AStA Adv Stat Anal (2019) 103:137–161
https://doi.org/10.1007/s10182-018-0327-6

ORIGINAL PAPER

SIMEX estimation for single-index model with covariate measurement error

Yiping Yang[1] · Tiejun Tong[2] · Gaorong Li[3]

Received: 3 April 2017 / Accepted: 27 April 2018 / Published online: 3 May 2018

Abstract In this paper, we consider the single-index measurement error model with mismeasured covariates in the nonparametric part. To solve the problem, we develop a simulation-extrapolation (SIMEX) algorithm based on the local linear smoother and the estimating equation. For the proposed SIMEX estimation, it is not needed to assume the distribution of the unobserved covariate. We transform the boundary of a unit ball in $\mathbb{R}^p$ to the interior of a unit ball in $\mathbb{R}^{p-1}$ by using the constraint $\|\beta\| = 1$. The proposed SIMEX estimator of the index parameter is shown to be asymptotically normal under some regularity conditions. We also derive the asymptotic bias and variance of the estimator of the unknown link function. Finally, the performance of the proposed method is examined by simulation studies and is illustrated by a real data example.

Keywords Single-index model · Measurement error · Local linear smoother · SIMEX · Estimating equation

Mathematics Subject Classification Primary 62G05 · 62G08; Secondary 62G20

✉ Gaorong Li
ligaorong@bjut.edu.cn

1 College of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing 400067, People's Republic of China

2 Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong

3 Beijing Institute for Scientific and Engineering Computing, Beijing University of Technology, Beijing 100124, People's Republic of China

1 Introduction

One major problem in fitting multivariate nonparametric regression models is the "curse of dimensionality." To overcome the problem, the single-index model has played an important role in the studies. In this paper, we consider the single-index model of the form

$$Y = g\left(\beta^{\mathrm{T}} X\right) + \varepsilon, \tag{1.1}$$

where Y is the response variable, X is a $p \times 1$ covariate vector, $g(\cdot)$ is an unknown link function, $\beta = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ is an unknown index parameter, and ε is a random error with $E(\varepsilon|X) = 0$ almost surely. We further assume the Euclidean norm $\|\beta\| = 1$ for the identifiability purpose. Model (1.1) reduces the covariate vector into an index which is a linear combination of covariates, and hence avoids the "curse of dimensionality."

Estimation for the index parameter and the unknown link function has attracted much attention. Duan and Li (1991) developed the sliced inverse regression method. Härdle and Tsybakov (1993) proposed the average derivative method to obtain a root-n consistent estimator of the index vector β. Carroll et al. (1997) used the local linear method to estimate the unknown parameters and the unknown link function for generalized partially linear single-index models. Naik and Tsai (2000) discussed the partial least-squares estimator for single-index models. Xue and Zhu (2006) and Zhu and Xue (2006) proposed the bias-corrected empirical likelihood method to construct the confidence intervals or regions of the parameters of interest. Liang et al. (2010) studied the semiparametrically efficient profile least-squares estimators of regression coefficients for partially linear single-index models. Cui et al. (2011) introduced the estimating function method to study the single-index models. Pang and Xue (2012) and Yang et al. (2014) investigated the single-index random effects models with longitudinal data. Li et al. (2014) constructed the simultaneous confidence bands for the nonparametric link function in single-index models. Li et al. (2015) proposed a penalized procedure combined with the bias-corrected GEE estimator and bias-correct QIF estimator.

The above studies have assumed that covariates can be directly observed. However, the measurement error models arise frequently in practice and are attracting attentions from medical and statistical research. For example, covariates such as the blood pressure (Carroll et al. 2006) and the CD4 count (Lin and Carroll 2000; Liang 2009) are often subject to measurement error. If one ignores these measurement errors, the estimators and inference may be biased. Hence, we are interested in estimating the index parameter β and the unknown link function $g(\cdot)$ in model (1.1) when the covariate vector X is measured with error. We assume an additive measurement error model as

$$W = X + U, \tag{1.2}$$

where W is the observed surrogate, U follows $N(0, \Sigma_u)$ and is independent of (X, Y). When U is zero, there is no measurement error. For simplicity, we consider only the case where the measurement error covariance matrix Σ_u is known. Otherwise, Σ_u need to be first estimated, e.g., by the replication experiments method in Carroll et al.

(2006). We refer to the models characterized by (1.1) and (1.2) as the single-index measurement error model.

To eliminate the effects of measurement error, Cook and Stefanski (1994) developed the SIMEX method to correct the estimates in the presence of additive measurement error. Carroll et al. (1996) further investigated the asymptotic distribution of the SIMEX estimator. Since then, the SIMEX method has become a standard tool for correcting the biases induced by measurement error in covariates for many complex models. Carroll et al. (1999) and Delaigle and Hall (2008) applied the SIMEX technique to local polynomial nonparametric regression and spline-based regression. Liang and Ren (2005) applied the SIMEX technique to the generalized partially linear models with the linear covariate being measured with additive error. Apanasovich et al. (2009) derived the limiting distribution of SIMEX in semiparameric measurement error models and gave computable asymptotically correct standard error estimates.

Note that the aforementioned SIMEX methods may not be able to handle the multivariate nonparametric measurement error regression models owing to the "curse of dimensionality." In view of this, Liang and Wang (2005) considered the partially linear single-index measurement error models with the linear part containing the measurement error, where they applied the correction for attenuation approach to obtain the efficient estimators of the parameters of interest. Their method, however, is not applicable for the occurrence with measurement errors in the nonparametric part. This motivates us to develop a new SIMEX method to solve this problem. Specifically, we combine the SIMEX method, the local linear approximation method, and the estimating equation to handle the single-index measurement error model. Our method has several desirable features. First, our proposed method can deal with multivariate nonparametric measurement error regression and avoids "curse of dimensionality" by introducing the index parameter. Second, we use the SIMEX technique to construct the unbiased estimation and reduce the bias of the estimator and do not assume the distribution of the unobservable X. Third, to transfer the restricted estimating equation with the constraint $\|\beta\| = 1$ to the unrestricted estimating equation, we regard the constraint $\|\beta\| = 1$ as a piece of prior information and adopt the "delete-one-component" method.

The remainder of the paper is organized as follows. In Sect. 2, we develop the SIMEX algorithm to obtain the estimators of the index parameter and the unknown link function and investigate their asymptotic properties. In Sect. 3, we present and compare the results from simulation studies and also apply the proposed method to a real data example for illustration. Some concluding remarks are given in Sect. 4, and the proofs of the main results are given in the Appendix.

2 Main results

2.1 Methodology

To conduct the unbiased estimation for β in the presence of covariate measurement error, Cook and Stefanski (1994) introduced the SIMEX algorithm. The SIMEX algorithm consists of the simulation step, the estimation step, and extrapolation steps. It aims to add additional variability to the observed W in order to establish the trend between the measurement error induced bias and the variance of induced measurement error and then extrapolate this trend back to the case without measurement error (Carroll et al. 2006). In this section, we use the SIMEX algorithm, the local linear smoother and the estimating equation to estimate β and $g(\cdot)$. First, we estimate $g(\cdot)$ as a function of β by using the local linear smoother. We then estimate the parametric part based on the estimating equation. The proposed algorithm is described as follows.

(I) Simulation step

For each $i = 1, \ldots, n$, we generate a sequence of variables

$$W_{is}(\lambda) = W_i + (\lambda \Sigma_u)^{1/2} U_{is}, \qquad s = 1, \ldots, \mathcal{S},$$

where $U_{is} \sim N(0, I_p)$, I_p is a $p \times p$ identity matrix, $\mathcal{S}$ is a given integer, $\lambda \geq 0$ and $\lambda \in \Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_M\}$ is the grid of λ in the extrapolation step. Here λ controls how much additional independent measurement error is added to the original W data. Simulation evidence suggests that the extrapolantation should be fitted for λ in a range of $[0, \lambda_M]$ with $1 \leq \lambda_M \leq 2$ (see Carroll et al. 2006).

(II) Estimation step Suppose that $g(\cdot)$ has a continuous second derivative. For t in a small neighborhood of t_0, $g(t)$ can be approximated as $g(t) \approx g(t_0) + g'(t_0)(t - t_0) \equiv a + b(t - t_0)$. With the simulated $W_{is}(\lambda)$, we first estimate $g(t_0)$ as a function of β by a local linear smoother, denoted by $\hat{g}(\beta, \lambda; t_0)$, in Step 1. We then propose a new estimator of $\beta(\lambda)$ in Steps 2 and 3, denoted by $\hat{\beta}(\lambda)$. The specific procedure is as follows.

Step 1 For each fixed t_0 and β, $\hat{g}(\beta, \lambda; t_0)$ and $\hat{g}'(\beta, \lambda; t_0)$ are estimated by minimizing

$$\sum_{i=1}^{n} \left\{ Y_i - a - b[\beta^{\mathrm{T}} W_{is}(\lambda) - t_0] \right\}^2 K_h(\beta^{\mathrm{T}} W_{is}(\lambda) - t_0), \tag{2.1}$$

with respect to a and b, where $K_h(\cdot) = h^{-1} K(\cdot/h)$, $K(\cdot)$ is a kernel function with the bandwidth h. Let $\hat{a}$ and $\hat{b}$ be the solutions to problem (2.1). Then, $\hat{g}(\beta, \lambda; t_0) = \hat{a}$ and $\hat{g}'(\beta, \lambda; t_0) = \hat{b}$. Let

$$\mathbb{M}_{ni}(\beta, \lambda; t_0) = \mathbb{U}_{ni}(\beta, \lambda; t_0) \Big/ \sum_{j=1}^{n} \mathbb{U}_{nj}(\beta, \lambda; t_0),$$

$$\widetilde{\mathbb{M}}_{ni}(\beta, \lambda; t_0) = \widetilde{\mathbb{U}}_{ni}(\beta, \lambda; t_0) \Big/ \sum_{j=1}^{n} \mathbb{U}_{nj}(\beta, \lambda; t_0),$$

where $\mathbb{U}_{ni}(\beta, \lambda; t_0) = K_h(\beta^{\mathrm{T}} W_{is}(\lambda) - t_0)\{S_{n,2}(\beta, \lambda; t_0) - [\beta^{\mathrm{T}} W_{is}(\lambda) - t_0] S_{n,1}(\beta, \lambda; t_0)\}$, $\widetilde{\mathbb{U}}_{ni}(\beta, \lambda; t_0) = K_h(\beta^{\mathrm{T}} W_{is}(\lambda) - t_0)\{[\beta^{\mathrm{T}} W_{is}(\lambda) - t] S_{n,0}(\beta, \lambda; t_0) - S_{n,1}(\beta, \lambda; t_0)\}$, and $S_{n,l}(\beta, \lambda; t_0) = \frac{1}{n} \sum_{i=1}^{n} (\beta^{\mathrm{T}} W_{is}(\lambda) - t_0)^l K_h(\beta^{\mathrm{T}} W_{is}(\lambda) - t_0)$ for $l = 0, 1, 2$. Simple calculation yields

$$\hat{g}(\beta, \lambda; t_0) = \sum\nolimits_{i=1}^{n} \mathbb{M}_{ni}(\beta, \lambda; t_0) Y_i, \tag{2.2}$$

$$\hat{g}'(\beta, \lambda; t_0) = \sum_{i=1}^{n} \widetilde{\mathbb{M}}_{ni}(\beta, \lambda; t_0) Y_i. \tag{2.3}$$

Chang et al. (2010) showed that the convergence rate of the estimator of $g'(t)$ is slower than that of $g(t)$ if the same bandwidth is used. Because of this, we have suggested another bandwidth h_1 to control the variability in the estimator of $g'(t)$. We use h_1 to replace h in $\hat{g}'(\beta, \lambda; t_0)$ and write it as $\hat{g}'_{h_1}(\beta, \lambda; t_0)$.

Step 2 To estimate β, we use the "delete-one-component" method in Zhu and Xue (2006) to transform the boundary of a unit ball in $\mathbb{R}^p$ to the interior of a unit ball in $\mathbb{R}^{p-1}$. Let $\beta^{(r)} = (\beta_1, \ldots, \beta_{r-1}, \beta_{r+1}, \ldots, \beta_p)^{\mathrm{T}}$ be a $(p-1)$-dimensional vector deleting the rth component β_r. Without loss of generality, we assume there is a positive component β_r; otherwise, we may consider $\beta_r = -(1 - \|\beta^{(r)}\|^2)^{1/2}$. Let

$$\beta = (\beta_1, \ldots, \beta_{r-1}, (1 - \|\beta^{(r)}\|^2)^{1/2}, \beta_{r+1}, \ldots, \beta_p)^{\mathrm{T}}.$$

Note that $\beta^{(r)}$ satisfies the constraint $\|\beta^{(r)}\| < 1$. We conclude that β is infinitely differentiable in a neighborhood of $\beta^{(r)}$ and the Jacobian matrix is $J_{\beta^{(r)}} = (\gamma_1, \ldots, \gamma_p)^{\mathrm{T}}$, where $\gamma_j (1 \le j \le p, j \ne r)$ is a $(p-1)$-dimensional vector with the sth component being 1, and $\gamma_r = -(1 - \|\beta^{(r)}\|^2)^{-\frac{1}{2}} \beta^{(r)}$. Given the estimators $\hat{g}(\beta, \lambda; t_0)$ and $\hat{g}'_{h_1}(\beta, \lambda; t_0)$ in (2.2) and (2.3), respectively, an estimator $\hat{\beta}_s^{(r)}(\lambda)$ of $\beta^{(r)}$ is obtained by solving the following equation:

$$Q_{ns}(\beta^{(r)}, \lambda) = \frac{1}{n} \sum_{i=1}^{n} \hat{\eta}_{is}(\beta^{(r)}, \lambda) = 0, \tag{2.4}$$

where

$$\hat{\eta}_{is}(\beta^{(r)}, \lambda) = \left[Y_i - \hat{g}\left(\beta, \lambda; \beta^{\mathrm{T}} W_{is}(\lambda)\right)\right] \hat{g}'_{h_1}\left(\beta, \lambda; \beta^{\mathrm{T}} W_{is}(\lambda)\right) J^{\mathrm{T}}_{\beta^{(r)}} W_{is}(\lambda),$$

$$\beta^{\mathrm{T}} W_{is}(\lambda) = \beta^{(r)T} W_{is}^{(r)}(\lambda) + \left(1 - \|\beta^{(r)}\|^2\right)^{1/2} W_{is,r}(\lambda),$$

$$W_{is}^{(r)}(\lambda) = \left(W_{is,1}(\lambda), \ldots, W_{is,(r-1)}(\lambda), W_{is,(r+1)}(\lambda), \ldots, W_{is,p}(\lambda)\right)^{\mathrm{T}}.$$

Next, we can obtain an estimator of β, say $\hat{\beta}_s(\lambda)$, by implementing the Fisher's method of scoring version of the Newton–Raphson algorithm to solve estimating Eq. (2.4). We summarize the iterative algorithm in what follows.

(1) Choose the initial values for β, denoted by $\widetilde{\beta}_s(\lambda)$, where $s = 1, \ldots, \mathcal{S}$.

(2) Compute

$$\hat{\beta}_s^*(\lambda) = \widetilde{\beta}_s(\lambda) + J_{\widetilde{\beta}_s^{(r)}} B_{ns}^{-1}(\widetilde{\beta}_s^{(r)}, \lambda) Q_{ns}(\widetilde{\beta}_s^{(r)}, \lambda),$$

where $B_{ns}(\beta^{(r)}, \lambda) = \frac{1}{n} \sum_{i=1}^{n} J^{\mathrm{T}}_{\beta^{(r)}} W_{is}(\lambda) \hat{g}'^2_{h_1}\left(\beta, \lambda; \beta^{\mathrm{T}} W_{is}(\lambda)\right) W_{is}^{\mathrm{T}}(\lambda) J_{\beta^{(r)}}$.

(3) Update $\widetilde{\beta}_s(\lambda)$ with $\widetilde{\beta}_s(\lambda) = \hat{\beta}_s^*(\lambda)/\|\hat{\beta}_s^*(\lambda)\|$.

(4) Repeat Step (2) and Step (3) until convergence.

In the iterative algorithm, the initial values of β are obtained by fitting a linear model with norm 1.

Similar to Cui et al. (2011), we discuss the solution of the estimating equation. In fact, the solution of the estimating equation $Q_{ns}(\beta^{(r)}, \lambda)$ is just the least-squares estimator of $\beta^{(r)}$. The least-squares objective function is defined by

$$G(\beta^{(r)}, \lambda) = \sum_{i=1}^{n} \left\{ Y_i - \hat{g}\left(\beta, \lambda; \beta^{\mathrm{T}} W_{is}(\lambda)\right) \right\}^2 .$$

The minimum of the objective function $G(\beta^{(r)}, \lambda)$ with respect to $\beta^{(r)}$ is the solution of the estimating equation $Q_{ns}(\beta^{(r)}, \lambda)$ because the estimating equation $Q_{ns}(\beta^{(r)}, \lambda)$ is the gradient vector of $G(\beta^{(r)}, \lambda)$. Note that $\{\|\beta^{(r)}\| < 1\}$ is an open and connected subset of $\mathbb{R}^{p-1}$. By the regularity condition (C2), we known that the least-squares objective function $G(\beta^{(r)}, \lambda)$ is twice continuously differentiable on $\{\|\beta^{(r)}\| < 1\}$ such that the global minimum of $G(\beta^{(r)}, \lambda)$ can be achieved at some points. By some simple calculations, we have

$$\frac{1}{n} \frac{\partial^2 G(\beta^{(r)}, \lambda)}{\partial \beta^{(r)} \beta^{(r)T}} = -\frac{\partial Q_{ns}(\beta^{(r)}, \lambda)}{\partial \beta^{(r)}} = \mathcal{A}(\beta(\lambda), \lambda) + o_p(1),$$

where $\mathcal{A}(\beta(\lambda), \lambda)$ is a positive definite matrix for $\lambda \in \Lambda$ defined in condition (C6). Note that the Hessian matrix $\frac{1}{n} \frac{\partial^2 G(\beta^{(r)}, \lambda)}{\partial \beta^{(r)} \beta^{(r)T}}$ is positive definite for all values of $\beta^{(r)}$ and $\lambda \in \Lambda$. Hence, estimating Eq. (2.4) has a unique solution.

Step 3 With the estimated values $\hat{\beta}_s(\lambda)$ over $s = 1, \ldots, \mathcal{S}$, we average them and obtain the final estimate of β as

$$\hat{\beta}(\lambda) = \frac{1}{\mathcal{S}} \sum_{s=1}^{\mathcal{S}} \hat{\beta}_s(\lambda).$$

(III) Extrapolation step For the extrapolant function, we consider the widely used quadratic function $\mathcal{G}(\lambda, \mathbf{\Gamma}) = \Gamma_1 + \Gamma_2\lambda + \Gamma_3\lambda^2$ with $\mathbf{\Gamma} = (\Gamma_1, \Gamma_2, \Gamma_3)^{\mathrm{T}}$ (Lin and Carroll 2000; Liang and Ren 2005). We fit a regression model of $\{\hat{\beta}(\lambda), \lambda \in \Lambda\}$ on $\{\lambda \in \Lambda\}$ based on $\mathcal{G}(\lambda, \mathbf{\Gamma})$, and denote $\hat{\mathbf{\Gamma}}$ as the estimated value of $\mathbf{\Gamma}$. The SIMEX estimator of β is then defined as $\hat{\beta}_{\mathrm{SIMEX}} = \mathcal{G}(-1, \hat{\mathbf{\Gamma}})$. When λ shrinks to 0, the SIMEX estimator reduces to the naive estimator, $\hat{\beta}_{\mathrm{Naive}} = \mathcal{G}(0, \hat{\mathbf{\Gamma}})$, that neglects the measurement error with a direct replacement of X by W.

The SIMEX estimator, $\hat{g}_{\mathrm{SIMEX}}(t_0)$, is obtained in the same way. β in Step 1 of the estimation step is replaced by $\hat{\beta}_{\mathrm{SIMEX}}$, and the estimator $\hat{g}_s(\lambda; t_0)$ is obtained with the bandwidth h_2. $\hat{g}_s(\lambda; t_0)$ over $s = 1, \ldots, \mathcal{S}$ is averaged; then, $\hat{g}(\lambda; t_0)$ is obtained by

$$\hat{g}(\lambda; t_0) = \frac{1}{\mathcal{S}} \sum_{s=1}^{\mathcal{S}} \hat{g}_s(\lambda; t_0).$$

The extrapolation step results in $\hat{\mathbb{A}}$, which minimizes $\sum_{\lambda \in \Lambda} \{\hat{g}(\lambda; t_0) - \mathcal{G}(\lambda; \mathbb{A})\}^2$ with respect to $\mathbb{A}$. The SIMEX estimator of $\hat{g}_{\text{SIMEX}}(t_0)$ is given by

$$\hat{g}_{\text{SIMEX}}(t_0) = \mathcal{G}(-1, \hat{\mathbb{A}}).$$

2.2 Asymptotic properties

To investigate the asymptotic properties of the estimators for the index parameter and the link function, we first present some regularity conditions.

(C1) The density function, $f(t)$, of $\beta^{\mathrm{T}} X$ is bounded away from zero. It also satisfies the Lipschitz condition of order 1 on $\mathcal{T} = \{t = \beta^{\mathrm{T}} x : x \in A\}$, where A is the bounded support set of X.

(C2) $g(\cdot)$ has a continuous second derivative on $\mathcal{T}$.

(C3) The kernel $K(\cdot)$ is a bounded and symmetric density function with a bounded support satisfying the Lipschitz condition of order 1 and $\int_{-\infty}^{\infty} u^2 K(u) du \neq 0$.

(C4) $\sup_x E(\varepsilon^4 | X = x) < \infty$.

(C5) $nh^2/(\log n)^2 \to \infty$, $nh^4 \log n \to 0$, $nhh_1^3/(\log n)^2 \to \infty$, and $\lim \sup_{n \to \infty} nh_1^5 < \infty$.

(C6) $\mathcal{A}(\beta(\lambda), \lambda)$ is a positive definite matrix for $\lambda \in \Lambda$, where

$$\mathcal{A}(\beta(\lambda), \lambda) = E\left\{\left[g'\left(\lambda; \beta^{\mathrm{T}}(\lambda) W_{is}(\lambda)\right)\right]^2 J^{\mathrm{T}}_{\beta^{(r)}(\lambda)} \widetilde{W}_{is}(\lambda) \widetilde{W}^{\mathrm{T}}_{is}(\lambda) J_{\beta^{(r)}(\lambda)}\right\}$$

with $\widetilde{W}_{is}(\lambda) = W_{is}(\lambda) - E[W_{is}(\lambda) | \beta^{\mathrm{T}}(\lambda) W_{is}(\lambda)]$.

(C7) The extrapolant function is theoretically exact.

Condition (C1) ensures that the density function of $\beta^{\mathrm{T}} X$ is positive. Condition (C2) is the standard condition in smoothness. Condition (C3) is the common assumption for the second-order kernels. Condition (C4) is a necessary condition for deriving the asymptotic normality for the proposed estimator. Condition (C5) is commonly used in nonparametric estimation (Pang and Xue 2012). Note that the range of bandwidth h does not contain the optimal bandwidth $O(n^{-1/5})$, undersmoothing is applied to eliminate the bias. Finally, condition (C6) ensures that there is asymptotic variance for the estimator $\hat{\beta}_{\text{SIMEX}}$, and condition (C7) is a common assumption for the SIMEX method (see Liang and Ren (2005)). Liang and Ren (2005) had pointed out the exact extrapolant function is known only in some special cases, and the quadratic or rational extrapolant for certain estimators is exact with normal measurement error.

To derive the theoretical results, we introduce some new definitions and notations. For the given $\Lambda = \{\lambda_1, \ldots, \lambda_M\}$, let $\hat{\beta}(\Lambda)$ be the vector of estimators $(\hat{\beta}(\lambda_1), \ldots, \hat{\beta}(\lambda_M))$, denoted by $\text{vec}\{\hat{\beta}(\lambda), \lambda \in \Lambda\}$. Let also $\mathbf{\Gamma} = (\Gamma_1^{\mathrm{T}}, \ldots, \Gamma_q^{\mathrm{T}})^{\mathrm{T}}$, where Γ_j is the parameter vector estimated in the extrapolation step for the jth component of $\hat{\beta}(\lambda)$ for $j = 1, \ldots, q$. We define $\mathcal{G}(\Lambda, \mathbf{\Gamma}) = \mathrm{vec}\{\mathcal{G}(\lambda_m, \Gamma_j), j = 1, \ldots, q, m = 1, \ldots, M\}$, $\mathrm{Res}(\mathbf{\Gamma}) = \hat{\beta}(\Lambda) - \mathcal{G}(\Lambda, \mathbf{\Gamma})$, $s^{\mathrm{T}}(\mathbf{\Gamma}) = \{\partial/\partial(\mathbf{\Gamma})^{\mathrm{T}}\}\mathrm{Res}(\mathbf{\Gamma})$, $D(\mathbf{\Gamma}) = s(\mathbf{\Gamma})s^{\mathrm{T}}(\mathbf{\Gamma})$,

$$\begin{aligned}
\eta_{i\mathcal{S}}(\beta(\lambda), \lambda) &= \frac{1}{\mathcal{S}} \sum_{s=1}^{\mathcal{S}} \Big[Y_i - g\Big(\lambda; \beta^{\mathrm{T}}(\lambda) W_{is}(\lambda)\Big)\Big] \\
&\qquad g'\Big(\lambda; \beta^{\mathrm{T}}(\lambda) W_{is}(\lambda)\Big) J^{\mathrm{T}}_{\beta^{(r)}(\lambda)} \widetilde{W}_{is}(\lambda), \\
\Psi_{i\mathcal{S}}\Big\{\beta(\Lambda), \Lambda\Big\} &= \mathrm{vec}\{\eta_{i\mathcal{S}}(\beta(\lambda), \lambda), \lambda \in \Lambda\}, \\
\mathcal{J}\Big\{\beta(\Lambda), \Lambda\Big\} &= \mathrm{diag}\{J_{\beta^{(r)}(\lambda)}, \lambda \in \Lambda\}, \\
\mathcal{A}_{11}\Big\{\beta(\Lambda), \Lambda\Big\} &= \mathrm{diag}\{\mathcal{A}(\beta(\lambda), \lambda), \lambda \in \Lambda\}
\end{aligned}$$

and

$$\begin{aligned}
\Sigma = \mathcal{J}\Big\{\beta(\Lambda), \Lambda\Big\} \mathcal{A}_{11}^{-1}\Big\{\beta(\Lambda), \Lambda\Big\} C_{11}\Big\{\beta(\Lambda), \Lambda\Big\} \\
\Big\{\mathcal{A}_{11}^{-1}\Big\{\beta(\Lambda), \Lambda\Big\}\Big\}^{\mathrm{T}} \mathcal{J}^{\mathrm{T}}\Big\{\beta(\Lambda), \Lambda\Big\}
\end{aligned}$$

with

$$C_{11}\Big\{\beta(\Lambda), \Lambda\Big\} = \mathrm{cov}\Big[\Psi_{i\mathcal{S}}\Big\{\beta(\Lambda), \Lambda\Big\}\Big].$$

Theorem 1 *Suppose that the regularity conditions* $(C1)$–$(C7)$ *hold. Then, as* $n \to \infty$*, we have*

$$\sqrt{n}(\hat{\beta}_{\mathrm{SIMEX}} - \beta) \xrightarrow{\mathcal{L}} N\{0, \mathcal{G}_{\mathbf{\Gamma}}(-1, \mathbf{\Gamma})\Sigma(\mathbf{\Gamma})\{\mathcal{G}_{\mathbf{\Gamma}}(-1, \mathbf{\Gamma})\}^{\mathrm{T}}\},$$

where $\xrightarrow{\mathcal{L}}$ *denotes the convergence in distribution,* $\mathcal{G}_{\mathbf{\Gamma}}(\lambda, \mathbf{\Gamma}) = \{\partial/\partial(\mathbf{\Gamma})^{\mathrm{T}}\}\mathcal{G}(\lambda, \mathbf{\Gamma})$, $\Sigma(\mathbf{\Gamma}) = D^{-1}(\mathbf{\Gamma})s(\mathbf{\Gamma})\Sigma s^{\mathrm{T}}(\mathbf{\Gamma})D^{-1}(\mathbf{\Gamma})$.

Theorem 1 indicates that $\hat{\beta}_{\mathrm{SIMEX}}$ is a root-n consistent estimator. Its asymptotic distribution is similar to that of the parametric estimator of β without measurement error, whereas the asymptotic covariance matrix of the resulting estimator is more complicated.

To apply Theorem 1 to construct the confidence interval of β, we give the consistent estimator of the asymptotic covariance matrix by the sandwich method. Take $\hat{C}_{11}(\cdot)$ to be the sample covariance matrix of the terms $\hat{\Psi}_{i\mathcal{S}}\Big\{\hat{\beta}(\Lambda), \Lambda\Big\}$ and $\hat{\mathcal{A}}_{11}(\cdot) = \mathrm{diag}\{\hat{\mathcal{A}}_m(\cdot)\}$ for $m = 1, \ldots, M$, where

$$\begin{aligned}
\widehat{\widehat{W}}_{is}(\lambda) &= W_{is}(\lambda) - \sum_{i=1}^{n} \mathbb{M}_{ni}\Big(\hat{\beta}(\lambda), \lambda; \hat{\beta}^{\mathrm{T}}(\lambda) W_{is}(\lambda)\Big) W_{is}(\lambda),\\
\hat{\eta}_{i\mathcal{S}}(\hat{\beta}(\lambda), \lambda) &= \frac{1}{\mathcal{S}} \sum_{s=1}^{\mathcal{S}} \Big[Y_i - \hat{g}\Big(\lambda; \hat{\beta}^{\mathrm{T}}(\lambda) W_{is}(\lambda)\Big)\Big]\\
&\quad \hat{g}'\Big(\lambda; \hat{\beta}^{\mathrm{T}}(\lambda) W_{is}(\lambda)\Big) J^{\mathrm{T}}_{\hat{\beta}^{(r)}(\lambda)} \widehat{\widehat{W}}_{is}(\lambda),\\
\hat{\Psi}_{i\mathcal{S}}\Big\{\hat{\beta}(\Lambda), \Lambda\Big\} &= \mathrm{vec}\{\hat{\eta}_{i\mathcal{S}}(\hat{\beta}(\lambda), \lambda), \lambda \in \Lambda\},\\
\hat{\mathcal{A}}_m(\cdot) &= \frac{1}{n\mathcal{S}} \sum_{i=1}^{n} \sum_{s=1}^{\mathcal{S}} \Big[\hat{g}'\Big(\lambda_m; \hat{\beta}^{\mathrm{T}}(\lambda_m) W_{is}(\lambda_m)\Big)\Big]^2\\
&\quad J^{\mathrm{T}}_{\hat{\beta}^{(r)}(\lambda_m)} \widehat{\widehat{W}}_{is}(\lambda_m) \widehat{\widehat{W}}^{\mathrm{T}}_{is}(\lambda_m) J_{\hat{\beta}^{(r)}(\lambda_m)}.
\end{aligned}$$

The estimator of the asymptotic covariance matrix of $\hat{\beta}_{\mathrm{SIMEX}}$ is defined as

$$\mathcal{G}_{\mathbf{\Gamma}}(-1, \hat{\mathbf{\Gamma}}) \hat{\Sigma}(\hat{\mathbf{\Gamma}}) \{\mathcal{G}_{\mathbf{\Gamma}}(-1, \hat{\mathbf{\Gamma}})\}^{\mathrm{T}},$$

where

$$\begin{aligned}
\hat{D}(\hat{\mathbf{\Gamma}}) &= s(\hat{\mathbf{\Gamma}}) s^{\mathrm{T}}(\hat{\mathbf{\Gamma}}),\\
\hat{\Sigma}(\hat{\mathbf{\Gamma}}) &= \hat{D}^{-1}(\hat{\mathbf{\Gamma}}) s(\hat{\mathbf{\Gamma}}) \hat{\Sigma} s^{\mathrm{T}}(\hat{\mathbf{\Gamma}}) \hat{D}^{-1}(\hat{\mathbf{\Gamma}}),\\
\hat{\Sigma} &= \hat{\mathcal{J}}(\cdot) \hat{\mathcal{A}}_{11}^{-1}(\cdot) \hat{C}_{11}(\cdot) \Big\{\hat{\mathcal{A}}_{11}^{-1}(\cdot)\Big\}^{\mathrm{T}} \hat{\mathcal{J}}^{\mathrm{T}}(\cdot)
\end{aligned}$$

with $\hat{\mathcal{J}}(\cdot) = \mathrm{diag}\{J_{\hat{\beta}^{(r)}(\lambda)}, \lambda \in \Lambda\}$.

Let $f_0(\cdot)$ be the density function of $\beta^{\mathrm{T}} W$, $\mu_l = \int t^l K(t) dt$ and $\nu_l = \int K^l(t) dt$ for $l = 1, 2$. Define

$$\begin{aligned}
\gamma(\lambda, \mathbb{A}) &= \{\partial / \partial(\mathbb{A})\} \mathcal{G}(\lambda, \mathbb{A}),\\
C(\Lambda, \mathbb{A}) &= \gamma^{\mathrm{T}}(-1, \mathbb{A}) \Big\{\sum_{\lambda \in \Lambda} \gamma(\lambda, \mathbb{A}) \gamma^{\mathrm{T}}(\lambda, \mathbb{A})\Big\}^{-1},
\end{aligned}$$

and $D = \gamma(0, \mathbb{A}) \gamma^{\mathrm{T}}(0, \mathbb{A})$.

Theorem 2 *Suppose that the regularity conditions* $(C1)$–$(C7)$ *hold, and assume that* $nh_2^5 = O(1)$. *Then, as* $n \to \infty$ *and* $\mathcal{S} \to \infty$, *the SIMEX estimator* $\hat{g}_{\mathrm{SIMEX}}(t_0)$ *is asymptotically equivalent to an estimator whose bias and variance are given, respectively, by*

$$C(\Lambda, \mathbb{A}) \sum_{\lambda \in \Lambda} \frac{1}{2} h_2^2 \mu_2 g''(\lambda; t_0) \gamma(\lambda, \mathbb{A})$$

and

$$[nh_2 f_0(t_0)]^{-1}\nu_2 \text{var}\left(Y|\beta^{\mathrm{T}}W = t_0\right) C(\Lambda, \mathbb{A})DC^{\mathrm{T}}(\Lambda, \mathbb{A}),$$

where $g(\lambda; t_0) = E(Y|\beta^{\mathrm{T}}W_s(\lambda) = t_0)$.

Theorem 2 implies that the $\hat{\beta}_{\text{SIMEX}}$ does not affect the estimator of $\hat{g}_{\text{SIMEX}}(t_0)$ because $\hat{\beta}_{\text{SIMEX}}$ is root-n consistent. The result can also be found in other measurement error contexts. For example, both Horowitz and Markatou (1996) and Horrace and Parmeter (2011) have found that the estimation of β in a parametric regression model does not impact density deconvolution of the error term. As pointed out in Carroll et al. (1999), the variance of $\hat{g}_{\text{SIMEX}}(t_0)$ is asymptotically the same as if the measurement error was ignored, but multiplied by a factor, $C(\Lambda, \mathbb{A})DC^{\mathrm{T}}(\Lambda, \mathbb{A})$, which is independent of the regression function.

Applying Theorem 2, the pointwise confidence intervals of $g(t_0)$ can be constructed, but we need to estimate the asymptotic bias and covariance of $\hat{g}_{\text{SIMEX}}(t_0)$. The bias of $\hat{g}_{\text{SIMEX}}(t_0)$ can be given by

$$\widehat{\text{bias}}\left(\hat{g}_{\text{SIMEX}}(t_0)\right) = C(\Lambda, \hat{\mathbb{A}})\frac{1}{\mathcal{S}}\sum_{s=1}^{\mathcal{S}}\sum_{\lambda\in\Lambda}\frac{1}{2}h_2^2\mu_2\hat{g}_s''(\lambda; t_0)\gamma(\lambda, \hat{\mathbb{A}}),$$

where $\hat{g}_s''(\lambda; t_0)$ is obtained by using the local cubic fit with an appropriate pilot bandwidth in estimation step. It is optimal for estimating $g_s''(\lambda; t_0)$ by choosing $h_* = O(n^{-1/7})$, and it can be chosen by the residual squares criterion (Fan and Gijbels (1996)). If the bandwidth $nh_2^5 \to 0$, the bias of $\hat{g}_{\text{SIMEX}}(t_0)$ is negligible. The variance of $\hat{g}_{\text{SIMEX}}(t_0)$ is n^{-1} times the sample variance of the terms (see Apanasovich et al. (2009))

$$C(\Lambda, \hat{\mathbb{A}})\sum_{\lambda\in\Lambda}\frac{\mathcal{S}^{-1}\sum_{s=1}^{\mathcal{S}}\left[Y_i - \hat{g}(\lambda; \hat{\beta}_{\text{SIMEX}}W_{is}(\lambda))\right]K_{h_2}\left(\hat{\beta}_{\text{SIMEX}}W_{is}(\lambda) - t_0\right)}{(n\mathcal{S})^{-1}\sum_{r=1}^{n}\sum_{s=1}^{\mathcal{S}}K_{h_2}\left(\hat{\beta}_{\text{SIMEX}}W_{is}(\lambda) - t_0\right)}\gamma(\lambda, \hat{\mathbb{A}}).$$

3 Numerical studies

3.1 Simulation study

In this section, we evaluate the finite sample performance of the proposed method via simulation studies. Consider the following model:

$$\begin{cases} Y_i = g(\beta^{\mathrm{T}}X_i) + \varepsilon_i, \\ W_i = X_i + U_i, \qquad i = 1, \ldots, n, \end{cases}$$

where $\beta = (\beta_1, \beta_2)^{\mathrm{T}} = (\sqrt{3}/3, \sqrt{6}/3)^{\mathrm{T}}$, X_i is a two-dimensional vector with independent $N(0, 1)$ components, the error ε_i is generated from $N(0, 0.2^2)$, Y_i is generated according to the model, U_i is generated from $N(0, \mathrm{diag}(\sigma_u^2, 0))$. We take $\sigma_u = 0.2, 0.4$, and 0.6 to represent different levels of measurement errors. Two different link functions are considered with $g_1(t) = 3\sin(\pi t/2)$ and $g_2(t) = -2(t-1)^2 + 1$. In simulation study, we compare the naive estimator $\hat{\beta}_{\mathrm{Naive}} = (\hat{\beta}_{1,\mathrm{Naive}}, \hat{\beta}_{2,\mathrm{Naive}})^{\mathrm{T}}$ that ignore measurement errors and the SIMEX estimator $\hat{\beta}_{\mathrm{SIMEX}} = (\hat{\beta}_{1,\mathrm{SIMEX}}, \hat{\beta}_{2,\mathrm{SIMEX}})^{\mathrm{T}}$ with quadratic extrapolation function. The sizes of the samples are $n = 50$, 100, and 150. For each setting, we simulate 500 times to assess the performance. Using the SIMEX algorithm, we take $\lambda = 0, 0.2, \ldots, 2$ and $\mathcal{S} = 50$. We use the Epanechnikov kernel $K(u) = 0.75(1-u^2)_+$. As pointed out in Liang and Wang (2005), the computation is quite expensive for the SIMEX method. In view of this, we apply a "rule of thumb" to select the bandwidths, which is the same in spirit as the selection method in Apanasovich et al. (2009). Specifically, the bandwidths h, h_1, and h_2 are taken to be $cn^{-1/4}(\log n)^{-1/2}$, $cn^{-1/5}$, and $cn^{-1/5}$, where c is the standard deviation of $\hat{\beta}_{\mathrm{int}}^{\mathrm{T}} W$, $\hat{\beta}_{\mathrm{int}}$ is obtained by a linear regression of Y on W with norm 1. Tables 1 and 2 report the biases of β, standard errors (SE), and the coverage probabilities (CP) of 95% confidence intervals obtained as $\hat{\beta} \pm 1.96\mathrm{SE}(\hat{\beta})$.

From Tables 1 and 2, we can see that the SIMEX estimators of β_1 and β_2 have smaller biases than the naive estimators. The coverage probabilities of the SIMEX method are closer to the nominal level than the naive method. However, the standard errors based on the SIMEX estimators are larger than those based on the naive estimators. On the other hand, we can also see that the bias and SE decrease as n increases and the estimators depend on the measurement error. Overall, the SIMEX method is better than the naive method in terms of bias reduction and coverage probabilities.

The estimators and standard errors of the link function $g(t)$ with $\sigma_u = 0.4$ are presented in Figs. 1 and 2, and other cases are similar. From Figs. 1 and 2, we see that the estimated SIMEX curves are closer to the real link function curves than the estimated naive curves. The SE of the SIMEX and naive estimators for the link function are not large, but the SE of the SIMEX estimators are slightly larger than the naive estimators.

Note that the SE based on the SIMEX estimators are larger than the naive estimators for the parameter β and the link function $g(\cdot)$. This can be intuitively illustrated with the linear model. Consider the linear model $Y = \beta_0 + \beta_x x + \epsilon$, where $E(\epsilon) = 0$ and $\mathrm{Var}(\epsilon) = \sigma_\epsilon^2$. If replacing x with $W + \sqrt{\lambda}\sigma_e e_b$, where $e_b \sim N(0, 1)$ and $W = x + e$ with e have mean 0 and variance σ_e^2, then $\hat{\beta}_x(b, \lambda)$ has the asymptotic variance $\{\sigma_\epsilon^2/[\sigma_x^2 + (1+\lambda)\sigma_e^2]\}$. If $\lambda = -1$, then $\beta_x(b, -1)$ is identical to the true parameter, with the asymptotic variance $\sigma_\epsilon^2/\sigma_x^2$. If $\lambda = 0$, $\beta_x(b, 0)$ is just the naive estimator, with the asymptotic variance $\sigma_\epsilon^2/(\sigma_x^2 + \sigma_e^2)$. Hence, it is easy to see that the SE of the naive estimators is smaller than that of the SIMEX estimators.

3.2 Real data analysis

We now analyze a data set from the Framingham Heart Study to illustrate the proposed method. The data set contains 5 variables with 1615 males, and it has been used by

Table 1 Biases, standard errors (SE), and the coverage probabilities (CP) of the SIMEX estimators and naive estimators with different sample sizes for $\sigma_u = 0.4$

$g(t)$	n	SIMEX						Naive					
		$\hat{\beta}_{1,\text{SIMEX}}$			$\hat{\beta}_{2,\text{SIMEX}}$			$\hat{\beta}_{1,\text{Naive}}$			$\hat{\beta}_{2,\text{Naive}}$		
		Bias	SE	CP	Bias	SE	CP	Bias	SE	CP	Bias	SE	CP
$g_1(t)$	50	− 0.0458	0.0814	0.874	0.0190	0.0535	0.878	− 0.0668	0.0458	0.496	0.0418	0.0267	0.454
	100	− 0.0371	0.0564	0.904	0.0178	0.0417	0.901	− 0.0635	0.0349	0.587	0.0414	0.0211	0.619
	150	− 0.0158	0.0332	0.925	0.0142	0.0204	0.925	− 0.0559	0.0298	0.760	0.0363	0.0182	0.769
$g_2(t)$	50	− 0.0404	0.0871	0.881	0.0172	0.0627	0.885	− 0.0762	0.0554	0.498	0.0547	0.0364	0.471
	100	− 0.0379	0.0576	0.917	0.0160	0.0321	0.914	− 0.0738	0.0359	0.635	0.0475	0.0228	0.644
	150	− 0.0171	0.0282	0.933	0.0119	0.0190	0.929	− 0.0493	0.0280	0.686	0.0327	0.0185	0.678

Table 2 Biases, standard errors (SE), and the coverage probabilities (CP) of the SIMEX estimators and naive estimators with different measurement errors for $n = 100$

$g(t)$	σ_u	SIMEX						Naive					
		$\hat{\beta}_{1,\text{SIMEX}}$			$\hat{\beta}_{2,\text{SIMEX}}$			$\hat{\beta}_{1,\text{Naive}}$			$\hat{\beta}_{2,\text{Naive}}$		
		Bias	SE	CP	Bias	SE	CP	Bias	SE	CP	Bias	SE	CP
$g_1(t)$	0.2	−0.0091	0.0398	0.933	0.0075	0.0282	0.928	−0.0158	0.0214	0.917	0.0108	0.0145	0.909
	0.4	−0.0371	0.0564	0.904	0.0178	0.0417	0.901	−0.0635	0.0349	0.587	0.0414	0.0211	0.619
	0.6	−0.0408	0.0740	0.874	0.0212	0.0620	0.879	−0.1210	0.0352	0.207	0.0724	0.0280	0.248
$g_2(t)$	0.2	−0.0082	0.0382	0.936	0.0076	0.0322	0.939	−0.0125	0.0222	0.909	0.0087	0.0133	0.909
	0.4	−0.0379	0.0576	0.917	0.0160	0.0341	0.914	−0.0738	0.0359	0.635	0.0475	0.0228	0.644
	0.6	−0.0395	0.0718	0.891	0.0209	0.0554	0.889	−0.1150	0.0376	0.446	0.0681	0.0364	0.380

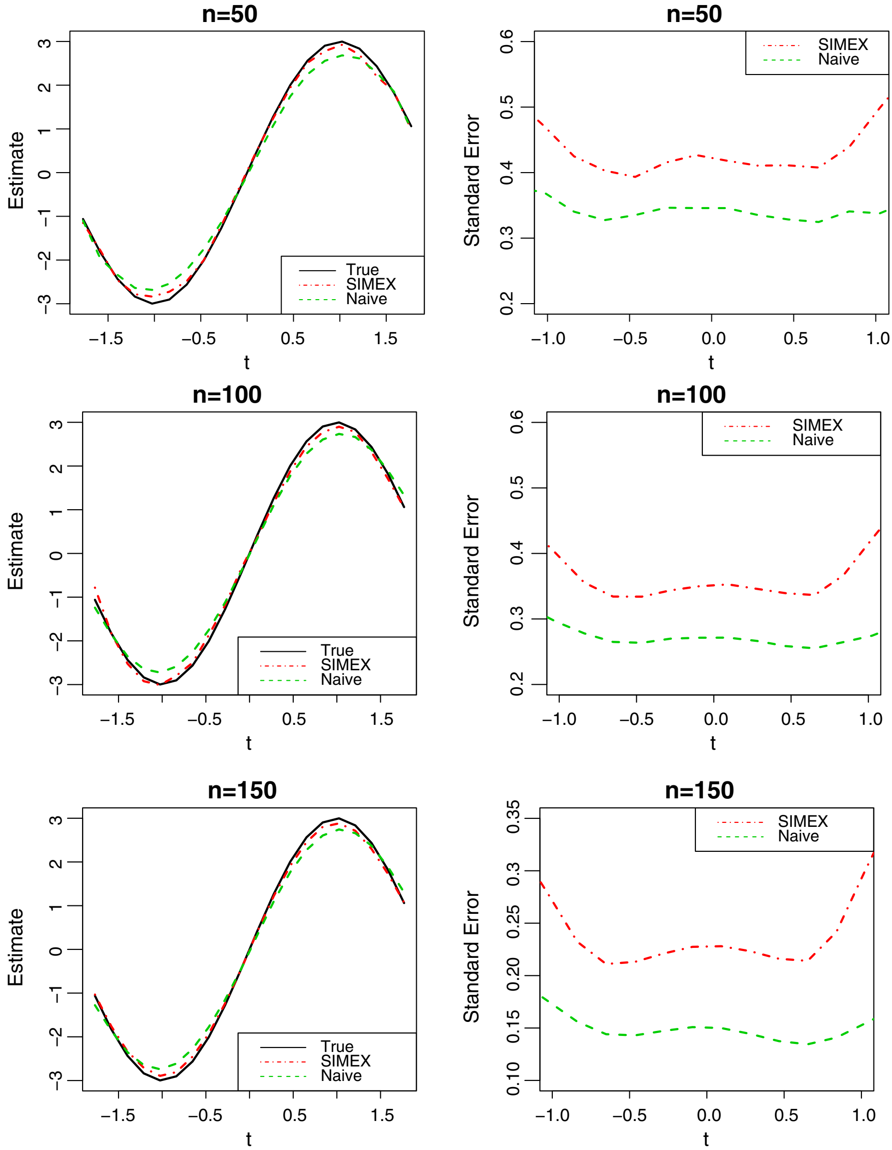

Fig. 1 Estimators and standard errors of the link function $g_1(t)$ with $\sigma_u = 0.4$ for different sample sizes. Dot-dashed lines: SIMEX method. Dashed lines: naive method

many authors to illustrate semiparametric partially linear models (see Liang et al. 1999; Wang et al. 2011). We are interested in whether the age and the serum cholesterol have an effect on the blood pressure. We use the proposed model to analyze the Framingham data to compare the SIMEX and naive estimators. We use the Epanechnikov kernel and the bandwidths $h = 0.0589$ and $h_1 = h_2 = 0.2309$. Let Y be their average blood pressure in a fixed two-year period, W_1 and W_2 be the standardized variable for the

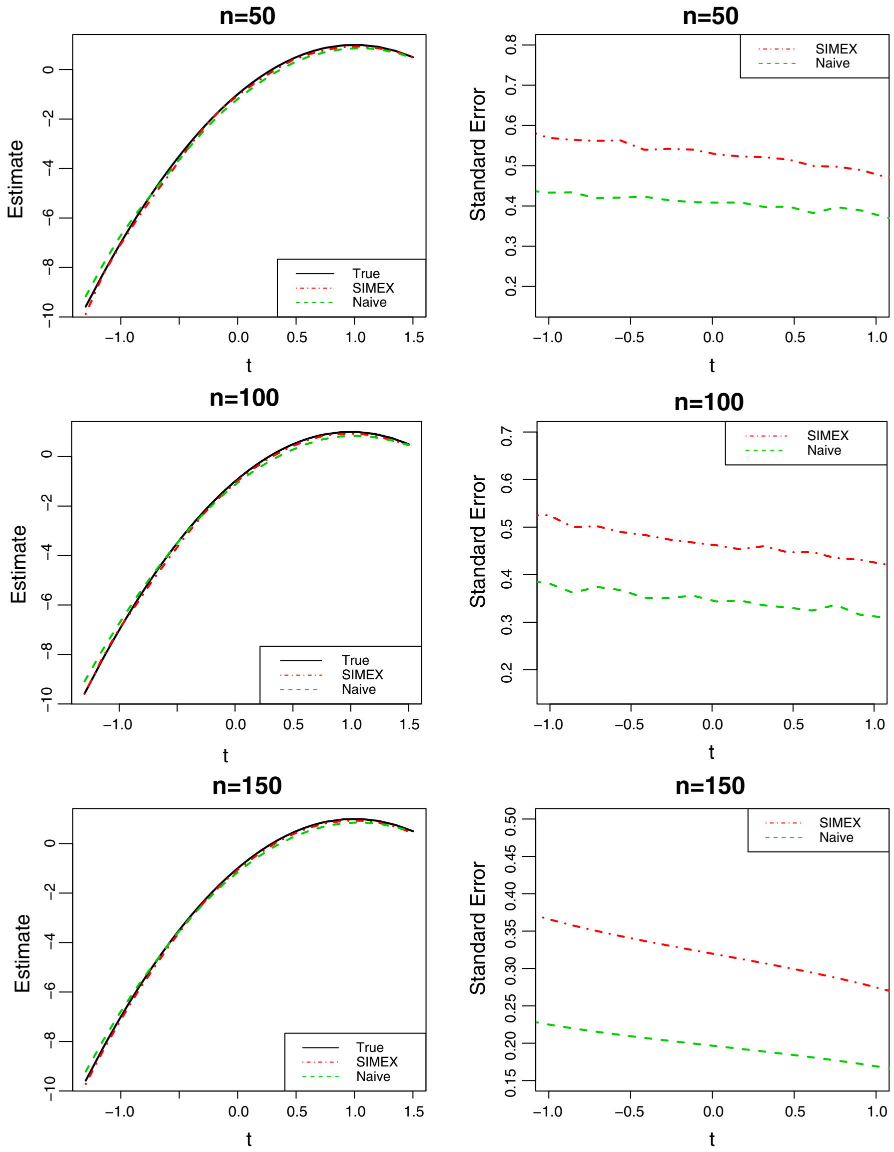

Fig. 2 Estimators and standard errors of the link function $g_2(t)$ with $\sigma_u = 0.4$ for different sample sizes. Dot-dashed lines: SIMEX method. Dashed lines: naive method

logarithm of the serum cholesterol level (log(SC)) and age, respectively. W_1 is subject to the measurement error U, and σ_u^2 is estimated to be 0.2632 by two replicates experiments. Figure 3 shows the duplicated serum cholesterol level measurements from 1615 males. The estimators and standard errors of β and $g(\cdot)$ based on the SIMEX, and naive methods are reported in Table 3, Figs. 4 and 5.

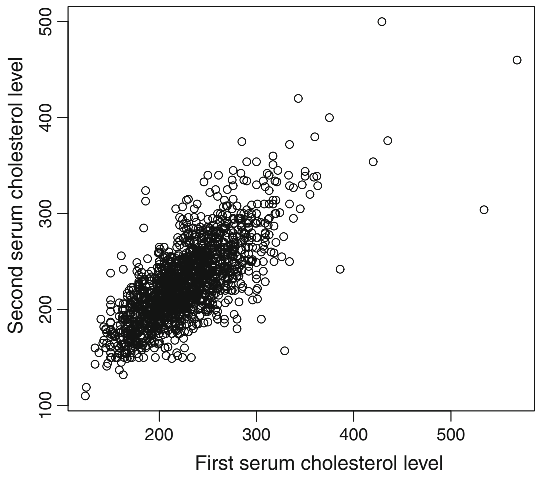

Fig. 3 Duplicated serum cholesterol level measurements from 1615 males in Framingham Heart Study

Table 3 The estimators (SE) of the parameters obtained by the SIMEX and naive methods for the Framingham data

Method	log(SC)	Age
SIMEX	0.5237(0.051)	0.8502(0.070)
Naive	0.4194(0.047)	0.9099(0.065)

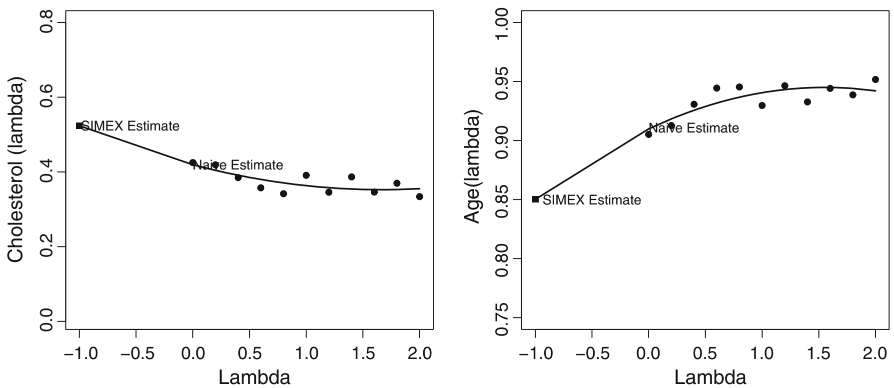

Fig. 4 Extrapolated point estimators for the Framingham data. The simulated estimates $\{\hat{\beta}(\lambda), \lambda\}$ are plotted (dots), and the fitted quadratic function (solid lines) is extrapolated to $\lambda = -1$. The extrapolation results are the SIMEX estimates (squares)

From Table 3, we can see that the SIMEX estimator of the index coefficient log(SC) is larger, while the SIMEX estimator of age is smaller than the naive estimators. The results also show that the serum cholesterol and the age are statistically significant. Figure 4 shows the trace of the extrapolation step for the SIMEX algorithm. The estimators of the two index coefficients for the different λ values are plotted. The SIMEX estimators of index coefficients correspond to -1 on the horizontal axis, while the naive estimators correspond to 0 on the horizontal axis. Figure 5 shows

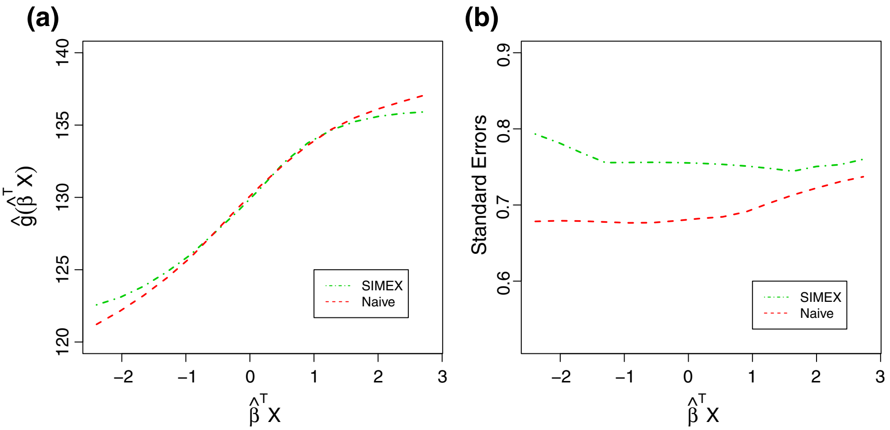

Fig. 5 **a** The link function estimators for the Framingham; **b** the pointwise standard errors of the link function for the Framingham data. Dot-dashed lines: SIMEX method. Dashed lines: naive method

that the estimators, and standard errors of $g(\cdot)$ are obtained by the SIMEX method and the naive method. The patterns of the two curves are similar. From Table 3 and Fig. 5, we find that the standard errors of the SIMEX method are larger than the naive method for the index coefficients and the link function. Table 3 and Fig. 5 also show that the age and the serum cholesterol have a positive association with the blood pressure. As expected, when the measurement error is taken into account, we find a somewhat stronger positive association between the serum cholesterol and the blood pressure. Liang et al. (1999) analyzed the relationship among the blood pressure, the age, and the logarithm of serum cholesterol level by the partially linear errors-in-variables model, where the logarithm of serum cholesterol level was the covariate of the corresponding parameter and the age was a scalar covariate of the corresponding unknown function. When they accounted for the measurement error, the estimator of the parameter was larger than that of ignoring the measurement error. It implied that the blood pressure and the serum cholesterol have a stronger positive correlation when considering the measurement error. The estimator of the unknown function shows that the age is positively associated with the blood pressure. Our findings basically agree with those in Liang et al. (1999). In addition, Liang et al. (1999) suggested that the serum cholesterol level and the blood pressure are a linear relationship. Figure 6 shows the scatter plot of the serum cholesterol level and the blood pressure. It may be more reasonable to use a nonlinear relationship between the serum cholesterol level and the blood pressure from Fig. 6. Hence, we analyze this data set by the single-index model. Figure 5a shows that the blood pressure increases as the index $\hat{\beta}^{\mathrm{T}} X$ increases.

4 Discussion

We propose the SIMEX estimation of the index parameter and the unknown link function for single-index models with covariate measurement error. The asymptotic normality of the estimator of the index parameter and the asymptotic bias and variance

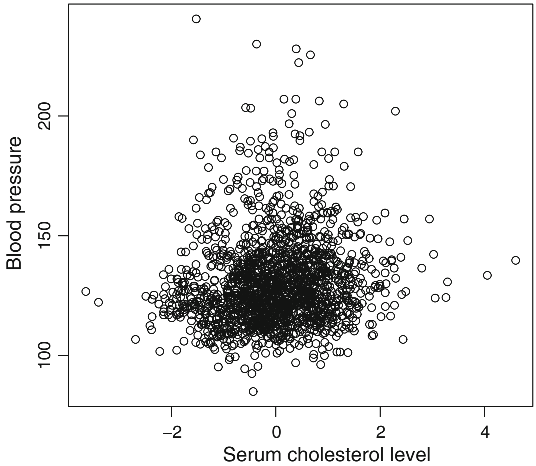

Fig. 6 Scatter plot of the serum cholesterol level and the blood pressure from 1615 males in Framingham Heart Study

of the estimator of the unknown link function are derived under some regularity conditions. The proposed index parameter estimator is root-n consistent, which is similar to that of the estimator of a parameter without measurement error, but the asymptotic covariance has a complicated form. The asymptotic variance of the estimator of the unknown link function is of order $(nh_2)^{-1}$. Our simulation studies indicate that the proposed method works well in practice.

However, it also exists some problems which are worth discussing. Firstly, to reduce the calculation time, we use the DPI bandwidth selection method in Ruppert et al. (1995). The bandwidth selector has not been considered. Carroll et al. (1999) suggested to use the empirical bias bandwidth (EBB) selection, and Staudenmayer and Ruppert (2004) developed another way to estimate the bandwidth. With the similar idea of Staudenmayer and Ruppert (2004), our future work will improve the proposed SIMEX estimation using better bandwidth selection. Secondly, in the estimation step, we transfer the restricted estimating equation with the constraint $\|\beta\| = 1$ to unrestricted estimating Eq. (2.4). Chang et al. (2010) pointed out that the unrestricted estimator of the index parameter is asymptotically more efficient than the restricted estimator. We only provide the asymptotic properties of $\hat{\beta}_{\text{SIMEX}}$ in Theorem 1. The semiparametric efficiency of $\hat{\beta}_{\text{SIMEX}}$ is, however, another important issue and deserves for future analysis. Thirdly, we assume that U is a normal variable (see Carroll et al. 1999; Liang and Ren 2005; Apanasovich et al. 2009), which is a super smooth measurement error. Fan and Truong (1993) studied the effect of errors in variables in nonparametric regression estimation by deconvolution and kernel estimators. They showed that the nonparametric regression with errors in variables depended strongly on the smoothness of error distribution. It is meaningful to discuss how to estimate the index parameter and the link function according to whether the measurement error is ordinary smooth or super smooth.

The proposed method can be extended to some other models, including partially linear single-index models with measurement error in nonparametric components and generalized single-index models with covariate measurement error. It can also be extended to single-index measurement error models with cluster data by assuming working independence in the estimating equations. Future study is needed to investigate how to take into account the within-cluster correlation for cluster data to improve the efficiency of the estimator of the index parameter for single-index measurement error models with cluster data.

Acknowledgements The authors would like to thank the Editor, the Associate Editor, and two reviewers for their insightful comments that led to a substantial improvement of an earlier manuscript. Gaorong Li's research was supported by the National Natural Science Foundation of China (11471029) and the Beijing Natural Science Foundation (1182003). Yiping Yang's research was supported by the National Natural Science Foundation of China (11301569), the Chongqing Research Program of Basic Theory and Advanced Technology (cstc2015jcyjA00023), Fifth batch of excellent talent support program for Chongqing Colleges and University and the Program for University Innovation Team of Chongqing (CXTDX201601026). Tiejun Tong's research was supported by the National Natural Science Foundation of China (11671338) and the Health and Medical Research Fund (04150476).

Appendix

The following notation will be used in the proofs of the lemmas and theorems. Set β_0 be true value, $\mathcal{B}_n = \{\beta : \|\beta\| = 1, \|\beta - \beta_0\| \leq c_1 n^{-1/2}\}$ for some positive constant c_1. Let $f_\lambda(\cdot)$ be the density function of $\beta^{\mathrm{T}} W_s(\lambda)$. Note that if $\lambda = 0$, $f_0(\cdot)$ is the density function of $\beta^{\mathrm{T}} W$.

Lemma 1 *Let $(\zeta_1, \eta_1), \ldots, (\zeta_n, \eta_n)$ be i.i.d. random vectors, where η_i's are scalar random variables. Assume further that $E|\eta_1|^r < \infty$, and $\sup_x \int |y|^r f(x, y)dy < \infty$, where $f(\cdot, \cdot)$ denotes the joint density of (ζ_1, η_1). Let $K(\cdot)$ be a bounded positive function with a bounded support, satisfying a Lipschitz condition. Then*

$$\sup_x \left| \frac{1}{n} \sum_{i=1}^{n} \{K_h(\zeta_i - x)\eta_i - E[K_h(\zeta_i - x)\eta_i]\} \right| = O_p\left(\left\{ \frac{\log(1/h)}{nh} \right\}^{1/2} \right),$$

provided that $n^{2\epsilon-1} h \to \infty$ for some $\epsilon < 1 - r^{-1}$.

Proof This follows immediately from the result that was obtained by Mack and Silverman (1982). □

Lemma 2 *Suppose that conditions (C1)–(C4) hold. Then*

$$\sup_{t \in \mathcal{T}, \beta \in \mathcal{B}_n} \left| \hat{g}(\beta, \lambda; t) - g(\lambda; t) \right| = O_p\left((nh/\log n)^{-1/2} + h^2 \right)$$

and

$$\sup_{t \in \mathcal{T}, \beta \in \mathcal{B}_n} \left| \hat{g}'(\beta, \lambda; t) - g'(\lambda; t) \right| = O_p\left((nh^3/\log n)^{-1/2} + h \right).$$

Proof By the theory of least squares, we have

$$(\hat{g}(\beta, \lambda; t), h\hat{g}'(\beta, \lambda; t))^{\mathrm{T}} = S_n^{-1}(\beta, \lambda; t)\xi_n(\beta, \lambda; t), \tag{A.1}$$

where

$$S_n(\beta, \lambda; t) = \begin{pmatrix} S_{n,0}(\beta, \lambda; t) & h^{-1} S_{n,1}(\beta, \lambda; t) \\ h^{-1} S_{n,1}(\beta, \lambda; t) & h^{-2} S_{n,2}(\beta, \lambda; t) \end{pmatrix}$$

and

$$\xi_n(\beta, \lambda; t) = (\xi_{n,0}(\beta, \lambda; t)), \xi_{n,1}(\beta, \lambda; t))^{\mathrm{T}}$$

with

$$\xi_{n,l}(\beta, \lambda; t) = \frac{1}{n} \sum_{i=1}^{n} Y_i \left(\frac{\beta^{\mathrm{T}} W_{is}(\lambda) - t}{h} \right)^l K_h(\beta^{\mathrm{T}} W_{is}(\lambda) - t)$$

for $l = 0, 1, 2$. A simple calculation yields, for $l = 0, 1, 2, 3$,

$$E[h^{-1} S_{n,l}(\beta, \lambda; t)] = f_\lambda(t) \mu_l + O(h). \tag{A.2}$$

By Lemma 1, we have

$$h^{-1} S_{n,l}(\beta, \lambda; t) - E[h^{-1} S_{n,l}(\beta, \lambda; t)] = O_p \left(\left\{ \frac{\log(1/h)}{nh} \right\}^{1/2} \right),$$

which, combining with (A.2), proves that, for $t \in \mathcal{T}$ and $\beta \in \mathcal{B}_n$,

$$h^{-1} S_{n,l}(\beta, \lambda; t) = f_\lambda(t) \mu_l + O_p \left(\left\{ \frac{\log(1/h)}{nh} \right\}^{1/2} + h \right), \quad l = 0, 1, 2, 3. \tag{A.3}$$

It can be obtained immediately that

$$S_n(\beta, \lambda; t) = S(\lambda; t) + O_p \left(\left\{ \frac{\log(1/h)}{nh} \right\}^{1/2} + h \right),$$

where $S(\lambda; t) = f_\lambda(t) \otimes \mathrm{diag}(1, \mu_2)$, and $\otimes$ is the Kronecker product.

Denote

$$\xi^*_{n,l}(\beta, \lambda; t) = \frac{1}{n} \sum_{i=1}^{n} [Y_i - g(\lambda; \beta^{\mathrm{T}} W_{is}(\lambda))] \left(\frac{\beta^{\mathrm{T}} W_{is}(\lambda) - t}{h} \right)^l K_h(\beta^{\mathrm{T}} W_{is}(\lambda) - t)$$

and

$$\xi^*_n(\beta, \lambda; t) = \left(\xi^*_{n,0}(\beta, \lambda; t), \xi^*_{n,1}(\beta, \lambda; t) \right)^{\mathrm{T}}.$$

Note that

$$E(\xi_n^*(\beta, \lambda; t)) = O(n^{-1/2}). \tag{A.4}$$

By Lemma 1 and (A.4), it can be shown that

$$\xi_n^*(\beta, \lambda; t) = O_p\left(\left\{\frac{\log(1/h)}{nh}\right\}^{1/2} + n^{-1/2}\right). \tag{A.5}$$

By applying Taylor's expansion for $g(\lambda; \beta^{\mathrm{T}} W_{is}(\lambda))$ at t, we can prove that

$$\begin{aligned}\xi_{n,0}(\beta, \lambda; t) - \xi_{n,0}^*(\beta, \lambda; t) &= S_{n,0}(\beta, \lambda; t)g(\lambda; t) + S_{n,1}(\beta, \lambda; t)hg'(\lambda; t)\\ &\quad + \frac{1}{2}h^2 S_{n,2}(\beta, \lambda; t)g''(\lambda; t) + o_p\{h^2 + (nh)^{-1/2}\}\end{aligned}$$

and

$$\begin{aligned}\xi_{n,1}(\beta, \lambda; t) - \xi_{n,1}^*(\beta, \lambda; t) &= S_{n,1}(\beta, \lambda; t)g(\lambda; t) + S_{n,2}(\beta, \lambda; t)hg'(\lambda; t)\\ &\quad + \frac{1}{2}h^2 S_{n,3}(\beta, \lambda; t)g''(\lambda; t) + o_p\{h^2 + (nh)^{-1/2}\}\end{aligned}$$

uniformly hold in $t \in \mathcal{T}$ and $\beta \in \mathcal{B}_n$. Hence

$$\begin{aligned}\xi_n(\beta, \lambda; t) - \xi_n^*(\beta, \lambda; t) = S_n(\beta, \lambda; t)\begin{pmatrix} g(\lambda; t) \\ hg'(\lambda; t)\end{pmatrix} + \frac{1}{2}h^2\begin{pmatrix} S_{n,2}(\beta, \lambda; t)g''(\lambda; t) \\ S_{n,3}(\beta, \lambda; t)g''(\lambda; t)\end{pmatrix}\\ + o_p\{h^2 + (nh)^{-1/2}\}.\end{aligned}$$

Combining this with (A.1)–(A.3) yields

$$\begin{aligned}\begin{pmatrix} \hat{g}(\lambda; t) - g(\lambda; t) \\ h\{\hat{g}'(\lambda; t) - g'(\lambda; t)\}\end{pmatrix} &= S^{-1}(\lambda; t)\xi_n^*((\beta, \lambda; t))\\ &\quad + \frac{1}{2}h^2\begin{pmatrix} \mu_2 g''(\lambda; t) \\ \frac{\mu_3}{\mu_2} g''(\lambda; t)\end{pmatrix} + o_p(h^2 + (nh)^{-1/2}).\end{aligned} \tag{A.6}$$

This together with (A.5) proves Lemma 2. □

Proof of Theorem 1 Assume $\beta(\lambda)$ is the true value based on the model $E[Y|\beta^{\mathrm{T}}(\lambda) W_s(\lambda)] = g(\beta^{\mathrm{T}}(\lambda)W_s(\lambda))$. Using Lemma 2 and the similar method in Theorem 1 of Chang et al. (2010), we have

$$\sqrt{n}\Big(\hat{\beta}_s(\lambda) - \beta(\lambda)\Big) = \sqrt{n} J_{\beta^{(r)}(\lambda)} A_n^{-1}(\beta(\lambda), \lambda) B_n(\beta(\lambda), \lambda) + o_p(1),$$

where

$$A_n(\beta(\lambda),\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left[g'\Big(\lambda;\beta^{\mathrm{T}}(\lambda)W_{is}(\lambda)\Big)\right]^2 J^{\mathrm{T}}_{\beta^{(r)}(\lambda)}\widetilde{W}_{is}(\lambda)\widetilde{W}^{\mathrm{T}}_{is}(\lambda)J_{\beta^{(r)}(\lambda)}$$

and

$$B_n(\beta(\lambda),\lambda) = \frac{1}{n}\sum_{i=1}^{n}\epsilon_{is}(\lambda)g'\Big(\lambda;\beta^{\mathrm{T}}(\lambda)W_{is}(\lambda)\Big)J^{\mathrm{T}}_{\beta^{(r)}(\lambda)}\widetilde{W}_{is}(\lambda)$$

with $\epsilon_{is}(\lambda) = Y_i - g\Big(\lambda;\beta^{\mathrm{T}}(\lambda)W_{is}(\lambda)\Big)$.

Extrapolation step deduces that

$$\sqrt{n}\Big(\hat{\beta}(\lambda)-\beta(\lambda)\Big) = J_{\beta^{(r)}(\lambda)}\mathcal{A}^{-1}(\beta(\lambda),\lambda)n^{-\frac{1}{2}}\sum_{i=1}^{n}\eta_{i\mathcal{S}}(\beta(\lambda),\lambda)+o_p(1), \tag{A.7}$$

where $\eta_{i\mathcal{S}}(\beta(\lambda),\lambda) = \frac{1}{\mathcal{S}}\sum_{s=1}^{\mathcal{S}}\epsilon_{is}(\lambda)g'\Big(\lambda;\beta^{\mathrm{T}}(\lambda)W_{is}(\lambda)\Big)J^{\mathrm{T}}_{\beta^{(r)}(\lambda)}\widetilde{W}_{is}(\lambda)$.

Then, using (A.7), the limit distribution of $\sqrt{n}\Big(\hat{\beta}(\Lambda)-\beta(\Lambda)\Big)$ is multivariate normal distribution with mean zero and covariance Σ.

$\hat{\mathbf{\Gamma}}$ in the extrapolation step is obtained by minimizing $\{\mathrm{Res}(\mathbf{\Gamma})\}\{\mathrm{Res}(\mathbf{\Gamma})\}^{\mathrm{T}}$. The estimating equation for $\hat{\mathbf{\Gamma}}$ is $0 = s(\mathbf{\Gamma})\mathrm{Res}(\mathbf{\Gamma})$, where $s^{\mathrm{T}}(\mathbf{\Gamma}) = \{\partial/\partial(\mathbf{\Gamma})^{\mathrm{T}}\}\mathrm{Res}(\mathbf{\Gamma})$. Then, we have

$$\sqrt{n}(\hat{\mathbf{\Gamma}}-\mathbf{\Gamma}) \xrightarrow{\mathcal{L}} N\{0,\Sigma(\mathbf{\Gamma})\}.$$

Because $\hat{\beta}_{\mathrm{SIMEX}} = \mathcal{G}(-1,\hat{\mathbf{\Gamma}})$, the SIMEX estimator is asymptotically normal with asymptotic variance

$$\mathcal{G}_{\mathbf{\Gamma}}(-1,\mathbf{\Gamma})\Sigma(\mathbf{\Gamma})\{\mathcal{G}_{\mathbf{\Gamma}}(-1,\mathbf{\Gamma})\}^{\mathrm{T}}.$$

□

Proof of Theorem 2 Note that $\|\hat{\beta}_{\mathrm{SIMEX}}-\beta\| = O_p(n^{-1/2})$, similar to the proof of (A.6), we have

$$\begin{aligned}&\hat{g}_s(\lambda;t_0)-g(\lambda;t_0)-\frac{1}{2}h_2^2\mu_2 g''(\lambda;t_0)\\&\quad=[f_\lambda(t_0)]^{-1}\frac{1}{n}\sum_{i=1}^{n}\Big\{[Y_i-g(\lambda;\beta^{\mathrm{T}}W_{is}(\lambda))]K_{h_2}(\beta^{\mathrm{T}}W_{is}(\lambda)-t_0)\Big\}\\&\qquad+o_p\{h_2^2+(nh_2)^{-1/2}\}.\end{aligned} \tag{A.8}$$

Using (A.8) and the decomposition of Carroll et al. (1996), since $\mathcal{S}$ is fixed and $\hat{g}(\lambda; t_0) = \mathcal{S}^{-1} \sum_{s=1}^{\mathcal{S}} \hat{g}_s(\lambda; t_0)$, we have

$$\begin{aligned} &\hat{g}(\lambda; t_0) - g(\lambda; t_0) - \frac{1}{2} h_2^2 \mu_2 g''(\lambda; t_0) \\ &\quad = [f_\lambda(t_0)]^{-1} \frac{1}{n} \sum_{i=1}^{n} \left\{ \mathcal{S}^{-1} \sum_{s=1}^{\mathcal{S}} [Y_i - g(\lambda; \beta^{\mathrm{T}} W_{is}(\lambda))] K_{h_2}(\beta^{\mathrm{T}} W_{is}(\lambda) - t_0) \right\} \\ &\qquad + o_p\{h_2^2 + (nh_2)^{-1/2}\}. \end{aligned} \tag{A.9}$$

If $\lambda = 0$, (A.9) becomes

$$\begin{aligned} &\hat{g}(0; t_0) - g(0; t_0) - \frac{1}{2} h_2^2 \mu_2 g''(0; t_0) \\ &\quad = [n f_0(t_0)]^{-1} \frac{1}{n} \sum_{i=1}^{n} [Y_i - g(0; \beta^{\mathrm{T}} W_i)] K_{h_2}(\beta^{\mathrm{T}} W_i - t_0) + o_p\{h_2^2 + (nh_2)^{-1/2}\}, \end{aligned}$$

which has mean zero and the following asymptotic variance

$$[nh_2 f_0(t_0)]^{-1} \mathrm{var}\left(Y | \beta^{\mathrm{T}} W = t_0\right) \nu_2. \tag{A.10}$$

For $\lambda > 0$, using the similar argument of (A8) in Carroll et al. (1999), we have

$$\mathrm{var}(\hat{g}(\lambda; t_0)) = O\left\{(nh_2 \mathcal{S})^{-1}\right\} + O\left(n^{-1}\right),$$

while for $\lambda = 0$,

$$\mathrm{var}(\hat{g}(\lambda; t_0)) = O\left\{(nh_2)^{-1}\right\}.$$

Then, for B sufficiently large, the variability of $\hat{g}(\lambda; \cdot)$ is negligible for $\lambda > 0$ compared to $\lambda = 0$. Hence, in what follows, we will ignore this variability by treating B as if it was equal to infinity.

We obtain $\hat{\mathbb{A}}$ by solving the following equation

$$0 = \sum_{\lambda \in \Lambda} \{\hat{g}(\lambda; t_0) - \mathcal{G}(\lambda, \mathbb{A})\} \gamma(\lambda, \mathbb{A}). \tag{A.11}$$

Applying a Taylor expansion to the right side of (A.11), we obtain

$$0 = \sum_{\lambda \in \Lambda} \{\hat{g}(\lambda; t_0) - \mathcal{G}(\lambda, \mathbb{A})\} \gamma(\lambda, \mathbb{A}) - \sum_{\lambda \in \Lambda} \gamma(\lambda, \mathbb{A}) \gamma^{\mathrm{T}}(\lambda, \mathbb{A})(\hat{\mathbb{A}} - \mathbb{A}),$$

Hence,

$$\hat{\mathbb{A}} - \mathbb{A} = \left\{ \sum_{\lambda \in \Lambda} \gamma(\lambda, \mathbb{A}) \gamma^{\mathrm{T}}(\lambda, \mathbb{A}) \right\}^{-1} \sum_{\lambda \in \Lambda} \{\hat{g}(\lambda; t_0) - \mathcal{G}(\lambda, \mathbb{A})\} \gamma(\lambda, \mathbb{A}). \quad \text{(A.12)}$$

The right side of (A.12) has approximate mean

$$\left\{ \sum_{\lambda \in \Lambda} \gamma(\lambda, \mathbb{A}) \gamma^{\mathrm{T}}(\lambda, \mathbb{A}) \right\}^{-1} \sum_{\lambda \in \Lambda} \frac{1}{2} h_2^2 \mu_2 g''(\lambda; t_0) \gamma(\lambda, \mathbb{A}),$$

and its approximate variance is given by

$$[n h_2 f_0(t_0)]^{-1} \nu_2 \mathrm{var}(Y | \beta^{\mathrm{T}} W = t_0) \left\{ \sum_{\lambda \in \Lambda} \gamma(\lambda, \mathbb{A}) \gamma^{\mathrm{T}}(\lambda, \mathbb{A}) \right\}^{-1}$$
$$D \left\{ \sum_{\lambda \in \Lambda} \gamma(\lambda, \mathbb{A}) \gamma^{\mathrm{T}}(\lambda, \mathbb{A}) \right\}^{-1}.$$

Because $\hat{g}_{\mathrm{SIMEX}}(t_0) = \mathcal{G}(-1, \hat{\mathbb{A}})$, its asymptotic bias is

$$C(\Lambda, \mathbb{A}) \sum_{\lambda \in \Lambda} \frac{1}{2} h_2^2 \mu_2 g''(\lambda; t_0) \gamma(\lambda, \mathbb{A}),$$

and its asymptotic variance is

$$[n h_2 f_0(t_0)]^{-1} \nu_2 \mathrm{var}\left(Y | \beta^{\mathrm{T}} W = t_0 \right) C(\Lambda, \mathbb{A}) D C^{\mathrm{T}}(\Lambda, \mathbb{A}).$$

This completes the proof. □

References

Apanasovich, T.V., Carroll, R.J., Maity, A.: Simex and standard error estimation in semiparametric measurement error models. Electron. J. Stat. **3**, 318–348 (2009)

Carroll, R.J., Lombard, F., Küchenhoff, H., Stefanski, L.A.: Asymptotics for the SIMEX estimator in structural measurement error models. J. Am. Stat. Assoc. **91**, 242–250 (1996)

Carroll, R.J., Fan, J., Gijbels, I., Wand, M.P.: Generalized partially linear single-index models. J. Am. Stat. Assoc. **92**, 477–489 (1997)

Carroll, R.J., Maca, J., Ruppert, D.: Nonparametric regression in the presence of measurement error. Biometrika **86**, 541–554 (1999)

Carroll, R.J., Ruppert, D., Stefanski, L.A., Crainiceanu, C.M.: Measurement Error in Nonlinear Model, 2nd edn. Chapman & Hall, London (2006)

Chang, Z.Q., Xue, L.G., Zhu, L.X.: On an asymptotically more efficient estimation of the single-index model. J. Multivar. Anal. **101**, 1898–1901 (2010)

Cook, J., Stefanski, L.A.: Simulation-extrapolation method in parametric measurement error models. J. Am. Stat. Assoc. **89**, 1314–1328 (1994)

Cui, X., Härdle, W., Zhu, L.X.: The EFM approach for single-index models. Ann. Stat. **39**, 1658–1688 (2011)
Delaigle, A., Hall, P.: Using SIMEX for smoothing parameter choice in errors-in-variables problems. J. Am. Stat. Assoc. **130**, 280–287 (2008)
Duan, N., Li, K.C.: Slicing regression: a link free regression method. Ann. Stat. **19**, 505–530 (1991)
Fan, J., Gijbels, I.: Local Polynomial Modeling and Its Applications. Chapman & Hall, London (1996)
Fan, J., Truong, Y.K.: Nonparametric regression with errors in variables. Ann. Stat. **21**, 1900–1925 (1993)
Härdle, W., Tsybakov, A.B.: How sensitive are average derivatives. J. Econom. **58**, 31–48 (1993)
Horowitz, J.L., Markatou, M.: Semiparametric estimation of regression models for panel data. Rev. Econ. Stud. **63**, 145–168 (1996)
Horrace, W.C., Parmeter, C.F.: Semiparametric deconvolution with unknown error variance. J. Product. Anal. **35**, 129–141 (2011)
Li, G.R., Peng, H., Dong, K., Tong, T.J.: Simultaneous confidence bands and hypothesis testing in single-index models. Stat. Sin. **24**, 937–955 (2014)
Li, G.R., Lai, P., Lian, H.: Variable selection and estimation for partially linear single-index models with longitudinal data. Stat. Comput. **25**, 579–593 (2015)
Liang, H.: Generalized partially linear mixed-effects models incorporating mismeasured covariates. Ann. Inst. Stat. Math. **61**, 27–46 (2009)
Liang, H., Ren, H.: Generalized partially linear measurement error models. J. Comput. Graph. Stat. **14**, 237–250 (2005)
Liang, H., Wang, N.: Partially linear single-index measurement error models. Stat. Sin. **15**, 99–116 (2005)
Liang, H., Härdle, W., Carroll, R.J.: Estimation in a semiparametric partially linear errors-in-variables model. Ann. Stat. **27**, 1519–1535 (1999)
Liang, H., Liu, X., Li, R.Z., Tsai, C.L.: Estimation and testing for partially linear single-index models. Ann. Stat. **38**, 3811–3836 (2010)
Lin, X., Carroll, R.J.: Nonparametric function estimation for clustered data when the predictor is measured without/with error. J. Am. Stat. Assoc. **95**, 520–534 (2000)
Mack, Y.P., Silverman, B.W.: Weak and strong uniform consistency of kernel regression estimates. Z. Wahrsch. verw. Gebiete **61**, 405–415 (1982)
Naik, P., Tsai, C.L.: Partial least squares estimator for single-index models. J. R. Stat. Soc. Series B **62**, 763–771 (2000)
Pang, Z., Xue, L.G.: Estimation for the single-index models with random effects. Comput. Stat. Data Anal. **56**, 1837–1853 (2012)
Ruppert, D., Sheather, S.J., Wand, M.P.: An effective bandwidth selector for local least squares regression. J. Am. Stat. Assoc. **90**, 1257–1270 (1995)
Staudenmayer, J., Ruppert, D.: Local polynomial regression and simulation-extrapolation. J. R. Stat. Soc. Series B **66**, 17–30 (2004)
Wang, L., Brown, L.D., Cai, T.T.: A difference based approach to the semiparametric partial linear model. Electron. J. Stat. **5**, 619–641 (2011)
Xue, L.G., Zhu, L.X.: Empirical likelihood for single-index models. J. Multivar. Anal. **97**, 1295–1312 (2006)
Yang, S.G., Xue, L.G., Li, G.R.: Simultaneous confidence bands for single-index random effects models with longitudinal data. Stat. Probab. Lett. **85**, 6–14 (2014)
Zhu, L.X., Xue, L.G.: Empirical likelihood confidence regions in a partially linear single-index model. J. R. Stat. Soc. Series B **68**, 549–570 (2006)