.5051	0.2037	−0.2046	0.5042	0.2035	−0.2041
Emp sd	0.1168	0.0677	0.1303	0.1162	0.0675	0.1298
Est sd	0.1147	0.0675	0.1282	0.1147	0.0675	0.1282
Emp cov	94.7%	94.9%	94.2%	95.0%	95.0%	94.3%

In Table 2, “emp sd” denotes the empirical standard deviation of the estimates, “est sd” denotes the estimated asymptotic standard deviation and “emp cov” denotes the empirical coverage of the estimated 95% confidence intervals.

6.2. Data analysis

In this section, we illustrate our method by analyzing the data from the Childhood Asthma Management Program (CAMP). CAMP is a longitudinal study designed to explore the long-term impact of several daily treatments for mild to moderate asthma in children.

We formed the outcome variable Y_i based on the average asthma symptoms (amsym) recorded in daily record of eight months. `amsym` is a binary variable indicating the severity of the asthma symptoms for a child (indexed by i). If the average `amsym` is greater than 0.5, then the outcome variable Y_i equals 1, implying moderate asthma symptoms. Otherwise, the outcome variable Y_i equals 0, implying mild asthma. The FEV1/FVC ratio (`preff`) is an important index used in diagnosis of asthma, which represents the proportion of a person's vital capability to expire in the first second of forced expiration to the full vital capacity. Four measurements of `preff` were recorded during the 8-month study for each child.

We let X_i be the unobserved `preff` and W_i be the average of four measurements with heteroscedastic measurement error. Other error-free variables $\mathbf{Z}_i$ are gender, age at baseline, and treatment group. Three treatment groups were included in the

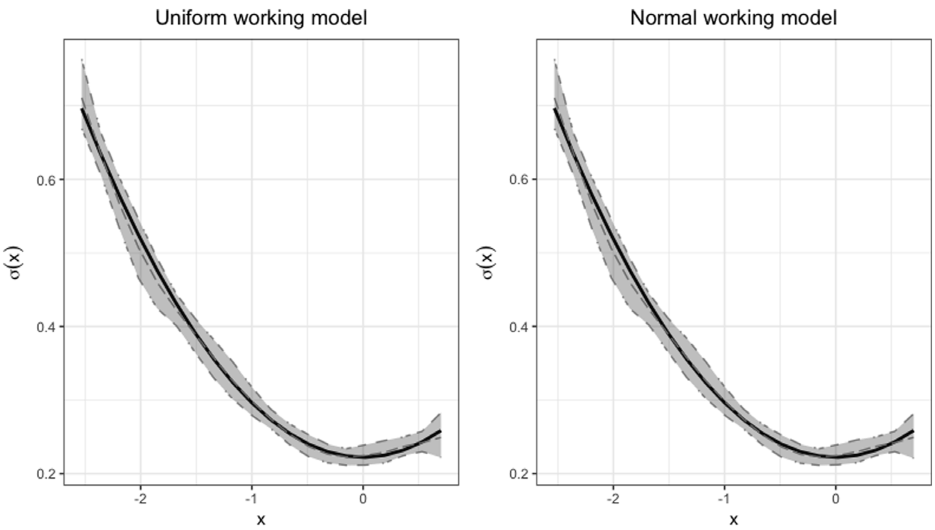

Fig. 2. Performance of the B-splines approximation of $\sigma(x)$ under two working models of $f_{X|\mathbf{Z}}$ in the first simulation study with heteroscedastic Laplace measurement errors.

Table 3
Results of the second simulation study.

	β_0		β_1		β_2		β_3	
Truth	−0.8		0.5		0.2		2.0	
	No measurement error							
Mean	−0.8201		0.4172		0.1611		2.0107	
Emp sd	0.1236		0.1648		0.0749		0.1436	
Est sd	0.1206		0.1675		0.0756		0.1441	
Emp cov	94.3%		92.8%		91.8%		94.1%	
	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3
Truth	−0.8	0.5	0.2	2.0	−0.8	0.5	0.2	2.0
	Homoscedastic measurement error							
	Uniform working model				Normal working model			
Mean	−0.8042	0.5294	0.2143	2.0151	−0.8042	0.5294	0.2144	2.0151
Emp sd	0.1263	0.2133	0.0998	0.1447	0.1263	0.2133	0.0998	0.1447
Est sd	0.9700	0.7301	0.1419	1.1875	1.0717	0.7767	0.1555	1.2770
Emp cov	95.6%	92.1%	83.2%	99.8%	95.7%	91.9%	83.6%	99.8%
	Heteroscedastic measurement error							
	Uniform working model				Normal working model			
Mean	−0.8047	0.5097	0.2044	2.0132	−0.8045	0.5109	0.2050	2.0136
Emp sd	0.1220	0.1990	0.0915	0.1436	0.1223	0.2004	0.0930	0.1420
Est sd	0.1249	0.2105	0.0970	0.1453	0.1249	0.2112	0.0973	0.1452
Emp cov	96.4%	96.3%	96.5%	94.8%	96.1%	96.4%	96.6%	94.9%

In Table 3, "emp sd" denotes the empirical standard deviation of the estimates, "est sd" denotes the estimated asymptotic standard deviation and "emp cov" denotes the empirical coverage of the estimated 95% confidence intervals.

study, and we coded them using two dummy variables `trt1` and `trt2`, where `trt1` $=1$ if the treatment is budesonide, `trt2` $=1$ if the treatment is nedocromil, and `trt1` = `trt2` $=0$ if the treatment is placebo.

The data set consisted of 737 children for whom $(Y_i, W_i, \mathbf{Z}_i)$ were measured. We considered the linear logistic regression model with heteroscedastic measurement error on X_i, viz.

$$\mathrm{logit}\{\Pr(Y_i = 1|X_i, Y_i)\} = \beta_0 + \beta_x X_i + \boldsymbol{\beta}_z \mathbf{Z}_i \quad \text{and} \quad W_i = X_i + \sigma(X_i)U_i,$$

where $\boldsymbol{\beta}_z = (\beta_{z1}, \beta_{z2}, \beta_{z3}, \beta_{z4})^\top$, $\mathbf{Z}_i$ is a vector of gender, age, `trt1`, `trt2` for the ith child and $U_i \sim \mathcal{N}(0, 1)$. We used a uniform working model for the distribution of $\mathbf{X}$ and adopted $L = 20$ discretization points in the implementation. We used quadratic splines with 6 knots to approximate $\sigma(X)$.

We further compared the results from our analysis with those of the "naive" logistic regression method, which ignores the existence of measurement error, and the estimating equations approach, which treats the measurement error variance as constant. The results from all three methods are summarized in Table 4. In the second method which assumes constant

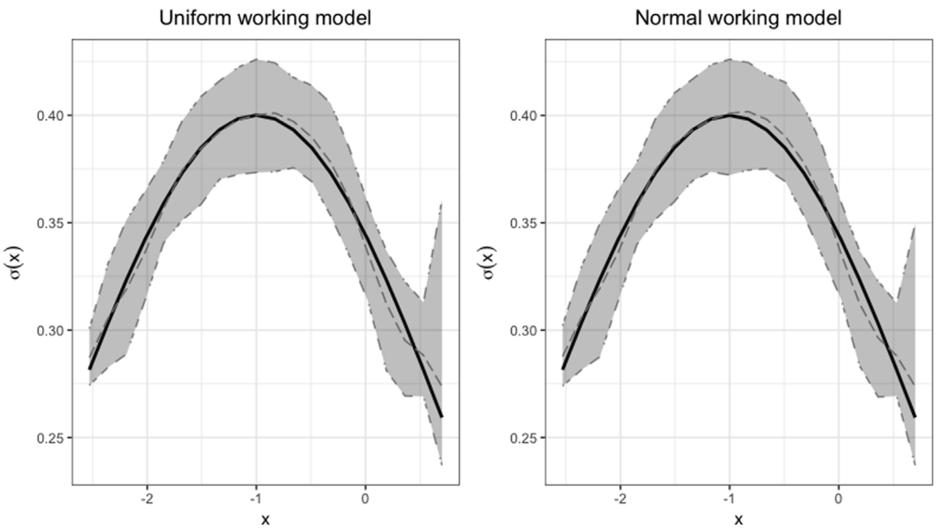

Fig. 3. Performance of the B-splines approximation of $\sigma(x)$ under two working models of $f_{X|\mathbf{Z}}$ in the second simulation study.

Table 4
Analysis of the CAMP data under uniform working model.

	β_0 intercept	β_x preff	β_{z1} gender	β_{z2} age	β_{z3} trt1	β_{z4} trt2
	No measurement error					
Est	0.3282	−2.8035	0.3774	0.0537	−0.7794	−0.3251
Est sd	0.5222	0.5297	0.1737	0.0411	0.2099	0.1978
p-value	0.5297	1.2e − 07	0.0298	0.1920	0.0002	0.1001
	Homoscedastic measurement error					
Est	0.4424	−2.9674	0.3859	0.0501	−0.7725	−0.3248
Est sd	0.4247	0.3518	3.8812	0.3296	0.5025	5.2613
p-value	0.2975	0.0000	0.9208	0.8791	0.1242	0.9508
	Heteroscedastic measurement error					
Est	0.4224	−2.9397	0.3846	0.0507	−0.7741	−0.3248
Est sd	0.1776	0.2655	3.2765	0.2263	0.4010	3.8781
p-value	0.0174	0.0000	0.9066	0.8226	0.0861	0.9332

measurement error variance, the estimated standard deviation is $\widehat{\sigma} = 0.0753$. The estimated heteroscedastic error standard deviation function $\widehat{\sigma}(x)$ in the third method is given in Fig. 4.

In the naive logistic regression model that ignores the measurement error, besides the error-prone variable `preff`, we also detect variables gender and `trt1` to significantly influence the severity of asthma at the significance level $\alpha = 0.05$. In the second and third methods considering measurement errors, only the error-prone variable `preff` is significant, with the same sign and slightly greater absolute value as the estimate. The variable `trt1` is almost significant with p-value < 0.1 in the heteroscedastic measurement error model, but gender is not significant at all. This difference indicates that ignoring the measurement error could lead to misleading results. Further, the B-spline approximation of $\sigma(x)$ fluctuates with respect to x and is not always close to $\widehat{\sigma}$. Hence the assumption about constant measurement error variance may be too restrictive for this data set. From Fig. 4, we can see that when x is close to 0 or 1, the measurement error variance tends to be larger.

We also tried using a normal working model for the distribution of X in the second and third methods. The estimates $\widehat{\boldsymbol{\beta}}$ are almost the same as those under the uniform working model. Specifically, with the normal working model, in the second method $\widehat{\boldsymbol{\beta}} = (0.4435, -2.9689, 0.3860, 0.0501, -0.7725, -0.3248)$, while in the third method $\widehat{\boldsymbol{\beta}} = (0.4244, -2.9416, 0.3847, 0.0507, 0.7741, -0.3248)$. However, due to the near singularity of the matrix $\widehat{\mathbf{Q}}$ in the sandwich formula, the inference results are inaccurate, so we did not pursue the normal working mode further.

7. Discussion

In this paper, we proposed a new method in the area of general measurement error model with heteroscedastic error. The method can be applied to any parametric model with an unspecified heteroscedastic measurement error variance structure. We have assumed the distribution of the error U to be known, but not restricted to the normal distribution. In practice, the

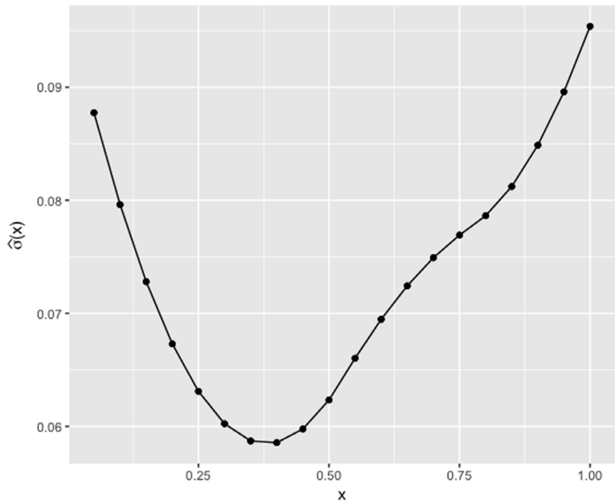

Fig. 4. B-spline approximation of $\sigma(x)$ under uniform working model.

density $f_U(u)$ needs to be determined using external information such as validation data. If $f_U(u)$ cannot be determined, one may be able to approximate it with $f_U(u) \approx \exp\{\mathbf{B}(u)^\top \boldsymbol{\alpha}\}/\int \exp\{\mathbf{B}(v)^\top \boldsymbol{\alpha}\}dv$ and estimate $\boldsymbol{\alpha}$ together with other parameters, provided that the problem is still identifiable. The identifiability issue in this case is difficult and warrants further research. Overall, identifiability in measurement error models is difficult and is often established case by case. We have proved the identifiability of the linear model with heteroscedastic normal and Laplace measurement errors rigorously. We have also shown the identifiability of $\boldsymbol{\beta}$ for a logistic model with heteroscedastic normal measurement errors. In addition, $f_{X|\mathbf{Z}}(x|\mathbf{z})$ is estimable [14]. Then $\sigma(X)$ is identifiable based on $W_1 - W_2 = \sigma(X)(U_1 - U_2)$, where the distribution of $W_1 - W_2$ is estimable and the distribution of $U_1 - U_2$ is known. Therefore, the whole problem for the logistic model with heteroscedastic normal measurement errors is identifiable.

We have assumed that the unobservable covariate X is a scalar for simplicity of presentation. If there is more than one unobservable covariate $\mathbf{X} = (X_1, \ldots, X_m)$ in the model and the measurement W_j given X_j is independent of other unobservable covariates, conceptually we can use $\mathbf{B}(X_j)^\top \boldsymbol{\gamma}_j$ to approximate the jth unknown function $\sigma_j(X_j)$, then append the estimating equations with these additional estimating equations obtained from the corresponding score functions for $\boldsymbol{\gamma}_j$, for $j \in \{1, \ldots, m\}$. The computation may be more challenging. Additionally, for simplicity, we used $\mathbf{B}(X)^\top \boldsymbol{\gamma}$ to approximate $\sigma(X)$ in our implementation. To ensure positivity of $\sigma(X)$, we could instead use $\exp\{\mathbf{B}(X)^\top \boldsymbol{\gamma}\}$ to approximate $\sigma(X)$. The theoretical properties of $\widehat{\boldsymbol{\beta}}_n$ would not change.

More generally, we would like to point out that multiple roots is a potential issue for estimating equations approaches. Choosing the correct estimator from multiple roots of the estimating equations may not be straightforward. Heyde and Morton [11] invented a criteria to discriminate the consistent estimator from multiple roots of estimating equations. More discussion can be found in [9] and Section 13.3 of [10]. In practice, using empirical knowledge or using estimates from more primitive but simpler methods to form starting values for our method can be a sensible option.

Acknowledgments

The authors are grateful to the Editor-in-Chief, an Associate Editor and the referees for comments and suggestions that led to significant improvements. The authors thank Ms. Amanda Applegate for her help on the preparation of this paper. This research was supported by National Science Foundation (NSF) grants DMS 1608540 and DMS 1820702, National Natural Science Foundation of China (NNSFC) grants 11690014 and 11690015, NIH grants HL138306-01, NS073671-01, National Institute on Drug Abuse (NIDA) grants P50 DA039838 and National Institute of Allergy and Infectious Diseases (NIAID) U19AI089672 The content is solely the responsibility of the authors and does not necessarily represent the official views of the NSF, the NNSFC, the NIDA, the NIAID or the NIH.

Appendix

A.1. Proof of the identifiability of the linear measurement error model

Computing the first and second order moments of Y, W_1 and W_2 given Z, we have

$$\mathrm{E}(Y|\mathbf{Z}) = \beta_0 + \beta_1 E(X|\mathbf{Z}) + \mathbf{Z}^\top \boldsymbol{\beta}_2, \quad \mathrm{E}(W_1|\mathbf{Z}) = \mathrm{E}(X|\mathbf{Z}),$$

$$\mathrm{var}(Y|\mathbf{Z}) = \beta_1^2 \mathrm{var}(X|\mathbf{Z}) + \sigma_\epsilon^2, \quad \mathrm{var}(W_1 + W_2|\mathbf{Z}) = 4\mathrm{var}(X|\mathbf{Z}) + 2E\{\sigma^2(X)|\mathbf{Z}\},$$
$$\mathrm{var}(W_1 - W_2|\mathbf{Z}) = 2\mathrm{E}\{\sigma^2(X)|\mathbf{Z}\}, \quad \mathrm{cov}(W_1, Y|\mathbf{Z}) = \beta_1 \mathrm{var}(X|\mathbf{Z}).$$

From the above equations, we get $\mathrm{var}(X|\mathbf{Z}) = \{\mathrm{var}(W_1 + W_2|\mathbf{Z}) - \mathrm{var}(W_1 - W_2|\mathbf{Z})\}/4$, hence $\mathrm{var}(X|\mathbf{Z})$ is identifiable. Subsequently, we have $\beta_1 = 4\mathrm{cov}(W_1, Y|\mathbf{Z})/\{\mathrm{var}(W_1 + W_2|\mathbf{Z}) - \mathrm{var}(W_1 - W_2|\mathbf{Z})\}$, hence β_1 is also identifiable. From

$$\sigma_\epsilon^2 = \mathrm{var}(Y|\mathbf{Z}) - \beta_1^2 \mathrm{var}(X|\mathbf{Z}) = \mathrm{var}(Y|\mathbf{Z}) - \frac{4\mathrm{cov}^2(W_1, Y|\mathbf{Z})}{\mathrm{var}(W_1 + W_2|\mathbf{Z}) - \mathrm{var}(W_1 - W_2|\mathbf{Z})},$$

we obtain the identifiability of σ_ϵ^2. Further, we have

$$\beta_0 + \mathbf{Z}^\top \boldsymbol{\beta}_2 = \mathrm{E}(Y|\mathbf{Z}) - \beta_1 \mathrm{E}(X|\mathbf{Z}) = \mathrm{E}(Y|\mathbf{Z}) - \frac{4\mathrm{cov}(W_1, Y|\mathbf{Z})\mathrm{E}(W_1|\mathbf{Z})}{\mathrm{var}(W_1 + W_2|\mathbf{Z}) - \mathrm{var}(W_1 - W_2|\mathbf{Z})}.$$

Then

$$\beta_0 = \mathrm{E}(Y|\mathbf{Z} = \mathbf{0}) - \frac{4\mathrm{cov}(W_1, Y|\mathbf{Z} = \mathbf{0})\mathrm{E}(W_1|\mathbf{Z} = \mathbf{0})}{\mathrm{var}(W_1 + W_2|\mathbf{Z} = \mathbf{0}) - \mathrm{var}(W_1 - W_2|\mathbf{Z} = \mathbf{0})},$$
$$\boldsymbol{\beta}_2 = \{\mathrm{E}(\mathbf{Z}\mathbf{Z}^\top)\}^{-1} \mathrm{E}\left\{\mathbf{Z}\mathrm{E}(Y|\mathbf{Z}) - \frac{4\mathbf{Z}\mathrm{cov}(W_1, Y|\mathbf{Z})\mathrm{E}(W_1|\mathbf{Z})}{\mathrm{var}(W_1 + W_2|\mathbf{Z}) - \mathrm{var}(W_1 - W_2|\mathbf{Z})} - \beta_0 \mathbf{Z}\right\}.$$

Therefore, β_0 and $\boldsymbol{\beta}_2$ are also identifiable.

Having obtained the identifiability of β_0, β_1, $\boldsymbol{\beta}_2$ and σ_ϵ, we proceed to prove the identifiability of the unknown function $\sigma(x)$ and the conditional density $f_{X|\mathbf{Z}}(x|\mathbf{z})$. We first consider the normal measurement errors. Given the observed data $(Y, W_1, W_2, \mathbf{Z})$ and the identifiability of β_0, β_1, $\boldsymbol{\beta}_2$ and σ_ϵ, assume the model is still not identifiable. Then, there exist $\{\sigma(x), f_{X|\mathbf{Z}}(x|\mathbf{z})\}$ and $\{\widetilde{\sigma}(x), \widetilde{f}_{X|\mathbf{Z}}(x|\mathbf{z})\}$ that satisfy

$$\int_{-\infty}^{+\infty} f_{Y|X,\mathbf{Z}}(y|x, \mathbf{z}, \beta_0, \beta_1, \boldsymbol{\beta}_2, \sigma_\epsilon) f_U\left\{\frac{w_1 - x}{\sigma(x)}\right\} f_U\left\{\frac{w_2 - x}{\sigma(x)}\right\} f_{X|\mathbf{Z}}(x|\mathbf{z}) \frac{1}{\sigma^2(x)}\, dx$$
$$= \int_{-\infty}^{+\infty} f_{Y|X,\mathbf{Z}}(y|x, \mathbf{z}, \beta_0, \beta_1, \boldsymbol{\beta}_2, \sigma_\epsilon) f_U\left\{\frac{w_1 - x}{\widetilde{\sigma}(x)}\right\} f_U\left\{\frac{w_2 - x}{\widetilde{\sigma}(x)}\right\} \widetilde{f}_{X|\mathbf{Z}}(x|\mathbf{z}) \frac{1}{\widetilde{\sigma}^2(x)}\, dx. \tag{A.1}$$

We rewrite (A.1) as a convolution, viz.

$$g \circ h = \int_{-\infty}^{+\infty} g(y - t)h(t)dt = \int_{-\infty}^{+\infty} g(y - t)\widetilde{h}(t)dt = g \circ \widetilde{h},$$

where $t = \beta_0 + x\beta_1 + \mathbf{z}^\top \boldsymbol{\beta}_2$,

$$g(y - t) = \exp\left\{-\frac{(y - t)^2}{2\sigma_\epsilon^2}\right\},$$
$$h(t) = \exp\left[-\frac{\{w_1 - (t - \beta_0 - \mathbf{z}^\top \boldsymbol{\beta}_2)/\beta_1\}^2 + \{w_2 - (t - \beta_0 - \mathbf{z}^\top \boldsymbol{\beta}_2)/\beta_1\}^2}{2\sigma^2\{(t - \beta_0 - \mathbf{z}^\top \boldsymbol{\beta}_2)/\beta_1\}}\right] \frac{f_{X|\mathbf{Z}}\{(t - \beta_0 - \mathbf{z}^\top \boldsymbol{\beta}_2)/\beta_1|\mathbf{z}\}}{\beta_1 \sigma^2\{(t - \beta_0 - \mathbf{z}^\top \boldsymbol{\beta}_2)/\beta_1\}},$$
$$\widetilde{h}(t) = \exp\left[-\frac{\{w_1 - (t - \beta_0 - \mathbf{z}^\top \boldsymbol{\beta}_2)/\beta_1\}^2 + \{w_2 - (t - \beta_0 - \mathbf{z}^\top \boldsymbol{\beta}_2)/\beta_1\}^2}{2\widetilde{\sigma}^2\{(t - \beta_0 - \mathbf{z}^\top \boldsymbol{\beta}_2)/\beta_1\}}\right] \frac{\widetilde{f}_{X|\mathbf{Z}}\{(t - \beta_0 - \mathbf{z}^\top \boldsymbol{\beta}_2)/\beta_1|\mathbf{z}\}}{\beta_1 \widetilde{\sigma}^2\{(t - \beta_0 - \mathbf{z}^\top \boldsymbol{\beta}_2)/\beta_1\}}.$$

By the convolution theorem of Fourier transforms, we have $\mathcal{F}(g)\mathcal{F}(h) = \mathcal{F}(g)\mathcal{F}(\widetilde{h})$, then $\mathcal{F}(h) = \mathcal{F}(\widetilde{h})$. Hence $h(t) = \widetilde{h}(t)$ for any $t \in \mathbb{R}$ via the inverse Fourier transformation. Because w_1, w_2 can be any values, this directly leads to $\sigma(x) = \widetilde{\sigma}(x)$ and $f_{X|\mathbf{Z}}(x|\mathbf{z}) = \widetilde{f}_{X|\mathbf{Z}}(x|\mathbf{z})$.

For Laplace measurement errors, we have

$$h(t) = \exp\left[-\frac{|w_1 - (t - \beta_0 - \mathbf{z}^\top \boldsymbol{\beta}_2)/\beta_1| + |w_2 - (t - \beta_0 - \mathbf{z}^\top \boldsymbol{\beta}_2)/\beta_1|}{\sqrt{1/2}\sigma\{(t - \beta_0 - \mathbf{z}^\top \boldsymbol{\beta}_2)/\beta_1\}}\right] \frac{f_{X|\mathbf{Z}}\{(t - \beta_0 - \mathbf{z}^\top \boldsymbol{\beta}_2)/\beta_1|\mathbf{z}\}}{\beta_1 \sigma^2\{(t - \beta_0 - \mathbf{z}^\top \boldsymbol{\beta}_2)/\beta_1\}},$$
$$\widetilde{h}(t) = \exp\left[-\frac{|w_1 - (t - \beta_0 - \mathbf{z}^\top \boldsymbol{\beta}_2)/\beta_1| + |w_2 - (t - \beta_0 - \mathbf{z}^\top \boldsymbol{\beta}_2)/\beta_1|}{\sqrt{1/2}\widetilde{\sigma}\{(t - \beta_0 - \mathbf{z}^\top \boldsymbol{\beta}_2)/\beta_1\}}\right] \frac{\widetilde{f}_{X|\mathbf{Z}}\{(t - \beta_0 - \mathbf{z}^\top \boldsymbol{\beta}_2)/\beta_1|\mathbf{z}\}}{\beta_1 \widetilde{\sigma}^2\{(t - \beta_0 - \mathbf{z}^\top \boldsymbol{\beta}_2)/\beta_1\}}.$$

Note that $\sigma(x) > 0$ and $\widetilde{\sigma}(x) > 0$ for any x. Similar to the proof of normal measurement errors, we have $\sigma(x) = \widetilde{\sigma}(x)$ and $f_{X|\mathbf{Z}}(x|\mathbf{z}) = \widetilde{f}_{X|\mathbf{Z}}(x|\mathbf{z})$. This completes the proof of the identifiability of all components in the model. □

A.2. The derivation of nuisance tangent space

For a parametric model, the nuisance tangent space is the linear space in $\mathcal{H}$ spanned by the nuisance score vector. For semiparametric models, in which the nuisance parameter is infinite-dimensional, the nuisance tangent space is defined as the mean squared closure of all parametric submodel nuisance tangent spaces. The parametric submodel is a true parametric model contained in the semiparametric model. In our original models (1) and (2), the nuisance score vector for the parametric submodel $f_{X|\mathbf{Z}}(x|\mathbf{z},\xi_1)$ is

$$\mathbf{S}_{f_{X|\mathbf{Z}}}(Y,W_1,W_2,\mathbf{Z},\boldsymbol{\beta},\sigma,\xi_1)=\frac{\int\left[\{\partial f_{X|\mathbf{Z}}(x|\mathbf{z},\xi_1)/\partial\xi_1\}/f_{X|\mathbf{Z}}(x)\right]f_{Y,W_1,W_2,X|\mathbf{Z}}(y,w_1,w_2,x,\mathbf{z},\boldsymbol{\beta},\sigma,\xi_1)dx}{\int f_{Y,W_1,W_2,X|\mathbf{Z}}(y,w_1,w_2,x,\mathbf{z},\boldsymbol{\beta},\sigma,\xi_1)dx}$$
$$=\mathrm{E}\left\{\frac{\partial}{\partial\xi_1}f_{X|\mathbf{Z}}(X|\mathbf{Z},\xi_1)\frac{1}{f_{X|\mathbf{Z}}(X|\mathbf{Z},\xi_1)}\,\middle|\,Y,W_1,W_2,\mathbf{Z}\right\},$$

where $\mathrm{E}[\{\partial f_{X|\mathbf{Z}}(x|\mathbf{z},\xi_1)/\partial\xi_1\}/f_{X|\mathbf{Z}}(x|\mathbf{z},\xi_1)|\mathbf{Z}]=0$. Thus the nuisance tangent space with respect to $f_{X|\mathbf{Z}}$ is $\Lambda_{f_{X|\mathbf{Z}}}=\{\mathrm{E}\{\mathbf{a}(X,\mathbf{Z})|Y,W_1,W_2,\mathbf{Z}\}:\mathrm{E}\{\mathbf{a}(X,\mathbf{Z})|\mathbf{Z}\}=\mathbf{0}\}$. Similarly, the nuisance score vector for the parametric submodel $\sigma(x,\xi_2)$ is given as

$$\mathbf{S}_{\sigma}(Y,W_1,W_2,\mathbf{Z},\boldsymbol{\beta},\xi_2,f_{X|\mathbf{Z}})=\frac{\int -V(u_1,u_2)[\{\partial\sigma(x,\xi_2)/\partial\xi_2\}/\sigma(x,\xi_2)]f_{Y,W_1,W_2,X|\mathbf{Z}}(y,w_1,w_2,x,\mathbf{z},\boldsymbol{\beta},\xi_2,f_{X|\mathbf{Z}})dx}{\int f_{Y,W_1,W_2,X|\mathbf{Z}}(y,w_1,w_2,x,\mathbf{z},\boldsymbol{\beta},\xi_2,f_{X|\mathbf{Z}})dx}$$
$$=\mathrm{E}\left\{-V(U_1,U_2)\frac{\partial}{\partial\xi_2}\sigma(X,\xi_2)\frac{1}{\sigma(X,\xi_2)}\,\middle|\,Y,W_1,W_2,\mathbf{Z}\right\},$$

where $\partial\ln\{\sigma(x,\xi_2)\}/\partial\xi_2$ can be any function of x. Thus the nuisance tangent space with respect to σ is $\Lambda_\sigma=\{\mathrm{E}\{V(U_1,U_2)\mathbf{b}(X)|Y,W_1,W_2,\mathbf{Z}\}:\forall\mathbf{b}(X)\}$. It is easy to verify that $\mathrm{E}\{V(U_1,U_2)\}=0$. Since U is independent of X and $\mathbf{Z}$, $\mathrm{E}\{V(U_1,U_2)\mathbf{b}(X)|\mathbf{Z}\}=\mathrm{E}[\mathrm{E}\{V(U_1,U_2)|X,\mathbf{Z}\}\mathbf{b}(X)|\mathbf{Z}]=\mathbf{0}$ for arbitrary $\mathbf{b}(X)$. Thus

$$\mathrm{E}\{V(U_1,U_2)\mathbf{b}(X)|Y,W_1,W_2,\mathbf{Z}\}\in\mathcal{H}.$$

A.3. Nontrivial orthogonal complement of the nuisance tangent space for logistic model with heteroscedastic normal measurement errors

Following the proof in [15], if we can find a nonzero function $h(Y,W_1,W_2,Z)$ such that $\mathrm{E}\{h(Y,W_1,W_2,Z)|X,Z\}=0$ and $\mathrm{E}\{h(Y,W_1,W_2,Z)V(U_1,U_2)|X,Z\}=0$, simultaneously, then the orthogonal complement of the nuisance tangent space is nontrivial. Hence, the root-n estimators exist, which means the parameters are estimable and the problem is identifiable.

Consider the logistic regression model $\operatorname{logit}\Pr(Y=1|X,Z)=\beta_0+\beta_1X+\beta_2Z$, and $W=X+\sigma(X)U$, where U is standard normal. Since Y is binary, any function of Y,W_1,W_2,Z can be written as $h(Y,W_1,W_2,Z)=Yh_1(W_1,W_2,Z)-h_2(W_1,W_2,Z)$. Since Y and W_j are conditionally independent given X and Z for $j\in\{1,2\}$, then $\mathrm{E}\{h(Y,W_1,W_2,Z)|X,Z\}=\mathrm{E}(Y|X,Z)\mathrm{E}\{h_1(W_1,W_2,Z)|X,Z\}-\mathrm{E}\{h_2(W_1,W_2,Z)|X,Z\}$. The conditional expectation $\mathrm{E}\{h(Y,W_1,W_2,Z)|X,Z\}=0$, if

$$\mathrm{E}\{h_1(W_1,W_2,Z)|X,Z\}=\{1+\exp(-\beta_0-\beta_1X-\beta_2Z)\}\mathrm{E}\{h_2(W_1,W_2,Z)|X,Z\}.$$

When the conditional distribution of W_j given $X=x,Z=z$ is $\mathcal{N}[x,\sigma^2(x)]$ for $j\in\{1,2\}$, then standard calculations for normal densities yield

$$\mathrm{E}\{\exp(\beta_1W)|X,Z\}=\exp(\beta_1X)\exp\{\beta_1^2\sigma^2(X)/2\}.$$

Further, since U_1 and U_2 are independent, then W_1-W_2 given $X=x$ and $Z=z$ is $\mathcal{N}[0,2\sigma^2(x)]$. Similarly, we have

$$\mathrm{E}[\exp\{\beta_1(W_1-W_2)/\sqrt{2}\}|X,Z]=\exp\{\beta_1^2\sigma^2(X)/2\}.$$

Let $h_2(W_1,W_2,Z)=\exp(\beta_0+\beta_2Z)\exp(\beta_1W_1)$. Then

$$\mathrm{E}\{h_1(W_1,W_2,Z)|X,Z\}=\mathrm{E}\{h_2(W_1,W_2,Z)|X,Z\}+\exp\{\beta_1^2\sigma^2(X)/2\}.$$

Hence, a nontrivial solution exists by choosing $h_1(W_1,W_2,Z)=h_2(W_1,W_2,Z)+\exp\{\beta_1(W_1-W_2)/\sqrt{2}\}$. It can be verified that this nontrivial solution $h(Y,W_1,W_2,Z)$ also satisfies $\mathrm{E}\{h(Y,W_1,W_2,Z)V(U_1,U_2)|X,Z\}=0$ based on standard calculations. Therefore, the orthogonal complement of the nuisance tangent space is nontrivial and $\boldsymbol{\beta}=(\beta_0,\beta_1,\beta_2)$ is identifiable.

A.4. The derivation of $\mathbf{S}_{a,\mathrm{eff},\boldsymbol{\gamma}}(Y,W_1,W_2,\boldsymbol{\beta},\boldsymbol{\gamma},f_{X|\mathbf{Z}})$

To estimate $\boldsymbol{\gamma}$ at each $\boldsymbol{\beta}$ in the approximate model (3), we treat $\boldsymbol{\beta},\boldsymbol{\gamma}$ as parameters of interest and $f_{X|\mathbf{Z}}$ as a nuisance. The nuisance tangent space is

$$\Lambda_{a,f_{X|\mathbf{Z}}}=\{\mathrm{E}_a\{\mathbf{c}(X,\mathbf{Z})|Y,W_1,W_2,\mathbf{Z}\}:\mathrm{E}_a\{\mathbf{c}(X,\mathbf{Z})|\mathbf{Z}\}=0\},$$

and its orthogonal complement is

$$\{h(Y,W_1,W_2,\mathbf{Z}):\mathrm{E}_a\{h(Y,W_1,W_2,\mathbf{Z})|X,\mathbf{Z}\}=\mathbf{0}\text{ almost surely}\}.$$

The score vector for $\boldsymbol{\gamma}$ is easily verified to be

$$\mathbf{S}_{a,\boldsymbol{\gamma}}(y,w_1,w_2,\mathbf{z},\boldsymbol{\beta},\boldsymbol{\gamma},f_{X|\mathbf{Z}})=\frac{\int -V(u_{a,1},u_{a,2})/\{\mathbf{B}(x)^\top\boldsymbol{\gamma}\}f_{a,Y,W_1,W_2,X|\mathbf{Z}}(y,w_1,w_2,x,\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\gamma},f_{X|\mathbf{Z}})\mathbf{B}(x)dx}{\int f_{a,Y,W_1,W_2,X|\mathbf{Z}}(y,w_1,w_2,x,\mathbf{Z},\boldsymbol{\beta},\boldsymbol{\gamma},f_{X|\mathbf{Z}})dx}.$$

Therefore the efficient score $\mathbf{S}_{a,\mathrm{eff},\boldsymbol{\gamma}}$ is the projection of $\mathbf{S}_{a,\boldsymbol{\gamma}}$ onto the orthogonal complement of the nuisance tangent space and is easily verified to be as in (4), where $\mathbf{c}(X,\mathbf{Z})$ satisfies (5).

A.5. Proof of Theorem 1

By the definitions of $\mathbf{S}^*_{\rm eff}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}, \sigma, f^*_{X|\mathbf{Z}})$ and $\mathbf{S}^*_{a,{\rm eff},\boldsymbol{\gamma}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, f^*_{X|\mathbf{Z}})$, we have

$$\mathrm{E}\{\mathbf{S}^*_{\rm eff}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}_0, \sigma, f^*_{X|\mathbf{Z}})|X, \mathbf{Z}_i\} = \mathbf{0} \quad \text{and} \quad \mathrm{E}_a\{\mathbf{S}^*_{a,{\rm eff},\boldsymbol{\gamma}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, f^*_{X|\mathbf{Z}})|X, \mathbf{Z}_i\} = \mathbf{0}.$$

Then the non-conditional expectations also equal $\mathbf{0}$ almost everywhere, i.e.,

$$\mathrm{E}\{\mathbf{S}^*_{\rm eff}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}_0, \sigma, f^*_{X|Z})\} = \mathbf{0} \quad \text{and} \quad \mathrm{E}_a\{\mathbf{S}^*_{a,{\rm eff},\boldsymbol{\gamma}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, f^*_{X|\mathbf{Z}})\} = \mathbf{0}.$$

This leads to

$$\mathrm{E}\{\mathbf{S}^*_{\rm eff}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, f^*_{X|\mathbf{Z}})\} = o_p(1) \quad \text{and} \quad \mathrm{E}\{\mathbf{S}^*_{a,{\rm eff},\boldsymbol{\gamma}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, f^*_{X|\mathbf{Z}})\} = o_p(1),$$

element-wise by Remark 3. Condition (C6) ensures that as a vector function of $\boldsymbol{\theta}$,

$$(\mathrm{E}\{\mathbf{S}^*_{\rm eff}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, f^*_{X|\mathbf{Z}})\}^\top, \mathrm{E}\{\mathbf{S}^*_{a,{\rm eff},\boldsymbol{\gamma}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, f^*_{X|\mathbf{Z}})\}^\top)^\top$$

is invertible near $\boldsymbol{\theta}^*$ and the first derivative of the inverse function is bounded in the neighborhood of its zero. Therefore, $\|\boldsymbol{\theta}^* - \boldsymbol{\theta}_0\|_2 = o_p(1)$. Moreover, since

$$\frac{1}{n}\sum_{i=1}^n \mathbf{S}^*_{\rm eff}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\gamma}}_n, f^*_{X|\mathbf{Z}}) = \mathbf{0} \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^n \mathbf{S}^*_{a,{\rm eff},\boldsymbol{\gamma}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\gamma}}_n, f^*_{X|\mathbf{Z}}) = \mathbf{0},$$

we have

$$\mathrm{E}\{\mathbf{S}^*_{\rm eff}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\gamma}}_n, f^*_{X|\mathbf{Z}})\} = o(1) \quad \text{and} \quad \mathrm{E}\{\mathbf{S}^*_{a,{\rm eff},\boldsymbol{\gamma}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\gamma}}_n, f^*_{X|\mathbf{Z}})\} = o(1)$$

element-wise. Using the same argument, we obtain $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*\|_2 = o_p(1)$. Hence $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|_2 = o_p(1)$. This concludes the proof of Theorem 1. □

A.6. Proof of Theorem 2

To prove asymptotic normality, we first expand the estimating function (7) as a function of $\boldsymbol{\beta}$ about $\boldsymbol{\beta}_0$ keeping function $\widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta})$ fixed, to obtain $\mathbf{T}_1 + \mathbf{T}_2(\widetilde{\boldsymbol{\beta}}_n)\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = \mathbf{0}$, where

$$\mathbf{T}_1 = \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{S}^*_{\rm eff}\{Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}_0, \widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}_0), f^*_{X|\mathbf{Z}}\} \quad \text{and} \quad \mathbf{T}_2(\boldsymbol{\beta}) = \mathbf{T}_{21}(\boldsymbol{\beta}) + \mathbf{T}_{22}(\boldsymbol{\beta})\,\frac{\partial}{\partial \boldsymbol{\beta}}\,\widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}).$$

Here

$$\mathbf{T}_{21}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\beta}^\top}\,\mathbf{S}^*_{\rm eff}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}, \widehat{\boldsymbol{\gamma}}_n, f^*_{X|\mathbf{Z}})$$

and

$$\mathbf{T}_{22}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial \widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta})^\top}\,\mathbf{S}^*_{\rm eff}\{Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}, \widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}), f^*_{X|\mathbf{Z}}\},$$

and $\widetilde{\boldsymbol{\beta}}_n$ is on the line connecting $\boldsymbol{\beta}_0$ and $\widehat{\boldsymbol{\beta}}_n$. Since $\widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta})$ satisfies $n^{-1}\sum_{i=1}^n \mathbf{S}^*_{a,{\rm eff},\boldsymbol{\gamma}}\{Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}, \widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}), f^*_{X|\mathbf{Z}}\} = \mathbf{0}$ for any $\boldsymbol{\beta}$,

$$\frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\beta}^\top}\,\mathbf{S}^*_{a,{\rm eff},\boldsymbol{\gamma}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}, \widehat{\boldsymbol{\gamma}}_n, f^*_{X|\mathbf{Z}}) + \frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial \widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta})^\top}\,\mathbf{S}^*_{a,{\rm eff},\boldsymbol{\gamma}}\{Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}, \widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}), f^*_{X|\mathbf{Z}}\}\,\frac{\partial}{\partial \boldsymbol{\beta}^\top}\,\widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}) = \mathbf{0}.$$

Then $\partial\widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^\top = -\{\mathbf{T}_{23}(\boldsymbol{\beta})\}^{-1}\mathbf{T}_{24}(\boldsymbol{\beta})$, where

$$\mathbf{T}_{23}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial \widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta})^\top}\,\mathbf{S}^*_{a,{\rm eff},\boldsymbol{\gamma}}\{Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}, \widehat{\boldsymbol{\gamma}}_n(\boldsymbol{\beta}), f^*_{X|\mathbf{Z}}\}$$

and

$$\mathbf{T}_{24}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\beta}^\top}\,\mathbf{S}^*_{a,{\rm eff},\boldsymbol{\gamma}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}, \widehat{\boldsymbol{\gamma}}_n, f^*_{X|\mathbf{Z}}).$$

Hence $\mathbf{T}_2(\widetilde{\boldsymbol\beta}_n) = \mathbf{T}_{21}(\widetilde{\boldsymbol\beta}_n) - \mathbf{T}_{22}(\widetilde{\boldsymbol\beta}_n)\ \{\mathbf{T}_{23}(\widetilde{\boldsymbol\beta}_n)\}^{-1}\mathbf{T}_{24}(\widetilde{\boldsymbol\beta}_n)$. We further expand $\mathbf{T}_1$ as a function of $\widehat{\boldsymbol\gamma}_n(\boldsymbol\beta_0)$ about $\boldsymbol\gamma_0(\boldsymbol\beta_0)$ to obtain $\mathbf{T}_1 = \mathbf{T}_{11} + \mathbf{T}_{12}\ \{\widetilde{\boldsymbol\gamma}_n(\boldsymbol\beta_0)\}\sqrt{n}\ \{\widehat{\boldsymbol\gamma}_n(\boldsymbol\beta_0) - \boldsymbol\gamma_0(\boldsymbol\beta_0)\}$, where

$$\mathbf{T}_{11} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{S}^*_{\mathrm{eff}}\{Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol\beta_0, \boldsymbol\gamma_0(\boldsymbol\beta_0), f^*_{X|\mathbf{Z}}\}$$

and

$$\mathbf{T}_{12}\ \{\boldsymbol\gamma(\boldsymbol\beta_0)\} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol\gamma(\boldsymbol\beta_0)^\top}\mathbf{S}^*_{\mathrm{eff}}\{Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol\beta_0, \boldsymbol\gamma(\boldsymbol\beta_0), f^*_{X|\mathbf{Z}}\},$$

and $\widetilde{\boldsymbol\gamma}_n(\boldsymbol\beta_0)$ is a value between $\widehat{\boldsymbol\gamma}_n(\boldsymbol\beta_0)$ and $\boldsymbol\gamma_0(\boldsymbol\beta_0)$.

By the consistency of $\mathbf{B}(x)^\top\widetilde{\boldsymbol\gamma}_n$ to $\sigma(x)$, for arbitrary $d_{\boldsymbol\gamma}\times p$ matrix $\mathbf{G}$ with $\|\mathbf{G}\|_2 = 1$, we have

$$\mathbf{T}_{12}\ \{\widetilde{\boldsymbol\gamma}_n(\boldsymbol\beta_0)\}\mathbf{G} = \mathrm{E}\left\{\frac{\partial}{\partial\boldsymbol\gamma^\top}\mathbf{S}^*_{\mathrm{eff}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol\beta_0, \boldsymbol\gamma, f^*_{X|\mathbf{Z}})\mathbf{G}\bigg|_{\mathbf{B}(X)^\top\boldsymbol\gamma=\sigma(X)}\right\}\ \{1+o_p(1)\},$$

where

$$\begin{aligned}
&\mathrm{E}\left\{\frac{\partial}{\partial\boldsymbol\gamma^\top}\mathbf{S}^*_{\mathrm{eff}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol\beta_0, \boldsymbol\gamma, f^*_{X|\mathbf{Z}})\mathbf{G}\bigg|_{\mathbf{B}(X)^\top\boldsymbol\gamma=\sigma(X)}\right\}\\
&= \int\left\{\frac{\partial}{\partial\boldsymbol\gamma^\top}\mathbf{S}^*_{\mathrm{eff}}(y_i, w_{i1}, w_{i2}, \mathbf{z}_i, \boldsymbol\beta_0, \boldsymbol\gamma, f^*_{X|\mathbf{Z}})\mathbf{G}\bigg|_{\mathbf{B}(X)^\top\boldsymbol\gamma=\sigma(X)}\right\}\times\\
&\qquad f_{Y|W_1,W_2,\mathbf{Z}}(y_i, w_{i1}, w_{i2}, \mathbf{z}_i, \boldsymbol\beta_0, \sigma, f_{X|\mathbf{Z}})f_{\mathbf{Z}}(\mathbf{z}_i)dy_i dw_{i1}dw_{i2}d\mathbf{z}_i.
\end{aligned}$$

By Remark 3, the latter is equal to

$$\begin{aligned}
&\int\left\{\frac{\partial}{\partial{\boldsymbol\gamma_0}^\top}\mathbf{S}^*_{\mathrm{eff}}(y_i, w_{i1}, w_{i2}, \mathbf{z}_i, \boldsymbol\beta_0, \boldsymbol\gamma_0, f^*_{X|\mathbf{Z}})\mathbf{G} + O_p(h_b^q)\right\}\\
&\times\left\{f_{a,Y,W_1,W_2|\mathbf{Z}}(y_i, w_{i1}, w_{i2}, \mathbf{z}_i, \boldsymbol\beta_0, \boldsymbol\gamma_0, f_{X|\mathbf{Z}})f_{\mathbf{Z}}(\mathbf{z}_i) + O_p(h_b^q)\right\}dy_i dw_{i1}dw_{i2}d\mathbf{z}_i
\end{aligned}$$

Given that $\|\partial\mathbf{S}^*_{\mathrm{eff}}(y_i, w_{i1}, w_{i2}, \mathbf{z}_i, \boldsymbol\beta_0, \boldsymbol\gamma_0, f^*_{X|\mathbf{Z}})/\partial{\boldsymbol\gamma_0}^\top\|_\infty$ is integrable by condition (C7) and in view of the fact that $f_{a,Y,W_1,W_2|\mathbf{Z}}$ $(y_i, w_{i1}, w_{i2}, \mathbf{z}_i, \boldsymbol\beta_0, \boldsymbol\gamma_0, f_{X|\mathbf{Z}})f_{\mathbf{Z}}(\mathbf{z}_i)$ is absolutely integrable, the latter expression is also equal to

$$\int\frac{\partial}{\partial{\boldsymbol\gamma_0}^\top}\mathbf{S}^*_{\mathrm{eff}}(y_i, w_{i1}, w_{i2}, \mathbf{z}_i, \boldsymbol\beta_0, \boldsymbol\gamma_0, f^*_{X|\mathbf{Z}})\mathbf{G}\times f_{a,Y,W_1,W_2|\mathbf{Z}}(y_i, w_{i1}, w_{i2}, \mathbf{z}_i, \boldsymbol\beta_0, \boldsymbol\gamma_0, f_{X|\mathbf{Z}})f_{\mathbf{Z}}(\mathbf{z}_i)dy_i dw_{i1}dw_{i2}d\mathbf{z}_i + O_p(h_b^q).$$

Calling again on Remark 3, we can rewrite this expression as

$$\begin{aligned}
&\frac{\partial}{\partial{\boldsymbol\gamma_0}^\top}\int\{\mathbf{S}^*_{\mathrm{eff}}(y_i, w_{i1}, w_{i2}, \mathbf{z}_i, \boldsymbol\beta_0, \sigma, f^*_{X|\mathbf{Z}}) + O_P(h_b^q)\}\mathbf{G}\times\\
&\{f_{Y,W_1,W_2|\mathbf{Z}}(y_i, w_{i1}, w_{i2}, \mathbf{z}_i, \beta_0, \sigma, f_{X|\mathbf{Z}})f_{\mathbf{Z}}(\mathbf{z}_i) + O_p(h_b^q)\}dy_i dw_{i1}dw_{i2}d\mathbf{z}_i\\
&-\int\{\mathbf{S}^*_{\mathrm{eff}}(y_i, w_{i1}, w_{i2}, \mathbf{z}_i, \boldsymbol\beta_0, \sigma, f^*_{X|\mathbf{Z}}) + O_P(h_b^q)\}\mathbf{G}\\
&\times\frac{\partial}{\partial{\boldsymbol\gamma_0}^\top}f_{a,Y,W_1,W_2|\mathbf{Z}}(y_i, w_{i1}, w_{i2}, \mathbf{z}_i, \boldsymbol\beta_0, \boldsymbol\gamma_0, f_{X|\mathbf{Z}})f_{\mathbf{Z}}(\mathbf{z}_i)dy_i dw_{i1}dw_{i2}d\mathbf{z}_i + O_P(h_b^q)
\end{aligned}$$

Given that $\mathrm{E}\{\mathbf{S}^*_{\mathrm{eff}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol\beta, \sigma, f^*_{X|\mathbf{Z}})\} = \mathbf{0}$, this expression reduces to

$$\begin{aligned}
&-\int\mathbf{S}^*_{\mathrm{eff}}(y_i, w_{i1}, w_{i2}, \mathbf{z}_i, \boldsymbol\beta_0, \sigma, f^*_{X|\mathbf{Z}})\ \{\mathbf{G}^\top\mathbf{S}_{a,\boldsymbol\gamma}(y_i, w_{i1}, w_{i2}, \mathbf{z}_i, \boldsymbol\beta_0, \boldsymbol\gamma_0, f_{X|\mathbf{Z}})\}^\top\times\\
&\times f_{Y,W_1,W_2|\mathbf{Z}}(y_i, w_{i1}, w_{i2}, \mathbf{z}_i, \boldsymbol\beta_0, \sigma, f_{X|\mathbf{Z}})f_{\mathbf{Z}}(\mathbf{z}_i)dy_i dw_{i1}dw_{i2}d\mathbf{z}_i + O_p(h_b^p) = O_p(h_b^q).
\end{aligned}$$

For the last equality, first note that for any $p\times d_{\boldsymbol\gamma}$ matrix $\mathbf{K}$, there exists a function $\mathbf{b}(X)$ such that $\mathbf{KB}(X) = \mathbf{b}(X)$. Then by Remark 3 and the definitions of Λ_σ and $\Lambda_{a,\boldsymbol\gamma}$, for any $d_{\boldsymbol\gamma}\times p$ matrix $\mathbf{G}$, there exists a function $\mathbf{g}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol\beta_0, \sigma)\in\Lambda_\sigma$ such that

$$\sup_X|\mathbf{G}^\top\mathbf{S}_{a,\boldsymbol\gamma}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol\beta_0, \boldsymbol\gamma_0, f_{X|\mathbf{Z}}) - \mathbf{g}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol\beta_0, \sigma)| = O_p(h_b^q).$$

Further, $\mathbf{S}^*_{\mathrm{eff}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol\beta_0, \sigma, f^*_{X|\mathbf{Z}})$ is orthogonal to any function in Λ_σ, thus the last equality holds. Hence, we obtain $\|\mathbf{T}_{12}\ \{\widetilde{\boldsymbol\gamma}(\boldsymbol\beta_0)\}\|_2 = O_p(h_b^q)$.

Based on the asymptotic results of Proposition 4 in [12], we have $\|\widehat{\boldsymbol\gamma}_n(\boldsymbol\beta_0) - \boldsymbol\gamma_0(\boldsymbol\beta_0)\|_2 = O_p\{(nh_b)^{-1/2}\}$. Then we have

$$\|\mathbf{T}_{12}\ \{\widetilde{\boldsymbol\gamma}_n(\boldsymbol\beta_0)\}\sqrt{n}\ \{\widehat{\boldsymbol\gamma}_n(\boldsymbol\beta_0) - \boldsymbol\gamma_0(\boldsymbol\beta_0)\}\|_2 = O_p(h_b^{q-1/2}).$$

Further, by Remark 3 we have

$$\mathbf{T}_{11} = n^{-1/2} \sum_{i=1}^{n} \mathbf{S}^{*}_{\text{eff}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}_0, \sigma, f^{*}_{X|\mathbf{Z}}) + O_p(n^{1/2} h_b^q).$$

Given that $h_b^{q-1/2} = o_p(n^{1/2} h_b^q)$, we have

$$\mathbf{T}_{1} = n^{-1/2} \sum_{i=1}^{n} \mathbf{S}^{*}_{\text{eff}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}_0, \sigma, f^{*}_{X|\mathbf{Z}}) + O_p(n^{1/2} h_b^q).$$

Note that $n^{1/2} h_b^q = o_p(1)$ by conditions (C4) and (C5).

Now consider $\mathbf{T}_2(\widetilde{\boldsymbol{\beta}}_n)$. By the consistency of $\widetilde{\boldsymbol{\beta}}_n$ to $\boldsymbol{\beta}_0$ and $\mathbf{B}(x)^\top \widehat{\boldsymbol{\gamma}}_n$ to $\sigma(x)$, we have

$$\mathbf{T}_{21}(\widetilde{\boldsymbol{\beta}}_n) = \mathrm{E}\left\{ \frac{\partial}{\partial \boldsymbol{\beta}_0{}^\top} \mathbf{S}^{*}_{\text{eff}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}_0, \boldsymbol{\gamma}, f^{*}_{X|\mathbf{Z}}) \bigg|_{\mathbf{B}(X)^\top \boldsymbol{\gamma} = \sigma(X)} \right\} \{1 + o_p(1)\}$$

and

$$\mathbf{T}_{24}(\widetilde{\boldsymbol{\beta}}_n) = \mathrm{E}\left\{ \frac{\partial}{\partial \boldsymbol{\beta}_0{}^\top} \mathbf{S}^{*}_{a,\text{eff},\boldsymbol{\gamma}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}_0, \boldsymbol{\gamma}, f^{*}_{X|\mathbf{Z}}) \bigg|_{\mathbf{B}(X)^\top \boldsymbol{\gamma} = \sigma(X)} \right\} \{1 + o_p(1)\}.$$

We also have

$$\mathbf{T}_{22}(\widetilde{\boldsymbol{\beta}}_n) = \mathrm{E}\left\{ \frac{\partial}{\partial \boldsymbol{\gamma}^\top} \mathbf{S}^{*}_{\text{eff}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}_0, \boldsymbol{\gamma}, f^{*}_{X|\mathbf{Z}}) \bigg|_{\mathbf{B}(X)^\top \boldsymbol{\gamma} = \sigma(X)} \right\} \{1 + o_p(1)\}.$$

We have already proved that $\mathrm{E}[\{\partial \mathbf{S}^{*}_{\text{eff}}(Y_i, W_{i1}, W_{i2}, \boldsymbol{\beta}_0, \boldsymbol{\gamma}, f^{*}_{X|\mathbf{Z}})/\partial \boldsymbol{\gamma}^\top\} \mathbf{G}|_{\mathbf{B}(X)^\top \boldsymbol{\gamma} = \sigma(X)}] = O_p(h_b^q)$ element-wise for any arbitrary $d_{\boldsymbol{\gamma}} \times p$ matrix $\mathbf{G}$ with $\|\mathbf{G}\|_2 = 1$ by showing that for any $d_{\boldsymbol{\gamma}} \times p$ matrix $\mathbf{G}$, there exists a function $\mathbf{g} \in \Lambda_\sigma$, which is orthogonal to $\mathbf{S}^{*}_{\text{eff}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}_0, \sigma, f^{*}_{X|\mathbf{Z}})$, such that $\sup_X |\mathbf{G}^\top \mathbf{S}_{a,\boldsymbol{\gamma}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, f_{X|\mathbf{Z}}) - \mathbf{g}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}_0, \sigma)| = O_P(h_b^q)$.

For $\mathbf{T}_{23}(\widetilde{\boldsymbol{\beta}}_n)$, based on the proof of Proposition 4 in [12], we have $\|\mathbf{T}_{23}(\widetilde{\boldsymbol{\beta}}_n)^{-1}\|_2 = O_p(h_b^{-1})$. Then we have $\mathbf{T}_{22}(\widetilde{\boldsymbol{\beta}}_n)\{\mathbf{T}_{23}(\widetilde{\boldsymbol{\beta}}_n)\}^{-1}\mathbf{T}_{24}(\widetilde{\boldsymbol{\beta}}_n) = O_p(h_b^{q-1})$, where $q > 1$ by condition (C2). Thus

$$\mathbf{T}_{2}(\widetilde{\boldsymbol{\beta}}_n) = \mathrm{E}\left\{ \frac{\partial}{\partial \boldsymbol{\beta}_0{}^\top} \mathbf{S}^{*}_{\text{eff}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}_0, \boldsymbol{\gamma}, f^{*}_{X|\mathbf{Z}}) \bigg|_{\mathbf{B}(X)^\top \boldsymbol{\gamma} = \sigma(X)} \right\} \{1 + o_p(1)\} + O_P(h_b^{q-1}).$$

Therefore,

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = -\left[\mathrm{E}\left\{ \frac{\partial}{\partial \boldsymbol{\beta}_0{}^\top} \mathbf{S}^{*}_{\text{eff}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}_0, \boldsymbol{\gamma}, f^{*}_{X|\mathbf{Z}}) \bigg|_{\mathbf{B}(X)^\top \boldsymbol{\gamma} = \sigma(X)} \right\} \right]^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{S}^{*}_{\text{eff}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}_0, \sigma, f^{*}_{X|\mathbf{Z}}) + o_p(1).$$

Since $n^{-1/2} \sum_{i=1}^{n} \mathbf{S}^{*}_{\text{eff}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}_0, \sigma, f^{*}_{X|\mathbf{Z}})$ is the sum of independent zero-mean random vectors, this will converge in distribution to a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix

$$[\mathrm{E}\{\mathbf{S}^{*}_{\text{eff}}(Y_i, W_{i1}, W_{i2}, \mathbf{Z}_i, \boldsymbol{\beta}_0, \sigma, f^{*}_{X|\mathbf{Z}})^{\otimes 2}\}]^{-1}.$$

This concludes the proof of Theorem 2. □

A.7. Estimating equations for subjects without replication of W

In the approximate model, the conditional density of (Y, W) given $\mathbf{Z}$ is

$$f^{(1)}_{a,Y,W|\mathbf{Z}}(y, w, \mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\gamma}, f_{X|\mathbf{Z}}) = \int \frac{f_{Y|X,|Z}(y|x, \mathbf{z}, \boldsymbol{\beta}) f_U\{(w - x)/\mathbf{B}(x)^\top \boldsymbol{\gamma}\} f_{X|\mathbf{Z}}(x|\mathbf{z})}{\mathbf{B}(x)^\top \boldsymbol{\gamma}}\, dx.$$

We use the superscript $^{(1)}$ to denote the corresponding quantities that are calculated when only one W is available. The corresponding scores for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are given as

$$\mathbf{S}^{(1)}_{a,\boldsymbol{\beta}}(y, w, \mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\gamma}, f_{X|\mathbf{Z}}) = \frac{\int \{\partial f_{Y|X,\mathbf{Z}}(y|x, \mathbf{z}, \boldsymbol{\beta})/\partial \boldsymbol{\beta}\} f_U\{(w - x)/\mathbf{B}(x)^\top \boldsymbol{\gamma}\} f_{X|\mathbf{Z}}(x|\mathbf{z})/\{\mathbf{B}(x)^\top \boldsymbol{\gamma}\} dx}{\int f_{Y|X,\mathbf{Z}}(y|x, \mathbf{z}, \boldsymbol{\beta}) f_U\{(w - x)/\mathbf{B}(x)^\top \boldsymbol{\gamma}\} f_{X|\mathbf{Z}}(x|\mathbf{z})/\{\mathbf{B}(x)^\top \boldsymbol{\gamma}\} dx},$$

and

$$\mathbf{S}^{(1)}_{a,\boldsymbol{\gamma}}(y, w, \mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\gamma}, f_{X|\mathbf{Z}}) = \frac{\int -V^{(1)}(u_a) f_{Y|X,\mathbf{Z}}(y|x, \mathbf{z}, \boldsymbol{\beta}) f_U\{(w-x)/\mathbf{B}(x)^\top\boldsymbol{\gamma}\} f_{X|\mathbf{Z}}(x|\mathbf{z}) \mathbf{B}(x)/\{\mathbf{B}(x)^\top\boldsymbol{\gamma}\}^2 dx}{\int f_{Y|X,\mathbf{Z}}(y|x, \mathbf{z}, \boldsymbol{\beta}) f_U\{(w-x)/\mathbf{B}(x)^\top\boldsymbol{\gamma}\} f_{X|\mathbf{Z}}(x|\mathbf{z})/\{\mathbf{B}(x)^\top\boldsymbol{\gamma}\} dx},$$

where $V^{(1)}(u_a) = u_a f_U'(u_a)/f_U(u_a) + 1$ and $u_a = (w-x)/\mathbf{B}(x)^\top\boldsymbol{\gamma}$. Following the derivations in Section 3, the approximate efficient score for $\boldsymbol{\beta}$ with working model $f^*_{X|\mathbf{Z}}$ is

$$\mathbf{S}^{*(1)}_{a,\mathrm{eff}}(Y, W, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\gamma}, f^*_{X|\mathbf{Z}}) = \mathbf{S}^{*(1)}_{a,\boldsymbol{\beta}}(Y, W, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\gamma}, f^*_{X|\mathbf{Z}}) - \mathrm{E}^{*(1)}_a\{\mathbf{a}(X, \mathbf{Z})|Y, W, \mathbf{Z}\} - \mathrm{E}^{*(1)}_a\{V^{(1)}(U_a)\mathbf{b}(X)|Y, W, \mathbf{Z}\},$$

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

$$\mathrm{E}^{(1)}_a\{\mathbf{S}^{*(1)}_{a,\boldsymbol{\beta}}(Y, W, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\gamma}, f^*_{X|\mathbf{Z}})|X, \mathbf{Z}\} = \mathrm{E}^{(1)}_a[\mathrm{E}^{*(1)}_a\{\mathbf{a}(X, \mathbf{Z})|Y, W, \mathbf{Z}\}|X, \mathbf{Z}] + \mathrm{E}^{(1)}_a[\mathrm{E}^{*(1)}_a\{V^{(1)}(U_a)\mathbf{b}(X)|Y, W, \mathbf{Z}\}|X, \mathbf{Z}]$$

and

$$\begin{aligned}&\mathrm{E}^{(1)}_a\{\mathbf{S}^{*(1)}_{a,\boldsymbol{\beta}}(Y, W, \mathbf{Z}, \boldsymbol{\beta}, \widehat{\boldsymbol{\gamma}}(\boldsymbol{\beta}), f^*_{X|\mathbf{Z}})V(U_a)|X, \mathbf{Z}\}\\ &\quad = \mathrm{E}^{(1)}_a[\mathrm{E}^{*(1)}_a\ \{\mathbf{a}(X, \mathbf{Z})|Y, W, \mathbf{Z}\}V(U_a)|X, \mathbf{Z}] + \mathrm{E}^{(1)}_a[\mathrm{E}^{*(1)}_a\ \{V(U_a)\mathbf{b}(X)|Y, W_1, W_2, \mathbf{Z}\}V(U_a)|X, \mathbf{Z}].\end{aligned}$$

The efficient score for $\boldsymbol{\gamma}$ is given as

$$\mathbf{S}^{*(1)}_{a,\mathrm{eff},\boldsymbol{\gamma}}(Y, W, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\gamma}, f^*_{X|\mathbf{Z}}) = \mathbf{S}^{*(1)}_{a,\boldsymbol{\gamma}}(Y, W, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\gamma}, f^*_{X|\mathbf{Z}}) - \mathrm{E}^{*(1)}_a\{\mathbf{c}(X, \mathbf{Z})|Y, W, \mathbf{Z}\},$$

where $\mathbf{c}(X, \mathbf{Z})$ satisfies

$$\mathrm{E}^{(1)}_a\{\mathbf{S}^{*(1)}_{a,\boldsymbol{\gamma}}(Y, W, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\gamma}, f^*_{X|\mathbf{Z}})|X, \mathbf{Z}\} = \mathrm{E}^{(1)}_a[\mathrm{E}^{*(1)}_a\ \{\mathbf{c}(X, \mathbf{Z})|Y, W, \mathbf{Z}\}|X, \mathbf{Z}].$$

The corresponding estimating equations can then be constructed based on

$$\mathbf{S}^{*(1)}_{a,\mathrm{eff}}(Y, W, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\gamma}, f^*_{X|\mathbf{Z}}) \quad \text{and} \quad \mathbf{S}^{*(1)}_{a,\mathrm{eff},\boldsymbol{\gamma}}(Y, W, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\gamma}, f^*_{X|\mathbf{Z}}).$$

References

[1] S.M. Berry, R.J. Carroll, D. Ruppert, Bayesian smoothing and regression splines for measurement error problems, J. Amer. Statist. Assoc. 97 (2002) 160–169.
[2] A. Bertrand, C. Legrand, D. Léonard, I. Van Keilegom, Robustness of estimation methods in a survival cure model with mismeasured covariates, Comput. Statist. Data Anal. 113 (2017) 3–18.
[3] R.J. Carroll, D. Ruppert, L.A. Stefanski, C.M. Crainiceanu, Measurement Error in Nonlinear Models: A Modern Perspective, CRC Press, 2006.
[4] C.L. Cheng, J. Riu, On estimating linear relationships when both variables are subject to heteroscedastic measurement errors, Technometrics 48 (2006) 511–519.
[5] C. De Boor, A Practical Guide to Splines, in: Applied Mathematical Sciences, vol. 27, Springer, 2001.
[6] V. Devanarayan, L.A. Stefanski, Empirical simulation extrapolation for measurement error models with replicate measurements, Statist. Probab. Lett. 59 (2002) 219–225.
[7] W.A. Fuller, Measurement Error Models, Vol. 305, John Wiley & Sons, 2009.
[8] Y. Guo, R.J. Little, Regression analysis with covariates that have heteroscedastic measurement error, Stat. Med. 30 (2011) 2278–2294.
[9] J.J. Hanfelt, K.Y. Liang, Approximate likelihood ratios for general estimating functions, Biometrika 82 (1995) 461–477.
[10] C.C. Heyde, Quasi-likelihood and its Application: A General Approach to Optimal Parameter Estimation, Springer Science & Business Media, 1997, 2008.
[11] C. Heyde, R. Morton, Multiple roots in general estimating equations, Biometrika 85 (1998) 954–959.
[12] F. Jiang, Y. Ma, A spline-assisted semiparametric approach to non-parametric measurement error models, 2018, arXiv:1804.00793.
[13] A. Sarkar, B.K. Mallick, R.J. Carroll, Bayesian semiparametric regression in the presence of conditionally heteroscedastic measurement and regression errors, Biometrics 70 (2014) 823–834.
[14] J. Staudenmayer, D. Ruppert, J.P. Buonaccorsi, Density estimation in the presence of heteroscedastic measurement error, J. Amer. Statist. Assoc. 103 (2008) 726–736.
[15] A.A. Tsiatis, Y. Ma, Locally efficient semiparametric estimators for functional measurement error models, Biometrika 91 (2004) 835–848.
[16] G.Y. Yi, Statistical Analysis with Measurement Error or Misclassification, Springer, 2016.