Electronic Journal of Statistics
Vol. 13 (2019) 5360–5414
ISSN: 1935-7524
https://doi.org/10.1214/19-EJS1654

Partial linear models with general distortion measurement errors*

Jun Zhang†

College of Mathematics and Statistics
Shenzhen University
Shenzhen, 118060, China
e-mail: zhangjunstat@gmail.com

Abstract: This paper considers partial linear regression models when neither the response variable nor the covariates can be directly observed, but are instead measured with both multiplicative and additive distortion measurement errors. We propose conditional variance estimation methods to calibrate the unobserved variables. A profile least-squares estimator associated with the asymptotic results and confidence intervals is then proposed. To do hypothesis testing of the parameters, a restricted estimator under the null hypothesis and a test statistic are proposed. The asymptotic properties of the estimator and the test statistic are also established. Further, we employ the smoothly clipped absolute deviation penalty to select relevant variables. The resulting penalized estimators are shown to be asymptotically normal and have the oracle property. Estimation, hypothesis testing, and variable selection are discussed under the scenario of multiplicative distortion alone. Simulation studies demonstrate the performance of the proposed procedure and a real example is analyzed to illustrate its applicability.

AMS 2000 subject classifications: Primary 62G05; secondary 62G08, 62G20.
Keywords and phrases: Multiplicative and additive distortion measurement errors, local linear smoothing, variable selection, restricted estimator.

Received February 2019.

Contents

*This work was supported by the the National Natural Science Foundation of China (Grant No. 11871411).

†Corresponding author. The author thanks the editor, the associate editor, and a referee for their constructive suggestions that helped us to improve the early manuscript.

1. Introduction

In many applications involving regression analysis, observations of the variables of interest may include with measurement errors. A general distortion errors-in-variables partial linear model can be written as

$$\begin{cases} Y = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}_0 + g(Z) + \epsilon, \\ \widetilde{Y} = \phi_M(U)Y + \phi_A(U), \\ \widetilde{\boldsymbol{X}} = \boldsymbol{\psi}_M(U)\boldsymbol{X} + \boldsymbol{\psi}_A(U), \end{cases} \tag{1.1}$$

where Y is an unobservable response variable, $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)^{\mathrm{T}}$ is an unobservable continuous covariate vector (throughout this paper, the superscript "T" denotes the transpose operator on a vector or a matrix), $\boldsymbol{\beta}_0$ is an unknown $p \times 1$ parameter vector on a compact parameter space $\Theta_{\boldsymbol{\beta}} \subset \mathbb{R}^p$, Z is a observed univariate covariate, and $g(\cdot)$ is an unknown smooth function. The model error ϵ satisfies $E(\epsilon|\boldsymbol{X}, Z) = 0$ and $E(\epsilon^2|\boldsymbol{X}, Z) < \infty$. The confounding variable $U \in \mathbb{R}^1$ is observable and independent of $(\boldsymbol{X}, Z, Y)$. For the distortion functions $(\boldsymbol{\psi}_M(\cdot), \boldsymbol{\psi}_A(\cdot))$, the multiplicative distortion function $\boldsymbol{\psi}_M(\cdot)$ is a $p \times p$-diagonal matrix given by $\mathrm{diag}\big(\psi_{M,1}(\cdot), \ldots, \psi_{M,p}(\cdot)\big)$, and the additive distortion function $\boldsymbol{\psi}_A(\cdot)$ is a p-dimensional vector given by $\big(\psi_{A,1}(\cdot), \ldots, \psi_{A,p}(\cdot)\big)^{\mathrm{T}}$. Moreover, we assume that $(\phi_M(\cdot), \phi_A(\cdot), \psi_{M,r}(\cdot), \psi_{A,r}(\cdot))$, $r = 1, \ldots, p$, are unknown continuous distortion functions. Note that $(\boldsymbol{\psi}_M(\cdot), \boldsymbol{\psi}_A(\cdot))$ and $(\phi_M(\cdot), \phi_A(\cdot))$ distort the unobserved $\boldsymbol{X}$ and Y in both multiplicative and additive relations.

To date, there has been little discussion on the coexistence of the two kinds of distortion measurement errors in partial linear models. [8] and [38] considered multiplicative distortion measurement errors ($\phi_A(U) \equiv 0, \psi_{A,r}(U) \equiv 0$), and [39] considered the additive distortion measurement errors ($\phi_M(U) \equiv 1, \psi_{M,r}(U) \equiv 1$). Multiplicative distortion measurement data usually occur in health-related studies or medical science research. For example, [7] numerically normalized the collected data according to body mass index (BMI) to study the relation between fibrinogen level and serum transferrin level among hemodialysis patients. This processing of the collected data [7] implies that there may exist a multiplicative relation between the unobserved primary variables and BMI, which is called the confounding variable. Unfortunately, the exact relation between the confounding variable and the primary variables is typically unknown, and simply dividing the confounding variable may lead to an inconsistent estimator of the parameter for a given statistical model. From another perspective, [25, 26] adopted some flexible multiplicative adjustments by introducing unknown smooth distortion functions $\phi_M(u)$ and $\psi_{M,r}(u)$ on the confounding variable. Recently, a number of researchers have studied multiplicative distortion measurement error models (see [1, 22, 21, 28, 25, 20, 27, 11, 41, 36]). The topic of additive distortion measurement errors was first considered in [25]. Later, [20] proposed some graphical techniques for assessing departures from or violations of assumptions regarding the type and form of the additive or multiplicative distortion.

Regarding additive distortion, [33] proposed a residual-based estimator of the correlation coefficient between two unobserved primary variables, and showed that the estimator is asymptotically efficient as if all the variables are observed exactly, i.e., without distortion. [34] studied the estimation and variable selection in partial linear single-index models when the response variable and some covariates are measured with additive distortion measurement errors, i.e., $\phi_M(u) \equiv 1$ and $\psi_{M,r}(u) \equiv 1$, $r = 1, \ldots, p$. Suppose that there are no multiplicative distortions errors ($\phi_M(u) = \psi_{M,r}(u) \equiv 1$). If we treat the nonparametric function $g(Z)$ as a single-index model $g(Z) = g(\theta Z)$ ($\theta = 1$), the estimation method proposed in [34] is not applicable to the partial linear model (1.1). Their leave-one-out component estimation method is not workable because of the identifiability problem for the single-index parameter. The leave-one-out component estimation method for a single-index parameter can theoretically achieve the semiparametric efficient bound, and the asymptotic covariance matrix of profile least-squares estimators is usually invertible, which can be used to construct asymptotic intervals and hypothesis testing for further statistical inference. In detail, to use this method, we need to transform the single-index parameter γ into $\gamma = (\sqrt{1 - \|\gamma_{(-1)}\|^2}, \gamma_{(-1)} = (\gamma_2, \ldots, \gamma_r)^{\mathrm{T}})^{\mathrm{T}}$, and we must also first estimate $\gamma_{(-1)}$. Obviously, the leave-one-out component estimation method can not be used in partial linear models, because $\theta = 1$ is a one-dimensional parameter and $\theta_{(-1)}$ is an empty set.

This paper intends discusses partial linear models that contain both multiplicative and additive distortion measurement errors. Because the model (1.1) contains the additive distortion functions $\phi_A(u)$ and $\psi_{A,r}(u)$, the calibration estimation proposed in this paper is different from that in [1] and [34]. In [1], the authors only considered the existence of multiplicative distortion measurement errors ($\phi_A(u) = \psi_{A,r}(u) \equiv 0$), and used the conditional mean calibration procedure to estimate $(\phi_M(u), Y, \psi_{M,r}(u), X_r)$: $Y = \frac{\widetilde{Y}}{\phi_M(u)}$, $\phi_M(u) = \frac{E(\widetilde{Y}|U=u)}{E(\widetilde{Y})}$, $X_r = \frac{\widetilde{X}_r}{\psi_{M,r}(u)}$, $\psi_{M,r}(u) = \frac{E(\widetilde{X}_r|U=u)}{E(\widetilde{X}_r)}$, $r = 1, \ldots, p$. In [34], the authors considered only additive distortion measurement errors ($\phi_M(u) = \psi_{M,r}(u) \equiv 1$), and used the conditional mean calibration procedure to obtain the relations $Y - E(Y) = \widetilde{Y} - E(\widetilde{Y}|U)$, $X - E(X_r) = \widetilde{X}_r - E(\widetilde{X}_r|U)$, $r = 1, \ldots, p$. The authors then used the regression "residuals" $\{\widetilde{Y}_i - \hat{E}(\widetilde{Y}_i|U_i), \widetilde{X}_{ri} - \hat{E}(\widetilde{X}_{ri}|U_i)\}$ to estimate the parameters in PLSiMs.

From model (1.1), $E(\widetilde{Y}|U = u) = \phi_M(u)E(Y) + \phi_A(u)$, $E(\widetilde{X}_r|U = u) = \psi_{M,r}(u)E(X_r) + \psi_{A,r}(u)$. Note that all the distortion functions $(\phi_M(u), \phi_A(u))$, $(\psi_{M,r}(u), \psi_{A,r}(u))$ are unknown, and the conditional mean calibration [1] and residual-based calibration procedures in [34] are no longer workable, because we can not estimate them through $E(\widetilde{Y}|U = u)$ or $E(\widetilde{X}_r|U = u)$ alone. Consequently, in this paper, we propose a new calibration procedure by coupling the conditional means $(E(\widetilde{Y}|U = u), E(\widetilde{X}_r|U = u))$ and conditional variances $(\mathrm{Var}(\widetilde{Y}|U = u), \mathrm{Var}(\widetilde{X}_r|U = u))$. Using these estimates $\hat{E}(\widetilde{Y}|U=u)$, $\widehat{\mathrm{Var}}(\widetilde{Y}|U=u)$, $\hat{E}(\widetilde{X}_r|U = u)$, $\widehat{\mathrm{Var}}(\widetilde{X}_r|U = u)$, we obtain $(\hat{\phi}_M(u), \hat{\phi}_A(u), \hat{\psi}_{M,r}(u), \hat{\psi}_{A,r}(u))$ and $\hat{Y} = \frac{\widetilde{Y} - \hat{\phi}_A(U)}{\hat{\phi}_M(U)}$, $\hat{X}_r = \frac{\widetilde{X}_r - \hat{\psi}_{M,r}(U)}{\hat{\psi}_{M,r}(U)}$. Note that the calibrated variables $(\hat{Y}, \hat{X}_r)$ and the asymptotic results obtained in this paper are all different from those in [1] and [34].

With these calibrated variables, we use a profile least-squares estimation to obtain a root-n consistent estimator of $\boldsymbol{\beta}_0$. Specifically, we consider the estimation efficiency of the proposed estimators in the case when $\phi_A(u) = \psi_{A,r}(u) \equiv 0$. In this setting, without additive distortion, we further propose a second estimator by using the conditional absolute mean technique [2, 41]. The normal approximation is derived by estimating the asymptotic covariance matrices and the empirical likelihood-based statistics are proposed to construct two different asymptotic confidence intervals of the parameter $\boldsymbol{\beta}_0$.

To make further inferences, we consider the problem of checking whether the linear combination $\boldsymbol{A}\boldsymbol{\beta}_0 = \boldsymbol{b}$ holds. A restricted profile least-squares estimator and a test statistic are proposed by introducing Lagrange multipliers under the null hypothesis. Under the null hypothesis, the limiting distribution of the test statistic is shown to be a standard chi-squared distribution. We also investigate the asymptotic properties of the estimator and the test statistic under the local alternative hypothesis. Finally, to perform variable selection, we propose a profile penalized least-squares method based on the smoothly clipped absolute deviation method [4, SCAD]. We demonstrate that the resulting SCAD-based solution is selection-consistent. Monte Carlo simulation experiments are conducted to examine the performance of the proposed estimation and test procedures.

The remainder of this paper is organized as follows. In Section 2, we propose the conditional variance calibration for the unobserved variables, present a profile least-squares estimator of the parameter, and derive the related asymptotic results. In Section 3, the confidence intervals the of parameter are proposed. Section 4 considers the problem of checking whether the linear restriction $\boldsymbol{A}\boldsymbol{\beta}_0 = \boldsymbol{b}$ holds. In Section 5, variable selection for parameter $\boldsymbol{\beta}_0$ is discussed. Section 6 covers two estimation procedures when only multiplicative distortion exists. Hypothesis testing, confidence intervals construction, and variable selection are also discussed. In Section 7, we report the results of simulation studies. In Section 8, we present statistical analysis results using real data. All technical proofs of the asymptotic results are given in the appendix.

2. Estimation method and asymptotic results

2.1. Calibration

We first calibrate unobserved Y and $\boldsymbol{X}$ by using the observed *i.i.d.* sample $\left\{\widetilde{Y}_i, \widetilde{\boldsymbol{X}}_i, U_i\right\}_{i=1}^n$. To ensure identifiability, it is assumed that

$$E[\phi_M(U)] = 1, \quad E[\phi_A(U)] = 0, \tag{2.1}$$

$$E[\psi_{M,r}(U)] = 1, \quad E[\psi_{A,r}(U)] = 0, \quad r = 1, \ldots, p. \tag{2.2}$$

The identifiability conditions (2.1)-(2.2) are introduced by [26, 25], and it is analogous to the classical additive measurement errors: $E(e) = 0$ for $W = X + e$, where W is error-prone and X is error-free [11, 31].

Define

$$m_{\widetilde{Y}}(u) = E(\widetilde{Y}|U = u), \quad \sigma_{\widetilde{Y}|U}(u) = \sqrt{\mathrm{Var}(\widetilde{Y}|U = u)},$$

$$m_{\widetilde{X}_r}(u) = E(\widetilde{X}_r|U = u), \quad \sigma_{\widetilde{X}_r|U}(u) = \sqrt{\mathrm{Var}(\widetilde{X}_r|U = u)}, \quad r = 1, \ldots, p.$$

Suppose that $\sigma_Y \prod_{r=1}^p \sigma_{X_r} > 0$, where $\sigma_Y = \sqrt{\mathrm{Var}(Y)}$, $\sigma_{X_r} = \sqrt{\mathrm{Var}(X_r)}$, $r = 1, \ldots, p$. Under the independence condition between U and $(Y, \boldsymbol{X})$, the identifiability conditions (2.1)-(2.2) and condition (C1) entail that:

$$\sigma_{\widetilde{Y}|U}(u) = \phi_M(u)\sigma_Y, \quad E\left(\sigma_{\widetilde{Y}|U}(U)\right) = \sigma_Y, \tag{2.3}$$

$$\sigma_{\widetilde{X}_r|U}(u) = \psi_{M,r}(u)\sigma_{X_r}, \quad E\left(\sigma_{\widetilde{X}_r|U}(U)\right) = \sigma_{X_r}, \quad r = 1, \ldots, p. \tag{2.4}$$

The relations (2.3)-(2.4) entail that

$$\phi_M(u) = \frac{\sigma_{\widetilde{Y}|U}(u)}{E\left(\sigma_{\widetilde{Y}|U}(U)\right)} = \frac{\sigma_{\widetilde{Y}|U}(u)}{\sigma_Y}, \tag{2.5}$$

$$\psi_{M,r}(u) = \frac{\sigma_{\widetilde{X}_r|U}(u)}{E\left(\sigma_{\widetilde{X}_r|U}(U)\right)} = \frac{\sigma_{\widetilde{X}_r|U}(u)}{\sigma_{X_r}}, \quad r = 1, \ldots, p. \tag{2.6}$$

Because the square root of variances σ_Y, σ_{X_r}'s are used in the denominators of (2.5)-(2.6), the condition $\sigma_Y \prod_{r=1}^p \sigma_{X_r} > 0$ should be imposed here. Equivalently, it is required that the covariates X_r's and response variable Y are non-constant variables.

Using (2.5)-(2.6), $m_{\widetilde{Y}}(u) = \phi_M(u)E(Y) + \phi_A(u) = \phi_M(u)E(\widetilde{Y}) + \phi_A(u)$, and $m_{\widetilde{X}_r}(u) = \psi_{M,r}(u)E(X_r) + \psi_{A,r}(u) = \psi_{M,r}(u)E(\widetilde{X}_r) + \psi_{A,r}(u)$, we have

$$\phi_A(u) = m_{\widetilde{Y}}(u) - \frac{\sigma_{\widetilde{Y}|U}(u)}{E\left(\sigma_{\widetilde{Y}|U}(U)\right)} E(\widetilde{Y}), \tag{2.7}$$

$$\psi_{A,r}(u) = m_{\widetilde{X}_r}(u) - \frac{\sigma_{\widetilde{X}_r|U}(u)}{E\left(\sigma_{\widetilde{X}_r|U}(U)\right)} E(\widetilde{X}_r). \tag{2.8}$$

Together with (2.5)-(2.8), we have

$$Y = \frac{\widetilde{Y} - \phi_A(U)}{\phi_M(U)} = \frac{E\left(\sigma_{\widetilde{Y}|U}(U)\right)}{\sigma_{\widetilde{Y}|U}(U)} \left\{\widetilde{Y} - m_{\widetilde{Y}}(U)\right\} + E(\widetilde{Y}), \tag{2.9}$$

$$X_r = \frac{\widetilde{X}_r - \psi_{A,r}(U)}{\psi_{M,r}(U)} = \frac{E\left(\sigma_{\widetilde{X}_r|U}(U)\right)}{\sigma_{\widetilde{X}_r|U}(U)} \left\{\widetilde{X}_r - m_{\widetilde{X}_r}(U)\right\} + E(\widetilde{X}_r). \tag{2.10}$$

Thus, the unobserved variables $\{Y, X_r, r = 1, \ldots, p\}$ can be obtained through (2.9)-(2.10) at the population level. We summarize the calibration procedure as follows.

- The Nadaraya-Watson estimators are used to estimate $\phi_M(u)$, $\phi_A(u)$, $\psi_{M,r}(u)$ and $\psi_{A,r}(u)$. Define $\hat{f}_U(u) = \frac{1}{nh}\sum_{j=1}^n K_h(U_j - u)$, here $K_h(\cdot) = h^{-1}K(\cdot/h)$, $K(\cdot)$ is a symmetric density function, h is a positive-valued bandwidth. Let

$$\hat{m}_{\widetilde{Y}}(u) = \frac{1}{nh\hat{f}_U(u)} \sum_{j=1}^n K_h(U_j - u)\widetilde{Y}_j,$$

$$\hat{\sigma}^2_{\widetilde{Y}|U}(u) = \frac{1}{nh\hat{f}_U(u)} \sum_{j=1}^n K_h(U_j - u)\left[\widetilde{Y}_j - \hat{m}_{\widetilde{Y}}(U_j)\right]^2,$$

$$\hat{m}_{\widetilde{X}_r}(u) = \frac{1}{nh\hat{f}_U(u)} \sum_{j=1}^n K_h(U_j - u)\widetilde{X}_{rj},$$

$$\hat{\sigma}^2_{\widetilde{X}_r|U}(u) = \frac{1}{nh\hat{f}_U(u)} \sum_{j=1}^n K_h(U_j - u)\left[\widetilde{X}_{rj} - \hat{m}_{\widetilde{X}_r}(U_j)\right]^2.$$

We obtain $\hat{\sigma}_{\widetilde{Y}|U}(u) = \sqrt{\hat{\sigma}^2_{\widetilde{Y}|U}(u)}$, $\hat{\sigma}_{\widetilde{X}_r|U}(u) = \sqrt{\hat{\sigma}^2_{\widetilde{X}_r|U}(u)}$, and

$$\hat{E}\left(\sigma_{\widetilde{Y}|U}(U)\right) = \hat{\sigma}_Y = \frac{1}{n}\sum_{i=1}^n \hat{\sigma}_{\widetilde{Y}|U}(U_i),$$

$$\hat{E}\left(\sigma_{\widetilde{X}_r|U}(U)\right) = \hat{\sigma}_{X_r} = \frac{1}{n}\sum_{i=1}^{n}\hat{\sigma}_{\widetilde{X}_r|U}(U_i).$$

Then, the distortion functions are estimated as

$$\hat{\phi}_M(u) = \frac{\hat{\sigma}_{\widetilde{Y}|U}(u)}{\hat{E}\left(\sigma_{\widetilde{Y}|U}(U)\right)}, \quad \hat{\psi}_{M,r}(u) = \frac{\hat{\sigma}_{\widetilde{X}_r|U}(u)}{\hat{E}\left(\sigma_{\widetilde{X}_r|U}(U)\right)}, \tag{2.11}$$

$$\hat{\phi}_A(u) = \hat{m}_{\widetilde{Y}}(u) - \frac{\hat{\sigma}_{\widetilde{Y}|U}(u)}{\hat{E}\left(\sigma_{\widetilde{Y}|U}(U)\right)}\overline{\widetilde{Y}}, \quad \overline{\widetilde{Y}} = \frac{1}{n}\sum_{i=1}^{n}\widetilde{Y}_i \tag{2.12}$$

$$\hat{\psi}_{A,r}(u) = \hat{m}_{\widetilde{X}_r}(u) - \frac{\hat{\sigma}_{\widetilde{X}_r|U}(u)}{\hat{E}\left(\sigma_{\widetilde{X}_r|U}(U)\right)}\overline{\widetilde{X}_r}, \quad \overline{\widetilde{X}_r} = \frac{1}{n}\sum_{i=1}^{n}\widetilde{X}_{ri}. \tag{2.13}$$

- Using (2.11)-(2.13), the calibrated variables for $\{Y_i, X_{ri}, r = 1, \ldots, p\}_{i=1}^{n}$ are defined as

$$\hat{Y}_i = \frac{\widetilde{Y}_i - \hat{\phi}_A(U_i)}{\hat{\phi}_M(U_i)}, \quad \hat{X}_{ri} = \frac{\widetilde{X}_{ri} - \hat{\psi}_{A,r}(U_i)}{\hat{\psi}_{M,r}(U_i)}. \tag{2.14}$$

2.2. A profile least squares estimator

In the following, we define $A^{\otimes 2} = AA^{\mathrm{T}}$ for any matrix or vector A. From model (1.1), we have

$$Y - E(Y|Z) = [\boldsymbol{X} - E(\boldsymbol{X}|Z)]^{\mathrm{T}}\boldsymbol{\beta}_0 + \epsilon. \tag{2.15}$$

Under the identifiability conditions (2.1)-(2.2) and the independence condition between U and Z, the model (2.15) is equivalent to

$$Y - E(\widetilde{Y}|Z) = \left[\boldsymbol{X} - E(\widetilde{\boldsymbol{X}}|Z)\right]^{\mathrm{T}}\boldsymbol{\beta}_0 + \epsilon. \tag{2.16}$$

Thus, a profile least squares estimator of $\boldsymbol{\beta}_0$ (at the population level) is obtained as

$$\boldsymbol{\beta}_0 = \left[E\left\{\left[\boldsymbol{X} - E(\widetilde{\boldsymbol{X}}|Z)\right]^{\otimes 2}\right\}\right]^{-1} E\left\{\left[\boldsymbol{X} - E(\widetilde{\boldsymbol{X}}|Z)\right]\left[Y - E(\widetilde{Y}|Z)\right]\right\}.$$

We define $S_Y(z)$ and $S_{\boldsymbol{X}}(z) = \left(s_{X_1}(z), \ldots, s_{X_p}(z)\right)^{\mathrm{T}}$ as

$$S_Y(z) = E(\widetilde{Y}|Z = z), \quad s_{X_r}(z) = E(\widetilde{X}_r|Z = z), \quad r = 1, \ldots, p.$$

To obtain the estimator of $\boldsymbol{\beta}_0$, we use local linear estimators to estimate $S_Y(z)$ and $s_{X_r}(z)$. These estimators are defined as

$$\hat{S}_Y(z) \quad = \quad \frac{Q_{n2}(z)M_{n0,\widetilde{Y}}(z) - Q_{n1}(z)M_{n1,\widetilde{Y}}(z)}{Q_{n2}(z)Q_{n0}(z) - [Q_{n1}(z)]^2}, \tag{2.17}$$

$$\hat{s}_{X_r}(z) = \frac{Q_{n2}(z)M_{n0,\widetilde{X}_r}(z) - Q_{n1}(z)M_{n1,\widetilde{X}_r}(z)}{Q_{n2}(z)Q_{n0}(z) - [Q_{n1}(z)]^2}, \tag{2.18}$$

where, $M_{n\delta,\widetilde{W}}(z) = \frac{1}{nh_1}\sum_{i=1}^{n}\left(\frac{Z_i-z}{h_1}\right)^{\delta} K\left(\frac{Z_i-z}{h_1}\right)\widetilde{W}_i$ with $\widetilde{W}_i = \widetilde{Y}_i$ or $\widetilde{W}_i = \widetilde{X}_{ri}$, $\delta = 0, 1$, $r = 1, \dots, p$, and $Q_{n\omega}(z) = \frac{1}{nh_1}\sum_{i=1}^{n}\left(\frac{Z_i-z}{h_1}\right)^{\omega} K\left(\frac{Z_i-z}{h_1}\right)$, $\omega = 0, 1, 2$.

Based on (2.16), the profile least squares estimator of $\boldsymbol{\beta}_0$ is obtained as

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \left\{\frac{1}{n}\sum_{i=1}^{n}\left[\hat{\boldsymbol{X}}_i - \hat{S}_{\boldsymbol{X}}(Z_i)\right]^{\otimes 2}\right\}^{-1} \\ &\quad \times \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\boldsymbol{X}}_i - \hat{S}_{\boldsymbol{X}}(Z_i)\right\}\left\{\hat{Y}_i - \hat{S}_Y(Z_i)\right\},\end{aligned} \tag{2.19}$$

where $\hat{\boldsymbol{X}}_i = (\hat{X}_{1i}, \dots, \hat{X}_{pi})^{\mathrm{T}}$ and $\hat{S}_{\boldsymbol{X}}(Z_i) = \left(\hat{s}_{X_1}(Z_i), \dots, \hat{s}_{X_p}(Z_i)\right)^{\mathrm{T}}$.

After obtaining estimator $\hat{\boldsymbol{\beta}}$, using the relation $g(z) = E(Y - \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}_0|Z = z) = E(\widetilde{Y} - \widetilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{\beta}_0|Z = z)$, the nonparametric part $g(z)$ is estimated by using the local linear estimator

$$(\hat{g}(z), \widehat{g'}(z)) = \arg\min_{a_z, b_z}\sum_{i=1}^{n}\left\{\widetilde{Y}_i - \widetilde{\boldsymbol{X}}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}} - a_z - b_z(Z_i - z)\right\}K_{h_2}(Z_i - z), \tag{2.20}$$

here, h_2 is the bandwidth. After simple calculation, we have

$$\hat{g}(z) = \hat{a}_z = \frac{T_{n2}(z)V_{n0}(z) - T_{n1}(z)V_{n1}(z)}{T_{n2}(z)T_{n0}(z) - [T_{n1}(z)]^2}, \tag{2.21}$$

where,

$$\begin{aligned}V_{n\delta}(z) &= \frac{1}{nh_2}\sum_{i=1}^{n}\left(\frac{Z_i - z}{h_2}\right)^{\delta} K\left(\frac{Z_i - z}{h_2}\right)[\widetilde{Y}_i - \widetilde{\boldsymbol{X}}_i^{\mathrm{T}}\hat{\boldsymbol{\beta}}] \\ T_{n\omega}(z) &= \frac{1}{nh_2}\sum_{i=1}^{n}\left(\frac{Z_i - z}{h_2}\right)^{\omega} K\left(\frac{Z_i - z}{h_2}\right), \delta = 0, 1, \omega = 0, 1, 2.\end{aligned}$$

In the following Theorem 2.1 and Theorem 2.2, we present the asymptotic results of estimators $\hat{\boldsymbol{\beta}}$ and $\hat{g}(z)$.

2.3. Asymptotic results

We now list the assumptions needed in the following theorems.

(C1) The distortion functions $\phi(u) > 0$ and $\psi_r(u) > 0$ for all $u \in [\mathcal{U}_L, \mathcal{U}_R]$, $r = 1, \dots, p$, where $[\mathcal{U}_L, \mathcal{U}_R]$ denotes the compact support of U. Moreover, the distortion functions $\phi_M(u)$, $\phi_A(u)$, $\psi_{M,r}(u)$'s and $\psi_{A,r}(u)$'s have third order continuous derivatives. The density function $f_U(u)$ of the random variable U is bounded away from 0 and satisfies the Lipschitz condition of order 1 on $[\mathcal{U}_L, \mathcal{U}_R]$.

(C2) For some $r \geq 4$, $E(|Y|^r) < \infty$, $E(|X_s|^r) < \infty$, $s = 1, \ldots, p$. The matrix $\mathbf{\Sigma}_0$ defined in Theorem 2.1 is a positive-definite matrix.

(C3) The density function of Z, $f_Z(z)$ is bounded away from zero on $\mathcal{Z}$, where $\mathcal{Z}$ is a compact set in $\mathcal{R}^1$. Moreover, $f_Z(z)$, $E(X_s|Z = z)$, $E(Y|Z = z)$ and $g(z)$ have bounded continuous second order derivatives on $\mathcal{Z}$.

(C4) The kernel function $K(\cdot)$ is a symmetric bounded density function supported on $[-A, A]$ satisfying a Lipschitz condition. $K(\cdot)$ also has second-order continuous bounded derivatives, satisfying $K^{(j)}(\pm A) = 0$, $\mu_2 = \int s^2 K(s)ds \neq 0$ and $\mu_{K^2} = \int K^2(s)ds > 0$.

(C5) As $n \to \infty$, the bandwidths h and h_1 satisfy $nh^4 \to 0$, $\frac{\log^2 n}{nh^2} \to 0$ and $nh_1^8 \to 0$ and $\frac{\log^2 n}{nh_1^2} \to 0$.

(C6) The tuning parameters λ_j $j = 1, \ldots, p$ satisfy $\lambda_j \to 0$, $\sqrt{n}\lambda_j \to \infty$ as $n \to \infty$, and

$$\text{liminf}_{n\to\infty}\text{liminf}_{u\to 0^+} p'_{\lambda_j}(u)/\lambda_j > 0.$$

These conditions are not restrictive and are satisfied in most practical situations. Condition (C1) is the typical smoothing assumptions in the distortion measurement errors literature, see also in [2, 32]. Conditions (C2)-(C3) are needed for the asymptotic normality of our statistics. See, for example [35]. Conditions (C4) is a common condition for kernel function $K(\cdot)$, and the Epanechnikov kernel satisfies this condition. This condition ensures the kernel smoothing estimators $\frac{1}{nh}\sum_{i=1}^n K(\frac{U_i-u}{h})$ and $\frac{1}{nh_1}\sum_{i=1}^n K(\frac{Z_i-z}{h_1})$ positive. [1] considered to use high-order kernel function $K^*(t) = \frac{15}{32}(3 - 7t^2)I\{|t| \leq 1\}$ such that $\int t^2 K^*(t)dt = 0$ but $\int t^4 K^*(t)dt > 0$. The high-order kernel function $K^*(t)$ has zero value and negative values when $|t| \geq \sqrt{3/7}$. For example, the involved estimators with $\frac{1}{nh}\sum_{i=1}^n K^*(\frac{U_i-u}{h})$ may produce negative values, and this is the drawback of the high-order kernel function. Condition (C5) is on bandwidths (h, h_1) in the nonparametric kernel smoothing. For bandwidth h_1, Condition (C5) requires that the "optimal" rate of order $n^{-1/5}$ can be used [35]. For bandwidth h, a under-smoothing condition $nh^4 \to 0$ is needed. The consequence of under-smoothing is that the biases of the nonparametric estimates are kept small and preclude the optimal bandwidth for h. Condition (C6) is the technique condition involved in SCAD [4].

In the following, we define

$$\begin{aligned}
&\mathbf{\Sigma}_0 = E\left\{[\boldsymbol{X} - E(\boldsymbol{X}|Z)]^{\otimes 2}\right\}, \quad \mathbf{\Sigma}_{0\epsilon} = E\left\{\epsilon^2\left[\boldsymbol{X} - E(\boldsymbol{X}|Z)\right]^{\otimes 2}\right\}, \\
&\boldsymbol{G}(\boldsymbol{X}, \boldsymbol{\psi}_M(U)) = \text{diag}\Bigg(\left[\frac{(X_1 - E(X_1))^2}{2\sigma_{X_1}^2} + \frac{1}{2}\right][\psi_{M,1}(U) - 1], \\
&\qquad \ldots, \left[\frac{(X_p - E(X_p))^2}{2\sigma_{X_p}^2} + \frac{1}{2}\right][\psi_{M,p}(U) - 1]\Bigg),
\end{aligned}$$

and

$$\mathbf{\Sigma}_{\phi_M,\boldsymbol{\psi}_M} = E\Bigg\{\Bigg[[\phi_M(U) - 1]\left[\frac{(Y - E(Y))^2}{2\sigma_Y^2} + \frac{1}{2}\right]\boldsymbol{\beta}_0$$

$$-\boldsymbol{G}(\boldsymbol{X}, \boldsymbol{\psi}_M(U))\boldsymbol{\beta}_0\Bigg]^{\otimes 2}\Bigg\}.$$

Theorem 2.1. *Suppose conditions (C1)-(C5) hold, we have*

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \xrightarrow{L} N\left(\mathbf{0}_p, \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_{0\varepsilon}\boldsymbol{\Sigma}_0^{-1} + \boldsymbol{\Sigma}_{\phi_M, \boldsymbol{\psi}_M}\right).$$

Remark. The first term $\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_{0\varepsilon}\boldsymbol{\Sigma}_0^{-1}$ is the usual asymptotic covariance matrix for the profile least squares estimator when data are exactly observed [6], i.e., $\phi_M(u) \equiv 1$, $\phi_A(u) \equiv 0$, $\psi_{A,r}(u) \equiv 1$ and $\psi_{A,r}(u) \equiv 0$, $r = 1, \ldots, p$. If the model error ϵ is further independent of $\boldsymbol{X}$, this term reduces to $E(\epsilon^2)\boldsymbol{\Sigma}_0^{-1}$. The second term $\boldsymbol{\Sigma}_{\phi_M, \boldsymbol{\psi}_M}$ is caused by the multiplicative and additive distortion measurement errors involved in the response variable and covariates. It is interesting to see that the additive distortions $\phi_A(u)$ and $\phi_{A,r}(u)$'s have no effect on the estimation of $\boldsymbol{\beta}_0$. If we further assumed that $\phi_M(u) = \psi_{M,r}(u) \equiv 1$, $r = 1, \ldots, p$, then the term $\boldsymbol{\Sigma}_{\phi,\boldsymbol{\psi}} = \mathbf{0}$. In this case, the estimator is efficient because the effect of additive distortions vanishes, which coincides with the asymptotic result of Theorem 1 in [39]. In other words, the profile least squares estimation procedure can automatically eliminate the effect induced by the additive distortions. And the profile least squares estimation procedure can also eliminate both the effect of multiplicative and additive distortions for estimating β_{0r} when $\beta_{0r} = 0$, i.e., $\mathrm{Avar}(\hat{\beta}_r) = e_r^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_{0\epsilon}\boldsymbol{\Sigma}_0^{-1}e_r$, where $\hat{\beta}_r$ is the r-th component of $\hat{\boldsymbol{\beta}}$, e_r is a p-dimensional vector with 1 in the r-th position and 0's elsewhere, $r = 1, \ldots, p$, and $\mathrm{Avar}(\hat{\beta}_r)$ stands for the asymptotic variance of $\hat{\beta}_r$ obtained in Theorem 2.1.

Theorem 2.2. *Suppose conditions (C1)-(C5) hold, as $h_2 \to 0$, $nh_2 \to \infty$,*

$$\sqrt{nh_2}\left(\hat{g}(z) - g(z) - \frac{\mu_2 h_2^2}{2}g''(z)\right) \xrightarrow{L} N\left(0, \frac{\mu_{K^2}\sigma^2(z)}{f_Z(z)}\right),$$

where $\sigma^2(z) = E\left\{\left[\widetilde{Y} - g(Z) - \widetilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{\beta}_0\right]^2 | Z = z\right\}$.

Remark. When the multiplicative distortions $\phi_M(u)$ and $\psi_{M,r}(u)$'s vanish ($\phi_M(u) = \psi_{M,r}(u) \equiv 1$), the asymptotic variance $\sigma^2(z) = E[\epsilon^2|Z = z] + \mathrm{Var}(\phi_A(U) - \boldsymbol{\psi}_A(U)^{\mathrm{T}}\boldsymbol{\beta}_0)$, which coincides with the asymptotic variance of Theorem 2 in [39]. Moreover, the estimator $\hat{g}(z)$ is asymptotically efficient when the additive distortion functions further satisfy $P(\phi_A(U) - \boldsymbol{\psi}_A(U)^{\mathrm{T}}\boldsymbol{\beta}_0 = 0) = 1$, i.e., the asymptotic bias and asymptotic variance of $\hat{g}(z)$ are the same as those obtained in [5] and [6].

3. Confidence intervals

3.1. Asymptotic normal approximation

According to Theorem 2.1, the $(1 - \alpha) \times 100\%$ $(0 < \alpha < 1)$ confidence interval for β_{0r} can be obtained by estimating the asymptotic covariance matrices. Let

$\hat{\epsilon}_i = \hat{Y}_i - \hat{\boldsymbol{\beta}}^{\mathrm{T}} \hat{\boldsymbol{X}}_i - \hat{g}(Z_i)$, $i = 1, \ldots, n$, we define

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^{n} \left[\hat{\boldsymbol{X}}_i - \hat{S}_{\boldsymbol{X}}(Z_i) \right]^{\otimes 2}, \quad \hat{\boldsymbol{\Sigma}}_{\epsilon} = \frac{1}{n} \sum_{i=1}^{n} \hat{\epsilon}_i^2 \left[\hat{\boldsymbol{X}}_i - \hat{S}_{\boldsymbol{X}}(Z_i) \right]^{\otimes 2},$$

$$\begin{aligned} \hat{\boldsymbol{G}}(\hat{\boldsymbol{X}}_i, \hat{\boldsymbol{\psi}}_M(U_i)) = \operatorname{diag} \Bigg(& \left[\frac{\left(\hat{X}_{1i} - \overline{\widetilde{X}_1}\right)^2}{2\left[\hat{E}\left(\sigma_{\widetilde{X}_1|U}(U)\right)\right]^2} + \frac{1}{2} \right] \left[\hat{\psi}_{M,1}(U_i) - 1 \right], \\ & \ldots, \left[\frac{\left(\hat{X}_{pi} - \overline{\widetilde{X}_p}\right)^2}{2\left[\hat{E}\left(\sigma_{\widetilde{X}_p|U}(U)\right)\right]^2} + \frac{1}{2} \right] \left[\hat{\psi}_{M,p}(U_i) - 1 \right] \Bigg), \end{aligned}$$

and

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}_{\phi_M, \boldsymbol{\psi}_M} = \frac{1}{n} \sum_{i=1}^{n} \Bigg\{ \Bigg[& \left[\hat{\phi}_M(U_i) - 1 \right] \left(\frac{\left(\hat{Y}_i - \overline{\widetilde{Y}}\right)^2}{2\left[\hat{E}\left(\sigma_{\widetilde{Y}|U}(U)\right)\right]^2} + \frac{1}{2} \right) \hat{\boldsymbol{\beta}} \\ & - \hat{\boldsymbol{G}}(\hat{\boldsymbol{X}}_i, \hat{\boldsymbol{\psi}}_M(U_i)) \hat{\boldsymbol{\beta}} \Bigg]^{\otimes 2} \Bigg\}. \end{aligned}$$

Moreover,

$$\hat{\sigma}_r^2 = e_r^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\Sigma}}_{\epsilon} \hat{\boldsymbol{\Sigma}}^{-1} e_r + e_r^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}_{\phi_M, \boldsymbol{\psi}_M} e_r. \tag{3.1}$$

Based on the estimator $\hat{\sigma}_r^2$, the $(1-\alpha) \times 100\%$ $(0 < \alpha < 1)$ confidence interval for β_{0r} is

$$\left(\hat{\beta}_r - \sqrt{\frac{\hat{\sigma}_r^2}{n}} z_{\alpha/2}, \quad \hat{\beta}_r + \sqrt{\frac{\hat{\sigma}_r^2}{n}} z_{\alpha/2} \right),$$

where $\hat{\beta}_r$ is the r-th component of $\hat{\boldsymbol{\beta}}$, $z_{\alpha/2}$ is the quantile satisfying $P(N(0,1) \geq z_{\alpha/2}) = \alpha/2$.

3.2. *Empirical likelihood method*

Empirical likelihood (EL) method proposed by [24] is another popular method to construct confidence intervals without estimating the asymptotic covariance matrix. The EL method is an appealing nonparametric approach for constructing confidence intervals (regions) for the parameter of interest. There is a large and growing literature extending empirical likelihood methods to many statistical problems. For example, [12, 10, 1, 17]. In the following, we construct confidence intervals of $\boldsymbol{\beta}_0$ based on the EL principle.

The EL method needs an auxiliary vector $\wp_{n,i}(\boldsymbol{\beta}') = (\wp_{n,i}^{[1]}(\boldsymbol{\beta}'), \ldots, \wp_{n,i}^{[p]}(\boldsymbol{\beta}'))^{\mathrm{T}}$ with the property of that $E\wp_{n,i}(\boldsymbol{\beta}') = 0$ when $\boldsymbol{\beta}' = \boldsymbol{\beta}_0$. Recalling that model

(2.16) is a linear regression model, the "ideal" auxiliary random vector can be constructed as

$$\wp_{n,i}(\boldsymbol{\beta}') = [\boldsymbol{X}_i - S_{\boldsymbol{X}}(Z_i)] \left(Y_i - S_Y(Z_i) - [\boldsymbol{X}_i - S_{\boldsymbol{X}}(Z_i)]^{\mathrm{T}} \boldsymbol{\beta}' \right).$$

Since $\{Y_i, \boldsymbol{X}_i\}_{i=1}^n$ are unavailable, we use the "calibrated" variables $\{\hat{Y}_i, \hat{\boldsymbol{X}}_i\}_{i=1}^n$. We now define the calibrated EL principle by plugging in $\{\hat{Y}_i - \hat{S}_Y(Z_i), \hat{\boldsymbol{X}}_i - \hat{S}_{\boldsymbol{X}}(Z_i)\}_{i=1}^n$ into $\wp_{n,i}(\boldsymbol{\beta}')$:

$$\hat{l}_n(\boldsymbol{\beta}') = -2 \max \Big\{ \sum_{i=1}^n \log(np_i) : p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{\wp}_{n,i}(\boldsymbol{\beta}') = 0 \Big\},$$

where

$$\hat{\wp}_{n,i}(\boldsymbol{\beta}') = \Big[\hat{\boldsymbol{X}}_i - \hat{S}_{\boldsymbol{X}}(Z_i)\Big] \left(\hat{Y}_i - \hat{S}_Y(Z_i) - \Big[\hat{\boldsymbol{X}}_i - \hat{S}_{\boldsymbol{X}}(Z_i)\Big]^{\mathrm{T}} \boldsymbol{\beta}' \right).$$

The Lagrange multiplier method entails $\hat{l}_n(\boldsymbol{\beta}') = 2\sum_{i=1}^n \log\{1 + \hat{\lambda}^{\mathrm{T}} \hat{\wp}_{n,i}(\boldsymbol{\beta}')\}$, where $\hat{\lambda}$ is determined by $\frac{1}{n}\sum_{i=1}^n \frac{\hat{\wp}_{n,i}(\boldsymbol{\beta}')}{1+\hat{\lambda}^{\mathrm{T}}\hat{\wp}_{n,i}(\boldsymbol{\beta}')} = 0$.

Theorem 3.1. *Suppose conditions in Theorem 2.1 hold, $\hat{l}_n(\boldsymbol{\beta}_0)$ asymptotically converges in distribution to χ^2_p, namely, a centered chi-squared distribution with p degrees of freedom.*

From Theorem 3.1, we can construct a confidence region of $\boldsymbol{\beta}_0$ by $I_\alpha = \{\boldsymbol{\beta}' : \hat{l}_n(\boldsymbol{\beta}') \leq c_\alpha\}$, where c_α denotes the α-quantile of the χ^2_p distribution.

4. Hypothesis testing

In the previous section, we consider the estimation and confidence intervals of $\boldsymbol{\beta}_0$. Another interesting topic is whether certain explanatory variables can significantly influence the response. In many important statistical applications, in addition to model information in model (1.1), let us give some prior information about $\boldsymbol{\beta}_0$ in the form of a set of linear restrictions as follows:

$$\mathcal{H}_0 : \boldsymbol{A}\boldsymbol{\beta}_0 = \boldsymbol{b}, \quad \mathcal{H}_1 : \boldsymbol{A}\boldsymbol{\beta}_0 \neq \boldsymbol{b}, \tag{4.1}$$

where $\boldsymbol{A}$ is a known $k \times p$ full-rank matrix, $\mathrm{rank}(\boldsymbol{A}) = k \leq p$ and $\boldsymbol{b}$ is a known k-vector constants. This hypothesis test is used to check the special structure of parameters $\boldsymbol{\beta}_0$ or the influence of the components of $\boldsymbol{X}$.

If the null hypothesis $\mathcal{H}_0$ is true, the condition $\boldsymbol{A}\boldsymbol{\beta}_0 = \boldsymbol{b}$ can be used to estimate $\boldsymbol{\beta}_0$. A restricted profile least squares estimation procedure by using Lagrange multiplier technique is proposed as:

$$\mathcal{W}_n(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \sum_{i=1}^n \left\{ \hat{Y}_i - \hat{S}_Y(Z_i) - \Big[\hat{\boldsymbol{X}}_i - \hat{S}_{\boldsymbol{X}}(Z_i)\Big]^{\mathrm{T}} \boldsymbol{\beta} \right\}^2 + 2\boldsymbol{\lambda}^{\mathrm{T}} (\boldsymbol{A}\boldsymbol{\beta} - \boldsymbol{b}),$$

where $\boldsymbol{\lambda}$ is a $k \times 1$ vector of the Lagrange multipliers. Differentiating $\mathcal{W}_n(\boldsymbol{\beta}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$,

$$\left\{ \begin{aligned} \frac{\partial \mathcal{W}_n(\boldsymbol{\beta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta}} &= 2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} - 2\sum_{i=1}^{n}\left[\hat{\boldsymbol{X}}_i - \hat{S}_{\boldsymbol{X}}(Z_i)\right] \\ &\qquad \times \left\{\hat{Y}_i - \hat{S}_Y(Z_i) - \left[\hat{\boldsymbol{X}}_i - \hat{S}_{\boldsymbol{X}}(Z_i)\right]^{\mathrm{T}}\boldsymbol{\beta}\right\} = \mathbf{0}, \\ \frac{\partial \mathcal{W}_n(\boldsymbol{\beta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} &= 2(\boldsymbol{A}\boldsymbol{\beta} - \boldsymbol{b}) = \mathbf{0}. \end{aligned} \right. \tag{4.2}$$

Using the estimator $\hat{\boldsymbol{\Sigma}}$ defined in subsection 3.1, the restricted estimator $\hat{\boldsymbol{\beta}}_{\mathrm{R}}$ of $\boldsymbol{\beta}_0$ derived from the first equation (4.2) satisfies

$$\sum_{i=1}^{n}\left[\hat{\boldsymbol{X}}_i - \hat{S}_{\boldsymbol{X}}(Z_i)\right]\left[\hat{Y}_i - \hat{S}_Y(Z_i)\right] = \hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\beta}}_{\mathrm{R}} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}. \tag{4.3}$$

Note that the profile least squares estimator $\hat{\boldsymbol{\beta}}$ in (2.19) satisfies

$$\sum_{i=1}^{n}\left[\hat{\boldsymbol{X}}_i - \hat{S}_{\boldsymbol{X}}(Z_i)\right]\left\{\hat{Y}_i - \hat{S}_Y(Z_i) - \left[\hat{\boldsymbol{X}}_i - \hat{S}_{\boldsymbol{X}}(Z_i)\right]^{\mathrm{T}}\hat{\boldsymbol{\beta}}\right\} = \mathbf{0}. \tag{4.4}$$

Together with (4.3)-(4.4), we have

$$\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{\beta}}_{\mathrm{R}} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}. \tag{4.5}$$

Then, equation (4.5) entails that

$$\hat{\boldsymbol{\beta}}_{\mathrm{R}} = \hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}. \tag{4.6}$$

Recalling that the estimator $\hat{\boldsymbol{\beta}}_{\mathrm{R}}$ in the second equation (4.2) satisfies $\boldsymbol{A}\hat{\boldsymbol{\beta}}_{\mathrm{R}} - \boldsymbol{b} = \mathbf{0}$, we multiply $\boldsymbol{A}$ on both sides in equation (4.6) and obtain

$$\boldsymbol{b} = \boldsymbol{A}\hat{\boldsymbol{\beta}}_{\mathrm{R}} = \boldsymbol{A}\hat{\boldsymbol{\beta}} + \boldsymbol{A}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}. \tag{4.7}$$

From (4.7), we obtain

$$\boldsymbol{\lambda} = -\left(\boldsymbol{A}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\left(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{b}\right). \tag{4.8}$$

We substitute expression $\boldsymbol{\lambda}$ in (4.8) to (4.6), and the restricted least squares estimator of $\boldsymbol{\beta}_0$ is obtained as

$$\hat{\boldsymbol{\beta}}_{\mathrm{R}} = \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}^{\mathrm{T}}\left(\boldsymbol{A}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\left(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{b}\right). \tag{4.9}$$

We now present the asymptotic normality of $\hat{\boldsymbol{\beta}}_{\mathrm{R}}$.

Theorem 4.1. *Let $\boldsymbol{\Omega}_{\boldsymbol{A}} = \boldsymbol{I}_p - \boldsymbol{\Sigma}_0^{-1}\boldsymbol{A}^{\mathrm{T}}\left(\boldsymbol{A}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\boldsymbol{A}$. Suppose conditions of Theorem 2.1 hold, under the null hypothesis $\mathcal{H}_0$, we have*

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{\mathrm{R}} - \boldsymbol{\beta}_0\right) \xrightarrow{L} N(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{A}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_{0\epsilon}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Omega}_{\boldsymbol{A}}^{\mathrm{T}} + \boldsymbol{\Omega}_{\boldsymbol{A}}\boldsymbol{\Sigma}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{\Omega}_{\boldsymbol{A}}^{\mathrm{T}}).$$

Remark. From the definition of $\boldsymbol{\Omega}_{\boldsymbol{A}}$, it is seen that $\boldsymbol{A}\boldsymbol{\Omega}_{\boldsymbol{A}} = \mathbf{0}$. Then, the asymptotic covariance matrix of $\boldsymbol{A}\hat{\boldsymbol{\beta}}_{\mathrm{R}} - \boldsymbol{A}\boldsymbol{\beta}_0$ under the null hypothesis $\mathcal{H}_0$ is a zero matrix, this is because the linear constrain $\boldsymbol{A}\hat{\boldsymbol{\beta}}_{\mathrm{R}} = \boldsymbol{b}$ holds true in (4.2) when we estimate $\boldsymbol{\beta}_0$.

To test hypothesis $\mathcal{H}_0$, we propose to use a weighted quadratic forms of $\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{b}$. Intuitively, if the null hypothesis $\mathcal{H}_0$ is false, i.e., $\boldsymbol{A}\boldsymbol{\beta}_0 \neq \boldsymbol{b}$, the value of $\|\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{b}\|$ should be significantly large. The test statistic for testing $\mathcal{H}_0$ is defined as

$$\mathcal{T}_n = n\left(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{b}\right)^{\mathrm{T}}\left(\boldsymbol{A}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Sigma}}_{\epsilon}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{A}\hat{\boldsymbol{\Sigma}}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\left(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{b}\right).$$

Theorem 4.2. *Suppose conditions in Theorem 2.1 hold, under the null hypothesis $\mathcal{H}_0$, we have*

$$\mathcal{T}_n \xrightarrow{L} \chi_k^2,$$

where χ_k^2 is a centered chi-squared distribution with degrees of freedom k.

Next, we consider the local alternative hypothesis

$$\mathcal{H}_{1n}: \boldsymbol{A}\boldsymbol{\beta}_0 = \boldsymbol{b} + n^{-1/2}\boldsymbol{c}, \quad \boldsymbol{c} \neq 0. \tag{4.10}$$

In the following, we present the asymptotic results of $\hat{\boldsymbol{\beta}}_{\mathrm{R}}$ and $\mathcal{T}_n$ under the local alternative hypothesis $\mathcal{H}_{1n}$.

Theorem 4.3. *Suppose conditions in Theorem 2.1 hold, under the local alternative hypothesis $\mathcal{H}_{1n}$, we have*

(a) *let $\eta_{\boldsymbol{c}} = -\boldsymbol{\Sigma}_0^{-1}\boldsymbol{A}^{\mathrm{T}}\left(\boldsymbol{A}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\boldsymbol{c}$,*

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{\mathrm{R}} - \boldsymbol{\beta}_0\right) \xrightarrow{L} N(\eta_{\boldsymbol{c}}, \boldsymbol{\Omega}_{\boldsymbol{A}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_{0\epsilon}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Omega}_{\boldsymbol{A}}^{\mathrm{T}} + \boldsymbol{\Omega}_{\boldsymbol{A}}\boldsymbol{\Sigma}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{\Omega}_{\boldsymbol{A}}^{\mathrm{T}}),$$

(b) *let $\pi_{\boldsymbol{c}} = \boldsymbol{c}^{\mathrm{T}}\left(\boldsymbol{A}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_{0\epsilon}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{A}\boldsymbol{\Sigma}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\boldsymbol{c}$,*

$$\mathcal{T}_n \xrightarrow{L} \chi_k^2(\pi_{\boldsymbol{c}}),$$

where $\chi_k^2(\pi_{\boldsymbol{c}})$ is the noncentral chi-squared distribution with degrees of freedom k, and $\pi_{\boldsymbol{c}}$ is the noncentrality parameter.

5. Variable selection

In the process of data analysis, the advent of modern technology allows many variables to be easily collected in scientific studies. Typically, many of them are included in the full model at the initial stage of modeling to reduce the model approximation error. It is of fundamental interest in statistical modeling to determine which variables should be selected and retained in the final statistical model. One popular variable selection method is the penalized least-squares method, which has been extensively studied over the past two decades. The least absolute shrinkage and selection operator [29, LASSO] and the smoothly clipped absolute deviation [4, SCAD] have been extensively discussed and are widely used.

Model (2.15) is a linear regression model with respect to $\boldsymbol{\beta}_0$. It is of interest to determine which covariates have nonzero effects on the response. There are a number of penalized variable selection methods for partial linear regression models (see, for example, [30, 18, 13]). In this section, we use the SCAD penalty function to select the nonzero component of $\boldsymbol{\beta}_0$. The SCAD penalty function $p_\zeta(\cdot)$ satisfies $p_\zeta(0) = 0$, $p'_\zeta(0+) > 0$, and its first order derivative is

$$p'_\zeta(\delta) = \zeta \left\{ I\{\delta \le \zeta\} + \frac{(a\zeta - \delta)}{(a-1)\zeta} I\{a\zeta > \delta\} I\{\delta > \zeta\} \right\},$$

where, a is some positive constant with $a > 2$ and ζ is a tuning parameter. From the perspective of Bayesian statistics, [4] suggests using $a = 3.7$, and so this value will be used throughout the remainder of this paper. For variable selection in multiplicative distortion measurement error models, [9] considered the use of Lasso-type penalty functions for simultaneous variable selection and parameter estimation in a linear regression model. There has been no discussion in the literature of the variable selection problem when both multiplicative and additive distortion exist in the partial linear model considered in this paper. To solve this problem, we propose the following SCAD penalized estimator:

$$\hat{\boldsymbol{\beta}}_P = \arg\min_{\boldsymbol{\beta}} \left\{ \frac{1}{2} \sum_{i=1}^{n} \left\{ \hat{Y}_i - \hat{S}_Y(Z_i) - \left[\hat{\boldsymbol{X}}_i - \hat{S}_{\boldsymbol{X}}(Z_i) \right]^{\mathrm{T}} \boldsymbol{\beta} \right\}^2 \right. \tag{5.1}$$
$$\left. + n \sum_{s=1}^{p} p_{\zeta_s}(|\beta_s|) \right\},$$

where $p_\zeta(\cdot)$ is the SCAD penalty function with the tuning parameter ζ.

We now study the sampling property of the resulting penalized least squares estimators. Without loss of generality, assume that $\boldsymbol{\beta}_0 = \left(\boldsymbol{\beta}_{0,\mathbf{1}}^{\mathrm{T}}, \boldsymbol{\beta}_{0,\mathbf{2}}^{\mathrm{T}}\right)^{\mathrm{T}}$, where $\boldsymbol{\beta}_{0,\mathbf{1}}$ denotes the $p_0 \times 1$ nonzero components of $\boldsymbol{\beta}_0$, and $\boldsymbol{\beta}_{0,\mathbf{2}}$ is a $(p - p_0) \times 1$ vector containing zeros. In addition, $\boldsymbol{X}_{\mathbf{1}}$ consists of the first p_0 components of $\boldsymbol{X}$ and $\boldsymbol{\psi}_{M,\mathbf{1}}(U)$ consists of the first p_0 components of $\boldsymbol{\psi}_M(U)$. Moreover, we define the following notation:

$$\boldsymbol{\Sigma}_{0,\mathbf{1}} = E\left\{ [\boldsymbol{X}_{\mathbf{1}} - E(\boldsymbol{X}_{\mathbf{1}}|Z)]^{\otimes 2} \right\}, \quad \boldsymbol{\Sigma}_{0\epsilon,\mathbf{1}} = E\left\{ \epsilon^2 [\boldsymbol{X}_{\mathbf{1}} - E(\boldsymbol{X}_{\mathbf{1}}|Z)]^{\otimes 2} \right\},$$

$$\boldsymbol{G_1}(\boldsymbol{X_1}, \boldsymbol{\psi}_{M,\mathbf{1}}(U)) = \text{diag}\Bigg(\left[\frac{(X_1 - E(X_1))^2}{2\sigma^2_{X_1}} + \frac{1}{2}\right][\psi_{M,1}(U) - 1], \\ \dots, \left[\frac{(X_{p_0} - E(X_{p_0}))^2}{2\sigma^2_{X_{p_0}}} + \frac{1}{2}\right][\psi_{M,p_0}(U) - 1] \Bigg),$$

$$\boldsymbol{\Sigma}_{\phi_M, \boldsymbol{\psi}_{M,\mathbf{1}}} = E\Bigg\{ \Bigg[[\phi_M(U) - 1]\left[\frac{(Y - E(Y))^2}{2\sigma^2_Y} + \frac{1}{2}\right]\boldsymbol{\beta}_{0,\mathbf{1}} \\ - \boldsymbol{G_1}(\boldsymbol{X_1}, \boldsymbol{\psi}_{M,\mathbf{1}}(U))\boldsymbol{\beta}_{0,\mathbf{1}} \Bigg]^{\otimes 2} \Bigg\},$$

$$\mathcal{R}_{\boldsymbol{\zeta_1}} = \Big\{ p'_{\zeta_1}(|\beta_{01}|)\text{sign}(\beta_{01}), \dots, p'_{\zeta_{p_0}}(|\beta_{0p_0}|)\text{sign}(\beta_{0p_0}) \Big\},$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\zeta_1}} = \text{diag}\Big\{ p''_{\zeta_1}(|\beta_{01}|), \dots, p''_{\zeta_{p_0}}(|\beta_{0p_0}|) \Big\}.$$

Theorem 5.1. *Under the conditions (C1)-(C6), the penalized estimator* $\hat{\boldsymbol{\beta}}_P = \left(\hat{\boldsymbol{\beta}}^{\mathrm{T}}_{P,\mathbf{1}}, \hat{\boldsymbol{\beta}}^{\mathrm{T}}_{P,\mathbf{2}}\right)^{\mathrm{T}}$ *satisfies:*

(a) (consistency) with probability tending to one, $\hat{\boldsymbol{\beta}}_{P,\mathbf{2}} = \mathbf{0}$*;*
(b) (asymptotic normality)

$$\sqrt{n}\left(\boldsymbol{\Sigma}_{0,\mathbf{1}} + \boldsymbol{\Sigma}_{\boldsymbol{\zeta_1}}\right)\left\{ \left(\hat{\boldsymbol{\beta}}_{P,\mathbf{1}} - \boldsymbol{\beta}_{0,\mathbf{1}}\right) - \left(\boldsymbol{\Sigma}_{0,\mathbf{1}} + \boldsymbol{\Sigma}_{\boldsymbol{\zeta_1}}\right)^{-1}\mathcal{R}_{\boldsymbol{\zeta_1}} \right\} \\ \xrightarrow{\mathcal{L}} N\left(\mathbf{0}_{p_0}, \boldsymbol{\Sigma}_{0\epsilon,\mathbf{1}} + \boldsymbol{\Sigma}_{0,\mathbf{1}}\boldsymbol{\Sigma}_{\phi_M,\boldsymbol{\psi}_{M,\mathbf{1}}}\boldsymbol{\Sigma}_{0,\mathbf{1}}\right).$$

Remark. The extra-bias $\sqrt{n}\mathcal{R}_{\boldsymbol{\zeta_1}}$ is induced by the SCAD penalty function. If we impose conditions $\sqrt{n}\mathcal{R}_{\boldsymbol{\zeta_1}} \to 0$ and $\boldsymbol{\Sigma}_{\boldsymbol{\zeta_1}} \to 0$, the asymptotic result of Theorem 5.1(b) is the same as Theorem 2.1 if the non-zero components of $\boldsymbol{\beta}_0$ were known beforehand. Moreover, the SCAD penalty also automatically shrinks the zero components of $\boldsymbol{\beta}_0$ to zeros. With an appropriate choice of the tuning parameter $\boldsymbol{\zeta}$, Theorem 5.1 indicates that the proposed variable selection procedure possesses the oracle property. We now discuss the choice of the tuning parameter.

We adopt the BIC selector to choose the regularization parameters ζ_j's [16] by reducing the p-dimensional regularization parameters $(\zeta_1, \dots, \zeta_p)$ to a single dimension. Let $\zeta_r = \zeta_0\hat{\sigma}_r$, $r = 1, \dots, p$, where $\hat{\sigma}_r$ is defined in (3.1). The BIC score for ζ_0 can be defined as

$$\text{BIC}(\zeta_0) = \log\{\text{MSE}(\zeta_0)\} + \frac{\log n}{n}N_{\zeta_0},$$

where $\text{MSE}(\zeta_0) = \frac{1}{n}\sum_{i=1}^n \left\{ \hat{Y}_i - \hat{S}_Y(Z_i) - \left[\hat{\boldsymbol{X}}_{i,\boldsymbol{\zeta}} - \hat{S}_{\boldsymbol{X},\boldsymbol{\zeta}}(Z_i)\right]^{\mathrm{T}}\hat{\boldsymbol{\beta}}_{P,\boldsymbol{\zeta}} \right\}^2$, $\hat{\boldsymbol{X}}_{i,\boldsymbol{\zeta}}$ and $\hat{S}_{\boldsymbol{X},\boldsymbol{\zeta}}(Z_i)$ consist of the components of $\hat{\boldsymbol{X}}_i$ and $\hat{S}_{\boldsymbol{X}}(Z_i)$ according to $\hat{\boldsymbol{\beta}}_{P,\boldsymbol{\zeta}}$, respectively. N_{ζ_0} is the number of nonzero coefficients of $\hat{\boldsymbol{\beta}}_{P,\boldsymbol{\zeta}}$, where $\hat{\boldsymbol{\beta}}_{P,\boldsymbol{\zeta}}$ is the resulting penalized estimator of $\boldsymbol{\beta}_0$ with tuning parameter $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_p)^{\mathrm{T}}$, where $\zeta_r = \zeta_0 \hat{\sigma}_r$. Thus, the minimization problem over ζ_j reduces to a one-dimensional minimization problem through ζ_0. The minimizer of the tuning parameter ζ_0 can be obtained by a grid search. Based on our experience in simulations, 30 grid points, evenly distributed over the range of ζ_0, are sufficient.

6. Comparison for no additive distortion scenario

6.1. Estimation

In the previous sections, we consider the coexistence of multiplicative and additive distortion measurement errors. In this section, we consider a special case of $\phi_A(U) = 0$, $\boldsymbol{\psi}_A(U) \equiv \mathbf{0}$, i.e., there is no additive distortions:

$$\widetilde{Y} = \phi_M(U)Y, \quad \widetilde{\boldsymbol{X}} = \boldsymbol{\psi}_M(U)\boldsymbol{X}. \tag{6.1}$$

We propose to use the recently studied conditional absolute mean calibration method [2, 41, CAMC] to estimate $\boldsymbol{\beta}_0$, and we discuss the asymptotic efficiency of estimator $\hat{\boldsymbol{\beta}}$ and the CAMC estimator.

If a random variable S satisfies $E(S) = 0$, it is easily seen that $\mathrm{Var}(S) = E[S - E(S)]^2 = ES^2$. Thus, under the multiplicative distortion setting (6.1), one can also use (2.8)-(2.10) and (2.14) to calibrate the unobserved response variable and covariates by the estimated $\hat{\phi}_A(u)$ and $\hat{\psi}_{A,r}(u)$'s. Another estimation procedure is to directly use the model assumption $\phi_A(U) = 0$ and $\boldsymbol{\psi}_A(U) = \mathbf{0}$ without estimating them. We now consider the later estimation procedure. According to (2.5)-(2.6) and (6.1), under the condition $\sigma_Y \prod_{r=1}^{p} \sigma_{X_r} > 0$, we have

$$Y = \frac{\widetilde{Y}}{\phi_M(U)}, \quad X_r = \frac{\widetilde{X}_r}{\psi_{M,r}(U)}, \quad r = 1, \ldots, p.$$

Use directly (2.11), the calibrated variables are defined as

$$\hat{Y}_{V,i} = \frac{\widetilde{Y}_i}{\hat{\phi}_M(U_i)}, \quad \hat{X}_{V,ri} = \frac{\widetilde{X}_{ri}}{\hat{\psi}_{M,r}(U_i)}, \quad r = 1, \ldots, p, \quad i = 1, \ldots, n.$$

Let $\hat{\boldsymbol{X}}_{V,i} = (\hat{X}_{V,1i}, \ldots, \hat{X}_{V,pi})^{\mathrm{T}}$, using (2.17) and (2.18), the parameter $\boldsymbol{\beta}_0$ is estimated as

$$\begin{aligned} \hat{\boldsymbol{\beta}}_V \quad = \quad & \left\{ \frac{1}{n} \sum_{i=1}^{n} \left[\hat{\boldsymbol{X}}_{V,i} - \hat{S}_{\boldsymbol{X}}(Z_i) \right]^{\otimes 2} \right\}^{-1} \\ & \times \frac{1}{n} \sum_{i=1}^{n} \left\{ \hat{\boldsymbol{X}}_{V,i} - \hat{S}_{\boldsymbol{X}}(Z_i) \right\} \left\{ \hat{Y}_{V,i} - \hat{S}_Y(Z_i) \right\}. \end{aligned} \tag{6.2}$$

The estimators (2.19) and (6.2) require the condition $\sigma_Y \prod_{r=1}^{p} \sigma_{X_r} > 0$, which is equivalent to $P(Y = E(Y)) + \sum_{r=1}^{p} P(X_r = E(X_r)) < 1$. In other words, none of the variables is a constant variable. The CAMC method proposed in [2] and [41] is by using

$$\phi_M(u) = \frac{E(|\widetilde{Y}|\ |U=u)}{E(|Y|)} = \frac{E(|\widetilde{Y}|\ |U=u)}{E(|\widetilde{Y}|)}, \tag{6.3}$$

$$\psi_{M,r}(u) = \frac{E(|\widetilde{X}_r|\ |U=u)}{E(|X_r|)} = \frac{E(|\widetilde{X}_r|\ |U=u)}{E(|\widetilde{X}_r|)}, \quad r=1,\ldots,p. \tag{6.4}$$

The equations (6.3) and (6.4) require the condition $E(|Y|)\prod_{r=1}^p E(|X_r|) > 0$, which is equivalent to $P(Y=0)+\sum_{r=1}^p P(X_r=0) < 1$. In other words, none of the variables $\{Y, X_r, r=1,\ldots,p\}$ is a zero variable under the condition $E(|Y|)\prod_{r=1}^p E(|X_r|) > 0$. Although $\{Y, X_r, r=1,\ldots,p\}$ is unobserved, $\widetilde{Y}\equiv 0$ if and only if $Y\equiv 0$ under the model assumption (6.1). Thus, the condition $E(|Y|)\prod_{r=1}^p E(|X_r|) > 0$ is much weaker than the condition $\sigma_Y\prod_{r=1}^p \sigma_{X_r} > 0$.

It is remarkable that the CAMC method is applicable for model (6.1) but not for model (1.1). When $\phi_A(u)\not\equiv 0$ and $\psi_{A,r}(u)\not\equiv 0$, $r=1,\ldots,p$, the CAMC method (6.3) is infeasible. It is seen that $E(|\widetilde{Y}|\ |U=u) = E(|\phi_M(u)Y+\phi_A(u)|)$, because $\phi_A(u)$ and $\phi_M(u)$ are two unknown functions, and only one equation (6.3) is not workable anymore. [2] proposed the CAMC method to estimate the conditional mean function $E(Y|X=x)$, and [41] used CAMC for model checking problem. We now propose another estimator of $\boldsymbol{\beta}_0$ based on the CAMC method. The Nadaraya-Watson estimators of $\phi_M(u)$ and $\psi_{M,r}(u)$ are defined as

$$\hat{\phi}_{M_{|\cdot|}}(u) = \frac{1}{n\hat{f}_U(u)\overline{|\widetilde{Y}|}}\sum_{i=1}^n K_h(U_i-u)|\widetilde{Y}_i|, \quad \overline{|\widetilde{Y}|} = \frac{1}{n}\sum_{i=1}^n |\widetilde{Y}_i| \tag{6.5}$$

$$\hat{\psi}_{M_{|\cdot|},r}(u) = \frac{1}{n\hat{f}_U(u)\overline{|\widetilde{X}_r|}}\sum_{i=1}^n K_h(U_i-u)|\widetilde{X}_{ri}|, \quad \overline{|\widetilde{X}_r|} = \frac{1}{n}\sum_{i=1}^n |\widetilde{X}_{ri}|. \tag{6.6}$$

Using (6.5)-(6.6), we obtain the CAMC calibrated variables $\{\hat{Y}_{C,i}, \hat{X}_{C,ri}, r = 1,\ldots,p\}_{i=1}^n$ as

$$\hat{Y}_{C,i} = \frac{\widetilde{Y}_i}{\hat{\phi}_{M_{|\cdot|}}(U_i)}, \quad \hat{X}_{C,ri} = \frac{\widetilde{X}_{ri}}{\hat{\psi}_{M_{|\cdot|},r}(U_i)}, \quad r=1,\ldots,p. \tag{6.7}$$

Let $\hat{\boldsymbol{X}}_{C,i} = (\hat{X}_{C,1i},\ldots,\hat{X}_{C,pi})^{\mathrm{T}}$, using (2.17) and (2.18), the parameter $\boldsymbol{\beta}_0$ is estimated as

$$\begin{aligned}\hat{\boldsymbol{\beta}}_C &= \left\{\frac{1}{n}\sum_{i=1}^n \left[\hat{\boldsymbol{X}}_{C,i} - \hat{S}_{\boldsymbol{X}}(Z_i)\right]^{\otimes 2}\right\}^{-1} \\ &\quad \times \frac{1}{n}\sum_{i=1}^n \left\{\hat{\boldsymbol{X}}_{C,i} - \hat{S}_{\boldsymbol{X}}(Z_i)\right\}\left\{\hat{Y}_{C,i} - \hat{S}_Y(Z_i)\right\}.\end{aligned}$$

We now present the asymptotic results for the estimators $\hat{\boldsymbol{\beta}}_V$ and $\hat{\boldsymbol{\beta}}_C$. Define the following notations:

$$\boldsymbol{F}(\boldsymbol{X}, \boldsymbol{\psi}_M(U)) = \operatorname{diag}\left(\frac{(\psi_{M,1}(U)-1)|X_1|}{E(|X_1|)}, \ldots, \frac{(\psi_{M,p}(U)-1)|X_p|}{E(|X_p|)} \right),$$

$$\boldsymbol{\Omega}_{\phi_M, \boldsymbol{\psi}_M} = E\left\{ \left[\frac{(\phi_M(U)-1)|Y|}{E(|Y|)} \boldsymbol{\beta}_0 - \boldsymbol{F}(\boldsymbol{X}, \boldsymbol{\psi}_M(U)) \boldsymbol{\beta}_0 \right]^{\otimes 2} \right\}.$$

Theorem 6.1. *Suppose conditions (C1)-(C5) hold, we have*

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_V - \boldsymbol{\beta}_0\right) \xrightarrow{L} N\left(\mathbf{0}_p, \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\Sigma}_{0\varepsilon} \boldsymbol{\Sigma}_0^{-1} + \boldsymbol{\Sigma}_{\phi_M, \boldsymbol{\psi}_M}\right),$$

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_C - \boldsymbol{\beta}_0\right) \xrightarrow{L} N\left(\mathbf{0}_p, \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\Sigma}_{0\varepsilon} \boldsymbol{\Sigma}_0^{-1} + \boldsymbol{\Omega}_{\phi_M, \boldsymbol{\psi}_M}\right).$$

Compared with Theorem 2.1, it is seen that estimators $\hat{\boldsymbol{\beta}}_V$ and $\hat{\boldsymbol{\beta}}$ have the same asymptotic mean and asymptotic covariance matrix. This is not surprising because the additive distortions $\phi_A(U)$ and $\boldsymbol{\psi}_A(U)$ has no effect on the profile least squares estimator $\hat{\boldsymbol{\beta}}$, and the distortion model (6.1) assumed that $\phi_A(U) = 0$ and $\boldsymbol{\psi}_A(U) \equiv \mathbf{0}$. As a result, it is natural that $\hat{\boldsymbol{\beta}}_V$ and $\hat{\boldsymbol{\beta}}$ have common asymptotic mean and asymptotic covariance matrix.

For the CAMC estimator $\hat{\boldsymbol{\beta}}_C$, it is seen that $\hat{\boldsymbol{\beta}}_C$ is more asymptotically efficient than $\hat{\boldsymbol{\beta}}_V$ when $\boldsymbol{\Sigma}_{\phi_M, \boldsymbol{\psi}_M} - \boldsymbol{\Omega}_{\phi_M, \boldsymbol{\psi}_M}$ is a positive definite matrix, and vice versa. In details, we denote the asymptotic variance of $\hat{\beta}_{V,r}$ (the r-th component of $\hat{\boldsymbol{\beta}}_V$) as $\operatorname{Avar}(\hat{\beta}_{V,r})$ and the asymptotic variance of $\hat{\beta}_{C,r}$ (the r-th component of $\hat{\boldsymbol{\beta}}_C$) as $\operatorname{Avar}(\hat{\beta}_{C,r})$. We have

$$\begin{aligned}
&\operatorname{Avar}(\hat{\beta}_{C,r}) - \operatorname{Avar}(\hat{\beta}_{V,r}) \\
&= \beta_{0r}^2 \operatorname{Var}(\phi_M(U)) \left\{ \frac{E(Y^2)}{[E(|Y|)]^2} - \frac{E[(Y - E(Y))^4]}{4\sigma_Y^4} - \frac{3}{4} \right\} \\
&\quad + \beta_{0r}^2 \operatorname{Var}(\psi_{M,r}(U)) \left\{ \frac{E(X_r^2)}{[E(|X_r|)]^2} - \frac{E[(X_r - E(X_r))^4]}{4\sigma_{X_r}^4} - \frac{3}{4} \right\} \\
&\quad - 2\beta_{0r}^2 \left\{ \frac{E(|Y X_r|)}{E(|Y|)E(|X_r|)} - \frac{E[(Y - E(Y))^2 (X_r - E(X_r))^2]}{4\sigma_Y^2 \sigma_{X_r}^2} - \frac{3}{4} \right\} \\
&\qquad \times \operatorname{Cov}(\phi_M(U), \psi_{M,r}(U)).
\end{aligned}$$

It is seen that if the response variable Y is exactly observed, i.e., $\phi_M(u) \equiv 1$, then $\operatorname{Var}(\phi_M(U)) = 0$ and $\operatorname{Cov}(\phi_M(U), \psi_{M,r}(U)) = 0$, the difference between the asymptotic variances $\operatorname{Avar}(\hat{\beta}_{V,r})$ and $\operatorname{Avar}(\hat{\beta}_{C,r})$ reduces to

$$\beta_{0r}^2 \operatorname{Var}(\psi_{M,r}(U)) \left\{ \frac{E(X_r^2)}{[E(|X_r|)]^2} - \frac{E[(X_r - E(X_r))^4]}{4\sigma_{X_r}^4} - \frac{3}{4} \right\}.$$

It is also seen that if the true parameter $\beta_{0r} = 0$, we have

$$\operatorname{Avar}(\hat{\beta}_{V,r}) = \operatorname{Avar}(\hat{\beta}_{C,r}) = e_r^{\mathrm{T}} \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\Sigma}_{0\epsilon} \boldsymbol{\Sigma}_0^{-1} e_r.$$

When $\beta_{0r}=0$, both $\hat{\boldsymbol{\beta}}_C$ and $\hat{\boldsymbol{\beta}}_V$ result in asymptotic efficient estimators, i.e., the profile least squares estimation with different calibration procedures eliminate the effect caused by the multiplicative distorting functions $\phi_M(u)$ and $\psi_{M,r}(u)$'s.

We directly substitute estimators $\hat{\boldsymbol{\beta}}$ with $\hat{\boldsymbol{\beta}}_V$ or $\hat{\boldsymbol{\beta}}_C$ in the (2.20) and (2.21), and obtain the estimators $\hat{g}_V(z)$ and $\hat{g}_C(z)$, respectively. In Theorem 6.1, the estimators $\hat{\boldsymbol{\beta}}_V$ and $\hat{\boldsymbol{\beta}}_C$ have root-n convergence rate, and local linear kernel smoothing estimators $\hat{g}_V(z)$ and $\hat{g}_C(z)$ have root-(nh_2) convergence rate, which is slow than the former one. So the asymptotic result of $\hat{g}_V(z)$ and $\hat{g}_C(z)$ are the same as those in Theorem 2.2, but the asymptotic variance $\sigma^2(z)$ is calculated under the model (6.1).

6.2. Confidence intervals

(1) Confidence intervals based on $\hat{\boldsymbol{\beta}}_V$.

According to Theorem 6.1, the $(1-\alpha)\times 100\%$ $(0<\alpha<1)$ confidence interval for β_{0r} can be obtained by estimating the asymptotic covariance matrices. Let $\hat{\epsilon}_{V,i}=\hat{Y}_{V,i}-\hat{\boldsymbol{\beta}}_V^{\mathrm{T}}\hat{\boldsymbol{X}}_{V,i}-\hat{g}_V(Z_i)$, $i=1,\dots,n$, here $\hat{g}_V(z)$ is obtained from (2.21) by substitute $\hat{\boldsymbol{\beta}}$ with $\hat{\boldsymbol{\beta}}_V$. We define

$$\hat{\boldsymbol{\Sigma}}_V=\frac{1}{n}\sum_{i=1}^{n}\left[\hat{\boldsymbol{X}}_{V,i}-\hat{S}_{\boldsymbol{X}}(Z_i)\right]^{\otimes 2},\quad \hat{\boldsymbol{\Sigma}}_{V,\epsilon}=\frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_{V,i}^2\left[\hat{\boldsymbol{X}}_{V,i}-\hat{S}_{\boldsymbol{X}}(Z_i)\right]^{\otimes 2},$$

and

$$\begin{aligned}\hat{\boldsymbol{G}}_V(\hat{\boldsymbol{X}}_{V,i},\hat{\boldsymbol{\psi}}_M(U_i))=\mathrm{diag}\Bigg(&\left[\frac{\big(\hat{X}_{V,1i}-\overline{\widetilde{X}_1}\big)^2}{2\big[\hat{E}\big(\sigma_{\widetilde{X}_1|U}(U)\big)\big]^2}+\frac{1}{2}\right]\left[\hat{\psi}_{M,1}(U_i)-1\right],\\ &\dots,\left[\frac{\big(\hat{X}_{V,pi}-\overline{\widetilde{X}_p}\big)^2}{2\big[\hat{E}\big(\sigma_{\widetilde{X}_p|U}(U)\big)\big]^2}+\frac{1}{2}\right]\left[\hat{\psi}_{M,p}(U_i)-1\right]\Bigg),\end{aligned}$$

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_{V,\phi_M,\boldsymbol{\psi}_M}=\frac{1}{n}\sum_{i=1}^{n}\Bigg\{\Bigg[&\left[\hat{\phi}_M(U_i)-1\right]\left(\frac{\big(\hat{Y}_{V,i}-\overline{\widetilde{Y}}\big)^2}{2\big[\hat{E}\big(\sigma_{\widetilde{Y}|U}(U)\big)\big]^2}+\frac{1}{2}\right)\hat{\boldsymbol{\beta}}_V\\ &-\hat{\boldsymbol{G}}_V(\hat{\boldsymbol{X}}_{V,i},\hat{\boldsymbol{\psi}}_M(U_i))\hat{\boldsymbol{\beta}}_V\Bigg]^{\otimes 2}\Bigg\}.\end{aligned}$$

Moreover,

$$\hat{\sigma}_{V,r}^2=e_r^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_V^{-1}\hat{\boldsymbol{\Sigma}}_{V,\epsilon}\hat{\boldsymbol{\Sigma}}_V^{-1}e_r+e_r^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{V,\phi_M,\boldsymbol{\psi}_M}e_r.$$

Based on the estimator $\hat{\sigma}_{V,r}^2$, the $(1-\alpha)\times 100\%$ $(0<\alpha<1)$ confidence interval for β_{0r} is

$$\left(\hat{\beta}_{V,r}-\sqrt{\frac{\hat{\sigma}_{V,r}^2}{n}}z_{\alpha/2},\quad \hat{\beta}_{V,r}+\sqrt{\frac{\hat{\sigma}_{V,r}^2}{n}}z_{\alpha/2}\right).$$

(2) Confidence intervals based on $\hat{\boldsymbol{\beta}}_C$.

Let $\hat{\epsilon}_{C,i} = \hat{Y}_{C,i} - \hat{\boldsymbol{\beta}}_C^{\mathrm{T}} \hat{\boldsymbol{X}}_{C,i} - \hat{g}_C(Z_i)$, $i = 1, \ldots, n$, here $\hat{g}_C(z)$ is obtained from (2.21) by substitute $\hat{\boldsymbol{\beta}}$ with $\hat{\boldsymbol{\beta}}_C$. We define

$$\hat{\boldsymbol{\Sigma}}_C = \frac{1}{n}\sum_{i=1}^{n}\Big[\hat{\boldsymbol{X}}_{C,i} - \hat{S}_{\boldsymbol{X}}(Z_i)\Big]^{\otimes 2}, \quad \hat{\boldsymbol{\Sigma}}_{C,\epsilon} = \frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_{C,i}^2\Big[\hat{\boldsymbol{X}}_{C,i} - \hat{S}_{\boldsymbol{X}}(Z_i)\Big]^{\otimes 2},$$

and

$$\hat{\boldsymbol{F}}(\hat{\boldsymbol{X}}_{C,i}, \hat{\boldsymbol{\psi}}_{M_{|\cdot|}}(U_i)) = \mathrm{diag}\left(\frac{|\widetilde{X}_{1i}| - |\hat{X}_{C,1i}|}{\overline{|\widetilde{X}_1|}}, \ldots, \frac{|\widetilde{X}_{pi}| - |\hat{X}_{C,pi}|}{\overline{|\widetilde{X}_p|}}\right),$$

$$\hat{\boldsymbol{\Omega}}_{\phi_M, \boldsymbol{\psi}_M} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{|\widetilde{Y}_i| - |\hat{Y}_{C,i}|}{\overline{|\widetilde{Y}|}}\hat{\boldsymbol{\beta}}_C - \hat{\boldsymbol{F}}(\hat{\boldsymbol{X}}_{C,i}, \hat{\boldsymbol{\psi}}_{M_{|\cdot|}}(U_i))\hat{\boldsymbol{\beta}}_C\right]^{\otimes 2}.$$

Moreover, we define

$$\hat{\sigma}_{C,r}^2 = e_r^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_C^{-1}\hat{\boldsymbol{\Sigma}}_{C,\epsilon}\hat{\boldsymbol{\Sigma}}_C^{-1}e_r + e_r^{\mathrm{T}}\hat{\boldsymbol{\Omega}}_{\phi_M, \boldsymbol{\psi}_M}e_r. \tag{6.8}$$

Based on the estimator $\hat{\sigma}_{C,r}^2$, the $(1-\alpha)\times 100\%$ $(0 < \alpha < 1)$ confidence interval for β_{0r} is

$$\left(\hat{\beta}_{C,r} - \sqrt{\frac{\hat{\sigma}_{C,r}^2}{n}}z_{\alpha/2}, \quad \hat{\beta}_{C,r} + \sqrt{\frac{\hat{\sigma}_{C,r}^2}{n}}z_{\alpha/2}\right).$$

(3) Empirical likelihood method

Analogous to Section 3.2, we now define two calibrated EL principles by plugging in $\{\hat{Y}_{V,i} - \hat{S}_Y(Z_i), \hat{\boldsymbol{X}}_{V,i} - \hat{S}_{\boldsymbol{X}}(Z_i)\}_{i=1}^n$ and $\{\hat{Y}_{C,i} - \hat{S}_Y(Z_i), \hat{\boldsymbol{X}}_{C,i} - \hat{S}_{\boldsymbol{X}}(Z_i)\}_{i=1}^n$ into $\wp_{n,i}(\boldsymbol{\beta}')$:

$$\begin{aligned}
\hat{l}_{V,n}(\boldsymbol{\beta}') &= -2\max\Big\{\sum_{i=1}^{n}\log(np_i) : p_i \geq 0, \sum_{i=1}^{n}p_i = 1, \sum_{i=1}^{n}p_i\hat{\wp}_{V,n,i}(\boldsymbol{\beta}') = 0\Big\},\\
\hat{l}_{C,n}(\boldsymbol{\beta}') &= -2\max\Big\{\sum_{i=1}^{n}\log(np_i) : p_i \geq 0, \sum_{i=1}^{n}p_i = 1, \sum_{i=1}^{n}p_i\hat{\wp}_{C,n,i}(\boldsymbol{\beta}') = 0\Big\},
\end{aligned}$$

where

$$\begin{aligned}
\hat{\wp}_{V,n,i}(\boldsymbol{\beta}') &= \Big[\hat{\boldsymbol{X}}_{V,i} - \hat{S}_{\boldsymbol{X}}(Z_i)\Big]\left(\hat{Y}_{V,i} - \hat{S}_Y(Z_i) - \Big[\hat{\boldsymbol{X}}_{V,i} - \hat{S}_{\boldsymbol{X}}(Z_i)\Big]^{\mathrm{T}}\boldsymbol{\beta}'\right),\\
\hat{\wp}_{C,n,i}(\boldsymbol{\beta}') &= \Big[\hat{\boldsymbol{X}}_{C,i} - \hat{S}_{\boldsymbol{X}}(Z_i)\Big]\left(\hat{Y}_{C,i} - \hat{S}_Y(Z_i) - \Big[\hat{\boldsymbol{X}}_{C,i} - \hat{S}_{\boldsymbol{X}}(Z_i)\Big]^{\mathrm{T}}\boldsymbol{\beta}'\right).
\end{aligned}$$

Similar to the proof of Theorem 3.1, both $\hat{l}_{V,n}(\boldsymbol{\beta}_0)$ and $\hat{l}_{C,n}(\boldsymbol{\beta}_0)$ asymptotically converge in distribution to χ_p^2. Thus, we can construct two empirical likelihood based confidence regions by $I_{V,\alpha} = \{\boldsymbol{\beta}' : \hat{l}_{V,n}(\boldsymbol{\beta}') \leq c_\alpha\}$ and $I_{C,\alpha} = \{\boldsymbol{\beta}' : \hat{l}_{C,n}(\boldsymbol{\beta}') \leq c_\alpha\}$.

6.3. A hypothesis testing

For the parameter hypothesis testing problem (4.1), the test statistics are defined as

$$\begin{aligned}\mathcal{T}_{V,n} &= n\left(\boldsymbol{A}\hat{\boldsymbol{\beta}}_V-\boldsymbol{b}\right)^{\mathrm{T}}\left(\boldsymbol{A}\hat{\boldsymbol{\Sigma}}_V^{-1}\hat{\boldsymbol{\Sigma}}_{V,\epsilon}\hat{\boldsymbol{\Sigma}}_V^{-1}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A}\hat{\boldsymbol{\Sigma}}_{V,\phi_M,\boldsymbol{\psi}_M}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\\ &\quad\times\left(\boldsymbol{A}\hat{\boldsymbol{\beta}}_V-\boldsymbol{b}\right),\\ \mathcal{T}_{C,n} &= n\left(\boldsymbol{A}\hat{\boldsymbol{\beta}}_C-\boldsymbol{b}\right)^{\mathrm{T}}\left(\boldsymbol{A}\hat{\boldsymbol{\Sigma}}_C^{-1}\hat{\boldsymbol{\Sigma}}_{C,\epsilon}\hat{\boldsymbol{\Sigma}}_C^{-1}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A}\hat{\boldsymbol{\Omega}}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\\ &\quad\times\left(\boldsymbol{A}\hat{\boldsymbol{\beta}}_C-\boldsymbol{b}\right).\end{aligned}$$

We have the following asymptotic results.

Theorem 6.2. *Suppose conditions in Theorem 1 hold, under the null hypothesis $\mathcal{H}_0$ of (4.1), we have*

$$\mathcal{T}_{V,n}\xrightarrow{L}\chi_k^2,\quad \mathcal{T}_{C,n}\xrightarrow{L}\chi_k^2.$$

Under the local null hypothesis $\mathcal{H}_{1n}$ of (4.10), we have

$$\mathcal{T}_{V,n}\xrightarrow{L}\chi_k^2(\pi_{\boldsymbol{c}}),\quad \mathcal{T}_{C,n}\xrightarrow{L}\chi_k^2(\pi_{C,\boldsymbol{c}}),$$

where $\chi_k^2(\pi_{C,\boldsymbol{c}})$ is the noncentral chi-squared distribution with degrees of freedom k, and $\pi_{C,\boldsymbol{c}}$ is the noncentrality parameter:

$$\pi_{C,\boldsymbol{c}}=\boldsymbol{c}^{\mathrm{T}}\left(\boldsymbol{A}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_{0\epsilon}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A}\boldsymbol{\Omega}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\boldsymbol{c}.$$

From Theorem 6.2, we can use two test statistics $\mathcal{T}_{V,n}$ and $\mathcal{T}_{C,n}$ to check the hypothesis $\mathcal{H}_0$ in (4.1). It is seen that if $\pi_{C,\boldsymbol{c}}>\pi_{\boldsymbol{c}}$, $\mathcal{T}_{C,n}$ performs asymptotically more powerful than $\mathcal{T}_{V,n}$ for detecting the local alternative hypothesis $\mathcal{H}_{1n}$; if $\pi_{C,\boldsymbol{c}}=\pi_{\boldsymbol{c}}$, both two statistics are asymptotically equivalent. If the local alternative hypothesis $\mathcal{H}_{1n}$ is given in advance, we can use the larger value of estimators $\hat{\pi}_{\boldsymbol{c}}$ and $\hat{\pi}_{C,\boldsymbol{c}}$ to decide which statistic is better.

6.4. Variable selection

Analogous to (5.1), the penalized estimators of $\boldsymbol{\beta}_0$ are defined as

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{P,V}=\arg\min_{\boldsymbol{\beta}}\Bigg\{&\frac{1}{2}\sum_{i=1}^n\left\{\hat{Y}_{V,i}-\hat{S}_Y(Z_i)-\left[\hat{\boldsymbol{X}}_{V,i}-\hat{S}_{\boldsymbol{X}}(Z_i)\right]^{\mathrm{T}}\boldsymbol{\beta}\right\}^2\\ &+n\sum_{s=1}^p p_{\zeta_s}(|\beta_s|)\Bigg\},\end{aligned}$$

$$\hat{\boldsymbol{\beta}}_{P,C} = \arg\min_{\boldsymbol{\beta}} \left\{ \frac{1}{2} \sum_{i=1}^{n} \left\{ \hat{Y}_{C,i} - \hat{S}_Y(Z_i) - \left[\hat{\boldsymbol{X}}_{C,i} - \hat{S}_{\boldsymbol{X}}(Z_i) \right]^{\mathrm{T}} \boldsymbol{\beta} \right\}^2 \right.$$
$$\left. + n \sum_{s=1}^{p} p_{\zeta_s}(|\beta_s|) \right\},$$

where $p_\zeta(\cdot)$ is the SCAD penalty function with tuning parameter ζ. Similar to Theorem 6.1, the asymptotic result of $\hat{\boldsymbol{\beta}}_{P,V}$ is the same as Theorem 5.1. Now we present the asymptotic results of $\hat{\boldsymbol{\beta}}_{P,C}$. In the following, we define

$$\begin{aligned} &\boldsymbol{F}_{\mathbf{1}}(\boldsymbol{X}_{\mathbf{1}}, \boldsymbol{\psi}_{M,\mathbf{1}}(U)) \\ &= \operatorname{diag}\left(\frac{(\psi_{M,1}(U)-1)|X_1|}{E(|X_1|)}, \dots, \frac{(\psi_{M,p_0}(U)-1)|X_{p_0}|}{E(|X_{p_0}|)} \right), \\ &\boldsymbol{\Omega}_{\phi_M, \boldsymbol{\psi}_{M,\mathbf{1}}} = E\left\{ \left[\frac{(\phi_M(U)-1)|Y|}{E(|Y|)} \boldsymbol{\beta}_{0,\mathbf{1}} - \boldsymbol{F}_{\mathbf{1}}(\boldsymbol{X}_{\mathbf{1}}, \boldsymbol{\psi}_{M,\mathbf{1}}(U)) \boldsymbol{\beta}_{0,\mathbf{1}} \right]^{\otimes 2} \right\}. \end{aligned}$$

Theorem 6.3. *Under the conditions (C1)-(C6), the penalized estimator* $\hat{\boldsymbol{\beta}}_{P,C} = \left(\hat{\boldsymbol{\beta}}_{P,C,\mathbf{1}}^{\mathrm{T}}, \hat{\boldsymbol{\beta}}_{P,C,\mathbf{2}}^{\mathrm{T}} \right)^{\mathrm{T}}$ *satisfies:*

(a) (consistency) with probability tending to one, $\hat{\boldsymbol{\beta}}_{P,C,\mathbf{2}} = \mathbf{0}$*;*
(b) (asymptotic normality)

$$\begin{aligned} &\sqrt{n} \left(\boldsymbol{\Sigma}_{0,\mathbf{1}} + \boldsymbol{\Sigma}_{\boldsymbol{\zeta}_{\mathbf{1}}} \right) \left\{ \left(\hat{\boldsymbol{\beta}}_{P,C,\mathbf{1}} - \boldsymbol{\beta}_{0,\mathbf{1}} \right) - \left(\boldsymbol{\Sigma}_{0,\mathbf{1}} + \boldsymbol{\Sigma}_{\boldsymbol{\zeta}_{\mathbf{1}}} \right)^{-1} \mathcal{R}_{\boldsymbol{\zeta}_{\mathbf{1}}} \right\} \\ &\xrightarrow{\mathcal{L}} N\left(\mathbf{0}_{p_0}, \boldsymbol{\Sigma}_{0\epsilon,\mathbf{1}} + \boldsymbol{\Sigma}_{0,\mathbf{1}} \boldsymbol{\Omega}_{\phi_M, \boldsymbol{\psi}_{M,\mathbf{1}}} \boldsymbol{\Sigma}_{0,\mathbf{1}} \right). \end{aligned}$$

To choose the regularization parameters ζ_j's for the penalized estimator $\hat{\boldsymbol{\beta}}_{P,C}$, we also adopt the BIC selector suggested by [16]. Let $\zeta_j = \zeta_0 \hat{\sigma}_{C,j}$, where $\hat{\sigma}_{C,j}$'s are defined in (6.8). The BIC score for ζ_0 can be defined as

$$\mathrm{BIC}(\zeta_0) = \log\{\mathrm{MSE}(\zeta_0)\} + \frac{\log n}{n} N_{\zeta_0},$$

where, $\mathrm{MSE}(\zeta_0) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \hat{Y}_{C,i} - \hat{S}_Y(Z_i) - \left[\hat{\boldsymbol{X}}_{C,i,\boldsymbol{\zeta}} - \hat{S}_{\boldsymbol{X},\boldsymbol{\zeta}}(Z_i) \right]^{\mathrm{T}} \hat{\boldsymbol{\beta}}_{P,C,\boldsymbol{\zeta}} \right\}^2$, $\hat{\boldsymbol{X}}_{C,i,\boldsymbol{\zeta}}$ and $\hat{S}_{\boldsymbol{X},\boldsymbol{\zeta}}(Z_i)$ consist of the components of $\hat{\boldsymbol{X}}_{C,i}$ and $\hat{S}_{\boldsymbol{X}}(Z_i)$ according to $\hat{\boldsymbol{\beta}}_{P,C,\boldsymbol{\zeta}}$, respectively. N_{ζ_0} is the number of nonzero coefficients of $\hat{\boldsymbol{\beta}}_{P,C,\boldsymbol{\zeta}}$, where $\hat{\boldsymbol{\beta}}_{P,C,\boldsymbol{\zeta}}$ is the resulting penalized estimator of $\boldsymbol{\beta}_0$ with tuning parameter $\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_p)^{\mathrm{T}}$, with $\zeta_j = \zeta_0 \hat{\sigma}_{C,j}$. Based on our experience in simulations, 30 grid points are set to be evenly distributed over the range of ζ_0.

7. Implementation

This section reports the results of simulation studies to demonstrate the performance of our proposed estimators. The Epanechnikov kernel $K(t) = 0.75(1 -$

$t^2)I\{|t|<1\}$ is used here. According to condition (C5), the bandwidth h_1 can be chosen as the optimal convergence rate, but the bandwidth h should be chosen for under-smoothing ($nh^4 \to 0$). The consequence of under-smoothing is that the biases of the non-parametric estimates are remain small and preclude the optimal bandwidth for h. The asymptotic variances of the proposed estimators for $\boldsymbol{\beta}_0$ depend on neither the bandwidths (h, h_1) nor the kernel function $K(t)$. Hence, we can use the rule of thumb: $h = \hat{\sigma}_U n^{-1/3}$, $h_1 = \hat{\sigma}_Z n^{-1/5}$, and with $\hat{\sigma}_U$ being the sample deviation of U and $\hat{\sigma}_Z$ being the sample deviation of Z. This method is fairly effective and easy to implement in practice. Our experience suggests that the numerical results are stable when we shift several values around the data-driven bandwidths.

Example 1. We consider the model

$$Y \quad = \quad \beta_{01}X_1 + \beta_{02}X_2 + \beta_{03}X_3 + 2\sin(\pi Z) + \epsilon. \tag{7.1}$$

A total of 1000 realizations are generated and sample sizes of $n = 300$, $n = 500$, and $n = 1000$ are considered. $\boldsymbol{\beta}_0 = (\beta_{01}, \beta_{02}, \beta_{03})^{\mathrm{T}} = (2, -0.1, 0)^{\mathrm{T}}$, $(\boldsymbol{X}, Z) \sim N_4(\mu_{\boldsymbol{X}}, \boldsymbol{\Sigma}_{\boldsymbol{X}})$ with $\mu_{\boldsymbol{X}} = \boldsymbol{0}_{4\times 1}$ and $\boldsymbol{\Sigma}_{\boldsymbol{X}} = (\sigma_{ij})_{1\le i,j\le 3}$, $\sigma_{ij} = (0.5)^{|i-j|}$. The model error ϵ is independent of $\boldsymbol{X}$ and generated as $N(0, 0.5^2)$. The variable U follows the uniform distribution $\mathrm{U}[0,1]$, and the distortion functions are chosen as $\phi_M(U) = \frac{12((U-0.5)^2+1)}{13}$, $\phi_A(U) = U^2 - \frac{1}{3}$, $\psi_{M,1}(U) = 1 + 0.5\sin(2\pi U)$, $\psi_{A,1}(U) = U - \frac{1}{2}$, $\psi_{M,2}(U) = 1 - \frac{U^2-1/3}{2}$, $\psi_{A,2}(U) = U^2 - \frac{1}{3}$ $\psi_{M,3}(U) = \frac{2(U+3)}{7}$ and $\psi_{A,3}(U) = (U - \frac{1}{2})^3$.

(1.1) Estimation of $\boldsymbol{\beta}_0$. In Table 1, we report the mean, standard errors and mean squared errors for the true estimator $\hat{\boldsymbol{\beta}}_T$ (the profile least-squares estimator [14] using the simulated dataset $\{Y_i, \boldsymbol{X}_i, Z_i\}_{i=1}^n$), the proposed estimator $\hat{\boldsymbol{\beta}}$ and the naive estimator $\hat{\boldsymbol{\beta}}_N$ (the profile least squares estimator using the dataset $\{\widetilde{Y}_i, \widetilde{\boldsymbol{X}}_i, Z_i\}_{i=1}^n$ without calibration). Unsurprisingly, $\hat{\boldsymbol{\beta}}_T$ performs better than $\hat{\boldsymbol{\beta}}$, because the MSE values for $\hat{\boldsymbol{\beta}}_T$ are all smaller than those for $\hat{\boldsymbol{\beta}}$. For the proposed estimator $\hat{\boldsymbol{\beta}}$, all the mean values are close to the true value $(2, -1, 0)^{\mathrm{T}}$, and the MSE values decrease as the sample size n increases. In Theorem 2.1, we show that the estimator $\hat{\beta}_r$ is asymptotically efficient when $\beta_{0r} = 0$. In Table 1, we see that MSE values for the estimator $\hat{\beta}_3$ are very close those for the true estimator $\hat{\beta}_{T3}$ when $n = 1000$. The naive estimator $\hat{\boldsymbol{\beta}}_N$ has a large bias, especially when estimating β_{01} and β_{02}. All MSE values for the naive estimator are greater than those for the true estimator and proposed estimator in this table. This indicates that ignoring the multiplicative distortion functions $\phi_M(U)$, $\phi_A(U)$, $\psi_{M,r}(U)$ and $\psi_{A,r}(U)$ increases the bias and results in an inconsistent estimator, even when the sample size n is large.

(1.2) Confidence intervals. We report the 95% normal approximation (NA) confidence intervals and empirical likelihood (EL) confidence intervals for β_{0s}, $s = 1, 2, 3$. The results are reported in Table 2. In Table 2, as the sample size n increases, we see that both the NA confidence intervals and the EL confidence intervals achieve satisfactory performance, both in terms of the average length of the confidence intervals and the coverage probabilities. The NA con-

fidence intervals are wider and have larger coverage probabilities than the EL confidence intervals. Note that the EL method does not need to estimate the asymptotic variances of estimators, whereas the NA method does. Generally, the NA asymptotic intervals and the EL method are both recommended when the sample size is large.

(1.3) Restriction Estimator. We consider the restricted estimator under two constraints $\boldsymbol{A}_1 = (2, -1, 0)$ (i.e., $2\beta_{01} - \beta_{02} = 5$) and $\boldsymbol{A}_2 = (-0.5, 1, 0)$ (i.e., $-0.5\beta_{01} + \beta_{02} = -2$). In Table 4, the MSE of the restricted estimator for β_{02} using $\boldsymbol{A}_1$ is much smaller than the value in Table 1, and the MSEs for β_{01} and β_{03} have improved slightly. This indicates that the restricted condition $\boldsymbol{A}_1$ can improve the estimation efficiency for β_{02} without sacrificing much estimation efficiency for β_{01} and β_{03}. For $\boldsymbol{A}_2$, the MSE of the restricted estimator for β_{0s}, $s = 1, 2, 3$, is much smaller than the value in Table 1, which again implies that the restricted condition $\boldsymbol{A}_2$ improves the estimation efficiency for $\boldsymbol{\beta}_0$.

(1.4) Hypothesis test. We consider the following test problem for model (7.1),

$$\mathcal{H}_0 : \boldsymbol{A}\boldsymbol{\beta}_0 = 0, \quad \mathcal{H}_1 : \boldsymbol{A}\boldsymbol{\beta}_0 = c, \quad c \neq 0, \tag{7.2}$$

where $\boldsymbol{A}_1 = (0, 0, 1)$ and $c = 0.05, 0.10, \ldots, 0.40$ for the alternative hypothesis $\mathcal{H}_1$. Under the null hypothesis $\mathcal{H}_0$, we set $\boldsymbol{\beta}_0 = (2, -1, 0)^{\mathrm{T}}$, and $\boldsymbol{\beta}_0 = (2, -1, c)^{\mathrm{T}}$ for $\mathcal{H}_1$. The simulation results for the test statistic $\mathcal{T}_n$ are reported in Table 4. In Table 4, as the value of c increases, the power function increases rapidly. The power function tends to 1 as the sample size n increases, which shows that the test statistic $\mathcal{T}_n$ is powerful for this test problem.

Table 1
Simulation results of Mean (M), Standard Error (SD) and Mean Squared Error (MSE) for true estimator $\hat{\boldsymbol{\beta}}_T$, the proposed estimator $\hat{\boldsymbol{\beta}}$, and the naive estimator $\hat{\boldsymbol{\beta}}_N$. MSE is in the scale of $\times 10^{-3}$.

	$n = 300$			$n = 500$			$n = 1000$		
	M	SD	MSE	M	SD	MSE	M	SD	MSE
$\hat{\beta}_{T,1}$	2.0002	0.0504	2.5417	2.0005	0.0378	1.4348	1.9989	0.0271	0.7317
$\hat{\beta}_1$	1.9689	0.0977	10.5076	1.9902	0.0640	4.1945	1.9966	0.0418	1.7603
$\hat{\beta}_{N,1}$	1.7459	0.0710	69.6039	1.7476	0.0546	66.6695	1.7420	0.0397	66.1176
$\hat{\beta}_{T,2}$	-1.0005	0.0534	2.8584	-0.9996	0.0441	1.9407	-0.9993	0.0312	0.9688
$\hat{\beta}_2$	-0.9761	0.0668	5.0392	-0.9870	0.0484	2.5142	-0.9924	0.0328	1.1337
$\hat{\beta}_{N,2}$	-0.8683	0.0684	22.0162	-0.8688	0.0528	19.9913	-0.8676	0.0391	19.0435
$\hat{\beta}_{T,3}$	-0.0006	0.0565	3.1923	0.0009	0.0447	1.9898	-0.0003	0.0308	0.9464
$\hat{\beta}_3$	-0.0006	0.0605	3.6625	0.0019	0.0471	2.2233	-0.0006	0.0312	0.9776
$\hat{\beta}_{N,3}$	0.0080	0.0783	6.1981	0.0071	0.0597	3.6161	0.0065	0.0407	1.7020

Example 2. We conduct 1000 simulations from model (1.1) by choosing $\boldsymbol{\beta}_0 = (2, -1.5, 0.5,\ 0, 0, \ldots, 0)^{\mathrm{T}}$ and $g(z) = 2\sin(\pi z)$, where the length of $\boldsymbol{\beta}_0$ is set to be 10, 20 and 30, i.e., the number of zero components of $\boldsymbol{\beta}_0$ is 7, 17 and 27, respectively. The covariate $(\boldsymbol{X}, Z)$ follows normal distribution $N(\boldsymbol{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = (\sigma_{ij})_{1\le i,j\le (p+1)}$, $\sigma_{ij} = (-0.5)^{|i-j|}$, $p = 10$, $p = 20$ and $p = 30$. The model error ϵ, the confounding variable U, and the distortion functions $\phi_M(U)$ and $\phi_A(U)$ are the same as those in Example 1. The distortion function for X_r is

TABLE 2
Simulation results of confidence intervals. "NA" stands for the normal approximation and "EL" stands for the empirical likelihood. "Lower" stands for the lower bound, "Upper" stands for upper bound, "AL" stands for average length, "CP" stands for the coverage probabilities.

		$n = 300$			$n = 500$			$n = 1000$		
		β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
NA	Lower	1.8073	-1.1158	-0.1351	1.8648	-1.0927	-0.1047	1.9120	-1.0683	-0.0739
	Upper	2.132‘	-0.8368	0.1347	2.1094	-0.8780	0.1045	2.0835	-0.9169	0.0745
	AL	0.3247	0.2789	0.2699	0.2446	0.2174	0.2093	0.1715	0.1514	0.1483
	CP	94.1%	96.3%	96.3%	95.5%	96.7%	97.0%	96.4%	95.9%	96.5%
EL	Lower	1.8070	-1.1058	-0.1214	1.8733	-1.0863	-0.0937	1.9184	-1.0592	-0.0661
	Upper	2.1313	-0.8490	0.1238	2.1021	-0.8879	0.0948	2.0766	-0.9278	0.0674
	AL	0.3243	0.2567	0.2452	0.2288	0.1984	0.1885	0.1581	0.1318	0.1336
	CP	94.6%	94.3%	95.5%	94.5%	95.9%	95.5%	94.7%	94.9%	95.3%

TABLE 3
Simulation results of Mean (M), Standard Error (SD) and Mean Squared Error (MSE) for $\hat{\boldsymbol{\beta}}_{\mathrm{R}}$ *with* $\boldsymbol{A}_1\boldsymbol{\beta}_0 = 5$ *and* $\boldsymbol{A}_2\boldsymbol{\beta}_0 = -2$. *All values of MSE are in the scale of* 10^{-3}.

	$\boldsymbol{A}_1 = (2, -1, 0)$			$\boldsymbol{A}_2 = (-0.5, 1, 0)$		
	M	SD	MSE	M	SD	MSE
			$n = 300$			
β_{01}	1.9971	0.0287	0.8340	1.9849	0.0661	4.5999
β_{02}	-1.0057	0.0575	3.3363	-1.0075	0.0330	1.1499
β_{03}	0.0071	0.0592	3.5538	0.0112	0.0572	3.4048
			$n = 500$			
β_{01}	2.0005	0.0239	0.5755	1.9968	0.0545	2.9828
β_{02}	-0.9989	0.0479	2.3020	-1.0015	0.0272	0.7457
β_{03}	0.0036	0.0470	2.2212	0.0059	0.0457	2.1302
			$n = 1000$			
β_{01}	2.0012	0.0135	0.1839	2.0015	0.0292	0.8551
β_{02}	-0.9975	0.0270	0.7358	-0.9992	0.0146	0.2137
β_{03}	0.0002	0.0311	0.9721	0.0011	0.0300	0.9045

TABLE 4
Simulation results for power calculations of $\mathcal{T}_n$ *in Example 1.*

	$n = 300$			$n = 500$			$n = 1000$		
Significant level	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
$c = 0.00$	0.012	0.057	0.108	0.011	0.053	0.106	0.011	0.051	0.102
$c = 0.05$	0.018	0.079	0.138	0.023	0.129	0.206	0.078	0.208	0.331
$c = 0.10$	0.101	0.299	0.432	0.197	0.427	0.578	0.524	0.795	0.875
$c = 0.15$	0.329	0.591	0.732	0.600	0.842	0.914	0.942	0.991	0.997
$c = 0.20$	0.621	0.860	0.915	0.902	0.981	0.992	1.000	1.000	1.000
$c = 0.25$	0.877	0.966	0.990	0.988	0.998	1.000	1.000	1.000	1.000
$c = 0.30$	0.970	0.994	0.999	0.998	0.999	1.000	1.000	1.000	1.000
$c = 0.35$	0.993	0.997	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$c = 0.40$	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

set to be $\psi_{M,r}(U) = \frac{1+(r+1)U^r}{2}$, $\psi_{A,r}(U) = U^r - \frac{1}{r+1}$, $r = 1, \ldots, p$. The sample size n in this example is chosen as $n = 300$, $n = 500$ and $n = 1000$.

To measure the selection and estimation accuracy, we define $\omega_{u,\boldsymbol{\beta}_0}$, $\omega_{c,\boldsymbol{\beta}_0}$ and $\omega_{o,\boldsymbol{\beta}_0}$ as the proportions of underfitted, correctly fitted and overfitted models. In the case of overfitted models, "1", "2", and "≥ 3" are the proportions of models including 1, 2, and more than 2 insignificant covariates. We denote the mean squared error $\|\hat{\boldsymbol{\beta}}_P - \boldsymbol{\beta}_0\|_2^2$ as $\mathrm{Mse}_{\boldsymbol{\beta}_0}$, where $\hat{\boldsymbol{\beta}}_P$ is the final penalized estimators. Moreover, "$\mathrm{C}_{\boldsymbol{\beta}_0}$" and "$\mathrm{IN}_{\boldsymbol{\beta}_0}$" denote the average number of the zero coefficients that were correctly set to be zero, and the average number of non-zero coefficients that were incorrectly set to be zero, respectively.

In Table 5, we report the true penalized estimator (using the true covariates $(Y, \boldsymbol{X}, Z)$), the penalized estimator $\hat{\boldsymbol{\beta}}_P$, and the naive penalized estimator (using the observed data $(\widetilde{Y}_i, \widetilde{\boldsymbol{X}}_i, Z_i)$ directly without calibration). We see that the values of "$\mathrm{C}_{\boldsymbol{\beta}_0}$" for the true penalized estimator and $\hat{\boldsymbol{\beta}}_P$ are close to the true values 7 ($p = 10$), 17 ($p = 20$), and 27 ($p = 30$), and that "$\mathrm{IN}_{\boldsymbol{\beta}_0}$" is close to 0. However, the naive penalized estimator falsely penalizes the non-zero components of $\boldsymbol{\beta}_0$ to zero, and the values of $\mathrm{IN}_{\boldsymbol{\beta}_0}$ are nonzero even for large sample sizes. For the true penalized estimator and the penalized estimator $\hat{\boldsymbol{\beta}}_P$, the proportion of models that are correctly fitted (column $\omega_{c,\boldsymbol{\beta}_0}$) is above 90% when $n = 300$, and 100% when $n = 1000$. The proportions of models that are underfitted (column $\omega_{u,\boldsymbol{\beta}_0}$) and overfitted (columns under $\omega_{o,\boldsymbol{\beta}_0}$) for the true penalized estimator and $\hat{\boldsymbol{\beta}}_P$ are about 0% and 10% when $n = 300$ and $n = 500$, respectively. In the overfitted case, the proportion of models including 1 insignificant covariate dominates the cases including 2 or more insignificant covariates, and the latter is close to 0% in most situations. This indicates that the true penalized estimator and $\hat{\boldsymbol{\beta}}_P$ are most likely to select a final model that is very close to the true model. Moreover, the mean squared errors $\mathrm{Mse}_{\boldsymbol{\beta}_0}$ for $\hat{\boldsymbol{\beta}}_P$ are much smaller than for the naive penalized estimator. The naive penalized estimator definitely ruins the oracle property of the SCAD penalty function, giving larger values of $\mathrm{Mse}_{\boldsymbol{\beta}_0}$. The large biases can not be eliminated even when the sample size n increases to 1000, which coincides with the simulation results reported in Table 1. When the sample size $n = 300$, the percentage of underfitted models (columns under $\omega_{u,\boldsymbol{\beta}_0}$) is about 8%, which implies that the naive penalized estimator eventually produces an incorrect model. This again indicates that ignoring the distortion measurement errors in the variable selection process will ruin the oracle property and result in a wrong model.

Example 3. We generate 1000 realizations from model (7.1). The sample size are chosen to $n = 300$, $n = 500$ and $n = 1000$. The variables $(\boldsymbol{X}, Z, \epsilon, U)$ are the same as those in Example 1. The multiplicative distortion functions are $\phi_M(U) = \frac{12((U-0.5)^2+1)}{13}$, $\psi_{M,1}(U) = 1 + 0.3\cos(2\pi U)$, $\psi_{M,2}(U) = U^2 + \frac{2}{3}$ and $\psi_{M,3}(U) = \frac{5}{4} - U^3$. The additive distortion functions are set to be zero functions, i.e., $\phi_A(U) = 0$, $\psi_{A,r}(U) = 0$, $r = 1, \ldots, p$. We compare the performance of the estimators $(\hat{\boldsymbol{\beta}}_V, \hat{\boldsymbol{\beta}}_C)$ and their confidence intervals, test statistics $(\mathcal{T}_{V,n}, \mathcal{T}_{C,n})$ and penalized estimators $(\hat{\boldsymbol{\beta}}_{P,V}, \hat{\boldsymbol{\beta}}_{P,C})$. Note that the estimation method proposed in [1] can not be used in this example because $E(X_r) = 0$, $r = 1, \ldots, p$.

In Table 6, it is not surprising that the true estimator performs better than the proposed estimator, because the MSE values for $\hat{\boldsymbol{\beta}}_T$ are all smaller than $\hat{\boldsymbol{\beta}}_V$ and $\hat{\boldsymbol{\beta}}_C$, and their mean values are close to the true value $(2, -1, 0)^{\mathrm{T}}$. The performance of $\hat{\boldsymbol{\beta}}_V$ is slightly better than that of $\hat{\boldsymbol{\beta}}_C$. The latter has a slightly larger MSE. In Theorem 6.1, we have shown that the estimator $\hat{\beta}_r$ is asymptotically efficient when $\beta_{0r} = 0$. In Table 6, we see that the MSE values for the

TABLE 5

Simulation results for Example 2. "T" stands for the true penalized estimator, "P" standards for the penalized estimator $\hat{\boldsymbol{\beta}}_P$, and "N" stands for the naive penalized estimator. All values of $\mathrm{Mse}_{\boldsymbol{\beta}_0}$ *are in the scale of* 10^{-3}.

				$\omega_{o,\boldsymbol{\beta}_0}$ (%)				No of zeros	
(p_0,q_0)		$\omega_{u,\boldsymbol{\beta}_0}$ (%)	$\omega_{c,\boldsymbol{\beta}_0}$ (%)	"1 (%)"	"2" (%)	"≥ 3"(%)	$\mathrm{Mse}_{\boldsymbol{\beta}_0}$	$\mathrm{C}_{\boldsymbol{\beta}_0}$	$\mathrm{IN}_{\boldsymbol{\beta}_0}$
				$n=300$					
$(3,7)$	T	0.0	98.1	1.9	0.0	0.0	5.3655	6.981	0.000
$(3,7)$	P	0.0	91.0	8.6	0.4	0.0	11.6584	6.906	0.000
$(3,7)$	N	8.4	88.8	2.7	0.1	0.0	208.6972	6.967	0.084
				$n=500$					
$(3,7)$	T	0.0	100	0.0	0.0	0.0	3.5145	7.000	0.000
$(3,7)$	P	0.0	99.9	0.1	0.0	0.0	7.2799	6.999	0.000
$(3,7)$	N	3.1	96.6	0.3	0.0	0.0	202.1424	6.997	0.031
				$n=1000$					
$(3,7)$	T	0.0	100	0.0	0.0	0.0	1.8867	7.000	0.000
$(3,7)$	P	0.0	100	0.0	0.0	0.0	3.4036	7.000	0.000
$(3,7)$	N	0.5	99.5	0.0	0.0	0.0	195.0048	7.000	0.005
				$n=300$					
$(3,17)$	T	0.0	98.7	1.2	0.1	0.0	6.0981	16.986	0.000
$(3,17)$	P	0.0	94.6	5.0	0.4	0.0	14.0470	16.942	0.000
$(3,17)$	N	7.2	87.3	5.2	0.3	0.0	205.6913	16.942	0.072
				$n=500$					
$(3,17)$	T	0.0	100	0.0	0.0	0.0	3.7067	17.000	0.000
$(3,17)$	P	0.0	99.7	0.3	0.0	0.0	7.2445	16.997	0.000
$(3,17)$	N	4.1	95.5	0.4	0.0	0.0	201.2426	16.996	0.041
				$n=1000$					
$(3,17)$	T	0.0	100	0.0	0.0	0.0	1.9170	17.000	0.000
$(3,17)$	P	0.0	100	0.0	0.0	0.0	3.5003	17.000	0.000
$(3,17)$	N	0.2	99.8	0.0	0.0	0.0	194.4349	17.000	0.002
				$n=300$					
$(3,27)$	T	0.0	99.1	0.0	0.0	0.0	5.6787	26.991	0.000
$(3,27)$	P	0.0	90.0	8.7	0.7	0.2	13.1403	26.893	0.000
$(3,27)$	N	7.8	83.6	7.7	0.9	0.0	208.6185	26.895	0.078
				$n=500$					
$(3,27)$	T	0.0	100	0.0	0.0	0.0	3.6449	27.000	0.000
$(3,27)$	P	0.0	99.2	0.8	0.0	0.0	7.3020	26.992	0.000
$(3,27)$	N	2.8	96.5	0.7	0.0	0.0	202.0266	26.993	0.028
				$n=1000$					
$(3,27)$	T	0.0	100	0.0	0.0	0.0	1.7407	27.000	0.000
$(3,27)$	P	0.0	100	0.0	0.0	0.0	3.3344	27.000	0.000
$(3,27)$	N	0.8	99.2	0.0	0.0	0.0	194.5244	27.000	0.008

estimators $\hat{\beta}_{V,3}$ and $\hat{\beta}_{C,3}$ become very close to those for the true estimator $\hat{\beta}_{T,3}$ as the sample size n increases. Moreover, their MSE values are also close to those in Table 1. This again implies that the multiplicative distortions and additive distortions asymptotically have no effect on estimating $\beta_{03}=0$, regardless of the choice of distortion functions.

In Table 7, we report the 95% NA confidence intervals based on the estimator $\hat{\boldsymbol{\beta}}_V$ associated with EL confidence intervals, and the NA confidence intervals based on the estimator $\hat{\boldsymbol{\beta}}_C$ associated with EL confidence intervals for β_{0s}, $s=1,2,3$. In Table 7, the NA confidence intervals according to $\hat{\boldsymbol{\beta}}_C$ are slightly wider than for $\hat{\boldsymbol{\beta}}_V$, but the coverage probabilities are slightly larger than the empirical likelihood confidence intervals. Additionally, the EL method produces more accurate coverage probabilities than the NA method (see also Table 2).

In Table 8, we compare the performance of the test statistics $\mathcal{T}_{V,n}$ and $\mathcal{T}_{C,n}$ for the hypothesis testing problem 7.2. Simulation results are similar to those in Table 4. As the value of c increases, the power functions of $\mathcal{T}_{V,n}$ and $\mathcal{T}_{C,n}$

increase to a rapidly as the sample size n increases. It is clear that $\mathcal{T}_{C,n}$ is more powerful than $\mathcal{T}_{V,n}$, and this coincides with the confidence intervals of β_{03} reported in Table 7. As the 95% confidence intervals based on $\hat{\boldsymbol{\beta}}_C$ are slightly wider than those for $\hat{\boldsymbol{\beta}}_V$, if one uses the complement sets of the 95% confidence intervals to reject the hypothesis 7.2, the complement sets of the confidence intervals according to $\hat{\boldsymbol{\beta}}_C$ are more powerful than those for $\hat{\boldsymbol{\beta}}_V$. The simulation results in Table 8 also reveal that $\mathcal{T}_{C,n}$ (based on $\hat{\boldsymbol{\beta}}_C$) is more powerful than $\mathcal{T}_{V,n}$ (based on $\hat{\boldsymbol{\beta}}_V$).

Table 6

Simulation results of Mean (M), Standard Error (SD) and Mean Squared Error (MSE) for true estimator $\hat{\boldsymbol{\beta}}_T$, the proposed estimators $\hat{\boldsymbol{\beta}}_V$ and $\hat{\boldsymbol{\beta}}_C$, and the naive estimator $\hat{\boldsymbol{\beta}}_N$. MSE is in the scale of $\times 10^{-3}$.

	$n=300$			$n=500$			$n=1000$		
	M	SD	MSE	M	SD	MSE	M	SD	MSE
$\hat{\beta}_{T,1}$	1.9972	0.0501	2.5169	2.0009	0.0403	1.6246	1.9989	0.0280	0.7858
$\hat{\beta}_{V,1}$	1.9822	0.0558	3.4256	1.9924	0.0449	2.0718	1.9958	0.0311	0.9859
$\hat{\beta}_{C,1}$	1.9690	0.0577	4.2963	1.9794	0.0457	2.5095	1.9833	0.0313	1.2614
$\hat{\beta}_{N,1}$	1.8835	0.0608	17.2547	1.8850	0.0482	15.5462	1.8835	0.0330	14.6514
$\hat{\beta}_{T,2}$	-0.9978	0.0556	3.1008	-1.0009	0.0445	1.9853	-1.0005	0.0321	1.0341
$\hat{\beta}_{V,2}$	-0.9784	0.0626	4.3797	-0.9891	0.0494	2.5585	-0.9928	0.0347	1.2603
$\hat{\beta}_{C,2}$	-0.9787	0.0632	4.4477	-0.9893	0.0499	2.6080	-0.9933	0.0350	1.2745
$\hat{\beta}_{N,2}$	-0.8497	0.0595	26.1239	-0.8520	0.0475	24.1497	-0.8514	0.0334	23.1959
$\hat{\beta}_{T,3}$	-0.0006	0.0547	2.9924	0.0002	0.0443	1.9649	0.0015	0.0305	0.9351
$\hat{\beta}_{V,3}$	-0.0052	0.0566	3.2321	-0.0033	0.0461	2.1416	-0.0018	0.0306	0.9447
$\hat{\beta}_{C,3}$	-0.0045	0.0590	3.5057	-0.0020	0.0472	2.2324	0.0004	0.0317	1.0093
$\hat{\beta}_{N,3}$	-0.0715	0.0583	8.5270	-0.0698	0.0471	7.0912	-0.0699	0.0324	5.9402

Table 7

Simulation results of confidence intervals. "NAV" stands for the normal approximation based on $\hat{\boldsymbol{\beta}}_V$, "ELV" stands for the corresponding empirical likelihood method. "NAC" stands for the normal approximation based on $\hat{\boldsymbol{\beta}}_C$, "ELV" stands for the corresponding empirical likelihood method.

		$n=300$			$n=500$			$n=1000$		
		β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
NAV	Lower	1.8567	-1.1215	-0.1351	1.8929	-1.0974	-0.1023	1.9257	-1.0701	-0.0729
	Upper	2.1151	-0.8428	0.1226	2.0926	-0.8806	0.0950	2.0673	-0.9159	0.0668
	AL	0.2584	0.2787	0.2542	0.1997	0.2167	0.1973	0.1416	0.1542	0.1398
	CP	96.3%	96.7%	96.5%	95.7%	96.3%	96.3%	95.8%	95.6%	96.4%
NAC	Lower	1.8414	-1.1239	-0.1363	1.8786	-1.0989	-0.1066	1.9124	-1.0714	-0.0753
	Upper	2.1044	-0.8451	0.1305	2.0870	-0.8795	0.1016	2.0549	-0.9156	0.0727
	AL	0.2629	0.2824	0.2669	0.2020	0.2194	0.2082	0.1425	0.1558	0.1481
	CP	96.3%	96.7%	96.7%	95.5%	96.7%	97.0%	95.5%	95.7%	96.1%
ELV	Lower	1.8637	-1.1153	-0.1201	1.8996	-1.0844	-0.0930	1.9321	-1.0585	-0.0658
	Upper	2.1070	-0.8499	0.1092	2.0859	-0.8939	0.0847	2.0609	-0.9275	0.0603
	AL	0.2432	0.2653	0.2294	0.1863	0.1905	0.1778	0.1287	0.1309	0.1263
	CP	96.0%	94.5%	95.2%	95.2%	94.9%	95.1%	94.7%	94.6%	95.4%
ELC	Lower	1.8504	-1.1173	-0.1250	1.8861	-1.0850	-0.0965	1.9193	-1.0591	-0.0679
	Upper	2.094	-0.8492	0.1164	2.0735	-0.8935	0.0909	2.0480	-0.9279	0.0654
	AL	0.2445	0.2680	0.2415	0.1873	0.1915	0.1875	0.1286	0.1312	0.1334
	CP	94.7%	94.5%	95.5%	94.6%	94.7%	95.7%	94.7%	94.8%	95.9%

TABLE 8
Simulation results for power calculations of $\mathcal{T}_{V,n}$ and $\mathcal{T}_{C,n}$ in Example 3.

	$\mathcal{T}_{V,n}$			$\mathcal{T}_{C,n}$		
Significant level	0.01	0.05	0.10	0.01	0.05	0.10
			$n = 300$			
$c = 0.00$	0.012	0.052	0.102	0.010	0.054	0.104
$c = 0.05$	0.018	0.078	0.150	0.016	0.073	0.137
$c = 0.10$	0.081	0.253	0.363	0.085	0.259	0.385
$c = 0.15$	0.272	0.544	0.658	0.311	0.575	0.713
$c = 0.20$	0.557	0.769	0.854	0.607	0.848	0.912
$c = 0.25$	0.783	0.904	0.942	0.852	0.954	0.974
$c = 0.30$	0.887	0.938	0.953	0.970	0.994	1.000
$c = 0.35$	0.929	0.963	0.970	0.997	1.000	1.000
$c = 0.40$	0.935	0.958	0.966	1.000	1.000	1.000
			$n = 500$			
$c = 0.00$	0.011	0.051	0.095	0.011	0.053	0.102
$c = 0.05$	0.034	0.109	0.196	0.027	0.105	0.196
$c = 0.10$	0.178	0.393	0.515	0.184	0.413	0.562
$c = 0.15$	0.540	0.764	0.852	0.595	0.826	0.917
$c = 0.20$	0.767	0.897	0.923	0.895	0.976	0.991
$c = 0.25$	0.928	0.959	0.969	0.992	0.997	1.000
$c = 0.30$	0.946	0.970	0.984	0.995	1.000	1.000
$c = 0.35$	0.966	0.998	1.000	1.000	1.000	1.000
$c = 0.40$	1.000	1.000	1.000	1.000	1.000	1.000
			$n = 1000$			
$c = 0.00$	0.010	0.049	0.101	0.012	0.051	0.104
$c = 0.05$	0.061	0.193	0.301	0.063	0.212	0.333
$c = 0.10$	0.413	0.686	0.780	0.499	0.763	0.849
$c = 0.15$	0.812	0.902	0.931	0.922	0.987	0.993
$c = 0.20$	0.923	0.952	0.962	0.997	0.999	1.000
$c = 0.25$	0.942	0.968	1.000	1.000	1.000	1.000
$c = 0.30$	0.987	1.000	1.000	1.000	1.000	1.000
$c = 0.35$	1.000	1.000	1.000	1.000	1.000	1.000
$c = 0.40$	1.000	1.000	1.000	1.000	1.000	1.000

Example 4. In this example, we conduct 1000 simulations from model (1.1) to examine the performance of the penalized estimators $\hat{\boldsymbol{\beta}}_{P,V}$ and $\hat{\boldsymbol{\beta}}_{P,C}$. The sample sizes are set to be $n = 300$, $n = 500$, and $n = 1000$. The parameter $\boldsymbol{\beta}_0$, the variables $(\boldsymbol{X}, Z, U, Y, \epsilon)$ and multiplicative distortion functions $\phi_M(U)$ and $\psi_{M,r}(U)$'s are the same as those in Example 2. The additive distortion functions are all set to be zero: $\phi_A(U) = 0$, $\psi_{A,r}(U) = 0$, $r = 1, \ldots, p$.

In Table 9, we report the true penalized estimator (using the true simulated data $\{Y_i, \boldsymbol{X}_i, Z_i\}_{i=1}^n$), the penalized estimators $\hat{\boldsymbol{\beta}}_{P,V}$, $\hat{\boldsymbol{\beta}}_{P,C}$, and the naive penalized estimator (using the distorted data $\{\widetilde{Y}_i, \widetilde{\boldsymbol{X}}_i, Z_i\}_{i=1}^n$ directly without calibration). The simulation results are all similar to those in Table 5. The values of "$\mathrm{C}_{\boldsymbol{\beta}_0}$" for $\hat{\boldsymbol{\beta}}_{P,V}$, $\hat{\boldsymbol{\beta}}_{P,C}$ are close to the true values 7 ($p = 10$), 17 ($p = 20$), and 27 ($p = 30$), and "$\mathrm{IN}_{\boldsymbol{\beta}_0}$" is close to 0. The values of $\omega_{c,\boldsymbol{\beta}_0}$ are all above 97%. For $\omega_{u,\boldsymbol{\beta}_0}$ (underfitted model) and $\omega_{o,\boldsymbol{\beta}_0}$ (overfitted model), the values are all close to 0. However, the naive penalized estimator falsely penalizes the nonzero components of $\boldsymbol{\beta}_0$ to zero because the values of $\mathrm{IN}_{\boldsymbol{\beta}_0}$ are nonzero, even for large sample sizes, and the values of $\omega_{u,\boldsymbol{\beta}_0}$ are also nonzero. The naive penalized estimator definitely ruins the oracle property of SCAD penalty function, giving larger values of $\mathrm{Mse}_{\boldsymbol{\beta}_0}$. The large biases can not be eliminated even when the sample size n increases to 1000, which coincides with the simulation results reported in Table 5. This again indicates that ignoring the multiplicative distortion measurement errors for the variable selection process will ruin the oracle property and result in a poor model.

Table 9

Simulation results for Example 3. "T" stands for the true penalized estimator, "V" standards for the penalized estimator $\hat{\boldsymbol{\beta}}_{P,V}$, "C" standards for the penalized estimator $\hat{\boldsymbol{\beta}}_{P,C}$ and "N" stands for the naive penalized estimator. All values of $\mathrm{Mse}_{\boldsymbol{\beta}_0}$ *are in the scale of* 10^{-3}.

				$\omega_{o,\boldsymbol{\beta}_0}$ (%)				No of zeros	
(p_0, q_0)		$\omega_{u,\boldsymbol{\beta}_0}$ (%)	$\omega_{c,\boldsymbol{\beta}_0}$ (%)	"1 (%)"	"2" (%)	"$\geq$ 3"(%)	$\mathrm{Mse}_{\boldsymbol{\beta}_0}$	$\mathrm{C}_{\boldsymbol{\beta}_0}$	$\mathrm{IN}_{\boldsymbol{\beta}_0}$
				$n = 300$					
(3, 7)	T	0.0	99.8	0.2	0.0	0.0	19.2945	6.998	0.000
(3, 7)	V	0.0	99.4	0.6	0.4	0.0	29.7618	6.994	0.000
(3, 7)	C	0.0	99.7	0.3	0.0	0.0	29.6891	6.997	0.000
(3, 7)	N	20.3	77.6	2.0	0.1	0.0	225.8709	6.972	0.203
				$n = 500$					
(3, 7)	T	0.0	100	0.0	0.0	0.0	14.5730	7.000	0.000
(3, 7)	V	0.0	99.9	0.1	0.0	0.0	19.8203	6.999	0.000
(3, 7)	C	0.0	100	0.0	0.0	0.0	19.4369	7.999	0.000
(3, 7)	N	14.0	85.7	0.3	0.0	0.0	220.4970	6.997	0.140
				$n = 1000$					
(3, 7)	T	0.0	100	0.0	0.0	0.0	11.1170	7.000	0.000
(3, 7)	V	0.0	100	0.0	0.0	0.0	13.8354	7.000	0.000
(3, 7)	C	0.0	100	0.0	0.0	0.0	13.8996	7.000	0.000
(3, 7)	N	6.6	93.4	0.0	0.0	0.0	215.5112	7.000	0.006
				$n = 300$					
(3, 17)	T	0.0	99.8	0.2	0.0	0.0	19.1307	16.998	0.000
(3, 17)	V	0.0	98.5	1.4	0.0	0.1	30.4570	16.981	0.000
(3, 17)	C	0.0	98.5	1.4	0.1	0.0	28.9576	16.984	0.000
(3, 17)	N	20.3	75.1	4.2	0.4	0.0	226.2938	16.935	0.203
				$n = 500$					
(3, 17)	T	0.0	100	0.0	0.0	0.0	14.9827	17.000	0.000
(3, 17)	V	0.0	99.8	0.2	0.0	0.0	20.9940	16.998	0.000
(3, 17)	C	0.0	99.9	0.1	0.0	0.0	19.5397	16.999	0.000
(3, 17)	N	14.4	85.3	0.3	0.0	0.0	221.4022	16.995	0.144
				$n = 1000$					
(3, 17)	T	0.0	100	0.0	0.0	0.0	11.6909	17.000	0.000
(3, 17)	V	0.0	100	0.0	0.0	0.0	14.0145	17.000	0.000
(3, 17)	C	0.0	100	0.0	0.0	0.0	13.6771	17.000	0.000
(3, 17)	N	7.0	93.0	0.0	0.0	0.0	218.1803	17.000	0.070
				$n = 300$					
(3, 27)	T	0.0	99.7	0.3	0.0	0.0	18.3902	26.997	0.000
(3, 27)	V	0.0	97.6	2.3	0.0	0.1	29.2157	26.974	0.000
(3, 27)	C	0.0	97.5	2.4	0.1	0.0	28.4769	26.974	0.000
(3, 27)	N	21.4	72.8	5.3	0.5	0.0	227.5474	26.911	0.214
				$n = 500$					
(3, 27)	T	0.0	100	0.0	0.0	0.0	16.3829	27.000	0.000
(3, 27)	V	0.0	100	0.0	0.0	0.0	22.1077	27.000	0.000
(3, 27)	C	0.0	99.8	0.2	0.0	0.0	20.7386	26.998	0.000
(3, 27)	N	14.8	84.4	0.8	0.0	0.0	219.5000	26.991	0.148
				$n = 1000$					
(3, 27)	T	0.0	100	0.0	0.0	0.0	11.6011	27.000	0.000
(3, 27)	V	0.0	100	0.0	0.0	0.0	15.1849	27.000	0.000
(3, 27)	C	0.0	100	0.0	0.0	0.0	15.1381	27.000	0.000
(3, 27)	N	7.2	92.8	0.0	0.0	0.0	214.3303	27.000	0.072

8. Real data analysis

As an illustration, we now apply our method to the analysis of bodyfat data (http://lib.stat.cmu.edu/datasets/bodyfat). The dataset contains the percentage of body fat determined by underwater weighing and various body circumference measurements for 252 men. In practice, the accurate measurement of body fat is inconvenient and costly. Thus, simple methods of estimating body fat that are neither inconvenient nor costly are desirable. We used the partial linear model (1.1) to investigate the relationship between Y-percent body fat, Z-age, X_1-weight, X_2-height, X_3-neck circumference, X_4-chest circumference, X_5-abdomen circumference, X_6-hip circumference, X_7-thigh circumference, X_8-knee circumference, X_9-ankle circumference, X_{10}-biceps circumference, X_{11}-forearm circumference, X_{12}-wrist circumference, and the confounding variable U-the body density determined from underwater weighing. We first present the patterns of $\hat{\phi}_M(u)$, $\hat{\phi}_A(u)$, $\hat{\psi}_{M,r}(u)$'s and $\hat{\psi}_{A,r}(u)$'s in Figures 1–4. The plots show that all the distortion functions are non-constant.

Corresponding to the covariates $(X_1, \dots, X_{12})^{\mathrm{T}}$, Table 10 presents the estimators of $\boldsymbol{\beta}_0$, standard errors, p-values, confidence intervals based on NA, confidence intervals based on EL, and the penalized estimators associated with their estimated standard errors. The standard errors of the penalized estimator $\hat{\boldsymbol{\beta}}_P$ are obtained using the plug-in estimators of the asymptotic covariance matrices obtained in Theorem 5.1. The estimator $\hat{\boldsymbol{\beta}}$ and its associated p-values show that only variable X_5-abdomen is significant; the remaining p-values are all greater than 0.2. The only 95% NA confidence intervals that does not contain zero is that of β_{05}. For the 95% EL confidence intervals, only those for β_{01}, β_{05} and β_{012} exclude zero, implying that these parameters should be significant at the 5% significant level. The penalized estimator $\hat{\boldsymbol{\beta}}_P$ indicates that X_2-height, X_5-abdomen and X_{12}-wrist are irrelevant to the response Y. The above analysis suggests that X_5-abdomen is the most important variable in model (1.1). Intuitively, it makes sense that the percentage of body fat will increase as X_5 abdomen circumference becomes larger. Finally, we show the pattern of $\hat{g}(z)$ in Figure 5. This figure reveals that $g(z)$ is nonlinear, and that the percentage of body fat increases from age 20–45, decreases slightly from age 45–60, and then increases again from age 60–80. Generally, the parameter estimation results in Table 10 and the plot in Figure 1 show that X_5-abdomen and Z-age can be used as the two principal variables for predicting the percentage of body fat in future health studies.

Appendix A: Appendix

A.1. Three technical lemmas

Lemma A.1. *Suppose $E(W|U=u) = m(u)$ and its derivatives up to second order are bounded for all $u \in [a_1, a_2]$. $E(|W|^3)$ exists and $\sup_u \int |w|^s f(u,w)dw < \infty$, where $f(u,w)$ is the joint density of (U,W). Suppose (U_i, W_i), $i = 1, 2, \dots n$*

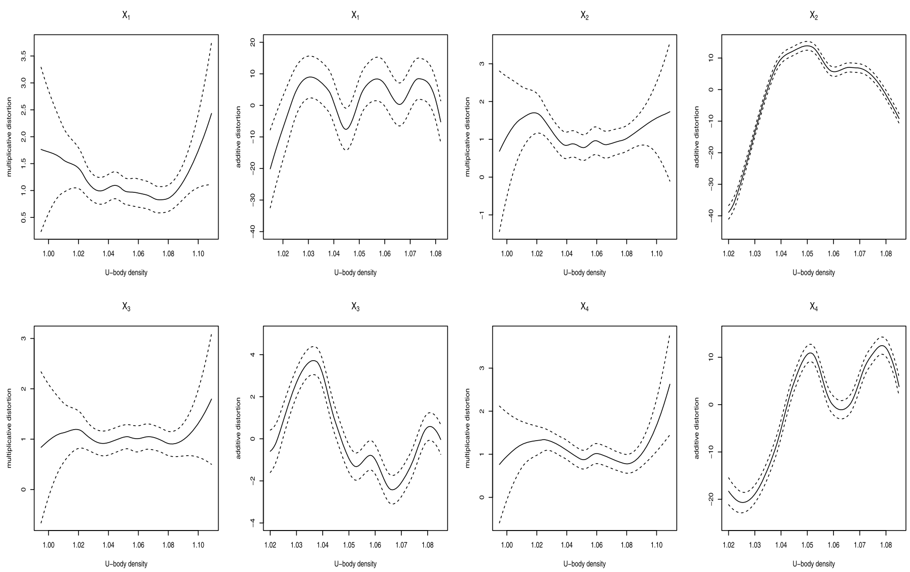

FIG 1. *Estimated curves of distorting functions* $\hat{\psi}_{M,r}(u)$ *and* $\hat{\psi}_{A,r}(u)$, $r = 1, \ldots, 4$, *against confounding variable U-the body density, associated 95% point-wise confidence intervals (dotted lines).*

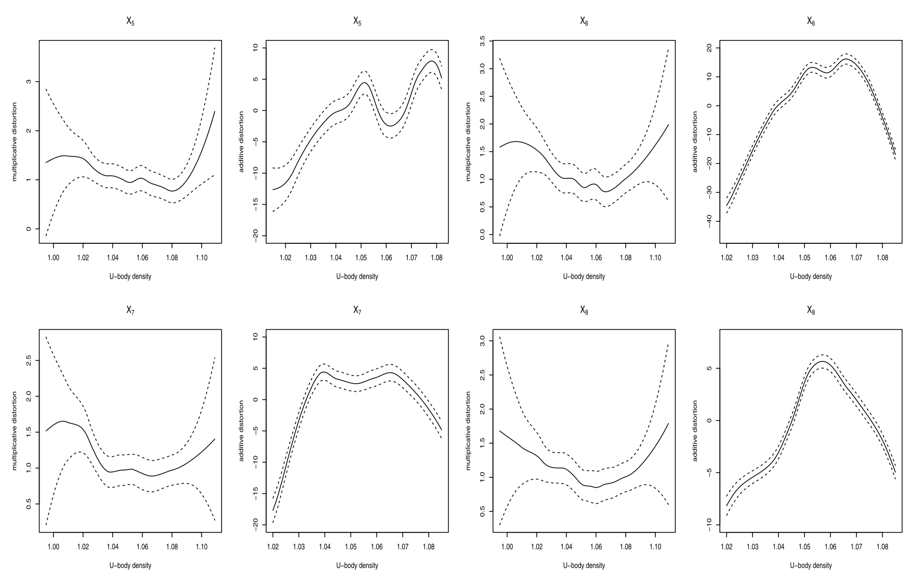

FIG 2. *Estimated curves of distorting functions* $\hat{\psi}_{M,r}(u)$ *and* $\hat{\psi}_{A,r}(u)$, $r = 5, \ldots, 8$, *against confounding variable U-the body density, associated 95% point-wise confidence intervals (dotted lines).*

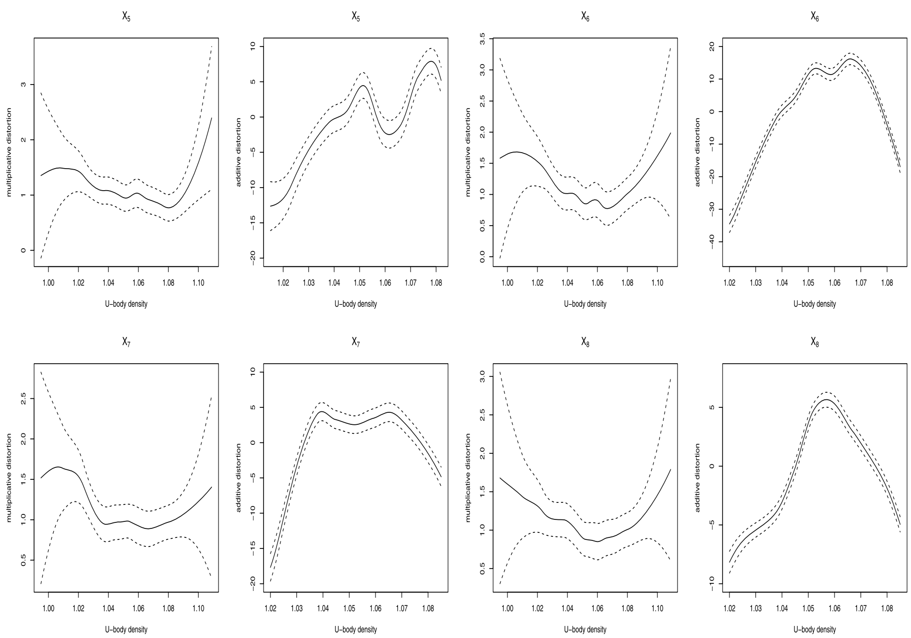

FIG 3. *Estimated curves of distorting functions $\hat{\psi}_{M,r}(u)$ and $\hat{\psi}_{A,r}(u)$, $r = 9, \ldots, 12$, against confounding variable U-the body density, associated 95% point-wise confidence intervals (dotted lines).*

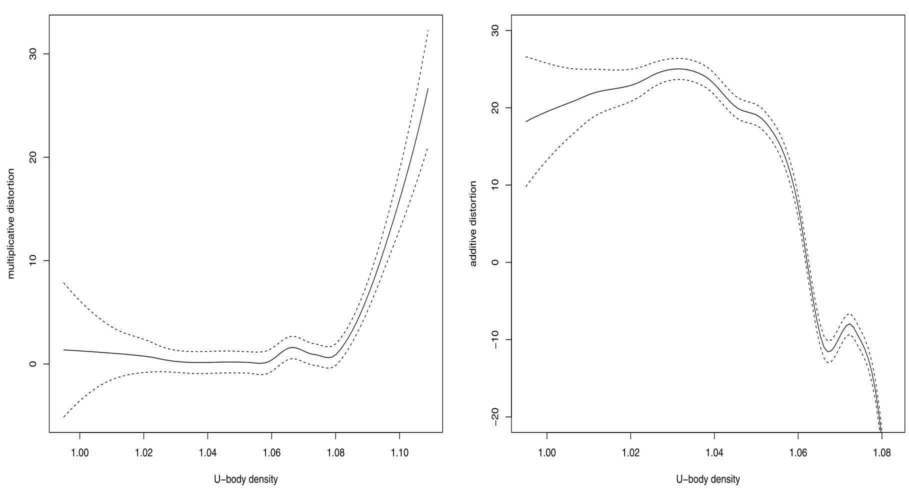

FIG 4. *Estimated curves of distorting functions $\hat{\phi}_M(u)$ (left panel) and $\hat{\phi}_A(u)$ (right panel) against confounding variable U-the body density, associated 95% point-wise confidence intervals (dotted lines).*

Table 10

Estimation results for the body fat data

	weight	height	neck	chest	abdomen	hip
$\hat{\boldsymbol{\beta}}$	-0.0213	-0.0836	-0.0634	-0.0082	0.2593	0.0422
$\text{Se}_{\hat{\boldsymbol{\beta}}}$	0.5032	1.0789	1.8082	0.8266	0.9379	1.2342
p-values	0.5001	0.2187	0.5777	0.8745	0.0001	0.5875
left NA	-0.0834	-0.2168	-0.2866	-0.1102	0.1435	-0.1102
right NA	0.0407	0.0496	0.1598	0.0938	0.3751	0.1945
left EL	-0.0343	-0.1765	-0.1964	-0.0552	0.2143	-0.0108
right EL	-0.0073	0.0024	0.0735	0.0417	0.3053	0.0961
$\hat{\boldsymbol{\beta}}_P$	0	-0.1147	0	0	0.2144	0
$\text{Se}_{\hat{\boldsymbol{\beta}}_P}$	(—)	1.0113	(—)	(—)	0.8845	(—)
	thigh	knee	ankle	biceps	forearm	wrist
$\hat{\boldsymbol{\beta}}$	-0.0336	-0.0523	0.0174	0.0285	0.0959	-0.3530
$\text{Se}_{\hat{\boldsymbol{\beta}}}$	1.1499	2.4373	2.0280	1.4388	1.5519	4.7116
p-values	0.6423	0.7329	0.8916	0.7534	0.3268	0.2342
left NA	-0.1756	-0.3533	-0.2329	-0.1491	-0.0957	-0.9347
right NA	0.1083	0.2485	0.2677	0.2061	0.2874	0.2286
left EL	-0.1036	-0.1933	-0.1536	-0.0845	-0.0531	-0.6710
right EL	0.0373	0.0946	0.1993	0.1444	0.2428	-0.0190
$\hat{\boldsymbol{\beta}}_P$	0	0	0	0	0	-0.4775
$\text{Se}_{\hat{\boldsymbol{\beta}}_P}$	(—)	(—)	(—)	(—)	(—)	3.9948

are independent and identically distributed (i.i.d.) samples from (U, W). If condition (C4) holds true for kernel function $K(u)$, and $n^{2\epsilon-1}h \to \infty$ for $\epsilon < 1 - s^{-1}$, we have

$$\sup_{u\in[a_1,a_2]}\left|\frac{1}{n}\sum_{i=1}^{n}K_h(U_i-u)W_i - f_U(u)m(u) - \frac{1}{2}[f_U(u)m(u)]''\mu_2h^2\right| \\ = O(\tau_{n,h}), a.s.$$

where $\mu_2 = \int K(u)u^2du$, and $\tau_{n,h} = h^3 + \sqrt{\log n/(nh)}$.

Proof. Lemma A.1 can be immediately proved from the result obtained by [19]. □

Lemma A.2. *Let $M(\boldsymbol{W})$ be a continuous function of $\boldsymbol{W} = (Y, \boldsymbol{X}, Z)$ satisfying $E\big[M^2(\boldsymbol{W})\big] < \infty$. Suppose Conditions (C1)-(C5) hold, we have*

$$\frac{1}{n}\sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)M(\boldsymbol{W}_i) \\ = \frac{1}{n}\sum_{i=1}^{n}\left\{\left\{\left[\frac{(Y_i-E(Y))^2}{2\sigma_Y^2}+\frac{1}{2}\right]\text{Cov}(Y, M(\boldsymbol{W})) + Y_iE[M(\boldsymbol{W})]\right\}\right. \\ \left.\times[\phi_M(U_i)-1]+\phi_A(U_i)E[M(\boldsymbol{W})]\right\} + o_P(n^{-1/2}).$$

For $r = 1, \ldots, p$, we have

$$\frac{1}{n}\sum_{i=1}^{n}\left(\hat{X}_{ri} - X_{ri}\right)M(\boldsymbol{W}_i)$$

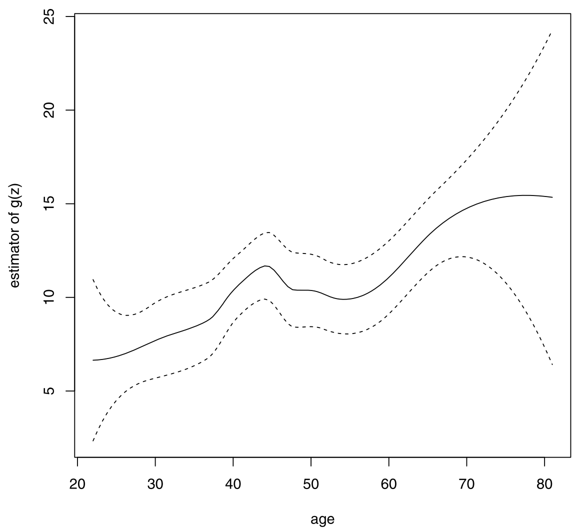

FIG 5. *The estimated curves of $\hat{g}(z)$ against Z-age, associated 95% pointwise confidence intervals (dotted lines).*

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{\left\{\left[\frac{(X_{ri}-E(X_r))^2}{2\sigma_{X_r}^2}+\frac{1}{2}\right]\mathrm{Cov}(X_r, M(\boldsymbol{W})) + X_{ri}E[M(\boldsymbol{W})]\right\}\right.$$
$$\left.\times\left[\psi_{M,r}(U_i)-1\right]+\psi_{A,r}(U_i)E[M(\boldsymbol{W})]\right\} + o_P(n^{-1/2}).$$

Proof. Lemma A.2 is the direct result of Theorem 1 in [37]. □

Lemma A.3. *Suppose that the Conditions (C1)-(C5) hold. Let $M(\boldsymbol{W})$ be a continuous function of $\boldsymbol{W} = (Y, \boldsymbol{X}, Z)$ satisfying $E\left[M^2(\boldsymbol{W})\right] < \infty$. Then,*

$$\frac{1}{n}\sum_{i=1}^{n}\left(\hat{Y}_{M,i}-Y_i\right)M(\boldsymbol{W}_i) = \frac{1}{n}\sum_{i=1}^{n}\left(|\widetilde{Y}_i|-|Y_i|\right)\frac{E\left[YM(\boldsymbol{W})\right]}{E(|Y|)} + o_P(n^{-1/2}).$$

For $r = 1, \ldots, p$, we have

$$\frac{1}{n}\sum_{i=1}^{n}\left(\hat{X}_{M,ri}-X_{ri}\right)M(\boldsymbol{W}_i)$$
$$= \frac{1}{n}\sum_{i=1}^{n}\left(|\widetilde{X}_{ri}|-|X_{ri}|\right)\frac{E\left[X_r M(\boldsymbol{W})\right]}{E(|X_r|)} + o_P(n^{-1/2}).$$

Proof. Lemma A.2 is the direct result of Lemma B.2 in [40]. See also the Lemma 1.1 in the on-line supplementary material of [41]. □

A.2. Proof of Theorem 1

Recalling that

$$\begin{aligned}\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0 &= \left\{\frac{1}{n}\sum_{i=1}^{n}\left[\hat{\boldsymbol{X}}_i-\hat{S}_{\boldsymbol{X}}(Z_i)\right]^{\otimes 2}\right\}^{-1} \\ &\quad\times\frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\boldsymbol{X}}_i-\hat{S}_{\boldsymbol{X}}(Z_i)\right\}\left\{\hat{Y}_i-\hat{S}_Y(Z_i)-\hat{\boldsymbol{X}}_i^{\mathrm{T}}\boldsymbol{\beta}_0+[\hat{S}_{\boldsymbol{X}}(Z_i)]^{\mathrm{T}}\boldsymbol{\beta}_0\right\} \\ &= \left\{\frac{1}{n}\sum_{i=1}^{n}\left[\hat{\boldsymbol{X}}_i-\hat{S}_{\boldsymbol{X}}(Z_i)\right]^{\otimes 2}\right\}^{-1}\left[\mathbb{D}_{n1}+\mathbb{D}_{n2}+\mathbb{D}_{n3}\right],\end{aligned} \tag{A.1}$$

where,

$$\mathbb{D}_{n1} = \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\boldsymbol{X}}_i-\hat{S}_{\boldsymbol{X}}(Z_i)\right\}\epsilon_i, \tag{A.2}$$

$$\mathbb{D}_{n2} = \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\boldsymbol{X}}_i-\hat{S}_{\boldsymbol{X}}(Z_i)\right\}\left\{\hat{Y}_i-Y_i-(\hat{\boldsymbol{X}}_i-\boldsymbol{X}_i)^{\mathrm{T}}\boldsymbol{\beta}_0\right\}, \tag{A.3}$$

$$\begin{aligned}\mathbb{D}_{n3} &= \frac{1}{n}\sum_{i=1}^{n}\left\{S_Y(Z_i)-\hat{S}_Y(Z_i)-(S_{\boldsymbol{X}}(Z_i)-\hat{S}_{\boldsymbol{X}}(Z_i))^{\mathrm{T}}\boldsymbol{\beta}_0\right\} \\ &\quad\times\left\{\hat{\boldsymbol{X}}_i-\hat{S}_{\boldsymbol{X}}(Z_i)\right\}.\end{aligned} \tag{A.4}$$

Step 1.1 For the expression $\mathbb{D}_{n1}$, we have

$$\begin{aligned}\mathbb{D}_{n1} &= \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\boldsymbol{X}}_i-\boldsymbol{X}_i\right\}\epsilon_i+\frac{1}{n}\sum_{i=1}^{n}\left\{\boldsymbol{X}_i-S_{\boldsymbol{X}}(Z_i)\right\}\epsilon_i \\ &\quad+\frac{1}{n}\sum_{i=1}^{n}\left\{S_{\boldsymbol{X}}(Z_i)-\hat{S}_{\boldsymbol{X}}(Z_i)\right\}\epsilon_i \\ &\stackrel{\text{def}}{=} \mathbb{D}_{n1}[1]+\mathbb{D}_{n1}[2]+\mathbb{D}_{n1}[3].\end{aligned} \tag{A.5}$$

Recalling $\epsilon_i=Y_i-\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0-g(Z_i)$ and $E(\epsilon_i|\boldsymbol{X}_i,Z_i)=0$. Using Lemma A.2, we have

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\left\{\hat{X}_{ri}-X_{ri}\right\}\epsilon_i \\ &=\frac{1}{n}\sum_{i=1}^{n}\left\{\left\{\left[\frac{(X_{ri}-E(X_r))^2}{2\sigma_{X_r}^2}+\frac{1}{2}\right]\mathrm{Cov}(X_r,\epsilon)+X_{ri}E(\epsilon)\right\}\right. \\ &\qquad\left.\times\left[\psi_{M,r}(U_i)-1\right]+\psi_{A,r}(U_i)E(\epsilon)\right\}+o_P(n^{-1/2})=o_P(n^{-1/2}).\end{aligned} \tag{A.6}$$

Based on (A.6), we have $\mathbb{D}_{n1}[1] = o_P(n^{-1/2})$. For argument $\mathbb{D}_{n1}[3]$, directly using Lemma A.1 and similar to the proof of Theorem 1 in [15], we have $\mathbb{D}_{n1}[3] = o_P(n^{-1/2})$. Thus, we obtain that

$$\mathbb{D}_{n1} = \frac{1}{n}\sum_{i=1}^{n}\left\{\boldsymbol{X}_i - S_{\boldsymbol{X}}(Z_i)\right\}\epsilon_i + o_P(n^{-1/2}). \tag{A.7}$$

Step 1.2 For the argument $\mathbb{D}_{n2}$, we have

$$\begin{aligned}
\mathbb{D}_{n2} &= \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\boldsymbol{X}}_i - \boldsymbol{X}_i\right\}\left\{\hat{Y}_i - Y_i - (\hat{\boldsymbol{X}}_i - \boldsymbol{X}_i)^{\mathrm{T}}\boldsymbol{\beta}_0\right\} \qquad \text{(A.8)}\\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\left\{\boldsymbol{X}_i - S_{\boldsymbol{X}}(Z_i)\right\}\left\{\hat{Y}_i - Y_i - (\hat{\boldsymbol{X}}_i - \boldsymbol{X}_i)^{\mathrm{T}}\boldsymbol{\beta}_0\right\}\\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\left\{S_{\boldsymbol{X}}(Z_i) - \hat{S}_{\boldsymbol{X}}(Z_i)\right\}\left\{\hat{Y}_i - Y_i - (\hat{\boldsymbol{X}}_i - \boldsymbol{X}_i)^{\mathrm{T}}\boldsymbol{\beta}_0\right\}\\
&\stackrel{\text{def}}{=} \mathbb{D}_{n2}[1] + \mathbb{D}_{n2}[2] + \mathbb{D}_{n2}[3].
\end{aligned}$$

Let $\hat{V}_i = \hat{Y}_i$, or $\hat{V}_i = \hat{X}_{ri}$, and $\hat{D}_i = \hat{Y}_i$, or $\hat{D}_i = \hat{X}_{ri}$, accordingly, $V_i = Y_i$, or $V_i = X_{ri}$ or $V_i = Z_i$, and $D_i = Y_i$, or $D_i = X_{ri}$ or $D_i = Z_i$. According to the proof of theorems in Zhang, Lin and Li (2019), as $nh^8 \to 0$ and $\frac{\log^2 n}{nh^2} \to 0$, Lemma A.1 entails that

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{V}_i - V_i)(\hat{D}_i - D_i) = O_P((n^{-1/2} + h^2 + \tau_{n,h})^2) = o_P(n^{-1/2}). \tag{A.9}$$

Using (A.9), we have $\mathbb{D}_{n2}[1] = o_P(n^{-1/2})$. For the argument $\mathbb{D}_{n2}[2]$, using $E[\boldsymbol{X} - S_{\boldsymbol{X}}(Z)|Z] = 0$ and $\mathrm{Cov}(Y, \boldsymbol{X} - S_{\boldsymbol{X}}(Z)) = \boldsymbol{\Sigma}_0\boldsymbol{\beta}_0$, Lemma A.2 entails that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left\{\boldsymbol{X}_i - S_{\boldsymbol{X}}(Z_i)\right\}\left\{\hat{Y}_i - Y_i\right\} \qquad \text{(A.10)}\\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{(Y_i - E(Y))^2}{2\sigma_Y^2} + \frac{1}{2}\right]\mathrm{Cov}(Y, \boldsymbol{X} - S_{\boldsymbol{X}}(Z))\left[\phi_M(U_i) - 1\right]\\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\left\{Y_i E[\boldsymbol{X} - S_{\boldsymbol{X}}(Z)]\left[\phi_M(U_i) - 1\right] + \phi_A(U_i)E[\boldsymbol{X} - S_{\boldsymbol{X}}(Z)]\right\}\\
&\quad + o_P(n^{-1/2})\\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{(Y_i - E(Y))^2}{2\sigma_Y^2} + \frac{1}{2}\right]\left[\phi_M(U_i) - 1\right]\boldsymbol{\Sigma}_0\boldsymbol{\beta}_0 + o_P(n^{-1/2}).
\end{aligned}$$

Similarly, we have

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\boldsymbol{X}_i - S_{\boldsymbol{X}}(Z_i)\right\}(\hat{\boldsymbol{X}}_i - \boldsymbol{X}_i)^{\mathrm{T}}\boldsymbol{\beta}_0 \tag{A.11}$$

$$
\begin{aligned}
&= \sum_{r=1}^{p} \left\{ \frac{1}{n} \sum_{i=1}^{n} \{ \boldsymbol{X}_i - S_{\boldsymbol{X}}(Z_i) \} (\hat{X}_{ri} - X_{ri}) \beta_{0r} \right\} \\
&= \sum_{r=1}^{p} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left[\frac{(X_{ri} - E(X_r))^2}{2\sigma_{X_r}^2} + \frac{1}{2} \right] \mathrm{Cov}(X_r, \boldsymbol{X} - S_{\boldsymbol{X}}(Z)) \right. \\
&\qquad\qquad \left. \times \left[\psi_{M,r}(U_i) - 1 \right] \beta_{0r} \right\} \\
&\quad + \sum_{r=1}^{p} \left\{ \frac{1}{n} \sum_{i=1}^{n} X_{ri} E[\boldsymbol{X} - S_{\boldsymbol{X}}(Z)] \left[\psi_{M,r}(U_i) - 1 \right] \beta_{0r} \right\} \\
&\quad + \sum_{r=1}^{p} \left\{ \frac{1}{n} \sum_{i=1}^{n} \psi_{A,r}(U_i) E[\boldsymbol{X} - S_{\boldsymbol{X}}(Z)] \beta_{0r} \right\} + o_P(n^{-1/2}) \\
&= \frac{1}{n} \sum_{i=1}^{n} \sum_{r=1}^{p} \left[\frac{(X_{ri} - E(X_r))^2}{2\sigma_{X_r}^2} + \frac{1}{2} \right] \left[\psi_{M,r}(U_i) - 1 \right] \boldsymbol{\Sigma}_0 e_r e_r^{\mathrm{T}} \boldsymbol{\beta}_0 + o_P(n^{-1/2}).
\end{aligned}
$$

Together with (A.10) and (A.11), we have

$$
\begin{aligned}
\mathbb{D}_{n2}[2] &= \frac{1}{n} \sum_{i=1}^{n} \left[\frac{(Y_i - E(Y))^2}{2\sigma_Y^2} + \frac{1}{2} \right] \left[\phi_M(U_i) - 1 \right] \boldsymbol{\Sigma}_0 \boldsymbol{\beta}_0 \\
&\quad - \frac{1}{n} \sum_{i=1}^{n} \sum_{r=1}^{p} \left[\frac{(X_{ri} - E(X_r))^2}{2\sigma_{X_r}^2} + \frac{1}{2} \right] \left[\psi_{M,r}(U_i) - 1 \right] \boldsymbol{\Sigma}_0 e_r e_r^{\mathrm{T}} \boldsymbol{\beta}_0 \\
&\quad + o_P(n^{-1/2}).
\end{aligned} \tag{A.12}
$$

Under the condition $nh_1^8 \to 0$ and $\frac{\log n}{nh_1^2} \to 0$, the conclusion of (A.1) in [15] entails that $\sup_{z \in \mathcal{Z}} |\hat{S}_Y(z) - S_Y(z)| = o_P(n^{-1/4})$, and $\sup_{z \in \mathcal{Z}} |\hat{s}_{X_r}(z) - s_{X_r}(z)| = o_P(n^{-1/4})$, $r = 1, \ldots, p$. According to the proof of Theorem 1 in [38], we can show that $\mathbb{D}_{n2}[3] = o_P(n^{-1/2})$, and also $\mathbb{D}_{n3} = o_P(n^{-1/2})$. Moreover,

$$
\frac{1}{n} \sum_{i=1}^{n} \left[\hat{\boldsymbol{X}}_i - \hat{S}_{\boldsymbol{X}}(Z_i) \right]^{\otimes 2} \xrightarrow{P} \boldsymbol{\Sigma}_0. \tag{A.13}
$$

Thus, together with (A.7), (A.12) and (A.13), we obtain that

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 &= \boldsymbol{\Sigma}_0^{-1} \left(\mathbb{D}_{n1} + \mathbb{D}_{n2} + \mathbb{D}_{n3} \right) + o_P(n^{-1/2}) \\
&= \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\Sigma}_0^{-1} \{ \boldsymbol{X}_i - S_{\boldsymbol{X}}(Z_i) \} \epsilon_i \\
&\quad + \frac{1}{n} \sum_{i=1}^{n} \left[\frac{(Y_i - E(Y))^2}{2\sigma_Y^2} + \frac{1}{2} \right] \left[\phi_M(U_i) - 1 \right] \boldsymbol{\beta}_0 \\
&\quad - \frac{1}{n} \sum_{i=1}^{n} \sum_{r=1}^{p} \left[\frac{(X_{ri} - E(X_r))^2}{2\sigma_{X_r}^2} + \frac{1}{2} \right] \left[\psi_{M,r}(U_i) - 1 \right] e_r e_r^{\mathrm{T}} \boldsymbol{\beta}_0
\end{aligned} \tag{A.14}
$$

$$+o_P(n^{-1/2}).$$

Using (A.14), we have completed the proof of Theorem 1.

A.3. Proof of Theorem 2

Note that

$$\begin{aligned}
\hat{g}(z) - g(z) &= \frac{T_{n2}(z)V_{n0}(z) - T_{n1}(z)V_{n1}(z)}{T_{n2}(z)T_{n0}(z) - [T_{n1}(z)]^2} - g(z) \qquad (A.15)\\
&= \frac{T_{n2}(z)[V_{n0}(z) - T_{n0}(z)g(z)]}{T_{n2}(z)T_{n0}(z) - [T_{n1}(z)]^2} - \frac{T_{n1}(z)[V_{n1}(z) - T_{n1}(z)g(z)]}{T_{n2}(z)T_{n0}(z) - [T_{n1}(z)]^2}\\
&= S_{n1}(z) - S_{n2}(z).
\end{aligned}$$

For the term $S_{n1}(z)$, using Lemma A.1 and Theorem 1, it is seen that

$$\begin{aligned}
S_{n1}(z) &= \frac{V_{n0}(z) - T_{n0}(z)g(z)}{T_{n0}(z) - [T_{n1}(z)]^2/T_{n2}(z)} \qquad (A.16)\\
&= \frac{1}{f_Z(z)nh_2}\sum_{i=1}^n K\left(\frac{Z_i - z}{h_2}\right)\left[\widetilde{Y}_i - g(Z_i) - \widetilde{\boldsymbol{X}}_i^{\mathrm{T}}\boldsymbol{\beta}_0\right]\\
&\quad + \frac{1}{f_Z(z)nh_2}\sum_{i=1}^n K\left(\frac{Z_i - z}{h_2}\right)\widetilde{\boldsymbol{X}}_i^{\mathrm{T}}\left(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}\right)\\
&\quad + \frac{\mu_2 h_2^2}{2}g''(z) + h_2^2\frac{f_Z'(z)g'(z)}{f_Z(z)} + o_P(h_2^2 + 1/\sqrt{nh_2})\\
&= \frac{1}{f_Z(z)nh_2}\sum_{i=1}^n K\left(\frac{Z_i - z}{h_2}\right)\left[\widetilde{Y}_i - g(Z_i) - \widetilde{\boldsymbol{X}}_i^{\mathrm{T}}\boldsymbol{\beta}_0\right] + \frac{\mu_2 h_2^2}{2}g''(z)\\
&\quad + h_2^2\frac{f_Z'(z)g'(z)}{f_Z(z)} + o_P(h_2^2 + 1/\sqrt{nh_2}).
\end{aligned}$$

Directly using Lemma A.1, similar to (A.16), we have

$$S_{n2}(z) = \frac{g'(z)f'(z)}{f_Z(z)}h_2^2\mu_2 + o_P(h_2^2 + 1/\sqrt{nh_2}). \qquad (A.17)$$

Together with (A.16) and (A.17), we have

$$\begin{aligned}
&\hat{g}(z) - g(z) - \frac{\mu_2 h_2^2}{2}g''(z) \qquad (A.18)\\
&= \frac{1}{f_Z(z)nh_2}\sum_{i=1}^n K\left(\frac{Z_i - z}{h_2}\right)\left[\widetilde{Y}_i - g(Z_i) - \widetilde{\boldsymbol{X}}_i^{\mathrm{T}}\boldsymbol{\beta}_0\right] + o_P(h_2^2 + 1/\sqrt{nh_2}).
\end{aligned}$$

The asymptotic result of Theorem 2 is directly obtained from (A.18), we have completed the proof of Theorem 2.

A.4. Proof of Theorem 3

We first consider the conditional mean calibration. For $1 \le r \le p$, let $\hat{\wp}^{[r]}_{n,i}(\boldsymbol{\beta}_0)$ be the r-component of $\hat{\wp}_{n,i}(\boldsymbol{\beta}_0)$. We decompose $\hat{\wp}^{[r]}_{n,i}(\boldsymbol{\beta}_0)$ into following terms:

$$\hat{\wp}^{[r]}_{n,i}(\boldsymbol{\beta}_0) = (Y_i - S_Y(Z_i) - [\boldsymbol{X}_i - S_{\boldsymbol{X}}(Z_i)]^{\mathrm{T}}\boldsymbol{\beta}_0)[X_{ri} - s_{X_r}(Z_i)] + \sum_{t=1}^{8} R^{[r]}_{n,it},$$

where,

$$\begin{aligned}
R^{[r]}_{n,i1} &= \{\hat{Y}_i - Y_i - [\hat{\boldsymbol{X}}_i - \boldsymbol{X}_i]^{\mathrm{T}}\boldsymbol{\beta}_0\}[X_{ri} - s_{X_r}(Z_i)],\\
R^{[r]}_{n,i2} &= \{\hat{Y}_i - Y_i - [\hat{\boldsymbol{X}}_i - \boldsymbol{X}_i]^{\mathrm{T}}\boldsymbol{\beta}_0\}[\hat{X}_{ri} - X_{ri}],\\
R^{[r]}_{n,i3} &= \{\hat{Y}_i - Y_i - [\hat{\boldsymbol{X}}_i - \boldsymbol{X}_i]^{\mathrm{T}}\boldsymbol{\beta}_0\}[s_{X_r}(Z_i) - \hat{s}_{X_r}(Z_i)],\\
R^{[r]}_{n,i4} &= \{Y_i - S_Y(Z_i) - [\boldsymbol{X}_i - S_{\boldsymbol{X}}(Z_i)]^{\mathrm{T}}\boldsymbol{\beta}_0\}[s_{X_r}(Z_i) - \hat{s}_{X_r}(Z_i)],\\
R^{[r]}_{n,i5} &= \{Y_i - S_Y(Z_i) - [\boldsymbol{X}_i - S_{\boldsymbol{X}}(Z_i)]^{\mathrm{T}}\boldsymbol{\beta}_0\}[\hat{X}_{ri} - X_{ri}],\\
R^{[r]}_{n,i6} &= \{S_Y(Z_i) - \hat{S}_Y(Z_i) - [S_{\boldsymbol{X}}(Z_i) - \hat{S}_{\boldsymbol{X}}(Z_i)]^{\mathrm{T}}\boldsymbol{\beta}_0\}[s_{X_r}(Z_i) - \hat{s}_{X_r}(Z_i)],\\
R^{[r]}_{n,i7} &= \{S_Y(Z_i) - \hat{S}_Y(Z_i) - [S_{\boldsymbol{X}}(Z_i) - \hat{S}_{\boldsymbol{X}}(Z_i)]^{\mathrm{T}}\boldsymbol{\beta}_0\}[\hat{X}_{ri} - X_{ri}],\\
R^{[r]}_{n,i8} &= \{S_Y(Z_i) - \hat{S}_Y(Z_i) - [S_{\boldsymbol{X}}(Z_i) - \hat{S}_{\boldsymbol{X}}(Z_i)]^{\mathrm{T}}\boldsymbol{\beta}_0\}[X_{ri} - s_{X_r}(Z_i)].
\end{aligned}$$

To prove Theorem 3, we need to show that

$$\max_{1\le i\le n} |\hat{\wp}^{[r]}_{n,it}| = o_P(n^{1/2}), \quad t = 1, \ldots, 8.$$

It is noted that for any sequence of $i.i.d$ random $\{V_i, 1 \le i \le n\}$ and $E[V^2] < \infty$, we have $\max_{1\le i\le n} \dfrac{|V_i|}{\sqrt{n}} \to 0, a.s.$. Then,

$$\max_{1\le i\le n} \left|(Y_i - S_Y(Z_i) - [\boldsymbol{X}_i - S_{\boldsymbol{X}}(Z_i)]^{\mathrm{T}}\boldsymbol{\beta}_0)[X_{ri} - s_{X_r}(Z_i)]\right| = o_P(n^{1/2}).$$

Next, for $R^{[r]}_{n,i1}$, according to the proof of Theorem 1 and Theorem 3 in [37],

$$\begin{aligned}
&\max_{1\le i\le n} |\{\hat{Y}_i - Y_i\}[X_{ri} - s_{X_r}(Z_i)]| \qquad (A.19)\\
&\le \max_{1\le i\le n} \left|\phi_M(U_i) - \hat{\phi}_M(U_i)\right| \left|\frac{Y_i[X_{ri} - s_{X_r}(Z_i)]}{\phi_M(U_i)}\right|\\
&+ \left|\phi_A(U_i) - \hat{\phi}_A(U_i)\right| \left|\frac{[X_{ri} - s_{X_r}(Z_i)]}{\phi_M(U_i)}\right| + O_P\left(h^4 + \frac{\log n}{nh}\right) O_P(n^{1/2})\\
&= o_P(n^{1/2}).
\end{aligned}$$

Similar to (A.19), we have

$$\max_{1\le i\le n} |R^{[r]}_{n,i1}| = o_P(n^{1/2}), \quad \max_{1\le i\le n} |R^{[r]}_{n,i5}| = o_P(n^{1/2}). \qquad (A.20)$$

For $R^{[r]}_{n,i2}$, similar to (A.19), we have

$$\begin{aligned}
&\max_{1\le i\le n}|\{\hat{Y}_i - Y_i\}[\hat{X}_{ri} - X_{ri}]| \qquad (A.21)\\
&\le \max_{1\le i\le n}\left|[\phi_M(U_i) - \hat{\phi}_M(U_i)][\psi_{M,r}(U_i) - \hat{\psi}_{M,r}(U_i)]\right|\\
&\qquad \times \max_{1\le i\le n}\left|\frac{Y_i X_{ri}}{\phi_M(U_i)\psi_{M,r}(U_i)}\right|\\
&+ \max_{1\le i\le n}\left|[\phi_M(U_i) - \hat{\phi}_M(U_i)][\psi_{A,r}(U_i) - \hat{\psi}_{A,r}(U_i)]\right|\\
&\qquad \times \max_{1\le i\le n}\left|\frac{Y_i}{\phi_M(U_i)\psi_{M,r}(U_i)}\right|\\
&+ \max_{1\le i\le n}\left|[\psi_{M,r}(U_i) - \hat{\psi}_{M,r}(U_i)][\phi_A(U_i) - \hat{\phi}_A(U_i)]\right|\\
&\qquad \times \max_{1\le i\le n}\left|\frac{X_{ri}}{\phi_M(U_i)\psi_{M,r}(U_i)}\right|\\
&+ \max_{1\le i\le n}\left|\frac{1}{\phi_M(U_i)\psi_{M,r}(U_i)}\right|\max_{1\le i\le n}\left|[\psi_{A,r}(U_i) - \hat{\psi}_{A,r}(U_i)][\phi_A(U_i) - \hat{\phi}_A(U_i)]\right|\\
&= O_P\left(h^4 + \frac{\log n}{nh} + n^{-1}\right)O_P(n^{1/2}) = o_P(n^{1/2}).
\end{aligned}$$

Thus, according to (A.21), we show that

$$\max_{1\le i\le n}|R^{[r]}_{n,i2}| = o_P(n^{1/2}). \qquad (A.22)$$

The conclusion of (A.1) in [15] entails that $\sup_{z\in\mathcal{Z}}|\hat{S}_Y(z) - S_Y(z)| = o_P(n^{-1/4})$, and $\sup_{z\in\mathcal{Z}}|\hat{s}_{X_r}(z) - s_{X_r}(z)| = o_P(n^{-1/4})$, $r = 1,\ldots,p$. Similar to (A.21)-(A.22), we have

$$\begin{aligned}
&\max_{1\le i\le n}|\{\hat{Y}_i - Y_i\}[s_{X_r}(Z_i) - \hat{s}_{X_r}(Z_i)]| \qquad (A.23)\\
&\le \max_{1\le i\le n}\left|\phi_M(U_i) - \hat{\phi}_M(U_i)\right|\left|\frac{Y_i}{\phi_M(U_i)}\right|\max_{1\le i\le n}|s_{X_r}(Z_i) - \hat{s}_{X_r}(Z_i)|\\
&+ \max_{1\le i\le n}\left|\phi_A(U_i) - \hat{\phi}_A(U_i)\right|\left|\frac{1}{\phi_M(U_i)}\right|\max_{1\le i\le n}|s_{X_r}(Z_i) - \hat{s}_{X_r}(Z_i)| = o_P(n^{1/2}).
\end{aligned}$$

From (A.23), we show that

$$\max_{1\le i\le n}|R^{[r]}_{n,i3}| = o_P(n^{1/2}). \qquad (A.24)$$

Similar to the proofs of $|R^{[r]}_{n,it}|$, $t = 1,2,3,5$, we have $\max_{1\le i\le n}|R^{[r]}_{n,it}| = o_P(n^{1/2})$ for $t = 4,6,7,8$. We omit the details. Followed the same argument in the proof

(2.14) in [23], we have $\hat\lambda = O_P(n^{1/2})$. Thus, $\max_{1\le i\le n} |\hat\lambda^{\mathrm{T}}\hat\wp_{n,i}(\boldsymbol{\beta}_0)| = o_P(1)$. Note that $\log(1+t) \approx t - \frac{1}{2}t^2$ for t sufficiently small, we have

$$\hat l(\boldsymbol{\beta}_0) = 2\sum_{i=1}^{n}\Big(\hat\lambda^{\mathrm{T}}\hat\wp_{n,i}(\boldsymbol{\beta}_0) - \frac{1}{2}\{\hat\lambda^{\mathrm{T}}\hat\wp_{n,i}(\boldsymbol{\beta}_0)\}^2\Big) + o_P(1). \tag{A.25}$$

Note that $\hat\lambda$ satisfies the following equation,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\hat\wp_{n,i}(\boldsymbol{\beta}_0)}{1+\hat\lambda^{\mathrm{T}}\hat\wp_{n,i}(\boldsymbol{\beta}_0)} = \mathbf{0}.$$

Furthermore,

$$\begin{aligned}\mathbf{0} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\hat\wp_{n,i}(\boldsymbol{\beta}_0)}{1+\hat\lambda^{\mathrm{T}}\hat\wp_{n,i}(\boldsymbol{\beta}_0)}\frac{1}{n}\sum_{i=1}^{n}\hat\wp_{n,i}(\boldsymbol{\beta}_0) - \frac{1}{n}\sum_{i=1}^{n}\hat\wp_{n,i}(\boldsymbol{\beta}_0)\hat\wp_{n,i}(\boldsymbol{\beta}_0)^{\mathrm{T}}\hat\lambda \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\frac{\hat\wp_{n,i}(\boldsymbol{\beta}_0)\{\hat\lambda^{\mathrm{T}}\hat\wp_{n,i}(\boldsymbol{\beta}_0)\}^2}{1+\hat\lambda^{\mathrm{T}}\hat\wp_{n,i}(\boldsymbol{\beta}_0)}. \end{aligned}\tag{A.26}$$

Above equation (A.26) and $\max_{1\le i\le n} |\hat\lambda^{\mathrm{T}}\hat\wp_{n,i}(\boldsymbol{\beta}_0)| = o_P(1)$ entail that

$$\hat\lambda = \Big(\frac{1}{n}\sum_{i=1}^{n}\hat\wp_{n,i}(\boldsymbol{\beta}_0)\hat\wp_{n,i}(\boldsymbol{\beta}_0)^{\mathrm{T}}\Big)^{-1}\frac{1}{n}\sum_{i=1}^{n}\hat\wp_{n,i}(\boldsymbol{\beta}_0) + o_P(n^{-1/2}). \tag{A.27}$$

Plugging the asymptotic expression (A.27) to (A.24), we have

$$\begin{aligned}\hat l(\boldsymbol{\beta}_0) &= n\Big(\frac{1}{n}\sum_{i=1}^{n}\hat\wp_{n,i}(\boldsymbol{\beta}_0)\Big)^{\mathrm{T}}\Big(\frac{1}{n}\sum_{i=1}^{n}\hat\wp_{n,i}(\boldsymbol{\beta}_0)\hat\wp_{n,i}(\boldsymbol{\beta}_0)^{\mathrm{T}}\Big)^{-1} \\ &\quad \times\Big(\frac{1}{n}\sum_{i=1}^{n}\hat\wp_{n,i}(\boldsymbol{\beta}_0)\Big) + o_P(1). \end{aligned}\tag{A.28}$$

According the proof Theorem 1, we can obtain that

$$\begin{aligned}\hat l(\boldsymbol{\beta}_0) &= n\Big(\frac{1}{n}\sum_{i=1}^{n}\kappa_{n,i}(\boldsymbol{\beta}_0)\Big)^{\mathrm{T}}\Big(\frac{1}{n}\sum_{i=1}^{n}\kappa_{n,i}(\boldsymbol{\beta}_0)\kappa_{n,i}(\boldsymbol{\beta}_0)^{\mathrm{T}}\Big)^{-1}\Big(\frac{1}{n}\sum_{i=1}^{n}\kappa_{n,i}(\boldsymbol{\beta}_0)\Big) \\ &\quad + o_P(1), \end{aligned}$$

where, $\kappa_{n,i}(\boldsymbol{\beta}_0) = (Y_i - S_Y(Z_i) - [\boldsymbol{X}_i - S_{\boldsymbol{X}}(Z_i)]^{\mathrm{T}}\boldsymbol{\beta}_0)[\boldsymbol{X}_i - S_{\boldsymbol{X}}(Z_i)]$ is independent and identically distributed p-dimensional random vector with zero mean. Theorem 3 for $\hat l(\boldsymbol{\beta}_0)$ follows from the central limit theorem and the Slutsky theorem.

A.5. Proof of Theorem 4 and Theorem 5

Step 1 Note that

$$\hat{\boldsymbol{\beta}}_{\mathrm{R}} = \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}^{\mathrm{T}}\left\{\boldsymbol{A}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}^{\mathrm{T}}\right\}^{-1}\left[\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{b}\right]. \tag{A.29}$$

Under the null hypothesis $\mathcal{H}_0$, we have $\boldsymbol{A}\boldsymbol{\beta}_0 = \boldsymbol{b}$. Using (A.29), it is seen that

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\mathrm{R}} - \boldsymbol{\beta}_0 &= \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) - \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}^{\mathrm{T}}\left\{\boldsymbol{A}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}^{\mathrm{T}}\right\}^{-1}\left[\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{A}\boldsymbol{\beta}_0\right] \\ &= \left[\boldsymbol{I}_p - \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}^{\mathrm{T}}\left\{\boldsymbol{A}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}^{\mathrm{T}}\right\}^{-1}\boldsymbol{A}\right]\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right).\end{aligned} \tag{A.30}$$

Together with (A.13) and (A.14), the equation (A.30) can be expressed as

$$\hat{\boldsymbol{\beta}}_{\mathrm{R}} - \boldsymbol{\beta}_0 = \left[\boldsymbol{I}_p - \boldsymbol{\Sigma}_0^{-1}\boldsymbol{A}^{\mathrm{T}}\left\{\boldsymbol{A}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{A}^{\mathrm{T}}\right\}^{-1}\boldsymbol{A}\right]\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) + o_P(n^{-1/2}). \tag{A.31}$$

Define $\boldsymbol{\Omega}_{\boldsymbol{A}} = \boldsymbol{I}_p - \boldsymbol{\Sigma}_0^{-1}\boldsymbol{A}^{\mathrm{T}}\left\{\boldsymbol{A}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{A}^{\mathrm{T}}\right\}^{-1}\boldsymbol{A}$, the expression (A.31) entails that

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{\mathrm{R}} - \boldsymbol{\beta}_0\right) \xrightarrow{L} N(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{A}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_{0\epsilon}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Omega}_{\boldsymbol{A}}^{\mathrm{T}} + \boldsymbol{\Omega}_{\boldsymbol{A}}\boldsymbol{\Sigma}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{\Omega}_{\boldsymbol{A}}^{\mathrm{T}}).$$

We have completed the proof of Theorem 4.

Step 2 Under the null hypothesis $\mathcal{H}_0 : \boldsymbol{A}\boldsymbol{\beta}_0 = \boldsymbol{b}$, using (A.14) and Theorem 1, we have

$$\begin{aligned}&\sqrt{n}\left(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{b}\right) = \sqrt{n}\boldsymbol{A}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \\ &\xrightarrow{L} N\left(\mathbf{0}, \boldsymbol{A}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_{0\epsilon}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{A}\boldsymbol{\Sigma}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{A}^{\mathrm{T}}\right).\end{aligned} \tag{A.32}$$

Similar to the analysis of (A.13), we have

$$\begin{aligned}&\boldsymbol{A}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Sigma}}_{\epsilon}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{A}\hat{\boldsymbol{\Sigma}}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{A}^{\mathrm{T}} \\ &\xrightarrow{P} \boldsymbol{A}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_{0\epsilon}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{A}\boldsymbol{\Sigma}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{A}^{\mathrm{T}}.\end{aligned} \tag{A.33}$$

The Slutsky theorem entails that

$$\begin{aligned}&\left[\boldsymbol{A}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Sigma}}_{\epsilon}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{A}\hat{\boldsymbol{\Sigma}}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{A}^{\mathrm{T}}\right]^{-1/2}\left[\sqrt{n}\left(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{b}\right)\right] \\ &\xrightarrow{L} N(\mathbf{0}, \boldsymbol{I}_k),\end{aligned} \tag{A.34}$$

where, $\boldsymbol{I}_k$ is a $k \times k$ dimensional identity matrix. Using (A.34), the continuous mapping theorem entails that

$$\mathcal{T}_n = n\left(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{b}\right)^{\mathrm{T}}\left[\boldsymbol{A}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Sigma}}_{\epsilon}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{A}\hat{\boldsymbol{\Sigma}}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{A}^{\mathrm{T}}\right]^{-1} \tag{A.35}$$

$$
\begin{aligned}
&\quad \times \left(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{b} \right) \\
\xrightarrow{L} &\quad \chi_k^2,
\end{aligned}
$$

where, χ_k^2 is the centered chi-squared distribution with degree of freedom k We have completed the proof of Theorem 5.

A.6. Proof of Theorem 6

Step 1 It is noted that $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{\beta}_0 - n^{-1/2}\boldsymbol{c}$ under the null hypothesis $\mathcal{H}_{1n}$. From (A.29), we have

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{\mathrm{R}} &= \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}^{\mathrm{T}} \left\{ \boldsymbol{A}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}^{\mathrm{T}} \right\}^{-1} \left[\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{b} \right] \\
&= \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}^{\mathrm{T}} \left\{ \boldsymbol{A}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}^{\mathrm{T}} \right\}^{-1} \left[\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{A}\boldsymbol{\beta}_0 + n^{-1/2}\boldsymbol{c} \right] \\
&= \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}^{\mathrm{T}} \left\{ \boldsymbol{A}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}^{\mathrm{T}} \right\}^{-1} \boldsymbol{A} \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right) \\
&\quad - n^{-1/2}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}^{\mathrm{T}} \left\{ \boldsymbol{A}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}^{\mathrm{T}} \right\}^{-1} \boldsymbol{c}.
\end{aligned} \tag{A.36}
$$

Using (A.30)-(A.31) and (A.36), we have

$$
\begin{aligned}
&\hat{\boldsymbol{\beta}}_{\mathrm{R}} - \boldsymbol{\beta}_0 \\
&= \boldsymbol{\Omega}_{\boldsymbol{A}} \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right) - n^{-1/2}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{A}^{\mathrm{T}} \left\{ \boldsymbol{A}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{A}^{\mathrm{T}} \right\}^{-1} \boldsymbol{c} + o_P(n^{-1/2}).
\end{aligned} \tag{A.37}
$$

According to Theorem 1, we have

$$
\begin{aligned}
&\sqrt{n} \left(\hat{\boldsymbol{\beta}}_{\mathrm{R}} - \boldsymbol{\beta}_0 \right) \\
&\xrightarrow{L} N(-\boldsymbol{\Sigma}_0^{-1}\boldsymbol{A}^{\mathrm{T}} \left\{ \boldsymbol{A}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{A}^{\mathrm{T}} \right\}^{-1} \boldsymbol{c}, \boldsymbol{\Omega}_{\boldsymbol{A}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_{\epsilon}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Omega}_{\boldsymbol{A}}^{\mathrm{T}} + \boldsymbol{\Omega}_{\boldsymbol{A}}\boldsymbol{\Sigma}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{\Omega}_{\boldsymbol{A}}^{\mathrm{T}}).
\end{aligned} \tag{A.38}
$$

Step 2 Under the local alternative hypothesis $\mathcal{H}_{1n} : \boldsymbol{A}\boldsymbol{\beta}_0 = \boldsymbol{b} + n^{-1/2}\boldsymbol{c}$, using Theorem 1, we have

$$
\begin{aligned}
&\sqrt{n} \left(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{b} \right) = \sqrt{n} \left(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{A}\boldsymbol{\beta}_0 + n^{-1/2}\boldsymbol{c} \right) \\
&= \sqrt{n}\boldsymbol{A} \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right) + \boldsymbol{c} \\
&\xrightarrow{L} N \left(\boldsymbol{c}, \boldsymbol{A}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_{0\epsilon}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{A}\boldsymbol{\Sigma}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{A}^{\mathrm{T}} \right).
\end{aligned} \tag{A.39}
$$

Using (A.33)-(A.34) and (A.39), we have

$$
\begin{aligned}
&\left[\boldsymbol{A}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Sigma}}_{\epsilon}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{A}\hat{\boldsymbol{\Sigma}}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{A}^{\mathrm{T}} \right]^{-1/2} \left[\sqrt{n} \left(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{b} \right) \right] \\
&\xrightarrow{L} N \left(\left[\boldsymbol{A}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_{0\epsilon}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{A}\boldsymbol{\Sigma}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{A}^{\mathrm{T}} \right]^{-1/2} \boldsymbol{c}, \boldsymbol{I}_k \right).
\end{aligned} \tag{A.40}
$$

Then, according to (A.40), the continuous mapping theorem entails that

$$\begin{aligned}\mathcal{T}_n &= n\left(\boldsymbol{A}\hat{\boldsymbol{\beta}}-\boldsymbol{b}\right)^{\mathrm{T}}\left[\boldsymbol{A}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Sigma}}_{\epsilon}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A}\hat{\boldsymbol{\Sigma}}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{A}^{\mathrm{T}}\right]^{-1} \\ &\quad\times\left(\boldsymbol{A}\hat{\boldsymbol{\beta}}-\boldsymbol{b}\right) \\ &\xrightarrow{L} \chi_k^2(\pi_{\boldsymbol{c}}),\end{aligned} \tag{A.41}$$

where, $\chi_k^2(\pi_{\boldsymbol{c}})$ is the noncentral chi-squared distribution with degree of freedom k, and $\pi_{\boldsymbol{c}}$ is the noncentrality parameter, defined as

$$\pi_{\boldsymbol{c}}=\boldsymbol{c}^{\mathrm{T}}\left[\boldsymbol{A}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_{0\epsilon}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A}\boldsymbol{\Sigma}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{A}^{\mathrm{T}}\right]^{-1}\boldsymbol{c}.$$

We have completed the proof of Theorem 6.

A.7. Proof of Theorem 7

Step 1 In this step, we establish the asymptotic expressions of minimizer estimator $\hat{\boldsymbol{\beta}}_P$. Define

$$\mathcal{L}_P(\boldsymbol{\beta})=\frac{1}{2}\sum_{i=1}^{n}\left\{\hat{Y}_i-\hat{S}_Y(Z_i)-\left[\hat{\boldsymbol{X}}_i-\hat{S}_{\boldsymbol{X}}(Z_i)\right]^{\mathrm{T}}\boldsymbol{\beta}\right\}^2+n\sum_{s=1}^{p}p_{\zeta_s}(|\beta_s|).$$

Let $\kappa_n = n^{-1/2}+a_n^*$ with $a_n^*=\max_{1\le j\le p}\{p'_{\zeta_j}(|\beta_{0j}|),\beta_{0j}\neq 0\}$, and $\boldsymbol{s}=(s_1,\ldots,s_p)^{\mathrm{T}}$ with $\|\boldsymbol{s}\|=C_0$. Moreover, we define $\boldsymbol{\beta}(n)=\boldsymbol{\beta}_0+\kappa_n\boldsymbol{s}$ and

$$\begin{aligned}\mathcal{F}_{n,1} &= \frac{1}{2}\sum_{i=1}^{n}\left\{\hat{Y}_i-\hat{S}_Y(Z_i)-\left[\hat{\boldsymbol{X}}_i-\hat{S}_{\boldsymbol{X}}(Z_i)\right]^{\mathrm{T}}\boldsymbol{\beta}(n)\right\}^2 \\ &\quad -\frac{1}{2}\sum_{i=1}^{n}\left\{\hat{Y}_i-\hat{S}_Y(Z_i)-\left[\hat{\boldsymbol{X}}_i-\hat{S}_{\boldsymbol{X}}(Z_i)\right]^{\mathrm{T}}\boldsymbol{\beta}_0\right\}^2, \\ \mathcal{F}_{n,2} &= -n\sum_{j=1}^{p_0}\{p_{\zeta_j}(|\beta_{0j}+\kappa_n s_j|)-p_{\zeta_j}(|\beta_{0j}|)\}.\end{aligned}$$

Using (A.13)-(A.14), we have

$$\begin{aligned}\mathcal{F}_{n,1} &= \frac{1}{2}\kappa_n^2\sum_{i=1}^{n}\boldsymbol{s}^{\mathrm{T}}\left[\hat{\boldsymbol{X}}_i-\hat{S}_{\boldsymbol{X}}(Z_i)\right]^{\otimes 2}\boldsymbol{s} \\ &\quad -\kappa_n\sum_{i=1}^{n}\boldsymbol{s}^{\mathrm{T}}\left[\hat{\boldsymbol{X}}_i-\hat{S}_{\boldsymbol{X}}(Z_i)\right]^{\mathrm{T}}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right) \\ &= \frac{n}{2}\kappa_n^2\boldsymbol{s}^{\mathrm{T}}\boldsymbol{\Sigma}_0\boldsymbol{s}-n\kappa_n\boldsymbol{s}^{\mathrm{T}}\boldsymbol{\Sigma}_0\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)+o_P(n\kappa_n^2C_0^2)+o_P(n^{1/2}\kappa_nC_0).\end{aligned} \tag{A.42}$$

As $a_n^*=O_P(n^{-1/2})$, we have $\kappa_n=O_P(n^{-1/2})$ and the asymptotic expression (A.42) entails that the first argument of $\mathcal{D}_{n,1}$ is positive and dominated by

$\frac{n}{2}\kappa_n^2 C_0^2$ in probability and the second argument of is dominated by $C_0 O_P(1)$. Taylor expansion and Cauchy-Schawz inequality entail that

$$|\mathcal{F}_{n,2}| \le n\sqrt{p_0}\kappa_n a_n^* \|\boldsymbol{s}\| + n\kappa_n^2 a_n^{**}\|\boldsymbol{s}\|^2 \le C_0 n\kappa_n^2\{\sqrt{p_0} + a_n^{**}C_0\},$$

where $a_n^{**} = \max_{1\le j\le p}\{p''_{\zeta_j}(|\beta_{0j}|), \beta_{0j} \ne 0\}$. Furthermore, $\mathcal{D}_{n,2}$ is bounded by $n\kappa_n^2 C_0^2$ in probability. Thus, as a_n^{**}, b_n^{**} tend to 0 and C_0 sufficiently large, $\mathcal{D}_{n,1}$ dominates $\mathcal{D}_{n,2}$. As a consequence, for any given $\delta > 0$, there exists a large constant C_0 such that

$$P\left\{\inf_{\mathcal{S}} \mathcal{L}_P(\boldsymbol{\beta}(n)) > \mathcal{L}_P(\boldsymbol{\beta}_0)\right\} \ge 1 - \delta,$$

where $\mathcal{S} = \{\boldsymbol{s} : \|\boldsymbol{s}\| = C_0\}$. We conclude that $\hat{\boldsymbol{b}}_P$ is $O_P(n^{-1/2})$.

Step 2. Let $\boldsymbol{\beta}_{\mathbf{1}}^*$ satisfies $\|\boldsymbol{\beta}_{\mathbf{1}}^* - \boldsymbol{\beta}_{\mathbf{01}}\| = O_P(n^{-1/2})$. Similar to the proof of Lemma 1 in [3], we can show that

$$\mathcal{L}_P\left((\boldsymbol{\beta}_{\mathbf{1}}^{*\mathrm{T}}, \mathbf{0}^{\mathrm{T}})^{\mathrm{T}}\right) = \min_{\mathcal{L}^*} \mathcal{L}_P\left((\boldsymbol{\beta}_{\mathbf{1}}^{*\mathrm{T}}, \boldsymbol{\beta}_{\mathbf{2}}^{*\mathrm{T}})^{\mathrm{T}}\right), \tag{A.43}$$

where, $\mathcal{L}^* = \{\|\boldsymbol{\beta}_{\mathbf{2}}^*\| \le L^* n^{-1/2}\}$ and L^* is a positive constant. We omit the details for the proof in this step.

Step 3. Denote that $\hat{\boldsymbol{\beta}}_{P,\mathbf{1}}$ is the penalized least squares estimator of $\boldsymbol{\beta}_{0,\mathbf{1}}$. In addition, we denote that $\hat{\boldsymbol{X}}_{i,\mathbf{1}}$ and $\hat{S}_{\boldsymbol{X},\mathbf{1}}(Z_i)$ consist of the first p_0 components of $\hat{\boldsymbol{X}}_i$ and $\hat{S}_{\boldsymbol{X}}(Z_i)$, respectively. Define $\mathcal{L}_P^*(\boldsymbol{\beta}_{\mathbf{1}}) = \mathcal{L}_P\left((\boldsymbol{\beta}_{\mathbf{1}}^{\mathrm{T}}, \mathbf{0}^{\mathrm{T}})^{\mathrm{T}}\right)$. Taylor expansion entails that

$$\begin{aligned}
\mathbf{0} &= \left.\frac{\partial \mathcal{L}_P^*(\boldsymbol{\beta}_{\mathbf{1}})}{\partial \boldsymbol{\beta}_{\mathbf{1}}}\right|_{\boldsymbol{\beta}_{\mathbf{1}} = \hat{\boldsymbol{\beta}}_{P,\mathbf{1}}} \qquad \text{(A.44)}\\
&= -\sum_{i=1}^n \left[\hat{\boldsymbol{X}}_{i,\mathbf{1}} - \hat{S}_{\boldsymbol{X},\mathbf{1}}(Z_i)\right]\left\{\hat{Y}_i - \left[\hat{\boldsymbol{X}}_{i,\mathbf{1}} - \hat{S}_{\boldsymbol{X},\mathbf{1}}(Z_i)\right]^{\mathrm{T}}\boldsymbol{\beta}_{0,\mathbf{1}}\right\}\\
&\quad + n\mathcal{R}_{\boldsymbol{\zeta}_{\mathbf{1}}} + \left(\sum_{i=1}^n \left[\hat{\boldsymbol{X}}_{i,\mathbf{1}} - \hat{S}_{\boldsymbol{X},\mathbf{1}}(Z_i)\right]^{\otimes 2} + n\boldsymbol{\Sigma}_{\boldsymbol{\zeta}_{\mathbf{1}}}\right)\left(\hat{\boldsymbol{\beta}}_{P,\mathbf{1}} - \boldsymbol{\beta}_{0,\mathbf{1}}\right) + O_P(\delta_n),
\end{aligned}$$

where, $\delta_n = n\|\hat{\boldsymbol{\beta}}_{P,\mathbf{1}} - \boldsymbol{\beta}_{\mathbf{01}}\|^2$. Similar to (A.14), we have that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^n \left[\hat{\boldsymbol{X}}_{i,\mathbf{1}} - \hat{S}_{\boldsymbol{X},\mathbf{1}}(Z_i)\right]\left\{\hat{Y}_i - \left[\hat{\boldsymbol{X}}_{i,\mathbf{1}} - \hat{S}_{\boldsymbol{X},\mathbf{1}}(Z_i)\right]^{\mathrm{T}}\boldsymbol{\beta}_{0,\mathbf{1}}\right\} \qquad \text{(A.45)}\\
&\xrightarrow{\mathcal{L}} N\left(\mathbf{0}_{p_0}, \boldsymbol{\Sigma}_{0\epsilon,\mathbf{1}} + \boldsymbol{\Sigma}_{0,\mathbf{1}}\boldsymbol{\Sigma}_{\phi_M,\boldsymbol{\psi}_{M,\mathbf{1}}}\boldsymbol{\Sigma}_{0,\mathbf{1}}\right),
\end{aligned}$$

where, $\boldsymbol{\Sigma}_{0\epsilon,\mathbf{1}}$, $\boldsymbol{\Sigma}_{0,\mathbf{1}}$ and $\boldsymbol{\Sigma}_{\phi_M,\boldsymbol{\psi}_{M,\mathbf{1}}}$ are defined in Theorem 7. The asymptotic expressions (A.44) and (A.45) entail that

$$\sqrt{n}\left(\boldsymbol{\Sigma}_{0,\mathbf{1}} + \boldsymbol{\Sigma}_{\boldsymbol{\zeta}_{\mathbf{1}}}\right)\left\{\left(\hat{\boldsymbol{\beta}}_{P,\mathbf{1}} - \boldsymbol{\beta}_{0,\mathbf{1}}\right) + \left(\boldsymbol{\Sigma}_{0,\mathbf{1}} + \boldsymbol{\Sigma}_{\boldsymbol{\zeta}_{\mathbf{1}}}\right)^{-1}\mathcal{R}_{\boldsymbol{\zeta}_{\mathbf{1}}}\right\}$$

$$\xrightarrow{\mathcal{L}} N\left(\mathbf{0}_{p_0}, \boldsymbol{\Sigma}_{0\epsilon,\mathbf{1}} + \boldsymbol{\Sigma}_{0,\mathbf{1}}\boldsymbol{\Sigma}_{\phi_M,\boldsymbol{\psi}_{M,\mathbf{1}}}\boldsymbol{\Sigma}_{0,\mathbf{1}}\right).$$

We have completed the proof of Theorem 7.

A.8. Proof of Theorem 8

The proof of asymptotic normality of $\hat{\boldsymbol{\beta}}_V$ can be directly obtained from the proof of Theorem 1 by treating $\phi_A(U) \equiv 0$, $\psi_{A,r}(U) \equiv 0$, $r = 1, \ldots, p$. We omit the details.

For the estimator $\hat{\boldsymbol{\beta}}_C$, similar to (A.1), we have

$$\begin{aligned}
&\hat{\boldsymbol{\beta}}_C - \boldsymbol{\beta}_0 \\
&= \left\{\frac{1}{n}\sum_{i=1}^n \left[\hat{\boldsymbol{X}}_{C,i} - \hat{S}_{\boldsymbol{X}}(Z_i)\right]^{\otimes 2}\right\}^{-1} \frac{1}{n}\sum_{i=1}^n \left\{\hat{\boldsymbol{X}}_{C,i} - \hat{S}_{\boldsymbol{X}}(Z_i)\right\} \\
&\qquad \times \left\{\hat{Y}_{C,i} - \hat{S}_Y(Z_i) - \hat{\boldsymbol{X}}_{C,i}^{\mathrm{T}}\boldsymbol{\beta}_0 + \hat{S}_{\boldsymbol{X}}^{\mathrm{T}}(Z_i)\boldsymbol{\beta}_0\right\} \\
&= \left\{\frac{1}{n}\sum_{i=1}^n \left[\hat{\boldsymbol{X}}_{C,i} - \hat{S}_{\boldsymbol{X}}(Z_i)\right]^{\otimes 2}\right\}^{-1} \left[\mathbb{K}_{n1} + \mathbb{K}_{n2} + \mathbb{K}_{n3}\right],
\end{aligned} \tag{A.46}$$

where,

$$\mathbb{K}_{n1} = \frac{1}{n}\sum_{i=1}^n \left\{\hat{\boldsymbol{X}}_{C,i} - \hat{S}_{\boldsymbol{X}}(Z_i)\right\}\epsilon_i, \tag{A.47}$$

$$\mathbb{K}_{n2} = \frac{1}{n}\sum_{i=1}^n \left\{\hat{\boldsymbol{X}}_{C,i} - \hat{S}_{\boldsymbol{X}}(Z_i)\right\}\left\{\hat{Y}_{C,i} - Y_i - (\hat{\boldsymbol{X}}_{C,i} - \boldsymbol{X}_i)^{\mathrm{T}}\boldsymbol{\beta}_0\right\}, \tag{A.48}$$

$$\begin{aligned}
\mathbb{K}_{n3} &= \frac{1}{n}\sum_{i=1}^n \left\{\hat{\boldsymbol{X}}_{C,i} - \hat{S}_{\boldsymbol{X}}(Z_i)\right\} \\
&\qquad \times \left\{S_Y(Z_i) - \hat{S}_Y(Z_i) - (S_{\boldsymbol{X}}(Z_i) - \hat{S}_{\boldsymbol{X}}(Z_i))^{\mathrm{T}}\boldsymbol{\beta}_0\right\}.
\end{aligned} \tag{A.49}$$

Step 8.1 For the expression $\mathbb{K}_{n1}$, we have

$$\begin{aligned}
\mathbb{K}_{n1} &= \frac{1}{n}\sum_{i=1}^n \left\{\hat{\boldsymbol{X}}_{C,i} - \boldsymbol{X}_i\right\}\epsilon_i + \frac{1}{n}\sum_{i=1}^n \left\{\boldsymbol{X}_i - S_{\boldsymbol{X}}(Z_i)\right\}\epsilon_i \\
&\quad \frac{1}{n}\sum_{i=1}^n \left\{S_{\boldsymbol{X}}(Z_i) - \hat{S}_{\boldsymbol{X}}(Z_i)\right\}\epsilon_i \\
&\stackrel{\text{def}}{=} \mathbb{K}_{n1}[1] + \mathbb{K}_{n1}[2] + \mathbb{K}_{n1}[3].
\end{aligned} \tag{A.50}$$

Recalling $\epsilon_i = Y_i - \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0 - g(Z_i)$ and $E(\epsilon_i|\boldsymbol{X}_i, Z_i) = 0$. Using Lemma A.3, we have

$$\frac{1}{n}\sum_{i=1}^n \left\{\hat{X}_{ri} - X_{ri}\right\}\epsilon_i \tag{A.51}$$

$$= \sum_{i=1}^{n} \left(|\widetilde{X}_{ri}| - |X_{ri}|\right) \frac{E\big[X_r \epsilon\big]}{E(|X_r|)} + o_P(n^{-1/2}) = o_P(n^{-1/2}).$$

Based on (A.51), we have $\mathbb{K}_{n1}[1] = o_P(n^{-1/2})$. From the proof of Theorem 1, $\mathbb{K}_{n1}[3] = \mathbb{D}_{n1}[3] = o_P(n^{-1/2})$. Thus, we obtain that

$$\mathbb{K}_{n1} = \frac{1}{n} \sum_{i=1}^{n} \{\boldsymbol{X}_i - S_{\boldsymbol{X}}(Z_i)\} \epsilon_i + o_P(n^{-1/2}). \tag{A.52}$$

Step 8.2 For the argument $\mathbb{K}_{n2}$, we have

$$\begin{aligned} \mathbb{K}_{n2} \quad = \quad & \frac{1}{n} \sum_{i=1}^{n} \left\{ \hat{\boldsymbol{X}}_{C,i} - \boldsymbol{X}_i \right\} \left\{ \hat{Y}_{C,i} - Y_i - (\hat{\boldsymbol{X}}_{C,i} - \boldsymbol{X}_i)^{\mathrm{T}} \boldsymbol{\beta}_0 \right\} \qquad \text{(A.53)} \\ & + \frac{1}{n} \sum_{i=1}^{n} \{\boldsymbol{X}_i - S_{\boldsymbol{X}}(Z_i)\} \left\{ \hat{Y}_{C,i} - Y_i - (\hat{\boldsymbol{X}}_{C,i} - \boldsymbol{X}_i)^{\mathrm{T}} \boldsymbol{\beta}_0 \right\} \\ & + \frac{1}{n} \sum_{i=1}^{n} \left\{ S_{\boldsymbol{X}}(Z_i) - \hat{S}_{\boldsymbol{X}}(\hat{Z}_i) \right\} \left\{ \hat{Y}_{C,i} - Y_i - (\hat{\boldsymbol{X}}_{C,i} - \boldsymbol{X}_i)^{\mathrm{T}} \boldsymbol{\beta}_0 \right\} \\ \stackrel{\text{def}}{=} \quad & \mathbb{K}_{n2}[1] + \mathbb{K}_{n2}[2] + \mathbb{K}_{n2}[3]. \end{aligned}$$

According the proof of Theorem 1 in [41], we have $\mathbb{K}_{n2}[1] = o_P(n^{-1/2})$.

For $\mathbb{K}_{n2}[2]$, using $E[\boldsymbol{X} - S_{\boldsymbol{X}}(Z)|Z] = 0$, Lemma A.3 entails that

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^{n} \{\boldsymbol{X}_i - S_{\boldsymbol{X}}(Z_i)\} \left\{ \hat{Y}_{C,i} - Y_i \right\} \qquad \text{(A.54)} \\ & = \frac{1}{n} \sum_{i=1}^{n} \left(|\widetilde{Y}_i| - |Y_i|\right) \frac{E\big[Y\{\boldsymbol{X} - S_{\boldsymbol{X}}(Z)\}\big]}{E(|Y|)} + o_P(n^{-1/2}) \\ & = \frac{1}{n} \sum_{i=1}^{n} \frac{|\widetilde{Y}_i| - |Y_i|}{E(|Y|)} \boldsymbol{\Sigma}_0 \boldsymbol{\beta}_0 + o_P(n^{-1/2}) \\ & = \frac{1}{n} \sum_{i=1}^{n} \frac{|Y_i|}{E(|Y|)} [\phi_M(U_i) - 1] \boldsymbol{\Sigma}_0 \boldsymbol{\beta}_0 + o_P(n^{-1/2}). \end{aligned}$$

Similarly, we have

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^{n} \{\boldsymbol{X}_i - S_{\boldsymbol{X}}(Z_i)\} (\hat{\boldsymbol{X}}_{C,i} - \boldsymbol{X}_i)^{\mathrm{T}} \boldsymbol{\beta}_0 \qquad \text{(A.55)} \\ & = \sum_{r=1}^{p} \left\{ \frac{1}{n} \sum_{i=1}^{n} \{\boldsymbol{X}_i - S_{\boldsymbol{X}}(Z_i)\} (\hat{X}_{C,ri} - X_{ri}) \beta_{0r} \right\} \\ & = \sum_{r=1}^{p} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(|\widetilde{X}_{ri}| - |X_{ri}|\right) \frac{E\big[X_r\{\boldsymbol{X} - S_{\boldsymbol{X}}(Z)\}\big]}{E(|X_r|)} \beta_{0r} \right\} + o_P(n^{-1/2}) \end{aligned}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{r=1}^{p}\frac{|\widetilde{X}_{ri}|-|X_{ri}|}{E(|X_r|)}\mathbf{\Sigma}_0 e_r e_r^{\mathrm{T}}\boldsymbol{\beta}_0 + o_P(n^{-1/2})$$
$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{r=1}^{p}\frac{|X_{ri}|}{E(|X_r|)}[\psi_{M,r}(U_i)-1]\mathbf{\Sigma}_0 e_r e_r^{\mathrm{T}}\boldsymbol{\beta}_0 + o_P(n^{-1/2}).$$

Together with (A.54) and (A.55), we have

$$\begin{aligned}\mathbb{K}_{n2}[2] &= \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i|}{E(|Y|)}[\phi_M(U_i)-1]\mathbf{\Sigma}_0\boldsymbol{\beta}_0 \\ &\quad -\frac{1}{n}\sum_{i=1}^{n}\sum_{r=1}^{p}\frac{|X_{ri}|}{E(|X_r|)}[\psi_{M,r}(U_i)-1]\mathbf{\Sigma}_0 e_r e_r^{\mathrm{T}}\boldsymbol{\beta}_0 + o_P(n^{-1/2}).\end{aligned} \tag{A.56}$$

The conclusion of (A.1) in [15] entails that $\sup_{z\in\mathcal{Z}}|\hat{S}_Y(z)-S_Y(z)| = o_P(n^{-1/4})$, and $\sup_{z\in\mathcal{Z}}|\hat{s}_{X_r}(z)-s_{X_r}(z)| = o_P(n^{-1/4})$, $r=1,\ldots,p$. According to the proof of Theorem 1 in [38] and the proof of Theorem 1 in [41], we can show that $\mathbb{K}_{n2}[3] = o_P(n^{-1/2})$, and also $\mathbb{K}_{n3} = o_P(n^{-1/2})$. Moreover,

$$\frac{1}{n}\sum_{i=1}^{n}\left[\hat{\boldsymbol{X}}_{C,i}-\hat{S}_{\boldsymbol{X}}(Z_i)\right]^{\otimes 2} \xrightarrow{P} \mathbf{\Sigma}_0. \tag{A.57}$$

Thus, together with (A.52), (A.56) and (A.57), we obtain that

$$\begin{aligned}\hat{\boldsymbol{\beta}}_C-\boldsymbol{\beta}_0 &= \mathbf{\Sigma}_0^{-1}\left(\mathbb{K}_{n1}+\mathbb{K}_{n2}+\mathbb{K}_{n3}\right)+o_P(n^{-1/2}) \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{\Sigma}_0^{-1}\left\{\boldsymbol{X}_i-S_{\boldsymbol{X}}(Z_i)\right\}\epsilon_i+\frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i|}{E(|Y|)}[\phi_M(U_i)-1]\boldsymbol{\beta}_0 \\ &\quad -\frac{1}{n}\sum_{i=1}^{n}\sum_{r=1}^{p}\frac{|X_{ri}|}{E(|X_r|)}[\psi_{M,r}(U_i)-1]e_r e_r^{\mathrm{T}}\boldsymbol{\beta}_0+o_P(n^{-1/2}).\end{aligned} \tag{A.58}$$

We have completed the proof of Theorem 8.

A.9. Proof of Theorem 9

The proof of the asymptotic results of $\mathcal{T}_{V,n}$ are similar to the proof of Theorem 5 and Theorem 6, we omit the details.

For the test statistic $\mathcal{T}_{C,n}$, under the null hypothesis $\mathcal{H}_0: \boldsymbol{A}\boldsymbol{\beta}_0 = \boldsymbol{b}$, using (A.58) and Theorem 1, we have

$$\begin{aligned}&\sqrt{n}\left(\boldsymbol{A}\hat{\boldsymbol{\beta}}_C-\boldsymbol{b}\right) = \sqrt{n}\boldsymbol{A}\left(\hat{\boldsymbol{\beta}}_C-\boldsymbol{\beta}_0\right) \\ &\xrightarrow{L} N\left(\mathbf{0}, \boldsymbol{A}\mathbf{\Sigma}_0^{-1}\mathbf{\Sigma}_{0\epsilon}\mathbf{\Sigma}_0^{-1}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A}\mathbf{\Omega}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{A}^{\mathrm{T}}\right).\end{aligned} \tag{A.59}$$

Similar to the analysis of (A.57), we have

$$\begin{aligned}&\boldsymbol{A}\hat{\boldsymbol{\Sigma}}_C^{-1}\hat{\boldsymbol{\Sigma}}_{C,\epsilon}\hat{\boldsymbol{\Sigma}}_C^{-1}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A}\hat{\boldsymbol{\Omega}}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{A}^{\mathrm{T}}\\&\xrightarrow{P}\boldsymbol{A}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_{0\epsilon}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A}\boldsymbol{\Omega}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{A}^{\mathrm{T}}.\end{aligned}\tag{A.60}$$

The Slutsky theorem entails that

$$\begin{aligned}&\left[\boldsymbol{A}\hat{\boldsymbol{\Sigma}}_C^{-1}\hat{\boldsymbol{\Sigma}}_{C,\epsilon}\hat{\boldsymbol{\Sigma}}_C^{-1}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A}\hat{\boldsymbol{\Omega}}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{A}^{\mathrm{T}}\right]^{-1/2}\left[\sqrt{n}\left(\boldsymbol{A}\hat{\boldsymbol{\beta}}_C-\boldsymbol{b}\right)\right]\\&\xrightarrow{L}N(\mathbf{0},\boldsymbol{I}_k).\end{aligned}\tag{A.61}$$

Using (A.61), the continuous mapping theorem entails that

$$\begin{aligned}\mathcal{T}_{C,n}&=n\left(\boldsymbol{A}\hat{\boldsymbol{\beta}}_C-\boldsymbol{b}\right)^{\mathrm{T}}\left[\boldsymbol{A}\hat{\boldsymbol{\Sigma}}_C^{-1}\hat{\boldsymbol{\Sigma}}_{C,\epsilon}\hat{\boldsymbol{\Sigma}}_C^{-1}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A}\hat{\boldsymbol{\Omega}}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{A}^{\mathrm{T}}\right]^{-1}\\&\quad\times\left(\boldsymbol{A}\hat{\boldsymbol{\beta}}_C-\boldsymbol{b}\right)\\&\xrightarrow{L}\chi_k^2.\end{aligned}\tag{A.62}$$

Under the local alternative hypothesis $\mathcal{H}_{1n}$, it is noted that $\boldsymbol{b}=\boldsymbol{A}\boldsymbol{\beta}_0-n^{-1/2}\boldsymbol{c}$ under the null hypothesis, we have

$$\begin{aligned}&\sqrt{n}\left(\boldsymbol{A}\hat{\boldsymbol{\beta}}_C-\boldsymbol{b}\right)=\sqrt{n}\left(\boldsymbol{A}\hat{\boldsymbol{\beta}}_C-\boldsymbol{A}\boldsymbol{\beta}_0+n^{-1/2}\boldsymbol{c}\right)\\&=\sqrt{n}\boldsymbol{A}\left(\hat{\boldsymbol{\beta}}_C-\boldsymbol{\beta}_0\right)+\boldsymbol{c}\\&\xrightarrow{L}N\left(\boldsymbol{c},\boldsymbol{A}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_{0\epsilon}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A}\boldsymbol{\Omega}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{A}^{\mathrm{T}}\right).\end{aligned}\tag{A.63}$$

Using (A.60)-(A.61) and (A.63), we have

$$\begin{aligned}&\left[\boldsymbol{A}\hat{\boldsymbol{\Sigma}}_C^{-1}\hat{\boldsymbol{\Sigma}}_{C,\epsilon}\hat{\boldsymbol{\Sigma}}_C^{-1}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A}\hat{\boldsymbol{\Omega}}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{A}^{\mathrm{T}}\right]^{-1/2}\left[\sqrt{n}\left(\boldsymbol{A}\hat{\boldsymbol{\beta}}_C-\boldsymbol{b}\right)\right]\\&\xrightarrow{L}N\left(\left[\boldsymbol{A}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_{0\epsilon}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A}\boldsymbol{\Omega}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{A}^{\mathrm{T}}\right]^{-1/2}\boldsymbol{c},\boldsymbol{I}_k\right).\end{aligned}\tag{A.64}$$

Then, according to (A.64), the continuous mapping theorem entails that

$$\begin{aligned}\mathcal{T}_{C,n}&=n\left(\boldsymbol{A}\hat{\boldsymbol{\beta}}_C-\boldsymbol{b}\right)^{\mathrm{T}}\left[\boldsymbol{A}\hat{\boldsymbol{\Sigma}}_C^{-1}\hat{\boldsymbol{\Sigma}}_{C,\epsilon}\hat{\boldsymbol{\Sigma}}_C^{-1}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A}\hat{\boldsymbol{\Omega}}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{A}^{\mathrm{T}}\right]^{-1}\\&\quad\times\left(\boldsymbol{A}\hat{\boldsymbol{\beta}}_C-\boldsymbol{b}\right)\\&\xrightarrow{L}\chi_k^2(\pi_{C,\boldsymbol{c}}),\end{aligned}\tag{A.65}$$

where, $\chi_k^2(\pi_{C,\boldsymbol{c}})$ is the noncentral chi-squared distribution with degree of freedom k, and $\pi_{C,\boldsymbol{c}}$ is the noncentrality parameter:

$$\pi_{C,\boldsymbol{c}}=\boldsymbol{c}^{\mathrm{T}}\left[\boldsymbol{A}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_{0\epsilon}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A}\boldsymbol{\Omega}_{\phi_M,\boldsymbol{\psi}_M}\boldsymbol{A}^{\mathrm{T}}\right]^{-1}\boldsymbol{c}.$$

We have completed the proof of Theorem 9.

A.10. Proof of Theorem 10

The proof of Theorem 10 (1) is the same as Step 1 and Step 2 in the proof of Theorem 7, we omit the details. Denote that $\hat{\boldsymbol{\beta}}_{P,C,\mathbf{1}}$ is the penalized least squares estimator of $\boldsymbol{\beta}_{0,\mathbf{1}}$. In addition, we denote that $\hat{\boldsymbol{X}}_{C,i,\mathbf{1}}$ and $\hat{S}_{\boldsymbol{X},\mathbf{1}}(Z_i)$ consist of the first p_0 components of $\hat{\boldsymbol{X}}_{C,i}$ and $\hat{S}_{\boldsymbol{X}}(Z_i)$, respectively. Define $\mathcal{L}_P^*(\boldsymbol{\beta}_{\mathbf{1}}) = \mathcal{L}_P\left((\boldsymbol{\beta}_{\mathbf{1}}^{\mathrm{T}}, \mathbf{0}^{\mathrm{T}})^{\mathrm{T}}\right)$. Taylor expansion entails that

$$\begin{aligned} \mathbf{0} &= \left.\frac{\partial \mathcal{L}_P^*(\boldsymbol{\beta}_{\mathbf{1}})}{\partial \boldsymbol{\beta}_{\mathbf{1}}}\right|_{\boldsymbol{\beta}_{\mathbf{1}}=\hat{\boldsymbol{\beta}}_{P,C,\mathbf{1}}} \qquad (\text{A.66}) \\ &= -\sum_{i=1}^{n}\left[\hat{\boldsymbol{X}}_{C,i,\mathbf{1}} - \hat{S}_{\boldsymbol{X},\mathbf{1}}(Z_i)\right]\left\{\hat{Y}_{C,i} - \left[\hat{\boldsymbol{X}}_{C,i,\mathbf{1}} - \hat{S}_{\boldsymbol{X},\mathbf{1}}(Z_i)\right]^{\mathrm{T}}\boldsymbol{\beta}_{0,\mathbf{1}}\right\} \\ &\quad + n\mathcal{R}_{\boldsymbol{\zeta}_{\mathbf{1}}} + \left(\sum_{i=1}^{n}\left[\hat{\boldsymbol{X}}_{C,i,\mathbf{1}} - \hat{S}_{\boldsymbol{X},\mathbf{1}}(Z_i)\right]^{\otimes 2} + n\boldsymbol{\Sigma}_{\boldsymbol{\zeta}_{\mathbf{1}}}\right)\left(\hat{\boldsymbol{\beta}}_{P,C,\mathbf{1}} - \boldsymbol{\beta}_{0,\mathbf{1}}\right) \\ &\quad + O_P(\delta_n), \end{aligned}$$

where $\delta_n = n\|\hat{\boldsymbol{\beta}}_{P,C,\mathbf{1}} - \boldsymbol{\beta}_{0\mathbf{1}}\|^2$. Similar to (A.14), we have that

$$\begin{aligned} &\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\hat{\boldsymbol{X}}_{C,i,\mathbf{1}} - \hat{S}_{\boldsymbol{X},\mathbf{1}}(Z_i)\right]\left\{\hat{Y}_{C,i} - \left[\hat{\boldsymbol{X}}_{C,i,\mathbf{1}} - \hat{S}_{\boldsymbol{X},\mathbf{1}}(Z_i)\right]^{\mathrm{T}}\boldsymbol{\beta}_{0,\mathbf{1}}\right\} \qquad (\text{A.67}) \\ &\xrightarrow{\mathcal{L}} N\left(\mathbf{0}_{p_0}, \boldsymbol{\Sigma}_{0\epsilon,\mathbf{1}} + \boldsymbol{\Sigma}_{0,\mathbf{1}}\boldsymbol{\Omega}_{\phi_M,\boldsymbol{\psi}_{M,\mathbf{1}}}\boldsymbol{\Sigma}_{0,\mathbf{1}}\right), \end{aligned}$$

where, $\boldsymbol{\Sigma}_{0\epsilon,\mathbf{1}}$, $\boldsymbol{\Sigma}_{0,\mathbf{1}}$ and $\boldsymbol{\Omega}_{\phi_M,\boldsymbol{\psi}_{M,\mathbf{1}}}$ are defined in Theorem 10. The asymptotic expressions (A.66) and (A.67) entail that

$$\begin{aligned} &\sqrt{n}\left(\boldsymbol{\Sigma}_{0,\mathbf{1}} + \boldsymbol{\Sigma}_{\boldsymbol{\zeta}_{\mathbf{1}}}\right)\left\{\left(\hat{\boldsymbol{\beta}}_{P,C,\mathbf{1}} - \boldsymbol{\beta}_{0,\mathbf{1}}\right) + \left(\boldsymbol{\Sigma}_{0,\mathbf{1}} + \boldsymbol{\Sigma}_{\boldsymbol{\zeta}_{\mathbf{1}}}\right)^{-1}\mathcal{R}_{\boldsymbol{\zeta}_{\mathbf{1}}}\right\} \\ &\xrightarrow{\mathcal{L}} N\left(\mathbf{0}_{p_0}, \boldsymbol{\Sigma}_{0\epsilon,\mathbf{1}} + \boldsymbol{\Sigma}_{0,\mathbf{1}}\boldsymbol{\Omega}_{\phi_M,\boldsymbol{\psi}_{M,\mathbf{1}}}\boldsymbol{\Sigma}_{0,\mathbf{1}}\right). \end{aligned}$$

We have completed the proof of Theorem 10.

References

[1] Cui, X., Guo, W., Lin, L. and Zhu, L. [2009]. Covariate-adjusted nonlinear regression, *The Annals of Statistics* **37**: 1839–1870. MR2533473

[2] Delaigle, A., Hall, P. and Zhou, W.-X. [2016]. Nonparametric covariate-adjusted regression, *The Annals of Statistics* **44**(5): 2190–2220. MR3546448

[3] Fan, J. and Li, R. [2001]. Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association* **96**(456): 1348–1360. MR1946581

[4] Fan, J. and Peng, H. [2004]. Nonconcave penalized likelihood with a diverging number of parameters, *The Annals of Statistics* **32**(3): 928–961. MR2065194

[5] Härdle, W. and Liang, H. [2007]. Partially linear models, *Statistical Methods for Biostatistics and Related Fields*, Springer, Berlin, pp. 87–103. MR2376405

[6] Härdle, W., Liang, H. and Gao, J. T. [2000]. *Partially Linear Models*, Springer Physica, Heidelberg. MR1787637

[7] Kaysen, G. A., Dubin, J. A., Müller, H.-G., Mitch, W. E., Rosales, L. M. and Nathan W. Levin the Hemo Study Group [2002]. Relationships among inflammation nutrition and physiologic mechanisms establishing albumin levels in hemodialysis patients, *Kidney International* **61**: 2240–2249.

[8] Li, F., Lin, L. and Cui, X. [2010]. Covariate-adjusted partially linear regression models, *Communications in Statistics. Theory and Methods* **39**(6): 1054–1074. MR2745361

[9] Li, F. and Lu, Y. [2018]. Lasso-type estimation for covariate-adjusted linear model, *Journal of Applied Statistics* **45**(1): 26–42. MR3736856

[10] Li, G., Lin, L. and Zhu, L. [2012]. Empirical likelihood for a varying coefficient partially linear model with diverging number of parameters, *Journal of Multivariate Analysis* **105**: 85–111. MR2877505

[11] Li, G., Zhang, J. and Feng, S. [2016]. *Modern Measurement Error Models*, Science Press, Beijing.

[12] Lian, H. [2012]. Empirical likelihood confidence intervals for nonparametric functional data analysis, *Journal of Statistical Planning and Inference* **142**(7): 1669–1677. MR2903379

[13] Lian, H., Liang, H. and Wang, L. [2014]. Generalized additive partial linear models for clustered data with diverging number of covariates using gee, *Statistica Sinica* **24**(1): 173–196. MR3183679

[14] Liang, H., Härdle, W. and Carroll, R. J. [1999]. Estimation in a semiparametric partially linear errors-in-variables model, *The Annals of Statistics* **27**(5): 1519–1535. MR1742498

[15] Liang, H. and Li, R. [2009]. Variable selection for partially linear models with measurement errors, *Journal of the American Statistical Association* **104**: 234–248. MR2504375

[16] Liang, H., Liu, X., Li, R. and Tsai, C. L. [2010]. Estimation and testing for partially linear single-index models, *The Annals of Statistics* **38**: 3811–3836. MR2766869

[17] Liang, H., Qin, Y., Zhang, X. and Ruppert, D. [2009]. Empirical likelihood-based inferences for generalized partially linear models, *Scandinavian Journal of Statistics. Theory and Applications* **36**(3): 433–443. MR2549703

[18] Liu, J., Lou, L. and Li, R. [2018]. Variable selection for partially linear models via partial correlation, *J. Multivariate Anal.* **167**: 418–434. MR3830655

[19] Mack, Y. P. and Silverman, B. W. [1982]. Weak and strong uniform consistency of kernel regression estimates, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **61**(3): 405–415. MR0679685

[20] Nguyen, D. V. and Şentürk, D. [2007]. Distortion diagnostics for covariate-

adjusted regression: Graphical techniques based on local linear modeling, *Journal of Data Science* **5**: 471–490. MR2474018

[21] Nguyen, D. V. and Şentürk, D. [2008]. Multicovariate-adjusted regression models, *Journal of Statistical Computation and Simulation* **78**: 813–827. MR2458498

[22] Nguyen, D. V., Şentürk, D. and Carroll, R. J. [2008]. Covariate-adjusted linear mixed effects model with an application to longitudinal data, *Journal of Nonparametric Statistics* **20**: 459–481. MR2446438

[23] Owen, A. [1991]. Empirical likelihood for linear models, *The Annals of Statistics* **19**: 1725–1747. MR1135146

[24] Owen, A. B. [2001]. *Empirical Likelihood*, Chapman and Hall/CRC, London.

[25] Şentürk, D. and Müller, H.-G. [2005]. Covariate adjusted correlation analysis via varying coefficient models, *Scandinavian Journal of Statistics. Theory and Applications* **32**(3): 365–383. MR2204625

[26] Şentürk, D. and Müller, H.-G. [2006]. Inference for covariate adjusted regression via varying coefficient models, *The Annals of Statistics* **34**: 654–679. MR2281880

[27] Şentürk, D. and Nguyen, D. V. [2006]. Estimation in covariate-adjusted regression, *Computational Statistics & Data Analysis* **50**(11): 3294–3310. MR2239669

[28] Şentürk, D. and Nguyen, D. V. [2009]. Partial covariate adjusted regression, *Journal of Statistical Planning and Inference* **139**(2): 454–468. MR2474018

[29] Tibshirani, R. [1996]. Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society, Series B* **58**: 267–288. MR1379242

[30] Wang, L., Xue, L., Qu, A. and Liang, H. [2014]. Estimation and model selection in generalized additive partial linear models for correlated data with diverging number of covariates, *Annals of Statistics* **42**(2): 592–624. MR3210980

[31] Yang, Y., Tong, T. and Li, G. [2019]. Simex estimation for single-index model with covariate measurement error, *AStA Advances in Statistical Analysis* **103**(1): 137–161. MR3922272

[32] Zhang, J. [2019]. Estimation and variable selection for partial linear single-index distortion measurement errors models, *Statistical Papers*. doi:10.1007/s00362-019-01119-6.

[33] Zhang, J., Chen, Q. and Zhou, N. [2017]. Correlation analysis with additive distortion measurement errors, *Journal of Statistical Computation and Simulation* **87**(4): 664–688. MR3590416

[34] Zhang, J. and Feng, Z. [2017]. Partial linear single-index models with additive distortion measurement errors, *Communications in Statistics – Theory and Methods* **46**(24): 12165–12193. MR3740858

[35] Zhang, J., Feng, Z. and Xu, P. [2015]. Estimating the conditional single-index error distribution with a partial linear mean regression, *Test* **24**: 61–83. MR3314574

[36] Zhang, J., Li, G. and Feng, Z. [2015]. Checking the adequacy for a distortion errors-in-variables parametric regression model, *Computational Statistics &*

Data Analysis **83**: 52–64. MR3281795

[37] Zhang, J., Lin, B. and Li, G. [2019]. Nonlinear regression models with general distortion measurement errors, *Journal of Statistical Computation and Simulation* **89**(8): 1482–1504. MR3929233

[38] Zhang, J., Zhou, N., Sun, Z., Li, G. and Wei, Z. [2016]. Statistical inference on restricted partial linear regression models with partial distortion measurement errors, *Statistica Neerlandica. Journal of the Netherlands Society for Statistics and Operations Research* **70**(4): 304–331. MR3565895

[39] Zhang, J., Zhou, Y., Lin, B. and Yu, Y. [2017]. Estimation and hypothesis test on partial linear models with additive distortion measurement errors, *Computational Statistics & Data Analysis* **112**: 114–128. MR3645593

[40] Zhang, J., Zhu, L. and Liang, H. [2012]. Nonlinear models with measurement errors subject to single-indexed distortion, *Journal of Multivariate Analysis* **112**: 1–23. MR2957282

[41] Zhao, J. and Xie, C. [2018]. A nonparametric test for covariate-adjusted models, *Statistics & Probability Letters* **133**: 65–70. MR3732354