Annals of the Institute of Statistical Mathematics (2020) 72:699–740
https://doi.org/10.1007/s10463-019-00706-6

Estimation and hypothesis test for partial linear single-index multiplicative models

Jun Zhang[1] · Xia Cui[2] · Heng Peng[3]

Received: 23 February 2018 / Revised: 7 November 2018 / Published online: 12 February 2019

Abstract

Estimation and hypothesis test for partial linear single-index multiplicative models are considered in this paper. To estimate unknown single-index parameter, we propose a profile least product relative error estimator coupled with a leave-one-component-out method. To test a hypothesis on the parametric components, a Wald-type test statistic is proposed. We employ the smoothly clipped absolute deviation penalty to select relevant variables. To study model checking problem, we propose a variant of the integrated conditional moment test statistic by using linear projection weighting function, and we also suggest a bootstrap procedure for calculating critical values. Simulation studies are conducted to demonstrate the performance of the proposed procedure and a real example is analyzed for illustration.

Keywords Local linear smoothing · Model checking · Profile least product relative error estimator · Single-index · Variable selection

---

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10463-019-00706-6) contains supplementary material, which is available to authorized users.

✉ Jun Zhang
zhangjunstat@gmail.com

Xia Cui
cuixia@gzhu.edu.cn

Heng Peng
hpeng@math.hkbu.edu.hk

1 College of Mathematics and Statistics, Shenzhen-Hong Kong Joint Research Center for Applied Statistical Sciences, Institute of Statistical Sciences, Shenzhen University, Shenzhen 518060, China

2 School of Economics and Statistics, Guangzhou University, Guangzhou 510006, China

3 Department of Mathematics, The Hong Kong Baptist University, Kowloon Tong, Hong Kong, China

1 Introduction

Let $(\boldsymbol{X}, \boldsymbol{Z}, Y)$ be a random vector, we assume that $(\boldsymbol{X}, \boldsymbol{Z})$ and Y satisfy the following partial linear single-index multiplicative model (PLSiMM):

$$Y = \exp\left(\boldsymbol{\alpha}_0^{\mathrm{T}}\boldsymbol{X} + g(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{Z})\right)\epsilon, \tag{1}$$

where Y is the response variable, $\boldsymbol{X} = (X_1, \ldots, X_q)^{\mathrm{T}} \in \mathbb{R}^q$, $\boldsymbol{Z} = (Z_1, \ldots, Z_p)^{\mathrm{T}} \in \mathbb{R}^p$, $g(\cdot)$ is an unknown smooth link function, $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$ are two unknown parameters. Both Y and ϵ considered in model (1) are positive variables, and the model error ϵ satisfies $E(\ln(\epsilon)|\boldsymbol{X}, \boldsymbol{Z}) = 0$ and also $E(\epsilon - \epsilon^{-1}|\boldsymbol{X}, \boldsymbol{Z}) = 0$. It is easily seen that $\ln(Y) = \boldsymbol{\alpha}_0^{\mathrm{T}}\boldsymbol{X} + g(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{Z}) + \ln(\epsilon)$. To make link function $g(\cdot)$ unique, the condition $E(\ln(\epsilon)|\boldsymbol{X}, \boldsymbol{Z}) = 0$ is used to identify the unknown link function $g(\cdot)$. The other condition $E(\epsilon - \epsilon^{-1}|\boldsymbol{X}, \boldsymbol{Z}) = 0$ is used for the least relative error estimation (Chen et al. 2016). The parameter $\boldsymbol{\beta}_0$ is an unknown index vector which belongs to the parameter space $\mathcal{B} = \left\{\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^{\mathrm{T}} \in \mathbb{R}^p, \|\boldsymbol{\beta}\| = 1, \beta_1 > 0\right\}$.

Model (1) is an important generalization of the multiplicative linear regression models or the accelerated failure model with an unknown single-index link function. When $\boldsymbol{\alpha}_0 = \mathbf{0}$, model (1) reduces to the single-index multiplicative model, which has been studied in Zhang et al. (2018), Liu and Xia (2018). By taking logarithmic transformation, model (1) with $Y^* = \ln(Y)$ and $\epsilon^* = \ln(\epsilon)$ becomes a classical partial linear single-index models or single-index models ($\boldsymbol{\alpha}_0 = \mathbf{0}$), see for example, Xia and Härdle (2006), Xia et al. (2002), Liang et al. (2010), Lian et al. (2015), Lian and Liang (2016), Li et al. (2015), Lai et al. (2013), Liang and Wang (2005), Li et al. (2014), Ichimura (1993), Cui et al. (2011), Peng and Huang (2011) and Bindele et al. (2018). Due to the theoretical and computational simplicity, such logarithmic transformation is reasonable in some cases. As Chen et al. (2010, 2016) claimed, a linear relationship in the transformed model is not linear in the original one, and one also needs to transform the analysis results back into the original measurement scale. So, it is more tenable to use the original model rather than the transformed model.

To estimate the parameter $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$ and link function $g(u)$, we propose the profile least product relative error criterion by considering to minimize the least product relative error criterion (Chen et al. 2016, LPRE) $\sum_{i=1}^n (|\epsilon_i^{-1} - 1| \times |\epsilon_i - 1|)$, which is equivalent to minimize $\sum_{i=1}^n (\epsilon_i^{-1} + \epsilon_i)$. We can see that the LPRE criterion function is infinitely differentiable and strictly convex. Chen et al. (2010) proposed the least absolute relative error (LARE) estimation by minimizing $\sum_{i=1}^n (|\epsilon_i^{-1} - 1| + |\epsilon_i - 1|)$. The LARE estimation enjoys the robustness and scale-free property; however, this criterion function is unsmooth and the computation is very complicated. The resulting LARE estimate of parameter can be asymptotically normal with a complex asymptotic covariance matrix, which involves the density of the error ϵ (Chen et al. 2010). Although the covariance matrix of the LARE estimator can be estimated, the estimation accuracy like confidence intervals of parameters would be affected when sample size is small or moderate. So, we consider to use least product relative error criterion and propose a profile least product relative error estimation for $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$, coupling with a leave-one-component-out method for single-index parameter. This profile least product relative error method will obtain efficient estimators than the logarithmic transformation method (Liang et al. 2010) in some cases. To estimate unknown link function $g(u)$, a local least product relative error estimation is proposed. After estimating unknown parameter, it is natural to construct hypothesis tests to assess the appropriateness of the linear constraint hypothesis. We test whether the true parameters satisfy some linear combinations or not. A Wald-type statistic under the null hypotheses is proposed. We show that the limiting distribution of the test statistic under the null hypothesis is a centered Chi-squared distribution. Moreover, a restricted estimator of parameter is proposed associated with its asymptotic properties.

In data analysis, the true model is often unknown; this allows the possibility of selecting an underfitted (or overfitted) model, which leads to the biased (or inefficient) estimators and predictions. Traditional variable selection methods, such as best subset variable selection or stepwise regression, do not work effectively for the semiparametric model. Variable selection for model (1) is challenging because it involves both nonparametric and parametric parts. The strictly convex of the least product relative error criterion motivates to use the penalized methods to shrink the estimated coefficients of superfluous variables to zero. In this paper, we adopt to use the smoothly clipped absolute deviation approach (Fan and Peng 2004, SCAD) that not only selects important variables consistently, but also produces parameter estimators as efficient as if the true model were known even in a high-dimensional setting, a property not possessed by the least absolute shrinkage and selection operator (Tibshirani 1996, lasso). We employ the least product relative error estimation approach to obtain the SCAD estimators for parameter $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$. We demonstrate that the resulting SCAD-based solution is selection consistent.

Lastly, we aim to develop a lack-of-fit test for checking the adequacy of partial linear single-index multiplicative models. We propose a variant of integrated conditional moment (ICM) test using linear projection weighting function but choosing the projection direction by fitting a single-index model from the estimated squared relative model error against all the covariates. It will be shown that the proposed method has good theoretical properties such as consistency, and has higher power than ICM test for the partial linear single-index model by taking logarithmic transformation on the response variable. Monte Carlo simulation experiments are conducted to examine the performance of the proposed test procedure.

This paper is organized as follows. In Sect. 2, we propose the least product relative error estimation procedure for parameter $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$ as well as the link function $g(u)$, introduce the algorithms and present the asymptotic results. In Sect. 3, we provide a Wald-type test statistic for the testing whether the true parameters satisfy some linear combinations, give a restricted estimator under the null hypothesis and obtain its theoretical properties. In Sect. 4, a penalized estimator based on SCAD penalty is proposed for variable selection. In Sect. 5, we develop a ICM test statistic for checking the adequacy of partial linear single-index multiplicative models and study theoretical properties of the test statistic. In Sect. 6, we report the results of simulation studies. In Sect. 7, statistical analysis of a real data is reported. All technical proofs of asymptotic results are given in the online "Supplementary Material."

2 Estimation methodology

2.1 Estimation of α_0, β_0 and $g(\cdot)$

For notation clarity, we rewrite the parameter space $\mathcal{B}$ as a $(p-1)$-dimensional space: $\mathcal{B}_{\phi} = \{\beta = \beta_{\phi} = (\sqrt{1-\|\phi\|^2}, \phi^{\mathrm{T}})^{\mathrm{T}} : \phi = (\beta_2, \ldots, \beta_p)^{\mathrm{T}}, \|\phi\|^2 = \sum_{l=2}^{p}\beta_l^2 < 1\}$, which is used to ensure that the single-index parameter β_0 and the link function $g(\cdot)$ can be uniquely defined. This is a commonly used assumption for single-index parameter. See, for example, Cui et al. (2011), Li et al. (2014). In the following, we use β_{ϕ} instead of β and use β_{ϕ_0} instead of β_0 for the true value. We also define $A^{\otimes 2} = AA^{\mathrm{T}}$ for any matrix or vector A.

Suppose that we have an i.i.d. sample $\{X_i, Z_i, Y_i\}_{i=1}^n$ from model (1), where $X_i = (X_{1i}, \ldots, X_{qi})^{\mathrm{T}}$ and $Z_i = (Z_{1i}, \ldots, Z_{pi})^{\mathrm{T}}$. The estimation procedure is summarized as follows.

(1) Transform model (1) into $\ln(Y) = \alpha_0^{\mathrm{T}} X + g(\beta_0^{\mathrm{T}} Z) + \ln(\epsilon)$. Initially, a local linear smoothing technique is used to estimate the nonparametric function $g(u)$. Approximate $g(u)$ by $g(u_*) + g'(u_*)(u - u_*)$ in a neighborhood of u. For given $\eta = (\alpha, \phi)$, the local linear estimator of $(g(u), g'(u))$ is obtained by minimizing (2) with respect to (b_0, d_0),

$$\sum_{i=1}^{n}\left\{\ln(Y_i) - \alpha^{\mathrm{T}} X_i - b_0 - d_0(\beta_{\phi}^{\mathrm{T}} Z_i - u)\right\}^2 K_h(\beta_{\phi}^{\mathrm{T}} Z_i - u), \tag{2}$$

where $K_h(\beta_{\phi}^{\mathrm{T}} Z_i - u) = h^{-1} K\left(\frac{\beta_{\phi}^{\mathrm{T}} Z_i - u}{h}\right)$ with $K(\cdot)$ being a kernel function and h being a bandwidth. Let $\hat{b}_0$ be the minimizer of (2), and denote it as $\hat{g}(u, \eta)$. Then, the estimator of $g(u)$ is obtained as

$$\begin{aligned}\hat{g}(u, \eta) &= \hat{b}_0 \\ &= \frac{T_{n,20}(u,\eta)T_{n,01}(u,\eta) - T_{n,10}(u,\eta)T_{n,11}(u,\eta)}{T_{n,00}(u,\eta)T_{n,20}(u,\eta) - T_{n,10}^2(u,\eta)},\end{aligned} \tag{3}$$

where $T_{n,l_1 l_2}(u,\eta) = \sum_{i=1}^{n} K_h(\beta_{\phi}^{\mathrm{T}} Z_i - u)(\beta_{\phi}^{\mathrm{T}} Z_i - u)^{l_1}[\ln(Y_i) - \alpha^{\mathrm{T}} X_i]^{l_2}$ for $l_1 = 0, 1, 2$, $l_2 = 0, 1$.

(2) To estimate the true value $\eta_0 = (\alpha_0^{\mathrm{T}}, \phi_0^{\mathrm{T}})^{\mathrm{T}}$, we propose the profile least product relative error estimation (PLPRE):

$$\begin{aligned}\hat{\eta} &= \left(\hat{\alpha}^{\mathrm{T}}, \hat{\phi}^{\mathrm{T}}\right)^{\mathrm{T}} \\ &= \arg\min_{\alpha \in \mathbb{R}^q, \phi^{\mathrm{T}}\phi < 1} \sum_{i=1}^{n}\left\{Y_i \exp\left(-\alpha^{\mathrm{T}} X_i - \hat{g}(\beta_{\phi}^{\mathrm{T}} Z_i, \eta)\right)\right. \\ &\quad \left. + Y_i^{-1}\exp\left(\alpha^{\mathrm{T}} X_i + \hat{g}(\beta_{\phi}^{\mathrm{T}} Z_i, \eta)\right)\right\}.\end{aligned} \tag{4}$$

Apply the equation $\beta_1 = \sqrt{1 - \left\|\boldsymbol{\phi}_0\right\|^2}$ to estimate β_1, $\hat{\beta}_1 = \sqrt{1 - \|\hat{\boldsymbol{\phi}}\|^2}$, and the estimator of $\boldsymbol{\beta}_{\boldsymbol{\phi}_0}$ is $\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}} = \left(\hat{\beta}_1, \hat{\boldsymbol{\phi}}^{\mathrm{T}}\right)^{\mathrm{T}}$.

(3) After obtaining $\left(\hat{\boldsymbol{\alpha}}^{\mathrm{T}}, \hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}^{\mathrm{T}}\right)^{\mathrm{T}}$, we re-estimate the unknown function $g(u)$ by using local least product relative error (LPLPRE) estimation:

$$\sum_{i=1}^{n} K_h\left(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}^{\mathrm{T}} \boldsymbol{Z}_i - u\right)\left\{Y_i \exp\left(-\hat{\boldsymbol{\alpha}}^{\mathrm{T}} \boldsymbol{X}_i - a_0 - a_1(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}^{\mathrm{T}} \boldsymbol{Z}_i - u)\right)\right. \\ \left. + Y_i^{-1} \exp\left(\hat{\boldsymbol{\alpha}}^{\mathrm{T}} \boldsymbol{X}_i + a_0 + a_1(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}^{\mathrm{T}} \boldsymbol{Z}_i - u)\right)\right\}. \tag{5}$$

The LPLPRE estimator of $g(u)$ is obtained as $\hat{g}_L(u, \hat{\boldsymbol{\eta}}) = \hat{a}_0$.

Algorithm

The minimization algorithm for (4) can be implemented by using the Newton–Raphson approximation. Meanwhile, a consistent initial value for $(\boldsymbol{\alpha}_0, \phi_0)$ can make the Newton–Raphson algorithm converge fast. One can use the dimension–reduction method for the transformed data $\{\boldsymbol{X}_i, \boldsymbol{Z}_i, \ln(Y_i)\}_{i=1}^{n}$, see for example Xia and Härdle (2006), Zhang et al. (2018). After we obtain the initial values, we then update $\boldsymbol{\eta}_0 = (\boldsymbol{\alpha}_0^{\mathrm{T}}, \phi_0^{\mathrm{T}})^{\mathrm{T}}$ by the Newton–Raphson approximation:

$$\hat{\boldsymbol{\eta}}_{\text{new}} = \hat{\boldsymbol{\eta}}_{\text{old}} - \left[\widetilde{\mathfrak{M}}_n\left(\hat{\boldsymbol{\eta}}_{\text{old}}\right)\right]^{-1} \widetilde{\mathfrak{N}}_n\left(\hat{\boldsymbol{\eta}}_{\text{old}}\right), \tag{6}$$

$$\widetilde{\mathfrak{M}}_n\left(\hat{\boldsymbol{\eta}}_{\text{old}}\right) = \sum_{i=1}^{n}\left[\hat{\epsilon}_{i,\text{old}}(\hat{\boldsymbol{\eta}}_{\text{old}}) + \hat{\epsilon}_{i,\text{old}}^{-1}(\hat{\boldsymbol{\eta}}_{\text{old}})\right]\left[\left.\begin{pmatrix} \boldsymbol{X}_i + \dfrac{\partial \hat{g}(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}} \boldsymbol{Z}_i, \boldsymbol{\eta})}{\partial \boldsymbol{\alpha}} \\ \dfrac{\partial \hat{g}(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}} \boldsymbol{Z}_i, \boldsymbol{\eta})}{\partial \boldsymbol{\phi}} \end{pmatrix}^{\otimes 2}\right|_{\eta=\hat{\eta}_{\text{old}}}\right],$$

$$\widetilde{\mathfrak{N}}_n\left(\hat{\boldsymbol{\eta}}_{\text{old}}\right) = \sum_{i=1}^{n}\left[-\hat{\epsilon}_{i,\text{old}}(\hat{\boldsymbol{\eta}}_{\text{old}}) + \hat{\epsilon}_{i,\text{old}}^{-1}(\hat{\boldsymbol{\eta}}_{\text{old}})\right]\left[\left.\begin{pmatrix} \boldsymbol{X}_i + \dfrac{\partial \hat{g}(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}} \boldsymbol{Z}_i, \boldsymbol{\eta})}{\partial \boldsymbol{\alpha}} \\ \dfrac{\partial \hat{g}(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}} \boldsymbol{Z}_i, \boldsymbol{\eta})}{\partial \boldsymbol{\phi}} \end{pmatrix}\right|_{\eta=\hat{\eta}_{\text{old}}}\right],$$

where $\hat{\epsilon}_{i,\text{old}}(\hat{\boldsymbol{\eta}}_{\text{old}}) = Y_i \exp\left(-\hat{\boldsymbol{\alpha}}_{\text{old}}^{\mathrm{T}} \boldsymbol{X}_i - \hat{g}(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}_{\text{old}}}^{\mathrm{T}} \boldsymbol{Z}_i, \hat{\boldsymbol{\eta}}_{\text{old}})\right)$, and $\hat{\epsilon}_{i,\text{old}}^{-1}(\hat{\boldsymbol{\eta}}_{\text{old}}) = Y_i^{-1} \exp\left(\hat{\boldsymbol{\alpha}}_{\text{old}}^{\mathrm{T}} \boldsymbol{X}_i + \hat{g}(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}_{\text{old}}}^{\mathrm{T}} \boldsymbol{Z}_i, \hat{\boldsymbol{\eta}}_{\text{old}})\right)$, $i = 1, \ldots, n$.

According to Li et al. (2011) and Cui et al. (2011), the two arguments $\widetilde{\mathfrak{M}}_n\left(\hat{\boldsymbol{\eta}}_{\text{old}}\right)$ and $\widetilde{\mathfrak{N}}_n\left(\hat{\boldsymbol{\eta}}_{\text{old}}\right)$ are equivalent to

$$\widetilde{\mathfrak{M}}_n\left(\hat{\boldsymbol{\eta}}_{\text{old}}\right) = \sum_{i=1}^{n}\left[\hat{\epsilon}_{i,\text{old}}(\hat{\boldsymbol{\eta}}_{\text{old}}) + \hat{\epsilon}_{i,\text{old}}^{-1}(\hat{\boldsymbol{\eta}}_{\text{old}})\right] \times \left[\left(\begin{array}{c} \boldsymbol{X}_i - \hat{\boldsymbol{m}}_{\boldsymbol{X}}(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}\boldsymbol{Z}_i) \\ \boldsymbol{J}_{\boldsymbol{\phi}}^{\mathrm{T}}\left[\boldsymbol{Z}_i - \hat{\boldsymbol{m}}_{\boldsymbol{Z}}(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}\boldsymbol{Z}_i)\right]\hat{g}'(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}\boldsymbol{Z}_i)\end{array}\right)^{\otimes 2}\Bigg|_{\eta=\hat{\eta}_{\text{old}}}\right], \tag{7}$$

$$\widetilde{\mathfrak{N}}_n\left(\hat{\boldsymbol{\eta}}_{\text{old}}\right) = \sum_{i=1}^{n}\left[-\hat{\epsilon}_{i,\text{old}}(\hat{\boldsymbol{\eta}}_{\text{old}}) + \hat{\epsilon}_{i,\text{old}}^{-1}(\hat{\boldsymbol{\eta}}_{\text{old}})\right] \times \left[\left(\begin{array}{c} \boldsymbol{X}_i - \hat{\boldsymbol{m}}_{\boldsymbol{X}}(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}\boldsymbol{Z}_i) \\ \boldsymbol{J}_{\boldsymbol{\phi}}^{\mathrm{T}}\left[\boldsymbol{Z}_i - \hat{\boldsymbol{m}}_{\boldsymbol{Z}}(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}\boldsymbol{Z}_i)\right]\hat{g}'(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}\boldsymbol{Z}_i)\end{array}\right)\Bigg|_{\eta=\hat{\eta}_{\text{old}}}\right], \tag{8}$$

where $\hat{g}'(u)$ is the local linear estimator $\hat{d}_0$ from (2), $\hat{\boldsymbol{m}}_{\boldsymbol{X}}(u)$ is the local linear estimator of $E[\boldsymbol{X}|\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z} = u]$ and it is defined as $\hat{\boldsymbol{m}}_{\boldsymbol{X}}(u) = \frac{\sum_{i=1}^{n} b_{n,i}(u)\boldsymbol{X}_i}{\sum_{i=1}^{n} b_{n,i}(u)}$, $\hat{\boldsymbol{m}}_{\boldsymbol{Z}}(u)$ is the local linear estimator of $E[\boldsymbol{Z}|\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z} = u]$ and it is defined as $\hat{\boldsymbol{m}}_{\boldsymbol{Z}}(u) = \frac{\sum_{i=1}^{n} b_{n,i}(u)\boldsymbol{Z}_i}{\sum_{i=1}^{n} b_{n,i}(u)}$, $b_{n,i}(u) = K_h(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}\boldsymbol{Z}_i - u)\big[T_{n,20}(u,\boldsymbol{\eta}) - (\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}\boldsymbol{Z}_i - u)T_{n,10}(u,\boldsymbol{\eta})\big]$. It is noted that the arguments $\widetilde{\mathfrak{M}}_n\left(\hat{\boldsymbol{\eta}}_{\text{old}}\right)$ and $\widetilde{\mathfrak{N}}_n\left(\hat{\boldsymbol{\eta}}_{\text{old}}\right)$ defined in (7)–(8) are different from Wang et al. (2017) for estimating the single-index parameter $\boldsymbol{\phi}_0$. In stead of $\boldsymbol{J}_{\boldsymbol{\phi}}^{\mathrm{T}}\left[\boldsymbol{Z}_i - \hat{\boldsymbol{m}}_{\boldsymbol{Z}}(\boldsymbol{X}_i|\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}\boldsymbol{Z}_i)\right]\hat{g}'(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}\boldsymbol{Z}_i)$ in (7)–(8), Wang et al. (2017) used $\boldsymbol{J}_{\boldsymbol{\phi}}^{\mathrm{T}}\boldsymbol{Z}_i\hat{g}'(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}\boldsymbol{Z}_i)$ in their Newton–Raphson algorithm. Li et al. (2011), Cui et al. (2011) proved that $\frac{\partial \hat{g}(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}z)}{\partial \boldsymbol{\phi}}$ converges to $-\boldsymbol{J}_{\boldsymbol{\phi}}^{\mathrm{T}}\{z - E[\boldsymbol{Z}|\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}\boldsymbol{Z} = \boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}z]\}g'(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}z)$. As a result, the argument $\boldsymbol{J}_{\boldsymbol{\phi}}^{\mathrm{T}}\left[\boldsymbol{Z}_i - \hat{\boldsymbol{m}}_{\boldsymbol{Z}}(\boldsymbol{X}_i|\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}\boldsymbol{Z}_i)\right]\hat{g}'(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}\boldsymbol{Z}_i)$ used in our algorithm (7)–(8) is more reasonable and tenable.

2.2 Asymptotic results

Here we list the conditions needed in asymptotic results.

- (C1) $E[X_s^4] < \infty$ for $s = 1, \ldots, q$, $E[Z_r^4] < \infty$ for $r = 1, \ldots, p$, and the matrices $\boldsymbol{\Gamma}_0$, $\boldsymbol{\Lambda}_0$ used in Theorem 1 and $\boldsymbol{\Gamma}_{L,0}$ in Proposition 1 are all positive definite and finite. Moreover, $E[|\ln(Y)|^r] < \infty$ for some $r > 3$.
- (C2) $E(\boldsymbol{X}\big|\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}\boldsymbol{Z} = u)$, $E(\boldsymbol{Z}\big|\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}\boldsymbol{Z} = u)$ and the density function $f_{\boldsymbol{\beta}_{\boldsymbol{\phi}}}(u)$ of the random variable $\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}\boldsymbol{Z}$ are twice continuously differentiable with respect to u. Their second-order derivatives are uniformly Lipschitz continuous on $\mathcal{C} = \{u = \boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}\boldsymbol{Z} : z \in \mathcal{Z} \subset \mathbb{R}^p, \boldsymbol{\beta}_{\boldsymbol{\phi}} \in \mathfrak{S}_{n,\boldsymbol{\beta}_{\boldsymbol{\phi}}}\}$, where $\mathcal{Z}$ is a compact support set, and $\mathfrak{S}_{n,\boldsymbol{\beta}_{\boldsymbol{\phi}}} = \{\boldsymbol{\beta}_{\boldsymbol{\phi}} \in \mathcal{B}_{\boldsymbol{\phi}} : \|\boldsymbol{\beta}_{\boldsymbol{\phi}} - \boldsymbol{\beta}_{\boldsymbol{\phi}_0}\| \le c_0 n^{-1/2+c_1}\}$ for some positive constant c_0 and $c_1 \in [0, 0.05)$. Moreover, $g(u)$ has two bounded and continuous derivatives on $u \in \mathcal{C}$ and $\inf_{u\in\mathcal{C}} f_{\boldsymbol{\beta}_{\boldsymbol{\phi}}}(u) > 0$.
- (C3) 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 contin-

uous bounded derivatives, satisfying $K^{(j)}(\pm A) = 0$ with $K^{(j)}(t) = \frac{d^j}{dt^j}K(t)$, $j = 0, 1, 2$, and $\int s^2 K(s)ds \neq 0$.

(C4) As $n \to \infty$, the bandwidth h satisfies $\frac{(\log n)^{2+2s_0}}{nh^2} \to 0$ for some $s_0 > 0$, and $nh^4 \to 0$.

(C5) For all ζ_j $j = 1, \ldots, p+q-1$, $\zeta_j \to 0$, $\sqrt{n}\zeta_j \to \infty$ as $n \to \infty$, moreover, $\text{liminf}_{n\to\infty}\text{liminf}_{u\to 0^+} p'_{\zeta_j}(u)/\zeta_j > 0$.

Condition (C1) is the moments of the covariates, and the technique condition of $(\boldsymbol{\Gamma}_0, \boldsymbol{\Lambda}_0)$ is imposed to ensure that it is invertible in Theorem 1. Condition (C2) is typical assumptions in the nonparametric smoothing literature, which are also found in Liang et al. (2010), Xia and Härdle (2006), Li et al. (2014), Boente and Rodriguez (2012). Condition (C3) is the common condition for the kernel function $K(t)$. The Epanechnikov kernel satisfies this condition. Condition (C4) is generally required for bandwidth h in single-index semiparametric models, see for example Liang et al. (2010), Peng and Huang (2011). The detailed bandwidth selection will be discussed in Theorem 2. Condition (C5) is a technique condition involved in the SCAD variable selection procedure (Fan and Peng 2004).

We define $\boldsymbol{J}_{\boldsymbol{\phi}_0} = \begin{pmatrix} -\boldsymbol{\phi}_0^{\mathrm{T}} \Big/ \sqrt{1 - \|\boldsymbol{\phi}_0\|^2} \\ \boldsymbol{I}_{p-1} \end{pmatrix}$, where $\boldsymbol{I}_{p-1}$ is an $(p-1)$-dimensional identity matrix, and $\breve{\boldsymbol{X}} = \boldsymbol{X} - E[\boldsymbol{X}|\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z}]$, $\breve{\boldsymbol{Z}} = (\boldsymbol{Z} - E[\boldsymbol{Z}|\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z}])g'(\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z})$,

$$\boldsymbol{\Lambda}_0 = E\left\{\begin{pmatrix} \breve{\boldsymbol{X}} \\ \boldsymbol{J}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\breve{\boldsymbol{Z}} \end{pmatrix}^{\otimes 2}\right\}, \quad \boldsymbol{\Gamma}_0 = E\left\{(\epsilon + \epsilon^{-1})\begin{pmatrix} \breve{\boldsymbol{X}} \\ \boldsymbol{J}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\breve{\boldsymbol{Z}} \end{pmatrix}^{\otimes 2}\right\},$$

$$\boldsymbol{\Sigma}_0 = E\left\{\left((\epsilon - \epsilon^{-1})\begin{pmatrix} \breve{\boldsymbol{X}} \\ \boldsymbol{J}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\breve{\boldsymbol{Z}} \end{pmatrix} - E\left[(\epsilon + \epsilon^{-1})\begin{pmatrix} \breve{\boldsymbol{X}} \\ \boldsymbol{J}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\breve{\boldsymbol{Z}} \end{pmatrix}\middle|\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z}\right]\ln(\epsilon)\right)^{\otimes 2}\right\}.$$

Theorem 1 *Under the conditions (C1)–(C4), as n goes to infinity, we have*

$$\sqrt{n}\left(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0\right) \xrightarrow{L} N\left(\mathbf{0}_{p+q-1}, \boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Sigma}_0\boldsymbol{\Gamma}_0^{-1}\right).$$

If ϵ is independent of $(\boldsymbol{X}^{\mathrm{T}}, \boldsymbol{Z}^{\mathrm{T}})^{\mathrm{T}}$, let $\boldsymbol{\kappa} = \dfrac{E[(\epsilon - \epsilon^{-1})^2]}{[E(\epsilon + \epsilon^{-1})]^2}$, we have

$$\sqrt{n}\left(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0\right) \xrightarrow{L} N\left(\mathbf{0}_{p+q-1}, \boldsymbol{\kappa}\boldsymbol{\Lambda}_0^{-1}\right).$$

After applying the logarithmic transformation on the response variable, we use the profile least squares method (Liang et al. 2010) to estimate $\boldsymbol{\eta}_0$. Denote that $\hat{\boldsymbol{\eta}}_{\mathrm{T}}$ is the transformation profile least squares estimator (TPLSE). Based on Theorem 1 in Liang et al. (2010), under the independence condition between $(\boldsymbol{X}^{\mathrm{T}}, \boldsymbol{Z}^{\mathrm{T}})^{\mathrm{T}}$ and ϵ, we have $\sqrt{n}(\hat{\boldsymbol{\eta}}_{\mathrm{T}} - \boldsymbol{\eta}_0) \xrightarrow{L} N\left(\mathbf{0}_{p+q-1}, \mathrm{Var}(\ln(\epsilon))\boldsymbol{\Lambda}_0^{-1}\right)$. Comparing with Theorem 1, we find that the PLPRE estimator $\hat{\boldsymbol{\eta}}$ is more efficient than $\hat{\boldsymbol{\eta}}_{\mathrm{T}}$ when $\boldsymbol{\kappa} < \mathrm{Var}(\ln(\epsilon))$ and vice versa. If ϵ follows from $\exp(N(0, \sigma^2))$, a direct calculation entails that

$$\boldsymbol{\kappa} = \frac{1}{2}\left\{\exp(\sigma^2) - \exp(-\sigma^2)\right\} = \sigma^2 + \sum_{s=1}^{\infty} \frac{\sigma^{4s+2}}{(2s+1)!}$$

Moreover, $\mathrm{Var}(\ln(\epsilon)) = E[\ln(\epsilon)]^2 = \sigma^2 < \boldsymbol{\kappa}$. This shows that the TPLSE estimator $\hat{\boldsymbol{\eta}}_{\mathrm{T}}$ is more asymptotically efficient than the PLPRE estimator $\hat{\boldsymbol{\eta}}$ when $\epsilon \sim \exp(N(0, \sigma^2))$ and ϵ is independent of $(\boldsymbol{X}, \boldsymbol{Z})$. If ϵ follows from $\exp(U[-a, a])$ for any $a \in (0, 1]$, where $U[-a, a]$ is a uniform distribution. Then,

$$\boldsymbol{\kappa} = \frac{\frac{a}{2}\{\exp(2a) - \exp(-2a)\} - 2a^2}{\exp(2a) + \exp(-2a) - 2} = a^2 \frac{\sum_{s=1}^{\infty} \frac{1}{(2s+1)!}(2a)^{2s}}{\sum_{s=1}^{\infty} \frac{1}{(2s)!}(2a)^{2s}}$$

A direct calculation shows that $\boldsymbol{\kappa} < \frac{a^2}{3} = E[\ln(\epsilon)]^2 = \mathrm{Var}(\ln(\epsilon))$ when $a \in (0, 1]$. Then, we have that the PLPRE estimator $\hat{\boldsymbol{\eta}}$ is more asymptotically efficient than the TPLSE estimator $\hat{\boldsymbol{\eta}}_{\mathrm{T}}$ when ϵ is independent of $(\boldsymbol{X}, \boldsymbol{Z})$ and $\epsilon \sim \exp(U[-a, a])$ for any $a \in (0, 1]$.

The asymptotic result of $\left(\hat{\boldsymbol{\alpha}}^{\mathrm{T}}, \hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}^{\mathrm{T}}\right)^{\mathrm{T}}$ can be obtained by using a simple application of the multivariate delta method. Let $\boldsymbol{I}_{\boldsymbol{J}_{\boldsymbol{\phi}_0}} = \begin{pmatrix} \boldsymbol{I}_q, & \boldsymbol{0}_{q\times(p-1)} \\ \boldsymbol{0}_{p\times q}, & \boldsymbol{J}_{\boldsymbol{\phi}_0} \end{pmatrix}$, where $\boldsymbol{I}_q$ is an q-dimensional identity matrix.

Theorem 2 *Under the conditions of Theorem 1, we have*

$$\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 \\ \hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}} - \boldsymbol{\beta}_{\boldsymbol{\phi}_0} \end{pmatrix} \xrightarrow{L} N\left(\boldsymbol{0}_{p+q}, \boldsymbol{I}_{\boldsymbol{J}_{\boldsymbol{\phi}_0}} \boldsymbol{\Gamma}_0^{-1} \boldsymbol{\Sigma}_0 \boldsymbol{\Gamma}_0^{-1} \boldsymbol{I}_{\boldsymbol{J}_{\boldsymbol{\phi}_0}}^{\mathrm{T}}\right).$$

Further, if ϵ is independent of $(\boldsymbol{X}^{\mathrm{T}}, \boldsymbol{Z}^{\mathrm{T}})^{\mathrm{T}}$, we have

$$\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 \\ \hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}} - \boldsymbol{\beta}_{\boldsymbol{\phi}_0} \end{pmatrix} \xrightarrow{L} N\left(\boldsymbol{0}_{p+q}, \boldsymbol{\kappa} \boldsymbol{I}_{\boldsymbol{J}_{\boldsymbol{\phi}_0}} \boldsymbol{\Lambda}_0^{-1} \boldsymbol{I}_{\boldsymbol{J}_{\boldsymbol{\phi}_0}}^{\mathrm{T}}\right).$$

Remark 1 It is worthwhile to point out that condition (C4) requires $nh^4 \to 0$. To meet this requirement, we use the order of $O(n^{-1/5}) \times n^{-2/15} = O(n^{-1/3})$ for the bandwidth h. In detail, we define the LPRE cross-validation score

$$\mathrm{CV}(h) = \sum_{i=1}^{n} \left\{ Y_i \exp\left(-\hat{\boldsymbol{\alpha}}_{(-i)}^{\mathrm{T}} \boldsymbol{X}_i - \hat{g}_{(-i)}(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}_{(-i)}}^{\mathrm{T}} Z_i, \hat{\boldsymbol{\eta}}_{(-i)})\right) \right.$$
$$\left. + Y_i^{-1} \exp\left(\hat{\boldsymbol{\alpha}}_{(-i)}^{\mathrm{T}} \boldsymbol{X}_i + \hat{g}_{(-i)}(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}_{(-i)}}^{\mathrm{T}} Z_i, \hat{\boldsymbol{\eta}}_{(-i)})\right) \right\},$$

where $\hat{\boldsymbol{\eta}}_{(-i)}$, $\hat{g}_{(-i)}(u, \hat{\boldsymbol{\eta}}_{(-i)})$ are computed from the data with the ith observation deleted. Let $\hat{h}_{\text{opt}} = \arg\min_{h} \text{CV}(h)$, then the bandwidth h is chosen as $h = \hat{h}_{\text{opt}} * n^{-2/15}$. This bandwidth choice is fairly effective and easy to implement in practice. Our numerical experience suggests that the numerical results in Sect. 6 were stable when we shifted several values around this data-driven bandwidth.

Next, we present the asymptotic results for the estimator $\hat{g}_L(u, \hat{\boldsymbol{\eta}})$.

Theorem 3 *Under the conditions (C1)–(C3), as $h \to 0$, $nh \to \infty$, we have*

$$\sqrt{nh}\left(\hat{g}_L(u, \hat{\boldsymbol{\eta}}) - g(u) - \frac{\mu_{K,2}h^2}{2} g''(u)\right) \xrightarrow{L} N\left(0, \frac{E[(\epsilon - \epsilon^{-1})^2 | \boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\text{T}} \boldsymbol{Z} = u]\mu_{K^2}}{\{E(\epsilon + \epsilon^{-1} | \boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\text{T}} \boldsymbol{Z} = u)\}^2 f_{\boldsymbol{\beta}_{\boldsymbol{\phi}_0}}(u)}\right).$$

If ϵ is independent of $(\boldsymbol{X}^{\text{T}}, \boldsymbol{Z}^{\text{T}})^{\text{T}}$, we have

$$\sqrt{nh}\left(\hat{g}_L(u, \hat{\boldsymbol{\eta}}) - g(u) - \frac{\mu_{K,2}h^2}{2} g''(u)\right) \xrightarrow{L} N\left(0, \frac{\kappa\mu_{K^2}}{f_{\boldsymbol{\beta}_{\boldsymbol{\phi}_0}}(u)}\right),$$

where $\mu_{K,2} = \int t^2 K(t)\text{d}t$ and $\mu_{K^2} = \int K^2(t)\text{d}t$.

The asymptotic results of Theorem 3 are the same as Theorem 3.1 obtained in Liu and Xia (2018). Directly using the asymptotic results of Theorem 1 in Li et al. (2014), we have

$$\sqrt{nh}\left(\hat{g}(u, \hat{\boldsymbol{\eta}}_{\text{T}}) - g(u) - \frac{\mu_{K,2}h^2}{2} g''(u)\right) \xrightarrow{L} N\left(0, \frac{v_0(u)\mu_{K^2}}{f_{\boldsymbol{\beta}_{\boldsymbol{\phi}_0}}(u)}\right),$$

where $v_0(u) = E\left([\ln(\epsilon)]^2 | \boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\text{T}} \boldsymbol{Z} = u\right)$. If ϵ is independent of $(\boldsymbol{X}^{\text{T}}, \boldsymbol{Z}^{\text{T}})^{\text{T}}$, the term $v_0(u)$ reduces to $\text{Var}(\ln(\epsilon))$. It is seen that the asymptotic biases of $\hat{g}_L(u)$ and $\hat{g}(u, \hat{\boldsymbol{\eta}}_{\text{T}})$ are the same, while the main difference is the asymptotic variances. Both two local linear estimators behave locally as well as the oracle estimate because TPLSE estimator $\hat{\boldsymbol{\eta}}_{\text{T}}$ and PLPRE estimator $\hat{\boldsymbol{\eta}}$ are both root-n consistent, faster than the root-(nh) convergence rate. The comparison of the efficiency between two estimators can be obtained locally through the asymptotical variance or globally through asymptotic integrated mean squared error (AIMSE). The asymptotic results of Theorem 3 can also be used to construct the simultaneous confidence bands of $g(u)$ for various inference tasks (Li et al. 2014). These deserve further study.

We estimate $\boldsymbol{\eta}_0$ in (4) by using the estimator $\hat{g}(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\text{T}} z, \boldsymbol{\eta})$, which is obtained by local linear estimate with the logarithmic transformation on the response variable, and the theoretical results of $\hat{\boldsymbol{\eta}}$ coincide with Liu and Xia (2018) and Zhang et al. (2018) if model error ϵ and $\boldsymbol{Z}$ are conditionally independent given $\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\text{T}} \boldsymbol{Z}$ and $\boldsymbol{\alpha}_0 = \boldsymbol{0}$. Moreover, we can also obtain the estimator $\hat{\boldsymbol{\eta}}_L$ of $\boldsymbol{\eta}_0$ by iterating the LPLPRE estimator $\hat{g}_L(u, \hat{\boldsymbol{\eta}})$ between (4) and (5), i.e.,

$$\begin{aligned}\hat{\boldsymbol{\eta}}_L &= \left(\hat{\boldsymbol{\alpha}}_L^{\mathrm{T}}, \hat{\boldsymbol{\phi}}_L^{\mathrm{T}}\right)^{\mathrm{T}} \\ &= \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^q, \boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\phi}<1} \sum_{i=1}^{n} \left\{ Y_i \exp\left(-\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{X}_i - \hat{g}_L(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}\boldsymbol{Z}_i, \boldsymbol{\eta})\right)\right. \\ &\quad \left. + Y_i^{-1}\exp\left(\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{X}_i + \hat{g}_L(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}\boldsymbol{Z}_i, \boldsymbol{\eta})\right)\right\}.\end{aligned} \tag{9}$$

Liu and Xia (2018) proposed to use minimization (9) by estimating $(\boldsymbol{\beta}_{\boldsymbol{\phi}_0}, g(u))$ in multiplicative single-index models ($\boldsymbol{\alpha} = \mathbf{0}$). It is noted that there are no explicit solutions for estimators $\hat{\boldsymbol{\eta}}_L$ and $\hat{g}_L(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}z, \boldsymbol{\eta})$ in minimization (9), and we can consider to iterative estimate $\boldsymbol{\eta}_0$ and $g(u)$ numerically.

In the following, we define $m_{\epsilon,\boldsymbol{X}}(u) = E\left[(\epsilon+\epsilon^{-1})\boldsymbol{X}\middle|\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z} = u\right]$, $m_{\epsilon,\boldsymbol{Z}}(u) = E\left[(\epsilon+\epsilon^{-1})\boldsymbol{Z}\middle|\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z} = u\right]$ and $m_{\epsilon}(u) = E\left[(\epsilon+\epsilon^{-1})\middle|\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z} = u\right]$. Similar to the Proposition 1 and Theorem 1 in Cui et al. (2011), we can have the following two propositions.

Proposition 1 *Under the conditions (C1)–(C3), as* $n \to 0$, $h \to 0$, $\frac{(\log n)^{1+s_0}}{nh^3} \to 0$ *for some* $s_0 > 0$, *we have*

$$\frac{\partial \hat{g}_L(\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}z, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}_0} = \begin{pmatrix} -\dfrac{m_{\epsilon,\boldsymbol{X}}(\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}z)}{m_{\epsilon}(\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}z)} \\ g'(\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}z)\boldsymbol{J}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\left(z - \dfrac{m_{\epsilon,\boldsymbol{Z}}(\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}z)}{m_{\epsilon}(\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}z)}\right)\end{pmatrix} + O_P\left(h^2 + \sqrt{\frac{(\log n)^{1+s_0}}{nh^3}}\right).$$

Let

$$\boldsymbol{\Gamma}_{L,0} = E\left\{(\epsilon+\epsilon^{-1})\begin{pmatrix} \boldsymbol{X} - \dfrac{m_{\epsilon,\boldsymbol{X}}(\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z})}{m_{\epsilon}(\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z})} \\ g'(\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z})\boldsymbol{J}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\left(\boldsymbol{Z} - \dfrac{m_{\epsilon,\boldsymbol{Z}}(\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z})}{m_{\epsilon}(\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z})}\right)\end{pmatrix}^{\otimes 2}\right\},$$

and

$$\boldsymbol{\Sigma}_{L,0} = E\left\{(\epsilon-\epsilon^{-1})^2\begin{pmatrix} \boldsymbol{X} - \dfrac{m_{\epsilon,\boldsymbol{X}}(\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z})}{m_{\epsilon}(\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z})} \\ g'(\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z})\boldsymbol{J}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\left(\boldsymbol{Z} - \dfrac{m_{\epsilon,\boldsymbol{Z}}(\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z})}{m_{\epsilon}(\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z})}\right)\end{pmatrix}^{\otimes 2}\right\}.$$

Proposition 2 *Under the conditions of Proposition 1, we have*

$$\sqrt{n}\left(\hat{\boldsymbol{\eta}}_L - \boldsymbol{\eta}_0\right) \stackrel{L}{\longrightarrow} N\left(\mathbf{0}_{p+q-1}, \boldsymbol{\Gamma}_{L,0}^{-1}\boldsymbol{\Sigma}_{L,0}\boldsymbol{\Gamma}_{L,0}^{-1}\right).$$

The asymptotic result of Proposition 2 is slightly different from Theorem 3.2 in Liu and Xia (2018). Liu and Xia (2018) take derivative at the point $\boldsymbol{\beta}_{\boldsymbol{\phi}_0}$ lying on the boundary of a unit ball $\|\boldsymbol{\beta}_{\boldsymbol{\phi}_0}\| = 1$, and Theorem 3.2 in Liu and Xia (2018) involves some generalized inverse matrices. We transform the boundary of a unit ball in $\mathbb{R}^p$ to the interior of a unit ball in $\mathbb{R}^{p-1}$ by taking derivative at the point $\boldsymbol{\phi}_0$ for the single-index parameter. If model error ϵ and $(\boldsymbol{X}^{\mathrm{T}}, \boldsymbol{Z}^{\mathrm{T}})$ are conditionally independent given $\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z}$, both estimators $\hat{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{\eta}}_L$ are asymptotic equivalent.

3 A hypothesis test

In previous section, we discuss the estimation of $\boldsymbol{\eta}_0 = (\boldsymbol{\alpha}_0^{\mathrm{T}}, \boldsymbol{\phi}_0^{\mathrm{T}})^{\mathrm{T}}$. Further interesting topic is to see whether certain explanatory variables influence the response significantly. Without loss of generality, we consider the linear hypothesis testing problem:

$$\mathcal{H}_0 : \boldsymbol{A}\boldsymbol{\eta}_0 = \boldsymbol{b}, \quad \text{vs} \quad \mathcal{H}_1 : \boldsymbol{A}\boldsymbol{\eta}_0 \neq \boldsymbol{b}, \tag{10}$$

where $\boldsymbol{A}$ is a known constant $k \times (p+q-1)$ matrix and $\boldsymbol{b}$ is a known constant k-vector. We shall also assume that $\mathrm{rank}(\boldsymbol{A}) = k \leq (p+q-1)$. For different purposes of testing components of $\boldsymbol{\alpha}_0$ and $\boldsymbol{\phi}_0$, we can set $\boldsymbol{A} = (\boldsymbol{A}_1, \mathbf{0})$ or $\boldsymbol{A} = (\mathbf{0}, \boldsymbol{A}_2)$ similarly.

3.1 A restricted estimation

If the null hypothesis $\mathcal{H}_0$ holds, the restriction condition $\boldsymbol{A}\boldsymbol{\eta}_0 = \boldsymbol{b}$ should be considered to obtain an estimator of $\boldsymbol{\eta}_0$. Following Wei and Wang (2012), we construct a restricted estimator by using Lagrange multiplier technique:

$$\begin{aligned}\hat{\boldsymbol{\eta}}_{\mathrm{R}} &= (\hat{\boldsymbol{\alpha}}_{\mathrm{R}}^{\mathrm{T}}, \hat{\boldsymbol{\phi}}_{\mathrm{R}}^{\mathrm{T}})^{\mathrm{T}} \\ &= \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^q, \boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\phi}<1} \left\{ \sum_{i=1}^n \left[Y_i \exp\left(-\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{X}_i - \hat{g}(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}\boldsymbol{Z}_i, \boldsymbol{\eta})\right)\right.\right. \\ &\quad \left.\left. + Y_i^{-1} \exp\left(\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{X}_i + \hat{g}(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}\boldsymbol{Z}_i, \boldsymbol{\eta})\right)\right] + \boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{\eta} - \boldsymbol{b}) \right\},\end{aligned} \tag{11}$$

where $\boldsymbol{\lambda}$ is a $k \times 1$ vector of the Lagrange multipliers. Differentiating quantity (11) with respect to $\boldsymbol{\eta}$ and $\boldsymbol{\lambda}$, the restricted estimator $\hat{\boldsymbol{\eta}}_{\mathrm{R}}$ is obtained as

$$\begin{aligned}\hat{\boldsymbol{\eta}}_{\mathrm{R}} &= \left(\hat{\boldsymbol{\alpha}}_{\mathrm{R}}^{\mathrm{T}}, \hat{\boldsymbol{\phi}}_{\mathrm{R}}^{\mathrm{T}}\right)^{\mathrm{T}} \\ &= \hat{\boldsymbol{\eta}} - \left[\widetilde{\mathfrak{M}}_n(\hat{\boldsymbol{\eta}})\right]^{-1} \boldsymbol{A}^{\mathrm{T}} \left[\boldsymbol{A}\left\{\widetilde{\mathfrak{M}}_n(\hat{\boldsymbol{\eta}})\right\}^{-1}\boldsymbol{A}^{\mathrm{T}}\right]^{-1} (\boldsymbol{A}\hat{\boldsymbol{\eta}} - \boldsymbol{b}),\end{aligned} \tag{12}$$

where $\widetilde{\mathfrak{M}}_n(\hat{\boldsymbol{\eta}})$ used in (12) is defined in (6) by substituting $\hat{\boldsymbol{\eta}}_{\text{old}}$ with $\hat{\boldsymbol{\eta}}$. For the single-index parameter $\boldsymbol{\beta}_{\boldsymbol{\phi}_0}$, the restricted estimator $\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}_{\text{R}}}$ under the null hypothesis $\mathcal{H}_0$ is defined as $\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}_{\text{R}}} = \left(\sqrt{1-\|\hat{\boldsymbol{\phi}}_{\text{R}}\|^2}, \hat{\boldsymbol{\phi}}_{\text{R}}^{\text{T}}\right)^{\text{T}}$.

In the following we present the asymptotic results for estimator $\hat{\boldsymbol{\eta}}_{\text{R}}$. Let $\boldsymbol{\Omega}_0 = \boldsymbol{I}_{p+q-1} - \boldsymbol{\Gamma}_0^{-1}\boldsymbol{A}^{\text{T}}\left(\boldsymbol{A}\boldsymbol{\Gamma}_0^{-1}\boldsymbol{A}^{\text{T}}\right)^{-1}\boldsymbol{A}$, where $\boldsymbol{I}_{p+q-1}$ is an $(p+q-1)$-dimensional identity matrix.

Theorem 4 *Suppose conditions in Theorem 2 hold, under the null hypothesis $\mathcal{H}_0$ in (10), we have*

$$\sqrt{n}\left(\hat{\boldsymbol{\eta}}_{\text{R}} - \boldsymbol{\eta}_0\right) \xrightarrow{L} N\left(\mathbf{0}_{p+q-1}, \boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Sigma}_0\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Omega}_0^{\text{T}}\right).$$

If ϵ is independent of $(\boldsymbol{X}^{\text{T}}, \boldsymbol{Z}^{\text{T}})^{\text{T}}$, let $\boldsymbol{\Pi}_0 = \boldsymbol{I}_{p+q-1} - \boldsymbol{\Lambda}_0^{-1}\boldsymbol{A}^{\text{T}}\left(\boldsymbol{A}\boldsymbol{\Lambda}_0^{-1}\boldsymbol{A}^{\text{T}}\right)^{-1}\boldsymbol{A}$, we have

$$\sqrt{n}\left(\hat{\boldsymbol{\eta}}_{\text{R}} - \boldsymbol{\eta}_0\right) \xrightarrow{L} N\left(\mathbf{0}_{p+q-1}, \kappa\boldsymbol{\Pi}_0\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Pi}_0^{\text{T}}\right).$$

3.2 A test statistic

To test null hypothesis $\mathcal{H}_0$ in (10), our test statistic is proposed as

$$\mathcal{T}_n = \left(\boldsymbol{A}\hat{\boldsymbol{\eta}} - \boldsymbol{b}\right)^{\text{T}}\left\{\boldsymbol{A}\left[\widetilde{\mathfrak{M}}_n(\hat{\boldsymbol{\eta}})\right]^{-1}\widetilde{\mathfrak{S}}_n(\hat{\boldsymbol{\eta}})\left[\widetilde{\mathfrak{M}}_n(\hat{\boldsymbol{\eta}})\right]^{-1}\boldsymbol{A}^{\text{T}}\right\}^{-1} \times \left(\boldsymbol{A}\hat{\boldsymbol{\eta}} - \boldsymbol{b}\right), \tag{13}$$

where $\widetilde{\mathfrak{S}}_n(\hat{\boldsymbol{\eta}})$ is defined as

$$\widetilde{\mathfrak{S}}_n(\hat{\boldsymbol{\eta}}) = \sum_{i=1}^{n}\left[(\hat{\epsilon}_i - \hat{\epsilon}_i^{-1})\begin{pmatrix}\boldsymbol{X}_i + \dfrac{\partial\hat{g}(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}^{\text{T}}\boldsymbol{Z}_i, \hat{\boldsymbol{\eta}})}{\partial\hat{\boldsymbol{\alpha}}} \\ \dfrac{\partial\hat{g}(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}^{\text{T}}\boldsymbol{Z}_i, \hat{\boldsymbol{\eta}})}{\partial\hat{\boldsymbol{\phi}}}\end{pmatrix} - \hat{M}(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}^{\text{T}}\boldsymbol{Z}_i)\ln(\hat{\epsilon}_i)\right]^{\otimes 2}.$$

where $\hat{\epsilon}_i = Y_i \exp\left(-\hat{\boldsymbol{\alpha}}^{\text{T}}\boldsymbol{X}_i - \hat{g}(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}^{\text{T}}\boldsymbol{Z}_i, \hat{\boldsymbol{\eta}})\right)$, $\frac{\partial\hat{g}(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}^{\text{T}}\boldsymbol{Z}_i,\hat{\boldsymbol{\eta}})}{\partial\hat{\boldsymbol{\alpha}}} = \left.\frac{\partial\hat{g}(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\text{T}}\boldsymbol{Z}_i,\boldsymbol{\eta})}{\partial\boldsymbol{\alpha}}\right|_{\boldsymbol{\eta}=\hat{\boldsymbol{\eta}}}$ and $\frac{\partial\hat{g}(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}^{\text{T}}\boldsymbol{Z}_i,\hat{\boldsymbol{\eta}})}{\partial\hat{\boldsymbol{\phi}}} = \left.\frac{\partial\hat{g}(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\text{T}}\boldsymbol{Z}_i,\boldsymbol{\eta})}{\partial\boldsymbol{\phi}}\right|_{\boldsymbol{\eta}=\hat{\boldsymbol{\eta}}}$, $i = 1, \ldots, n$. And, $\hat{M}(u)$ is the local linear estimator of $E\left[(\epsilon+\epsilon^{-1})(\breve{\boldsymbol{X}}^{\text{T}}, \breve{\boldsymbol{Z}}^{\text{T}}\boldsymbol{J}_{\boldsymbol{\phi}_0})^{\text{T}}|\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\text{T}}\boldsymbol{Z} = u\right]$, which is estimated as

$$\hat{M}(u) = \frac{\sum_{i=1}^{n} b_{n,i}(u)\hat{\boldsymbol{L}}_i}{\sum_{i=1}^{n} b_{n,i}(u)}, \quad \hat{\boldsymbol{L}}_i = (\hat{\epsilon}_i + \hat{\epsilon}_i^{-1}) \begin{pmatrix} \boldsymbol{X}_i + \dfrac{\partial \hat{g}(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}^{\mathrm{T}} \boldsymbol{Z}_i, \hat{\boldsymbol{\eta}})}{\partial \hat{\boldsymbol{\alpha}}} \\ \dfrac{\partial \hat{g}(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}^{\mathrm{T}} \boldsymbol{Z}_i, \hat{\boldsymbol{\eta}})}{\partial \hat{\boldsymbol{\phi}}} \end{pmatrix}.$$

If ϵ is independent with $(\boldsymbol{X}^{\mathrm{T}}, \boldsymbol{Z}^{\mathrm{T}})^{\mathrm{T}}$, the test statistic for $\mathcal{H}_0$ is proposed as

$$\mathcal{T}_{\mathrm{id},n} = n\hat{\boldsymbol{\kappa}}^{-1} \left(\boldsymbol{A}\hat{\boldsymbol{\eta}} - \boldsymbol{b}\right)^{\mathrm{T}} \left(\boldsymbol{A}\widehat{\boldsymbol{\Lambda}}^{-1}\boldsymbol{A}^{\mathrm{T}}\right)^{-1} \left(\boldsymbol{A}\hat{\boldsymbol{\eta}} - \boldsymbol{b}\right), \tag{14}$$

where $\hat{\boldsymbol{\kappa}}^{-1} = c_{n,1}/c_{n,2}$, and $c_{n,1}$, $c_{n,2}$ and $\widehat{\boldsymbol{\Lambda}}$ are defined as

$$c_{n,1} = \left\{\frac{1}{n}\sum_{i=1}^{n}(\hat{\epsilon}_i + \hat{\epsilon}_i^{-1})\right\}^2, \quad c_{n,2} = \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\epsilon}_i - \hat{\epsilon}_i^{-1}\right\}^2$$

and

$$\widehat{\boldsymbol{\Lambda}} = \frac{1}{n}\sum_{i=1}^{n}\left[\begin{pmatrix} \boldsymbol{X}_i + \dfrac{\partial \hat{g}(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}^{\mathrm{T}} \boldsymbol{Z}_i, \hat{\boldsymbol{\eta}})}{\partial \hat{\boldsymbol{\alpha}}} \\ \dfrac{\partial \hat{g}(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}^{\mathrm{T}} \boldsymbol{Z}_i, \hat{\boldsymbol{\eta}})}{\partial \hat{\boldsymbol{\phi}}} \end{pmatrix}^{\otimes 2}\right].$$

Theorem 5 *Suppose conditions in Theorem 2 hold, under the null hypothesis in (10), we have* $\mathcal{T}_n \xrightarrow{L} \chi_k^2$. *If* ϵ *is independent with* $(\boldsymbol{X}^{\mathrm{T}}, \boldsymbol{Z}^{\mathrm{T}})^{\mathrm{T}}$*, we have* $\mathcal{T}_{\mathrm{id},n} \xrightarrow{L} \chi_k^2$.

To study power, we consider the local alternative hypothesis

$$\mathcal{H}_{1n} : \boldsymbol{A}\boldsymbol{\eta}_0 = \boldsymbol{b} + n^{-1/2}\boldsymbol{c}, \quad \boldsymbol{c} \neq 0. \tag{15}$$

We have the following asymptotic results.

Theorem 6 *Suppose conditions in Theorem 2 hold, under local alternative hypothesis (15), let* $\boldsymbol{\pi} = -\boldsymbol{\Gamma}_0^{-1}\boldsymbol{A}^{\mathrm{T}}\left(\boldsymbol{A}\boldsymbol{\Gamma}_0^{-1}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\boldsymbol{c}$*, we have*

$$\sqrt{n}\left(\hat{\boldsymbol{\eta}}_{\mathrm{R}} - \boldsymbol{\eta}_0\right) \xrightarrow{L} N\left(\boldsymbol{\pi}, \boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Sigma}_0\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Omega}_0^{\mathrm{T}}\right).$$

And if ϵ *is independent of* $(\boldsymbol{X}^{\mathrm{T}}, \boldsymbol{Z}^{\mathrm{T}})^{\mathrm{T}}$*, we have*

$$\sqrt{n}\left(\hat{\boldsymbol{\eta}}_{\mathrm{R}} - \boldsymbol{\eta}_0\right) \xrightarrow{L} N\left(\boldsymbol{\pi}, \boldsymbol{\kappa}\boldsymbol{\Pi}_0\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Pi}_0^{\mathrm{T}}\right).$$

Theorem 7 *Suppose conditions in Theorem 2 hold, under the local alternative hypothesis* $\mathcal{H}_{1n}$ *in (15), we have* $\mathcal{T}_n \xrightarrow{L} \chi_k^2(\tau_0)$*, where* $\chi_k^2(\tau_0)$ *is a noncentral Chi-squared distributions with k degrees of freedom and noncentrality parameter* $\tau_0 = \boldsymbol{c}^{\mathrm{T}} \left(\boldsymbol{A}\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Sigma}_0\boldsymbol{\Gamma}_0^{-1}\boldsymbol{A}^{\mathrm{T}} \right)^{-1} \boldsymbol{c}$*. Furthermore, if* ϵ *is independent with* $(\boldsymbol{X}^{\mathrm{T}}, \boldsymbol{Z}^{\mathrm{T}})^{\mathrm{T}}$*, we have* $\mathcal{T}_{\mathrm{id},n} \xrightarrow{L} \chi_k^2(\rho_0)$*, where* $\chi_k^2(\rho_0)$ *is a noncentral Chi-squared distributions with k degrees of freedom and noncentrality parameter* $\rho_0 = \boldsymbol{\kappa}^{-1}\boldsymbol{c}^{\mathrm{T}} \left(\boldsymbol{A}\boldsymbol{\Lambda}_0^{-1}\boldsymbol{A}^{\mathrm{T}} \right)^{-1} \boldsymbol{c}$*.*

4 Variable selection

As we described in introduction, the LPRE criteria are strictly convex. We adopt the convex property to propose the following penalized PLPRE function for variable selection. The penalized estimator $\hat{\boldsymbol{\eta}}_P = \left(\hat{\boldsymbol{\alpha}}_P^{\mathrm{T}}, \hat{\boldsymbol{\phi}}_P^{\mathrm{T}} \right)^{\mathrm{T}}$ of $\boldsymbol{\eta}_0$ is defined as

$$\hat{\boldsymbol{\eta}}_P = \arg \min_{\boldsymbol{\alpha}\in\mathbb{R}^q, \boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\phi}<1} \left\{ \sum_{i=1}^{n} \left[Y_i \exp\left(-\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{X}_i - \hat{g}(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}\boldsymbol{Z}_i, \boldsymbol{\eta}) \right) + Y_i^{-1} \exp\left(\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{X}_i + \hat{g}(\boldsymbol{\beta}_{\boldsymbol{\phi}}^{\mathrm{T}}\boldsymbol{Z}_i, \boldsymbol{\eta}) \right) \right] + n \sum_{s=1}^{p+q-1} p_{\zeta_s}(|\eta_s|) \right\}, \tag{16}$$

where $p_\zeta(\cdot)$ is a penalty function with a tuning parameter ζ. For different purposes of selecting nonzero components of $\boldsymbol{X}$ and $\boldsymbol{Z}$, if we let $p_{\zeta_j}(\cdot) = 0$ for $j = 1, \ldots, q$, we aim at selecting $\boldsymbol{Z}$-variables only; if we let $p_{\zeta_j}(\cdot) = 0$ for $j = q+1, \ldots, p+q-1$, we select covariate $\boldsymbol{X}$.

There have been various penalty functions for variable selection problems. For example, the L_2-penalty $p_\zeta(|t|) = \zeta|t|^2$ results in a ridge regression. The L_1-penalty $p_\zeta(|t|) = \zeta|t|$ yields the least absolute shrinkage and selection operator (lasso) method (Tibshirani 1996). Fan and Peng (2004) further proposed the smoothly clipped absolute deviation (SCAD) method in a high-dimensional setting. The SCAD penalty function $p_\zeta(\cdot)$ satisfies $p_\zeta(0) = 0$, $p'_\zeta(0+) > 0$, and its first-order derivative is

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

where $(s)_+ = sI(s > 0)$ is the hinge loss function; a is some positive constant with $a > 2$. From Bayesian statistical point of view, Fan and Peng (2004) suggested to use $a = 3.7$ and this value will be used throughout this paper.

Next, we study the sampling property of the resulting penalized PLPRE estimators. Without loss of generality, assume that $\boldsymbol{\eta}_0 = \left(\boldsymbol{\eta}_{0,\mathbf{1}}, \boldsymbol{\eta}_{0,\mathbf{2}} \right)$, $\boldsymbol{\eta}_{0,\mathbf{1}} = (\boldsymbol{\alpha}_{0,\mathbf{1}}, \boldsymbol{\phi}_{0,\mathbf{1}})$, $\boldsymbol{\alpha}_{0,\mathbf{1}}$ is $q_0 \times 1$ nonzero components of $\boldsymbol{\alpha}_0$, $\boldsymbol{\phi}_{0,\mathbf{1}}$ is $(p_0 - 1) \times 1$ nonzero components of $\boldsymbol{\phi}_0$.

Moreover, $\boldsymbol{\eta}_{0,\mathbf{2}} = (\boldsymbol{\alpha}_{0,\mathbf{2}}, \boldsymbol{\phi}_{0,\mathbf{2}})$, $\boldsymbol{\alpha}_{0,\mathbf{2}}$ is a $(q - q_0) \times 1$ vector with zeros, and $\boldsymbol{\phi}_{0,\mathbf{2}}$ is a $(p - p_0) \times 1$ vector with zeros. Note that $\boldsymbol{\beta}_{0,1} > 0$ by the identifiability condition of the single-index parameter and thus covariate Z_1 is included in the final model. We denote that $\boldsymbol{X}_{\mathbf{1}}$ consists of the first q_0 components of $\boldsymbol{X}$ corresponding to $\boldsymbol{\alpha}_{0,\mathbf{1}}$, and $\boldsymbol{Z}_{\mathbf{1}}$ consists of the first p_0 components of $\boldsymbol{Z}$ corresponding to $(\beta_{0,1}, \boldsymbol{\phi}_{0,\mathbf{1}}^{\mathrm{T}})^{\mathrm{T}}$. We define the following notations

$$\begin{aligned}
\boldsymbol{\Sigma}_{0,\boldsymbol{\zeta}_{\mathbf{1}}} &= \operatorname{diag}\Big(p''_{\zeta_1}(|\alpha_{0,1}|), \ldots, p''_{\zeta_{q_0}}(|\alpha_{0,q_0}|), p''_{\zeta_{q_0+1}}(|\phi_{0,1}|), \ldots, \\
&\qquad p''_{\zeta_{q_0+p_0-1}}(|\phi_{0,p_0-1}|)\Big), \\
\mathcal{R}_{\boldsymbol{\alpha}_{0,\mathbf{1}}} &= \Big(p'_{\zeta_1}(|\alpha_{0,1}|)\operatorname{sign}(\alpha_{0,1}), \ldots, p'_{\zeta_{q_0}}(|\alpha_{0,q_0}|)\operatorname{sign}(\alpha_{0,q_0})\Big), \\
\mathcal{R}_{\boldsymbol{\phi}_{0,\mathbf{1}}} &= \Big(p'_{\zeta_{q_0+1}}(|\phi_{0,1}|)\operatorname{sign}(\phi_{0,1}), \ldots, p'_{\zeta_{q_0+p_0-1}}(|\phi_{0,p_0-1}|)\operatorname{sign}(\phi_{0,p_0-1})\Big),
\end{aligned}$$

and $\boldsymbol{J}_{\boldsymbol{\phi}_{0,\mathbf{1}}} = \begin{pmatrix} -\boldsymbol{\phi}_{0,\mathbf{1}}^{\mathrm{T}} \Big/ \sqrt{1 - \|\boldsymbol{\phi}_{0,\mathbf{1}}\|^2} \\ \boldsymbol{I}_{p_0-1} \end{pmatrix}$, where $\boldsymbol{I}_{p_0-1}$ is an $(p_0 - 1)$-dimensional identity matrix. Moreover, we define $\breve{\boldsymbol{X}}_{\mathbf{1}} = \boldsymbol{X}_{\mathbf{1}} - E[\boldsymbol{X}_{\mathbf{1}}|\boldsymbol{\beta}_{\boldsymbol{\phi}_{0,1}}^{\mathrm{T}}\boldsymbol{Z}_{\mathbf{1}}]$, $\breve{\boldsymbol{Z}}_{\mathbf{1}} = (\boldsymbol{Z}_{\mathbf{1}} - E[\boldsymbol{Z}_{\mathbf{1}}|\boldsymbol{\beta}_{\boldsymbol{\phi}_{0,1}}^{\mathrm{T}}\boldsymbol{Z}_{\mathbf{1}}])g'(\boldsymbol{\beta}_{\boldsymbol{\phi}_{0,1}}^{\mathrm{T}}\boldsymbol{Z}_{\mathbf{1}})$,

$$\begin{aligned}
\boldsymbol{\Lambda}_{0,\mathbf{1}} &= E\left\{\begin{pmatrix} \breve{\boldsymbol{X}}_{\mathbf{1}} \\ \boldsymbol{J}_{\boldsymbol{\phi}_{0,1}}^{\mathrm{T}}\breve{\boldsymbol{Z}}_{\mathbf{1}} \end{pmatrix}^{\otimes 2}\right\}, \quad \boldsymbol{\Gamma}_{0,\mathbf{1}} = E\left\{(\epsilon + \epsilon^{-1})\begin{pmatrix} \breve{\boldsymbol{X}}_{\mathbf{1}} \\ \boldsymbol{J}_{\boldsymbol{\phi}_{0,1}}^{\mathrm{T}}\breve{\boldsymbol{Z}}_{\mathbf{1}} \end{pmatrix}^{\otimes 2}\right\}, \\
\boldsymbol{\Sigma}_{0,\mathbf{1}} &= E\Bigg[\Bigg\{(\epsilon - \epsilon^{-1})\begin{pmatrix} \breve{\boldsymbol{X}}_{\mathbf{1}} \\ \boldsymbol{J}_{\boldsymbol{\phi}_{0,1}}^{\mathrm{T}}\breve{\boldsymbol{Z}}_{\mathbf{1}} \end{pmatrix} \\
&\qquad - E\left[(\epsilon + \epsilon^{-1})\begin{pmatrix} \breve{\boldsymbol{X}}_{\mathbf{1}} \\ \boldsymbol{J}_{\boldsymbol{\phi}_{0,1}}^{\mathrm{T}}\breve{\boldsymbol{Z}}_{\mathbf{1}} \end{pmatrix}\Bigg|\boldsymbol{\beta}_{\boldsymbol{\phi}_{0,1}}^{\mathrm{T}}\boldsymbol{Z}_{\mathbf{1}}\right]\ln(\epsilon)\Bigg\}^{\otimes 2}\Bigg].
\end{aligned}$$

Theorem 8 *Denote the penalized estimator* $\hat{\boldsymbol{\eta}}_P = \left(\hat{\boldsymbol{\eta}}_{P,\mathbf{1}}^{\mathrm{T}}, \hat{\boldsymbol{\eta}}_{P,\mathbf{2}}^{\mathrm{T}}\right)^{\mathrm{T}}$. *Under conditions (C1)–(C5), the estimator* $\hat{\boldsymbol{\eta}}_P$ *satisfies:*

(a) with probability tending to one, $\hat{\boldsymbol{\eta}}_{P,\mathbf{2}} = \mathbf{0}_{p+q-p_0-q_0}$;

(b) let $\mathcal{R}_{\boldsymbol{\zeta}_{0,\mathbf{1}}} = \left(\mathcal{R}_{\boldsymbol{\alpha}_{0,\mathbf{1}}}^{\mathrm{T}}, \mathcal{R}_{\boldsymbol{\phi}_{0,\mathbf{1}}}^{\mathrm{T}}\right)^{\mathrm{T}}$,

$$\begin{aligned}
&\sqrt{n}\left(\boldsymbol{\Gamma}_{0,\mathbf{1}} + \boldsymbol{\Sigma}_{0,\boldsymbol{\zeta}_{\mathbf{1}}}\right)\left\{\left(\hat{\boldsymbol{\eta}}_{P,\mathbf{1}} - \boldsymbol{\eta}_{0,\mathbf{1}}\right) + \left(\boldsymbol{\Gamma}_{0,\mathbf{1}} + \boldsymbol{\Sigma}_{0,\boldsymbol{\zeta}_{\mathbf{1}}}\right)^{-1}\mathcal{R}_{\boldsymbol{\zeta}_{0,\mathbf{1}}}\right\} \\
&\quad \xrightarrow{L} N\left(\mathbf{0}_{q_0+p_0-1}, \boldsymbol{\Sigma}_{0,\mathbf{1}}\right).
\end{aligned}$$

(c) If ϵ *is independent of* $(\boldsymbol{X}^{\mathrm{T}}, \boldsymbol{Z}^{\mathrm{T}})^{\mathrm{T}}$, *let* $\boldsymbol{\kappa}_1 = E(\epsilon + \epsilon^{-1})$ *and* $\boldsymbol{\kappa}_2 = E(\epsilon - \epsilon^{-1})^2$, *we have*

$$\sqrt{n}\left(\boldsymbol{\kappa}_1\boldsymbol{\Lambda}_{0,\mathbf{1}}+\boldsymbol{\Sigma}_{0,\zeta_\mathbf{1}}\right)\left\{\left(\hat{\boldsymbol{\eta}}_{P,\mathbf{1}}-\boldsymbol{\eta}_{0,\mathbf{1}}\right)+\left(\boldsymbol{\kappa}_1\boldsymbol{\Lambda}_{0,\mathbf{1}}+\boldsymbol{\Sigma}_{0,\zeta_\mathbf{1}}\right)^{-1}\mathfrak{R}_{\zeta_{0,\mathbf{1}}}\right\}$$
$$\xrightarrow{L} N\left(\mathbf{0}_{q_0+p_0-1},\boldsymbol{\kappa}_2\boldsymbol{\Lambda}_{0,\mathbf{1}}\right).$$

Remark 2 The SCAD procedure automatically shrinkages zero components of $\boldsymbol{\eta}$ to zeros and selects out nonzero components of $\boldsymbol{\eta}_0$ and retains $\sqrt{n}$-normality with an extra-bias $\sqrt{n}\mathfrak{R}_{0,\zeta_\mathbf{1}}$ which is caused by the SCAD penalty function. If we impose the condition $\sqrt{n}\mathfrak{R}_{0,\zeta_\mathbf{1}}\to 0$, the asymptotic results of Theorem 8 (b)–(c) are in accordance with Theorem 1 once we had known those nonzero components beforehand. Theorem 8 also indicates that the proposed variable selection procedure processes the oracle property with proper choice of tuning parameters $\boldsymbol{\zeta}$. We discuss the selection of tuning parameter in the following.

Suggested by Liang et al. (2010), we adopt the BIC selector to choose the regularization parameters ζ_j's. We use the approach as follows: let $\zeta_j=\zeta_0\text{se}(\hat{\boldsymbol{\eta}}_j)$, where $\text{se}(\hat{\boldsymbol{\eta}}_j)$'s are the standard errors of the unpenalized PLPRE estimators for $\boldsymbol{\eta}_j$ with $j=1,\ldots,p+q-1$. The BIC score for ζ_0 can be defined as

$$\text{BIC}(\zeta_0)=\ln\{\text{LPRE}(\zeta_0)\}+\frac{\ln n}{n}N_{\zeta_0}, \tag{17}$$

and

$$\text{LPRE}(\zeta_0)=\frac{1}{n}\sum_{i=1}^{n}\left\{Y_i\exp\left(-\hat{\boldsymbol{\alpha}}_{P,\zeta}^{\mathrm{T}}\boldsymbol{X}_i-\hat{g}(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}_{P,\zeta}}^{\mathrm{T}}\boldsymbol{Z}_i,\hat{\boldsymbol{\eta}}_{P,\zeta})\right)\right.$$
$$\left.+Y_i^{-1}\exp\left(\hat{\boldsymbol{\alpha}}_{P,\zeta}^{\mathrm{T}}\boldsymbol{X}_i+\hat{g}(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}_{P,\zeta}}^{\mathrm{T}}\boldsymbol{Z}_i,\hat{\boldsymbol{\eta}}_{P,\zeta})\right)\right\},$$

where N_{ζ_0} is the number of nonzero coefficients of $\hat{\boldsymbol{\eta}}_{P,\zeta}$, $\hat{\boldsymbol{\eta}}_{P,\zeta}$ is the resulting penalized estimator of $\boldsymbol{\eta}_0$ with tuning parameter $\boldsymbol{\zeta}=(\zeta_1,\ldots,\zeta_{q+p-1})^{\mathrm{T}}$, $\zeta_j=\zeta_0\text{se}(\hat{\boldsymbol{\eta}}_j)$, and $\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}_{P,\zeta}}=\left(\sqrt{1-\|\hat{\boldsymbol{\phi}}_{P,\zeta}\|^2},\hat{\boldsymbol{\phi}}_{P,\zeta}^{\mathrm{T}}\right)^{\mathrm{T}}$. The standard error $\text{se}(\hat{\boldsymbol{\eta}}_j)$ is obtained as the square root of $n^{-1}e_j^{\mathrm{T}}\left[\widetilde{\mathfrak{M}}_n(\hat{\boldsymbol{\eta}})\right]^{-1}\widetilde{\mathfrak{S}}_n(\hat{\boldsymbol{\eta}})\left[\widetilde{\mathfrak{M}}_n(\hat{\boldsymbol{\eta}})\right]^{-1}e_j$, where e_j is the $(p+q-1)$-dimensional vector with j-th position 1 and 0 elsewhere, $j=1,\ldots,p+q-1$. Thus, the minimization problem over ζ_j's will reduce to a one-dimensional minimization problem on ζ_0. The minimizer of the tuning parameter ζ_0 can be obtained by a grid search. Thirty grid points are set to be evenly distributed over the range of ζ_0. In detail, we set 30 grid points as $\{\zeta_{0,1}<\zeta_{0,2}<\cdots<\zeta_{0,30}\}$. According to (17), we calculated $\text{BIC}(\zeta_{0,s})$, $s=1,\ldots,30$ and obtained $\zeta_{0,\min}=\arg\min\{\text{BIC}(\zeta_{0,s}),s=1,\ldots,30\}$. Then, the final tuning parameters are obtained as $\zeta_j=\zeta_{0,\min}\text{se}(\hat{\beta}_j)$, $j=1,\ldots,p+q-1$. The gird number 30 is based on experience from simulations. In practice, the range of ζ_0 shall be selected to be wide enough so that the minimizer of $\{\text{BIC}(\zeta_{0,s})\}_{s=1}^{k}$ can be approximately at the center of the range, and k grid points are set over the range of ζ_0.

5 Model checking

For model checking, we consider

$$\mathcal{H}_0^*: Y = \exp\left(\boldsymbol{\alpha}_0^{\mathrm{T}}\boldsymbol{X} + g(\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z})\right)\epsilon \quad a.s. \quad \text{for some } g(\cdot), \boldsymbol{\alpha}_0, \boldsymbol{\beta}_{\boldsymbol{\phi}_0}. \tag{18}$$

Let $\boldsymbol{S} = (\boldsymbol{X}^{\mathrm{T}}, \boldsymbol{Z}^{\mathrm{T}})^{\mathrm{T}}$, under the null hypothesis $\mathcal{H}_0^*$, we have $E(\epsilon - \epsilon^{-1}|\boldsymbol{S}) = 0$. This conditional expectation motivates us to use the integrated conditional moment (ICM) tests, which transform the conditional expectation of the null hypothesis $E(\epsilon - \epsilon^{-1}|\boldsymbol{S}) = 0$ into uncountable many moments $E\{(\epsilon - \epsilon^{-1})w(\boldsymbol{s}, \boldsymbol{S})\} = 0$ for any $\boldsymbol{s}$ with the weighting function satisfying the equivalence of the conditional expectation and infinite unconditional moments. There is an amount of the literature on the ICM tests with different weight functions, for example, exponential weighting function $\exp(\sqrt{-1}\boldsymbol{s}^{\mathrm{T}}\boldsymbol{S})$ (Escanciano 2006; Bierens 1982), logistic weighting function $\{1 + \exp(\boldsymbol{s}^{\mathrm{T}}\boldsymbol{S})\}^{-1}$ (Lee et al. 2001), simple indicator weighting function $I(\boldsymbol{S} \le \boldsymbol{s})$ (Stute 1997; Lin et al. 2002) and also the linear indicator weighting function $I(\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{S} \le u)$ for $u \in \mathbb{R}^1$ (Xia et al. 2004; Stute and Zhu 2002; Escanciano 2006). These weighting functions all lead to consistency model checking methods with different power properties. However, as noted in Ma et al. (2014), Bierens and Ploberger (1997), no best weighting function in terms of power is possible because all these weighting functions lead to asymptotic admissible tests.

In this section, we adopt to use the linear indicator weighting function $I(\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{S} \le u)$ for $u \in \mathbb{R}^1$. Escanciano (2006) gave the rationality of the linear indicator weighing function. That is, for random variables ε and $\boldsymbol{W}$ with $E\|\boldsymbol{W}\| < \infty$, $E(\varepsilon|\boldsymbol{W}) = 0$ if and only if $E(\varepsilon|\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{W}) = 0$ almost sure for any unit vector $\boldsymbol{\delta}$ (Jones 1987). So, the departure between $\epsilon - \epsilon^{-1}$ and $\boldsymbol{S}$ can be detected by a projection of the function along a certain direction.

For unit vector $\boldsymbol{\delta}$, the sample version of $E\{(\epsilon - \epsilon^{-1})I(\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{S} \le t)\}$ is defined as

$$\begin{aligned}\mathcal{R}_n(u) = \frac{1}{n}\sum_{i=1}^{n}\Big\{&Y_i \exp\left(-\hat{\boldsymbol{\alpha}}^{\mathrm{T}}\boldsymbol{X}_i - \hat{g}(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}^{\mathrm{T}}\boldsymbol{Z}_i, \hat{\boldsymbol{\eta}})\right) \\ &-Y_i^{-1}\exp\left(\hat{\boldsymbol{\alpha}}^{\mathrm{T}}\boldsymbol{X}_i + \hat{g}(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}^{\mathrm{T}}\boldsymbol{Z}_i, \hat{\boldsymbol{\eta}})\right)\Big\}I(\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{S}_i \le t).\end{aligned} \tag{19}$$

To study the asymptotic properties of $\mathcal{R}_n(u)$, we introduce the following notations

$$\boldsymbol{\Delta}(u) = E\left\{(\epsilon + \epsilon^{-1})I\{\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{S} \le u\}\begin{pmatrix}\breve{\boldsymbol{X}} \\ \boldsymbol{J}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\breve{\boldsymbol{Z}}\end{pmatrix}\right\},$$

$$\boldsymbol{\Xi}(u) = I\{\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{S} \le u\} - [\boldsymbol{\Delta}(u)]^{\mathrm{T}}\boldsymbol{\Gamma}_0^{-1}\begin{pmatrix}\breve{\boldsymbol{X}} \\ \boldsymbol{J}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\breve{\boldsymbol{Z}}\end{pmatrix},$$

$$\aleph(u) = \boldsymbol{\Delta}^{\mathrm{T}}(u)\boldsymbol{\Gamma}_0^{-1}E\left[(\epsilon + \epsilon^{-1})\begin{pmatrix}\breve{\boldsymbol{X}} \\ \boldsymbol{J}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\breve{\boldsymbol{Z}}\end{pmatrix}\middle|\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z}\right] - E\left[(\epsilon + \epsilon^{-1})I\{\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{S} \le u\}\middle|\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z}\right].$$

Theorem 9 *Under conditions (C1)–(C5) and the null hypothesis* $\mathcal{H}_0^*$, $\sqrt{n}\mathcal{R}_n(u)$ *converges to* $\mathcal{R}(u)$ *in the Skorohod space* $D[-\infty, \infty]^{p+q}$, *where* $\mathcal{R}(u)$ *is a centered Gaussian process with covariance function*

$$\begin{aligned}
&\mathrm{Cov}\{\mathcal{R}(u_1), \mathcal{R}(u_2)\} \\
&\quad = E\left\{(\epsilon - \epsilon^{-1})^2 \boldsymbol{\Xi}(u_1)\boldsymbol{\Xi}(u_2)\right\} + E\left\{[\ln(\epsilon)]^2 \aleph(u_1)\aleph(u_2)\right\} \\
&\qquad + E\left\{(\epsilon - \epsilon^{-1})\ln(\epsilon)[\boldsymbol{\Xi}(u_1)\aleph(u_2) + \boldsymbol{\Xi}(u_2)\aleph(u_1)]\right\}.
\end{aligned}$$

If we take $\boldsymbol{\delta}$ as a random variable and denote the distribution of $\boldsymbol{\delta}$ by $F_{\boldsymbol{\delta}}(\delta)$. Let $u \in \mathrm{supp}(\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{S})$, where $\mathrm{supp}(\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{S})$ is the support of $\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{S}$, $E\left[(\epsilon - \epsilon^{-1})I\{\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{S} \le u\}\right] = 0$ is equivalent to

$$E_{\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{S}_*}\left\{E\left[(\epsilon - \epsilon^{-1})I\{\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{S} \le \boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{S}_*\}\right]^2 \Big| \boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{S}_*\right\} = 0,$$

where $\boldsymbol{S}_*$ an independent copy of $\boldsymbol{S}$, and $E_{\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{S}_*}(\cdot)$ stands for taking expectation of $\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{S}_*$. Then, our test statistic is defined as

$$\mathcal{T}_n^* = \int \left\{\sqrt{n}\mathcal{R}_n(\delta^{\mathrm{T}}\boldsymbol{S}_i)\right\}^2 dF_{\boldsymbol{\delta}}(\delta). \tag{20}$$

Based on Theorem 9 and the continuous mapping theorem, we have

$$\mathcal{T}_n^* \xrightarrow{P} \int \left\{\mathcal{R}(\delta^{\mathrm{T}}\boldsymbol{s})\right\}^2 dF_{\delta^{\mathrm{T}}\boldsymbol{S}}(\delta^{\mathrm{T}}\boldsymbol{s})F_{\boldsymbol{\delta}}(\delta),$$

where $F_{\delta^{\mathrm{T}}\boldsymbol{S}}(\delta^{\mathrm{T}}\boldsymbol{s})$ is the distribution function of $\delta^{\mathrm{T}}\boldsymbol{s}$ for a given δ.

To investigate the sensitivity of the proposed test, we consider the alternative hypothetical models

$$\mathcal{H}_{1n}^* : Y = \exp\left(\boldsymbol{\alpha}_0^{\mathrm{T}}\boldsymbol{X} + g(\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z})\right)\epsilon + n^{-1/2}\exp\left(m(\boldsymbol{X}, \boldsymbol{Z})\right)\epsilon$$

with some arbitrary bounded measurable function $m(\cdot, \cdot)$. In the following, we define $M(\boldsymbol{X}, \boldsymbol{Z}) = m(\boldsymbol{X}, \boldsymbol{Z}) - \boldsymbol{\alpha}_0^{\mathrm{T}}\boldsymbol{X} - g(\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z})$, $\breve{M}(\boldsymbol{X}, \boldsymbol{Z}) = \exp(M(\boldsymbol{X}, \boldsymbol{Z})) - E\Big[\exp(M(\boldsymbol{X}, \boldsymbol{Z}))\Big|\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z}\Big]$, $\digamma_0 = E\left\{(\epsilon + \epsilon^{-1})\breve{M}(\boldsymbol{X}, \boldsymbol{Z})\begin{pmatrix}\breve{\boldsymbol{X}} \\ \boldsymbol{J}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\breve{\boldsymbol{Z}}\end{pmatrix}\right\}$; moreover, $\boldsymbol{\Upsilon}_0 = E\left\{\breve{M}(\boldsymbol{X}, \boldsymbol{Z})\begin{pmatrix}\breve{\boldsymbol{X}} \\ \boldsymbol{J}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\breve{\boldsymbol{Z}}\end{pmatrix}\right\}$.

Theorem 10 *Suppose conditions (C1)–(C5) hold, under the local alternative hypothesis* $\mathcal{H}_{1n}^*$, *we have*

$$\sqrt{n}\left(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0\right) \xrightarrow{L} N\left(\boldsymbol{\Gamma}_0^{-1}\digamma_0, \boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Sigma}_0\boldsymbol{\Gamma}_0^{-1}\right).$$

Further, if ϵ *is independent of* $(\boldsymbol{X}^{\mathrm{T}}, \boldsymbol{Z}^{\mathrm{T}})^{\mathrm{T}}$, *we have*

$$\sqrt{n}\left(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0\right) \xrightarrow{L} N\left(\boldsymbol{\Lambda}_0^{-1}\boldsymbol{\Upsilon}_0, \boldsymbol{\kappa}\boldsymbol{\Lambda}_0^{-1}\right).$$

Let $\boldsymbol{\varpi}(u) = E\left((\epsilon + \epsilon^{-1})\breve{M}(\boldsymbol{X}, \boldsymbol{Z})I\{\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{S} \le u\}\right)$, *we have* $\sqrt{n}\mathcal{R}_n(u)$ *converges to* $\mathcal{R}(u) + \mu(u)$ *in the Skorohod space* $D[-\infty, \infty]^{p+q}$, *where* $\mu(u) = \boldsymbol{\varpi}(u) - [\boldsymbol{\Delta}(u)]^{\mathrm{T}}\boldsymbol{\Gamma}_0^{-1}\digamma_0$ *and* $\mathcal{R}(u)$ *is defined in Theorem 9. Moreover,*

$$\mathcal{T}_n^* \xrightarrow{P} \int \left\{\mathcal{R}(\delta^{\mathrm{T}}\boldsymbol{s}) + \mu(\delta^{\mathrm{T}}\boldsymbol{s})\right\}^2 dF_{\delta^{\mathrm{T}}\boldsymbol{S}}(\delta^{\mathrm{T}}\boldsymbol{s})F_{\boldsymbol{\delta}}(\delta).$$

Next, we follow the bootstrap method proposed by Ma et al. (2014) to mimic the distribution of the test statistic $\mathcal{T}_n^*$.

Step 1: Compute the estimated projection direction $\hat{\boldsymbol{\delta}}$ by fitting a single-index model with "synthesis" data $\{(\hat{\epsilon}_i - \hat{\epsilon}_i^{-1})^2, \boldsymbol{S}_i\}_{i=1}^n$, $\hat{\epsilon}_i = Y_i \exp(-\hat{\boldsymbol{\alpha}}^{\mathrm{T}}\boldsymbol{X}_i - \hat{g}(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}^{\mathrm{T}}\boldsymbol{Z}_i, \hat{\boldsymbol{\eta}}))$, and $\hat{\epsilon}_i^{-1} = Y_i^{-1}\exp(\hat{\boldsymbol{\alpha}}^{\mathrm{T}}\boldsymbol{X}_i + \hat{g}(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}^{\mathrm{T}}\boldsymbol{Z}_i, \hat{\boldsymbol{\eta}}))$, $i = 1, \ldots, n$.

Step 2: Compute the test statistic $\mathcal{T}_n^* = \frac{1}{n^2}\sum_{r=1}^n \left\{\sum_{i=1}^n (\hat{\epsilon}_i - \hat{\epsilon}_i^{-1})I\{\hat{\boldsymbol{\delta}}^{\mathrm{T}}\boldsymbol{S}_i \le \hat{\boldsymbol{\delta}}^{\mathrm{T}}\boldsymbol{S}_r\}\right\}^2$.

Step 3: Generate B times positive random variables sequence $\{\varepsilon_{ib}\}_{i=1}^n$, $b = 1, \ldots, B$ from two-point distribution which, respectively, takes values $\frac{1\mp\sqrt{5}}{2}$ with probability $\frac{5\pm\sqrt{5}}{10}$ and variance 1 and compute the following arguments for each b: $Y_i^{[b]} = \exp(\hat{\boldsymbol{\alpha}}^{\mathrm{T}}\boldsymbol{X}_i + \hat{g}(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}^{\mathrm{T}}\boldsymbol{Z}_i, \hat{\boldsymbol{\eta}}) + \varepsilon_{ib} * \ln(\hat{\epsilon}_i))$, $i = 1, \ldots, n$.

Step 4: For each b, we calculate the bootstrap residuals $\hat{\epsilon}_i^{[b]}$, and we further define the bootstrap test statistics

$$\mathcal{T}_n^{*[b]} = \frac{1}{n^2}\sum_{r=1}^n \left\{\sum_{i=1}^n \left[\hat{\epsilon}_i^{[b]} - (\hat{\epsilon}_i^{[b]})^{-1}\right] I\left(\hat{\boldsymbol{\delta}}^{\mathrm{T}}\boldsymbol{S}_i \le \hat{\boldsymbol{\delta}}^{\mathrm{T}}\boldsymbol{S}_r\right)\right\}^2,$$

where $\hat{\epsilon}_i^{[b]} = Y_i^{[b]}\exp(-\hat{\boldsymbol{\alpha}}^{[b]\mathrm{T}}\boldsymbol{X}_i - \hat{g}^{[b]}(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}^{[b]\mathrm{T}}\boldsymbol{Z}_i, \hat{\boldsymbol{\eta}}^{[b]}))$, $\{\hat{\boldsymbol{\eta}}^{[b]}, \hat{g}^{[b]}(\cdot, \hat{\boldsymbol{\eta}}^{[b]})\}$ are obtained from Sect. 2.1 by using the bootstrap samples $\{Y_i^{[b]}, \boldsymbol{X}_i, \boldsymbol{Z}_i\}_{i=1}^n$.

Step 5: We calculate the $1-\kappa$ quantile of the bootstrap test statistic $\mathcal{T}_n^{*[b]}$ as the κ-level critical value.

6 Implementation

In this section, we report simulation results to evaluate the performance of the proposed estimators. In the following simulations, the Epanechnikov kernel $K(t) = 0.75(1-t^2)^+$ is used. The bandwidth h is selected according to the remark in Theorem 2.

Example 1 We generate 1000 realizations and choose the sample size to be $n = 100$, $n = 300$ and $n = 500$ from model (1). We choose $\boldsymbol{\alpha}_0 = (1, -3, 1.5, -2, -1.5)^{\mathrm{T}}$, $\boldsymbol{\beta}_{\boldsymbol{\phi}_0} = (2, 1, 2, 1)^{\mathrm{T}}/\sqrt{10}$ in this example. Covariate $(\boldsymbol{X}, \boldsymbol{Z})$ is generated from $N(\boldsymbol{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = (\sigma_{ij})_{1\le i,j\le 9}, \sigma_{ij} = 0.5^{|i-j|}$. The link function is considered as $g(\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z}) = 0.5(1+\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z})^2$ and $g(\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z}) = \exp(0.5\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z} - 0.25)$. For the model error, we considered two cases.

Case 1 (Homoscedasticity). The model error ϵ is independent of $(\boldsymbol{X}^{\mathrm{T}}, \boldsymbol{Z}^{\mathrm{T}})^{\mathrm{T}}$ and generated from $\exp(U)$, where U follows uniform distribution $\mathrm{U}[-1, 1]$.

Case 2 (Heteroscedasticity) The model error ϵ is generated from $\exp(U_{Z_1})$, where U_{Z_1} follows a uniform distribution $\mathrm{U}(-|0.2Z_1+1|, |0.2Z_1+1|)$ conditional on covariate Z_1.

Simulation results for $(\hat{\boldsymbol{\alpha}}^{\mathrm{T}}, \hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}^{\mathrm{T}})^{\mathrm{T}}$ are reported in Tables 1 and 2. In the tables, we compare PLPRE estimator $(\hat{\boldsymbol{\alpha}}^{\mathrm{T}}, \hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}^{\mathrm{T}})^{\mathrm{T}}$ and the transformation profile least squares estimator (TPLS). The TPLS estimator for $(\boldsymbol{\alpha}_0^{\mathrm{T}}, \boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}})^{\mathrm{T}}$ is obtained by using the profile least squares estimation method Liang et al. (2010) with logarithmic transformed data $\{\ln(Y_i), \boldsymbol{X}_i, \boldsymbol{Z}_i\}_{i=1}^n$. From Tables 1, 2, we see that values of PLPRE estimator and TPLS estimator are close to the true values of $(\boldsymbol{\alpha}_0^{\mathrm{T}}, \boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}})^{\mathrm{T}}$, respectively, and the values of $\mathrm{MSE}(\hat{\boldsymbol{\alpha}}, \boldsymbol{\alpha}_0)$ and $\mathrm{MSE}(\hat{\boldsymbol{\alpha}}_{\hat{\boldsymbol{\phi}}}, \boldsymbol{\beta}_{\boldsymbol{\phi}_0})$ decrease with sample size n increasing. The angles (in radians) of $\arccos(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}, \boldsymbol{\beta}_{\boldsymbol{\phi}_0})$ become closer to zero when sample size n increases to 500. We also note that the performance of PLPRE estimator is better than the TPLS estimator. This simulation shows that PLPRE estimation method is more efficient than the transformation-based method. This phenomenon is also revealed in Chen et al. (2016).

In Tables 3 and 4, we consider the restricted estimator $(\hat{\boldsymbol{\alpha}}_{\mathrm{R}}^{\mathrm{T}}, \hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}_{\mathrm{R}}}^{\mathrm{T}})^{\mathrm{T}}$ by considering the condition $\boldsymbol{A}_1 = (1, 1, 0, -1, 0, 0, 0, 0)^{\mathrm{T}}$ for $\boldsymbol{\alpha}_0$, and the condition $\boldsymbol{A}_2 = (0, 0, 0, 0, 0, 1, 0, -1)^{\mathrm{T}}$ for $\boldsymbol{\phi}_0$ (that is, $\alpha_{0,1} + \alpha_{0,2} - \alpha_{0,4} = 0$, $\beta_{0,2} - \beta_{0,4} = 0$). We find that the restricted estimators are close to the true values. As the sample size n increases, the values of MSE for the restricted estimators decrease. Both two restriction conditions $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ decrease the values of MSE, especially for the estimation of $(\alpha_{0,1}, \alpha_{0,2}, \alpha_{0,4})$ in Table 3 and the estimation of $(\beta_{0,2}, \beta_{0,4})$ in Table 4. Moreover, restricted condition $\boldsymbol{A}_2$ also decreases the values of MSE of $\arccos(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}_{\mathrm{R}}}, \boldsymbol{\beta}_{\boldsymbol{\phi}_0})$ in Table 4. This again reveals that the restricted condition for single-index parameter can improve its estimation efficiency. Based on the numerical studies reported in Tables 1, 4, the PLPRE estimation procedure for parameter $(\boldsymbol{\alpha}_0^{\mathrm{T}}, \boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}})^{\mathrm{T}}$ performs well.

Further, we shall do linear hypothesis test:

$$\mathcal{H}_0 : \beta_{0,2} = \beta_{0,4}, \quad \mathcal{H}_1 : \beta_{0,2} = \beta_{0,4} + c,$$

where $c = 0.04, 0.08, 0.12, 0.16, 0.2$ for five alternative hypothesis $\mathcal{H}_1$. Under the null hypothesis $\mathcal{H}_0$, simulation results for the rejection probabilities of test statistics $\mathcal{T}_n$ and $\mathcal{T}_{\mathrm{id},n}$ under homoscedasticity are reported in Table 5. In Table 6, we report the simulation results of test statistic $\mathcal{T}_n$ under heteroscedasticity. From Tables 5, 6, we can see that as the value of c increases, the power function increases rapidly. We can also see that as sample size n increases, the power function tends to be one, which shows that the test statistic $\mathcal{T}_n$ is powerful for the test problem. In Table 5, the model error ϵ is independent with $(\boldsymbol{X}^{\mathrm{T}}, \boldsymbol{Z}^{\mathrm{T}})^{\mathrm{T}}$, and the test statistic $\mathcal{T}_{\mathrm{id},n}$ is more powerful than $\mathcal{T}_n$ for this test problem.

The performance of estimator $\hat{g}(u)$ of $g(u)$ is evaluated using the average squared error (ASE) and the average absolute error (AAE)

$$\mathrm{ASE} = n_0^{-1} \sum_{v=1}^{n_0} \left[\hat{g}(u_v) - g(u_v)\right]^2, \quad \mathrm{AAE} = n_0^{-1} \sum_{s=1}^{n_0} \left|\hat{g}(u_v) - g(u_v)\right|,$$

where $\{u_1, \ldots, u_{n_0}\}$ are the given grid points, and $n_0 = 200$ is the number of grid points. We compare the LPLPRE estimator $\hat{g}_L(u, \hat{\boldsymbol{\eta}})$ and the TPLS estimator $\hat{g}(u, \hat{\boldsymbol{\eta}}_{\mathrm{T}})$ and report the

Table 1 Mean (M), standard error (SD) and means squared error (MSE) of $\hat{\boldsymbol{\alpha}}$

	PLPRE					TPLS				
	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$
Homoscedasticity, $g(u) = 0.5(1+u)^2$										
$n = 100$										
M	1.0034	− 3.0000	1.5024	− 1.9960	− 1.4960	1.0015	− 3.0026	1.4998	− 1.9976	− 1.4986
SD	0.0636	0.0741	0.0731	0.0702	0.0702	0.0717	0.0839	0.0830	0.0788	0.0794
MSE	4.0647	5.4934	5.3158	4.9410	4.9455	5.1467	7.0527	6.8872	6.2221	6.3147
$n = 300$										
M	1.0022	− 2.9987	1.5019	− 2.0003	− 1.4964	1.0011	− 2.9997	1.5008	− 2.0015	− 1.4976
SD	0.0358	0.0395	0.0407	0.0400	0.0382	0.0396	0.0436	0.0450	0.0441	0.0424
MSE	1.2875	1.5636	1.6593	1.6050	1.4743	1.5708	1.9018	2.0316	1.9503	1.8029
$n = 500$										
M	0.9995	− 2.9989	1.5020	− 1.9992	− 1.4974	0.9985	− 2.9998	1.5012	− 2.0000	− 1.4984
SD	0.0277	0.0312	0.0306	0.0324	0.0278	0.0305	0.0343	0.0336	0.0357	0.0306
MSE	0.7692	0.9758	0.9414	1.0542	0.7802	0.9361	1.1788	1.1306	1.2765	0.9418
Homoscedasticity, $g(u) = \exp(0.5u - 0.25)$										
$n = 100$										
M	1.0015	− 2.9991	1.4998	− 1.9981	− 1.4973	1.0002	− 3.0008	1.4981	− 1.9992	− 1.4984
SD	0.0446	0.0493	0.0504	0.0499	0.0484	0.0494	0.0546	0.0549	0.0554	0.0531
MSE	1.9930	2.4336	2.5441	2.4974	2.3502	2.4408	2.9850	3.0179	3.0680	2.8207
$n = 300$										
M	1.0004	− 2.9996	1.5018	− 1.9987	− 1.4994	0.9994	− 3.0010	1.5006	− 1.9999	− 1.5001
SD	0.0362	0.0408	0.0399	0.0413	0.0384	0.0401	0.0452	0.0440	0.0458	0.0423
MSE	1.3107	1.6688	1.5997	1.7136	1.4787	1.6105	2.0455	1.9384	2.0965	1.7927

Table 1 continued

	PLPRE					TPLS				
	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$
$n = 500$										
M	1.0010	− 2.9990	1.5013	− 2.0003	− 1.4985	1.0000	− 3.0000	1.5006	− 2.0013	− 1.4993
SD	0.0271	0.0304	0.0318	0.0290	0.0293	0.0297	0.0334	0.0349	0.0319	0.0322
MSE	0.7348	0.9291	1.0180	0.8420	0.8635	0.8853	1.1199	1.2240	1.0231	1.0372
Heteroscedasticity, $g(u) = 0.5(1+u)^2$										
$n = 100$										
M	1.0052	− 3.0002	1.5024	− 1.9959	− 1.4996	1.0031	− 3.0030	1.5000	− 1.9984	− 1.5030
SD	0.0640	0.0730	0.0737	0.0706	0.0690	0.0723	0.0833	0.0840	0.0805	0.0775
MSE	4.1296	5.3302	5.4367	5.0045	4.7622	5.2388	6.9506	7.0613	6.4873	6.0113
$n = 300$										
M	1.0006	− 3.0007	1.5010	− 1.9975	− 1.5012	0.9991	− 3.0023	1.4998	− 1.9987	− 1.5028
SD	0.0362	0.0408	0.0395	0.0404	0.0394	0.0394	0.0446	0.0436	0.0441	0.0429
MSE	1.3128	1.6687	1.5668	1.6379	1.5523	1.5592	2.0001	1.8999	1.9488	1.8480
$n = 500$										
M	1.0005	− 2.9986	1.4999	− 1.9982	− 1.4990	0.9994	− 2.9996	1.4987	− 1.9991	− 1.5001
SD	0.0284	0.0321	0.0307	0.0321	0.0291	0.0309	0.0350	0.0334	0.0352	0.0317
MSE	0.8073	1.0326	0.9416	1.0339	0.8469	0.9584	1.2279	1.1223	1.2449	1.0053
Heteroscedasticity, $g(u) = \exp(0.5u - 0.25)$										
$n = 100$										
M	1.0025	− 2.9969	1.5002	− 2.0000	− 1.4980	1.0010	− 2.9979	1.4980	− 2.0016	− 1.4993
SD	0.0457	0.0502	0.0507	0.0510	0.0502	0.0503	0.0551	0.0561	0.0561	0.0552
MSE	2.1007	2.5362	2.5762	2.6058	2.5257	2.5317	3.0446	3.1484	3.1484	3.0474

Table 1 continued

	PLPRE					TPLS				
	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$
$n = 300$										
M	1.0002	− 2.9989	1.5024	− 1.9982	− 1.4995	0.9990	− 3.0003	1.5014	− 1.9993	− 1.5008
SD	0.0348	0.0397	0.0413	0.0423	0.0413	0.0382	0.0433	0.0451	0.0460	0.0450
MSE	1.2159	1.5798	1.7148	1.7942	1.7123	1.4607	1.8738	2.0343	2.1158	2.0302
$n = 500$										
M	1.0012	− 2.9989	1.5000	− 1.9987	− 1.4982	1.0003	− 3.0000	1.4991	− 1.9995	− 1.4991
SD	0.0280	0.0317	0.0315	0.0319	0.0317	0.0304	0.0345	0.0342	0.0349	0.0342
MSE	0.7850	1.0078	0.9937	1.0215	1.0125	0.9239	1.1946	1.1734	1.2182	1.1716

MSE is in the scale of $\times 10^{-3}$

Table 2 Mean (M), standard error (SD) and means squared error (MSE) of $\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}$ and arccos $(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}, \boldsymbol{\beta}_{\boldsymbol{\phi}_0})$ (denoted as Arc)

	PLPRE					TPLS				
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	Arc	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	Arc
Homoscedasticity, $g(u) = 0.5(1+u)^2$										
$n = 100$										
M	0.6198	0.3189	0.6371	0.3175	0.0776	0.6301	0.3159	0.6299	0.3124	0.0843
SD	0.0346	0.0511	0.0402	0.0420	0.0373	0.0368	0.0570	0.0425	0.0461	0.0384
MSE	1.3611	2.6249	1.6441	1.7654	7.4261	1.3648	3.2533	1.8173	2.1434	8.5907
$n = 300$										
M	0.6276	0.3169	0.6347	0.3177	0.0394	0.6313	0.3155	0.6323	0.3160	0.0424
SD	0.0168	0.0260	0.0204	0.0214	0.0177	0.0182	0.0284	0.0210	0.0237	0.0188
MSE	0.3067	0.6776	0.4240	0.4639	1.8728	0.3350	0.8103	0.4430	0.5650	2.1541
$n = 500$										
M	0.6289	0.3168	0.6341	0.3173	0.0321	0.6316	0.3157	0.6324	0.3161	0.0331
SD	0.0136	0.0199	0.0160	0.0165	0.0131	0.0148	0.0217	0.0164	0.0182	0.0139
MSE	0.1994	0.3994	0.2612	0.2747	1.2020	0.2209	0.4715	0.2697	0.3311	1.2935
Homoscedasticity, $g(u) = \exp(0.5u - 0.25)$										
$n = 100$										
M	0.5933	0.3182	0.6448	0.3206	0.1445	0.6232	0.3175	0.6181	0.3151	0.1572
SD	0.0880	0.0961	0.0716	0.0771	0.0872	0.0995	0.1029	0.0802	0.0852	0.0700
MSE	9.2692	9.2420	5.2789	5.9693	28.4856	9.9941	10.5974	6.6436	7.2667	29.6309
$n = 300$										
M	0.6121	0.3116	0.6424	0.3152	0.1153	0.6118	0.3105	0.6261	0.3119	0.1252
SD	0.0602	0.0726	0.0559	0.0623	0.0564	0.0623	0.0805	0.0624	0.0690	0.0527
MSE	4.0405	5.2959	3.2264	3.8804	16.4913	4.1816	6.5190	3.9379	4.7811	18.4670

Table 2 continued

	PLPRE					TPLS				
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	Arc	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	Arc
$n = 500$										
M	0.6163	0.3168	0.6410	0.3158	0.0877	0.6273	0.3171	0.6294	0.3142	0.0956
SD	0.0419	0.0584	0.0412	0.0467	0.0412	0.0468	0.0646	0.0478	0.0524	0.0420
MSE	2.0190	3.4100	1.7700	2.1841	9.3973	2.2211	4.1744	2.2964	2.7521	10.9114
Heteroscedasticity, $g(u) = 0.5(1+u)^2$										
$n = 100$										
M	0.6198	0.3141	0.6410	0.3131	0.0814	0.6288	0.3160	0.6297	0.3126	0.0935
SD	0.0417	0.0531	0.0390	0.0448	0.0414	0.0446	0.0614	0.0475	0.0520	0.0415
MSE	1.8963	2.8253	1.6000	2.0213	8.3581	2.0113	3.7747	2.2676	2.7175	10.4723
$n = 300$										
M	0.6283	0.3160	0.6348	0.3161	0.0428	0.6314	0.3177	0.6297	0.3172	0.0496
SD	0.0189	0.0280	0.0207	0.0244	0.0188	0.0209	0.0323	0.0257	0.0280	0.0219
MSE	0.3768	0.7832	0.4360	0.5957	2.1926	0.4389	1.0453	0.6678	0.7889	2.9423
$n = 500$										
M	0.6292	0.3150	0.6359	0.3147	0.0331	0.6312	0.3164	0.6325	0.3157	0.0374
SD	0.0159	0.0210	0.0152	0.0188	0.0145	0.0169	0.0242	0.0181	0.0216	0.0162
MSE	0.2651	0.4460	0.2455	0.3561	1.3130	0.2879	0.5874	0.3281	0.4663	1.6702
Heteroscedasticity, $g(u) = \exp(0.5u - 0.25)$										
$n = 100$										
M	0.6217	0.3162	0.6129	0.3269	0.1205	0.6225	0.3080	0.6151	0.3111	0.1312
SD	0.0966	0.0994	0.0885	0.0908	0.0943	0.1024	0.1088	0.0978	0.0962	0.0982
MSE	9.4487	9.8811	8.2111	8.3548	23.4189	10.5893	11.9117	9.8632	9.2839	26.8657

Table 2 continued

	PLPRE					TPLS				
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	Arc	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	Arc
$n = 300$										
M	0.6157	0.3173	0.6148	0.3150	0.0966	0.6075	0.3146	0.6253	0.3190	0.1089
SD	0.0923	0.0866	0.0669	0.0738	0.0886	0.0943	0.0918	0.0730	0.0796	0.0891
MSE	8.7990	7.4944	4.7867	5.4429	17.1885	9.5127	8.4311	5.3881	6.3875	19.8126
$n = 500$										
M	0.6235	0.3164	0.6453	0.3169	0.1038	0.6251	0.3155	0.6301	0.3138	0.1093
SD	0.0489	0.0655	0.0495	0.0539	0.0522	0.0502	0.0714	0.0552	0.0584	0.0470
MSE	2.4700	4.2932	2.6206	2.9055	13.5098	2.5788	5.0984	3.0530	3.4213	14.1802

MSE is in the scale of $\times 10^{-3}$

Table 3 Mean (M), standard error (SD) and means squared error (MSE) of $\hat{\boldsymbol{\alpha}}_{\mathrm{R}}$, $\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}_{\mathrm{R}}}$ and $\arccos(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}_{\mathrm{R}}}, \boldsymbol{\beta}_{\boldsymbol{\phi}_0})$ (denoted as Arc) under the restriction condition A_1

	$\hat{\alpha}_{\mathrm{R1}}$	$\hat{\alpha}_{\mathrm{R2}}$	$\hat{\alpha}_{\mathrm{R3}}$	$\hat{\alpha}_{\mathrm{R4}}$	$\hat{\alpha}_{\mathrm{R5}}$	$\hat{\beta}_{\mathrm{R1}}$	$\hat{\beta}_{\mathrm{R2}}$	$\hat{\beta}_{\mathrm{R3}}$	$\hat{\beta}_{\mathrm{R4}}$	Arc
Homoscedasticity, $g(u) = 0.5(1+u)^2$										
$n = 100$										
M	1.0033	− 2.9995	1.5023	− 1.9961	− 1.4963	0.6200	0.3187	0.6372	0.3173	0.0771
SD	0.0610	0.0651	0.0728	0.0514	0.0670	0.0347	0.0508	0.0401	0.0418	0.0372
MSE	3.7363	4.2379	5.3049	2.6588	4.5040	1.3626	2.5853	1.6332	1.7470	7.3370
$n = 300$										
M	1.0013	− 2.9997	1.5018	− 1.9983	− 1.4972	0.6276	0.3169	0.6347	0.3176	0.0393
SD	0.0341	0.0354	0.0404	0.0271	0.0368	0.0167	0.0259	0.0204	0.0214	0.0177
MSE	1.1691	1.2550	1.6412	0.7370	1.3655	0.3047	0.6749	0.4219	0.4615	1.8636
$n = 500$										
M	0.9995	− 2.9989	1.5020	− 1.9993	− 1.4974	0.6289	0.3168	0.6341	0.3174	0.0309
SD	0.0262	0.0282	0.0305	0.0223	0.0268	0.0136	0.0199	0.0160	0.0165	0.0131
MSE	0.6889	0.7982	0.9347	0.4973	0.7290	0.1991	0.3984	0.2608	0.2744	1.1330
Homoscedasticity, $g(u) = \exp(0.5u - 0.25)$										
$n = 100$										
M	1.0015	− 2.9994	1.4997	− 1.9978	− 1.4974	0.5932	0.3185	0.6450	0.3204	0.1433
SD	0.0417	0.0448	0.0493	0.0339	0.0453	0.0877	0.0958	0.0714	0.0767	0.0870
MSE	1.7402	2.0128	2.4355	1.1557	2.0655	9.2379	9.1807	5.2652	5.9040	28.1199
$n = 300$										
M	1.0005	− 2.9995	1.5017	− 1.9990	− 1.4993	0.6123	0.3115	0.6424	0.3152	0.1149
SD	0.0344	0.0371	0.0393	0.0298	0.0370	0.0599	0.0722	0.0558	0.0622	0.0561
MSE	1.1840	1.3810	1.5879	0.8882	1.3730	3.9891	5.2334	3.2219	3.8660	16.3578

Table 3 continued

	$\hat{\alpha}_{R1}$	$\hat{\alpha}_{R2}$	$\hat{\alpha}_{R3}$	$\hat{\alpha}_{R4}$	$\hat{\alpha}_{R5}$	$\hat{\beta}_{R1}$	$\hat{\beta}_{R2}$	$\hat{\beta}_{R3}$	$\hat{\beta}_{R4}$	Arc
$n = 500$										
M	1.0005	− 2.9997	1.5013	− 1.9992	− 1.4990	0.6165	0.3167	0.6408	0.3159	0.0875
SD	0.0250	0.0278	0.0318	0.0214	0.0282	0.0419	0.0582	0.0411	0.0466	0.0411
MSE	0.6268	0.7740	1.0145	0.4607	0.7988	2.0078	3.3923	1.7626	2.1784	9.3553
Heteroscedasticity, $g(u) = 0.5(1 + u)^2$										
$n = 100$										
M	1.0050	− 3.0009	1.5029	− 1.9958	− 1.4997	0.6198	0.3140	0.6410	0.3132	0.0815
SD	0.0614	0.0648	0.0729	0.0502	0.0661	0.0416	0.0530	0.0390	0.0448	0.0412
MSE	3.7968	4.2072	5.3243	2.5438	4.3756	1.8930	2.8214	1.5975	2.0158	8.3430
$n = 300$										
M	1.0012	− 3.0000	1.5011	− 1.9988	− 1.5006	0.6282	0.3161	0.6349	0.3161	0.0427
SD	0.0344	0.0375	0.0395	0.0272	0.0373	0.0189	0.0279	0.0206	0.0243	0.0188
MSE	1.1874	1.4056	1.5658	0.7444	1.3975	0.3761	0.7821	0.4335	0.5914	2.1839
$n = 500$										
M	1.0004	− 2.9986	1.4999	− 1.9982	− 1.4991	0.6292	0.3150	0.6359	0.3148	0.0331
SD	0.0269	0.0287	0.0306	0.0229	0.0276	0.0159	0.0210	0.0152	0.0188	0.0145
MSE	0.7251	0.8286	0.9394	0.5290	0.7625	0.2642	0.4455	0.2451	0.3551	1.3103
Heteroscedasticity, $g(u) = \exp(0.5u - 0.25)$										
$n = 100$										
M	1.0015	− 2.9989	1.5004	− 1.9973	− 1.4993	0.6192	0.3161	0.6219	0.3267	0.1158
SD	0.0434	0.0462	0.0507	0.0349	0.0471	0.0957	0.0989	0.0902	0.0906	0.0997
MSE	1.8918	2.1397	2.5666	1.2304	2.2249	9.3366	9.7974	8.2479	8.3170	23.3585

Table 3 continued

	$\hat{\alpha}_{R1}$	$\hat{\alpha}_{R2}$	$\hat{\alpha}_{R3}$	$\hat{\alpha}_{R4}$	$\hat{\alpha}_{R5}$	$\hat{\beta}_{R1}$	$\hat{\beta}_{R2}$	$\hat{\beta}_{R3}$	$\hat{\beta}_{R4}$	Arc
$n = 300$										
M	1.0002	− 2.9987	1.5024	− 1.9984	− 1.4996	0.6148	0.3174	0.6148	0.3150	0.0946
SD	0.0332	0.0361	0.0413	0.0281	0.0395	0.0912	0.0864	0.0669	0.0737	0.0880
MSE	1.1022	1.3073	1.7109	0.7957	1.5621	8.7431	7.4717	4.7840	5.4319	16.6935
$n = 500$										
M	1.0010	− 2.9992	1.5000	− 1.9982	− 1.4984	0.6248	0.3163	0.6452	0.3169	0.1036
SD	0.0264	0.0289	0.0314	0.0220	0.0300	0.0483	0.0654	0.0495	0.0538	0.0521
MSE	0.6994	0.8382	0.9908	0.4882	0.9068	2.3914	4.2730	2.6177	2.9008	13.4528

MSE is in the scale of $\times 10^{-3}$

Table 4 Mean (M), standard error (SD) and means squared error (MSE) of $\hat{\boldsymbol{\alpha}}_{\mathrm{R}}$, $\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}_{\mathrm{R}}}$ and arccos $(\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}_{\mathrm{R}}}, \boldsymbol{\beta}_{\boldsymbol{\phi}_0})$ (denoted as Arc) under the restriction condition A_2

	$\hat{\alpha}_{\mathrm{R1}}$	$\hat{\alpha}_{\mathrm{R2}}$	$\hat{\alpha}_{\mathrm{R3}}$	$\hat{\alpha}_{\mathrm{R4}}$	$\hat{\alpha}_{\mathrm{R5}}$	$\hat{\beta}_{\mathrm{R1}}$	$\hat{\beta}_{\mathrm{R2}}$	$\hat{\beta}_{\mathrm{R3}}$	$\hat{\beta}_{\mathrm{R4}}$	Arc
Homoscedasticity, $g(u) = 0.5(1+u)^2$										
$n = 100$										
M	1.0032	− 2.9998	1.5021	− 1.9961	− 1.4960	0.6218	0.3181	0.6371	0.3181	0.0612
SD	0.0633	0.0736	0.0727	0.0701	0.0695	0.0333	0.0333	0.0398	0.0333	0.0361
MSE	4.0200	5.4245	5.2860	4.9308	4.8545	1.2227	1.1133	1.6050	1.1133	5.0597
$n = 300$										
M	1.0022	− 2.9986	1.5020	− 2.0004	− 1.4963	0.6280	0.3173	0.6346	0.3173	0.0315
SD	0.0357	0.0394	0.0406	0.0398	0.0382	0.0163	0.0171	0.0204	0.0171	0.0174
MSE	1.2838	1.5585	1.6569	1.5850	1.4713	0.2852	0.2962	0.4228	0.2962	1.3008
$n = 500$										
M	0.9995	− 2.9989	1.5020	− 1.9993	− 1.4974	0.6291	0.3171	0.6341	0.3171	0.0249
SD	0.0276	0.0311	0.0305	0.0324	0.0278	0.0132	0.0132	0.0160	0.0132	0.0132
MSE	0.7662	0.9732	0.9390	1.0536	0.7787	0.1854	0.1754	0.2589	0.1754	0.7955
Homoscedasticity, $g(u) = \exp(0.5u - 0.25)$										
$n = 100$										
M	1.0014	− 2.9992	1.4998	− 1.9981	− 1.4973	0.5985	0.3199	0.6458	0.3199	0.1209
SD	0.0443	0.0492	0.0495	0.0495	0.0481	0.0814	0.0642	0.0707	0.0642	0.0821
MSE	1.9715	2.4207	2.4512	2.4560	2.3209	7.7853	4.1330	5.1786	4.1330	21.3730
$n = 300$										
M	1.0004	− 2.9996	1.5018	− 1.9987	− 1.4993	0.6150	0.3139	0.6428	0.3139	0.0926
SD	0.0361	0.0407	0.0398	0.0412	0.0384	0.0569	0.0500	0.0551	0.0500	0.0561
MSE	1.3069	1.6586	1.5873	1.6983	1.4773	3.5375	2.5118	3.1460	2.5118	11.7377

Table 4 continued

	$\hat{\alpha}_{R1}$	$\hat{\alpha}_{R2}$	$\hat{\alpha}_{R3}$	$\hat{\alpha}_{R4}$	$\hat{\alpha}_{R5}$	$\hat{\beta}_{R1}$	$\hat{\beta}_{R2}$	$\hat{\beta}_{R3}$	$\hat{\beta}_{R4}$	Arc
$n = 500$										
M	1.0010	− 2.9990	1.5013	− 2.0003	− 1.4985	0.6186	0.3163	0.6410	0.3163	0.0699
SD	0.0270	0.0304	0.0318	0.0289	0.0293	0.0385	0.0388	0.0408	0.0388	0.0393
MSE	0.7338	0.9253	1.0171	0.8398	0.8624	1.6747	1.5113	1.7361	1.5113	6.4414
Heteroscedasticity, $g(u) = 0.5(1 + u)^2$										
$n = 100$										
M	1.0052	− 3.0006	1.5029	− 1.9967	− 1.4995	0.6218	0.3135	0.6411	0.3135	0.0657
SD	0.0638	0.0723	0.0732	0.0702	0.0681	0.0395	0.0361	0.0386	0.0361	0.0395
MSE	4.0947	5.2300	5.3688	4.9423	4.6390	1.6787	1.3151	1.5687	1.3151	5.8867
$n = 300$										
M	1.0007	− 3.0008	1.5010	− 1.9975	− 1.5012	0.6287	0.3160	0.6350	0.3160	0.0343
SD	0.0360	0.0406	0.0395	0.0403	0.0393	0.0179	0.0197	0.0204	0.0197	0.0189
MSE	1.3018	1.6554	1.5619	1.6333	1.5474	0.3368	0.3882	0.4239	0.3882	1.5378
$n = 500$										
M	1.0005	− 2.9986	1.4999	− 1.9983	− 1.4991	0.6295	0.3148	0.6359	0.3148	0.0267
SD	0.0283	0.0320	0.0306	0.0320	0.0290	0.0155	0.0146	0.0151	0.0146	0.0144
MSE	0.8055	1.0285	0.9400	1.0319	0.8416	0.2483	0.2170	0.2413	0.2170	0.9238
Heteroscedasticity, $g(u) = \exp(0.5u - 0.25)$										
$n = 100$										
M	1.0025	− 2.9970	1.5001	− 2.0001	− 1.4980	0.6149	0.3224	0.6124	0.3224	0.1070
SD	0.0455	0.0501	0.0503	0.0509	0.0501	0.0953	0.0748	0.0882	0.0748	0.0963
MSE	2.0826	2.5249	2.5359	2.5895	2.5156	9.3885	5.6380	8.1814	5.6380	20.7244

Table 4 continued

	$\hat{\alpha}_{R1}$	$\hat{\alpha}_{R2}$	$\hat{\alpha}_{R3}$	$\hat{\alpha}_{R4}$	$\hat{\alpha}_{R5}$	$\hat{\beta}_{R1}$	$\hat{\beta}_{R2}$	$\hat{\beta}_{R3}$	$\hat{\beta}_{R4}$	Arc
$n = 300$										
M	1.0001	− 2.9989	1.5025	− 1.9982	− 1.4996	0.6195	0.3162	0.6152	0.3162	0.0934
SD	0.0348	0.0395	0.0412	0.0421	0.0412	0.0861	0.0569	0.0660	0.0569	0.0781
MSE	1.2138	1.5632	1.7030	1.7819	1.6986	7.5796	3.2394	4.6524	3.2394	14.8316
$n = 500$										
M	1.0011	− 2.9989	1.5000	− 1.9987	− 1.4983	0.6113	0.3166	0.6453	0.3166	0.0839
SD	0.0279	0.0316	0.0314	0.0318	0.0317	0.0531	0.0442	0.0494	0.0442	0.0525
MSE	0.7834	1.0013	0.9907	1.0189	1.0121	2.0006	1.9580	2.6080	1.9580	9.8067

MSE is in the scale of $\times 10^{-3}$

Table 5 Power calculations of $\mathcal{T}_n$ and $\mathcal{T}_{\mathrm{id},n}$ for homoscedasticity in Example 1

Significant level	$\mathcal{T}_n$			$\mathcal{T}_{\mathrm{id},n}$		
	0.01	0.05	0.10	0.01	0.05	0.10
	$n=100,\ g(u)=0.5(1+u)^2$					
$c=0.00$	0.004	0.035	0.074	0.011	0.047	0.091
$c=0.04$	0.021	0.068	0.147	0.028	0.099	0.189
$c=0.08$	0.033	0.145	0.246	0.121	0.267	0.378
$c=0.12$	0.106	0.309	0.454	0.252	0.474	0.583
$c=0.16$	0.290	0.568	0.701	0.514	0.727	0.818
$c=0.20$	0.510	0.751	0.855	0.705	0.878	0.936
	$n=300,\ g(u)=0.5(1+u)^2$					
$c=0.00$	0.009	0.042	0.086	0.010	0.050	0.101
$c=0.04$	0.034	0.115	0.215	0.082	0.216	0.314
$c=0.08$	0.219	0.497	0.655	0.399	0.664	0.779
$c=0.12$	0.691	0.899	0.949	0.849	0.949	0.974
$c=0.16$	0.951	0.984	0.997	0.977	0.997	0.998
$c=0.20$	0.997	1.000	1.000	1.000	1.000	1.000
	$n=500,\ g(u)=0.5(1+u)^2$					
$c=0.00$	0.011	0.051	0.110	0.010	0.048	0.099
$c=0.04$	0.062	0.207	0.339	0.144	0.339	0.459
$c=0.08$	0.531	0.760	0.861	0.704	0.860	0.932
$c=0.12$	0.923	0.982	0.995	0.970	0.994	0.999
$c=0.16$	0.997	1.000	1.000	0.999	1.000	1.000
$c=0.20$	1.000	1.000	1.000	1.000	1.000	1.000
	$n=100,\ g(u)=\exp(0.5u-0.25)$					
$c=0.00$	0.007	0.038	0.071	0.009	0.042	0.090
$c=0.04$	0.019	0.063	0.125	0.021	0.085	0.177
$c=0.08$	0.028	0.129	0.216	0.114	0.249	0.355
$c=0.12$	0.094	0.279	0.427	0.231	0.450	0.564
$c=0.16$	0.275	0.531	0.669	0.501	0.704	0.800
$c=0.20$	0.497	0.722	0.836	0.681	0.859	0.913
	$n=300,\ g(u)=\exp(0.5u-0.25)$					
$c=0.00$	0.011	0.047	0.092	0.012	0.052	0.106
$c=0.04$	0.021	0.089	0.160	0.054	0.172	0.297
$c=0.08$	0.174	0.462	0.635	0.366	0.637	0.744
$c=0.12$	0.662	0.864	0.919	0.801	0.915	0.939
$c=0.16$	0.912	0.934	0.969	0.957	0.979	0.990
$c=0.20$	0.968	0.997	1.000	0.998	1.000	1.000
	$n=500,\ g(u)=\exp(0.5u-0.25)$					
$c=0.00$	0.010	0.052	0.104	0.011	0.053	0.107
$c=0.04$	0.051	0.177	0.303	0.131	0.321	0.436

Table 5 continued

Significant level	$\mathcal{T}_n$			$\mathcal{T}_{\mathrm{id},n}$		
	0.01	0.05	0.10	0.01	0.05	0.10
$c = 0.08$	0.499	0.714	0.809	0.652	0.803	0.874
$c = 0.12$	0.869	0.934	0.969	0.914	0.992	1.000
$c = 0.16$	0.956	1.000	1.000	0.987	1.000	1.000
$c = 0.20$	1.000	1.000	1.000	1.000	1.000	1.000

Table 6 Power calculations of $\mathcal{T}_n$ for heteroscedasticity in Example 1

Significant level	$g(u) = 0.5(1+u)^2$			$g(u) = \exp(0.5u - 0.25)$		
	0.01	0.05	0.10	0.01	0.05	0.10
	$n = 100$					
$c = 0.00$	0.011	0.043	0.079	0.011	0.047	0.091
$c = 0.04$	0.023	0.098	0.199	0.020	0.062	0.131
$c = 0.08$	0.095	0.295	0.475	0.041	0.173	0.289
$c = 0.12$	0.250	0.588	0.736	0.124	0.389	0.532
$c = 0.16$	0.536	0.802	0.897	0.352	0.571	0.696
$c = 0.20$	0.648	0.882	0.943	0.545	0.714	0.866
	$n = 300$					
$c = 0.00$	0.009	0.045	0.092	0.013	0.054	0.091
$c = 0.04$	0.122	0.540	0.786	0.036	0.078	0.173
$c = 0.08$	0.813	0.963	0.990	0.054	0.216	0.366
$c = 0.12$	0.986	0.998	1.000	0.169	0.508	0.690
$c = 0.16$	0.999	1.000	1.000	0.413	0.751	0.858
$c = 0.20$	1.000	1.000	1.000	0.681	0.880	0.925
	$n = 500$					
$c = 0.00$	0.011	0.051	0.110	0.010	0.048	0.099
$c = 0.04$	0.568	0.930	0.980	0.129	0.269	0.381
$c = 0.08$	0.991	0.999	1.000	0.369	0.684	0.858
$c = 0.12$	1.000	1.000	1.000	0.641	0.892	0.941
$c = 0.16$	1.000	1.000	1.000	0.880	0.947	0.999
$c = 0.20$	1.000	1.000	1.000	0.967	1.000	1.000

simulation results in Table 7. We also see that the performance of PLPRE estimator $\hat{g}_L(u)$ is better than the TPLS estimator, since the values of mean, median and standard errors for PLPRE estimator are all smaller than those of TPLS estimator $\hat{g}(u, \hat{\boldsymbol{\eta}}_{\mathrm{T}})$. This simulation shows that PLPRE estimation method is more efficient than the transformation-based method even both estimators have root-(nh) convergence rate, slowly than the parametric convergence rate.

Example 2 In this example, we conduct 1000 simulations from model (1) by choosing $\boldsymbol{\alpha}_0 = (1, -3, 0, \ldots, 0)^{\mathrm{T}}$, $\boldsymbol{\beta}_{\boldsymbol{\phi}_0} = (1, -1, 1, 0, \ldots, 0)^{\mathrm{T}}/\sqrt{3}$. The length of $(\boldsymbol{\alpha}_0^{\mathrm{T}}, \boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}})^{\mathrm{T}}$ is set to be $p = q = 10$, $p = q = 15$ and $p = q = 20$, i.e., the number of zero

Table 7 The mean (M), median (Me) and standard errors (SD) for ASE and AAE

	$g(u) = 0.5(1+u)^2$				$g(u) = \exp(0.5u - 0.25)$			
	$\hat{g}_L(u, \hat{\boldsymbol{\eta}})$		$\hat{g}(u, \hat{\boldsymbol{\eta}})$		$\hat{g}_L(u, \hat{\boldsymbol{\eta}})$		$\hat{g}(u, \hat{\boldsymbol{\eta}})$	
	ASE	AAE	ASE	AAE	ASE	AAE	ASE	AAE
$n = 100$, homoscedasticity								
M	10.8691	83.7565	11.1910	84.5284	4.6983	34.1574	6.8281	38.6234
Me	8.5185	81.0115	8.7159	82.5541	4.3113	32.7373	6.4169	37.2098
SD	8.6678	38.1268	9.1889	38.5451	3.9934	14.1847	6.2112	16.1167
$n = 300$, homoscedasticity								
M	3.2581	46.6536	3.5286	48.0776	1.6975	18.2267	2.3927	20.5583
Me	2.8202	45.7154	2.8917	46.5989	1.4558	16.9987	2.1905	18.3754
SD	2.1202	17.8256	2.7611	21.0759	1.3135	6.4443	1.7476	7.9763
$n = 500$, homoscedasticity								
M	2.0962	36.7859	2.1866	37.0921	1.1214	13.3983	1.5539	15.2087
Me	1.5772	33.1865	1.6170	34.7934	1.0876	12.9467	1.4789	14.8876
SD	1.3582	13.5519	1.9136	17.6228	0.9791	5.7543	1.1191	6.3961
$n = 100$, heteroscedasticity								
M	12.3345	92.2456	14.1984	97.9456	5.4456	36.2890	7.5873	41.8347
Me	9.6980	86.4432	11.5432	94.2920	4.9947	34.2886	6.9928	39.8877
SD	9.9392	45.3547	11.3909	47.9111	5.1114	16.4647	6.8967	18.4490
$n = 300$, heteroscedasticity								
M	4.8830	52.3986	5.9676	55.2255	2.2234	21.7789	3.1219	22.9875
Me	4.3535	51.2160	2.8917	46.5989	2.0080	19.1213	2.8765	20.6690
SD	3.3321	19.1417	3.8980	24.1954	1.8689	7.2180	2.1418	8.4172
$n = 500$, heteroscedasticity								
M	3.2909	41.3323	4.0011	44.5431	1.6362	16.1410	1.9357	18.2446
Me	3.0008	39.4674	3.8182	42.8090	1.4341	15.7666	1.8221	16.9934
SD	1.5573	14.2829	2.1315	16.4546	1.2554	6.3310	1.5117	7.4521

The values in this table are in the scale of $\times 10^{-3}$

components of $\boldsymbol{\alpha}_0$ is 8, 18 and 28, and the number of zero components of $\boldsymbol{\phi}_0$ is 7, 17 and 27, respectively. The covariate $(\boldsymbol{X}^{\mathrm{T}}, \boldsymbol{Z}^{\mathrm{T}})^{\mathrm{T}}$ follows normal distribution $N(\mathbf{0}_{p+q}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = (\sigma_{ij})_{1 \le i,j \le (p+q)}$ with $\sigma_{ij} = (-0.5)^{|i-j|}$. The link function $g(u) = 0.5(1+u)^2$ is used in this example. Setting of model error ϵ is the same as Example 1.

To measure the selection and estimation accuracy, we define $\omega_{u,\boldsymbol{\alpha}_0}$, $\omega_{c,\boldsymbol{\alpha}_0}$ and $\omega_{o,\boldsymbol{\alpha}_0}$ as the proportions of underfitted, correctly fitted and overfitted models for the parameter $\boldsymbol{\alpha}_0$, and $\omega_{u,\boldsymbol{\phi}_0}$, $\omega_{c,\boldsymbol{\phi}_0}$ and $\omega_{o,\boldsymbol{\phi}_0}$ as the proportions of underfitted, correctly fitted and overfitted models for the parameter $\boldsymbol{\phi}_0$. In the case of overfitted, the labeled "1," "2" and "$\geq$ 3" are the proportions of models including 1, 2 and more than 2 insignificant covariates. Denote $(\mathrm{C}_{\boldsymbol{\alpha}_0}, \mathrm{C}_{\boldsymbol{\phi}_0})$ and $(\mathrm{IN}_{\boldsymbol{\alpha}_0}, \mathrm{IN}_{\boldsymbol{\phi}_0})$ as the average number of the zero coefficients that were correctly set to be zero, and the average number of the nonzero coefficients that were incorrectly set to be zero, respectively.

In Table 8, we report the performance of the penalized estimator $\hat{\boldsymbol{\eta}}_P$. Values of $(\mathrm{C}_{\boldsymbol{\alpha}_0}, \mathrm{C}_{\boldsymbol{\phi}_0})$ are close to the true values (8, 7) ($p = q = 8$), (13, 12) ($p = q = 15$) and (18, 17) ($p = q = 20$), and $(\mathrm{IN}_{\boldsymbol{\alpha}_0}, \mathrm{IN}_{\boldsymbol{\phi}_0})$ are close to 0. The proportion of correct model fitting (column $(\omega_{c,\boldsymbol{\alpha}_0}, \omega_{c,\boldsymbol{\phi}_0})$) is above 96% when the sample size $n \geq 300$. The proportions of underfitting (column $(\omega_{u,\boldsymbol{\alpha}_0}, \omega_{u,\boldsymbol{\phi}_0})$) and overfitted (columns under $(\omega_{o,\boldsymbol{\alpha}_0}, \omega_{o,\boldsymbol{\phi}_0})$) are at most 0.5% and at most 3% when the sample size $n \geq 300$. In overfitted case, proportion of models including one insignificant covariate dominates those including two or more insignificant covariates. The latter is nearly 0% in most situations when the sample size n is larger than or equal to 300. This indicates that the penalized estimator $\hat{\boldsymbol{\eta}}_P$ most likely selects model that is very close to the true one, which again implies the proposed penalized estimation method will result in a final correct model in most situations.

Example 3 In this example, we consider model checking problem. We generate 1000 experiments with sample size $n = 300$ and $n = 500$ from the following model

$$Y = \exp\left(\boldsymbol{\alpha}_0^{\mathrm{T}}\boldsymbol{X} + 0.5\left(1 + \boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z}\right)^2\right)\epsilon + c\exp\left(\sum_{r=1}^{q}|X_r| + |\boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}}\boldsymbol{Z}|\right)\epsilon.$$

Here, $c = 0$ corresponds to the null hypothesis $\mathcal{H}_0^*$ and $c \neq 0$ corresponds to the alternative hypothesis. We set $c = 0.25, 0.50, \ldots, 1.25$ for the alternative hypothesis. Settings of parameter $(\boldsymbol{\alpha}_0^{\mathrm{T}}, \boldsymbol{\beta}_{\boldsymbol{\phi}_0}^{\mathrm{T}})^{\mathrm{T}}$, covariates $(\boldsymbol{X}^{\mathrm{T}}, \boldsymbol{Z}^{\mathrm{T}})^{\mathrm{T}}$ and model error ϵ are the same as Example 1. In each simulation for the power calculation, 1000 bootstrap samples were generated. We also compared our results with integrated conditional moment test statistic proposed by Ma et al. (2014) by using the transformed data $\{\ln(Y_i), \boldsymbol{X}_i, \boldsymbol{Z}_i\}_{i=1}^n$.

The simulation results are reported in Table 9. Power functions are calculated at the nominal level $\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.10$. It is clear that all empirical levels obtained by the test statistic $\mathcal{T}_n^*$ and ICM test statistic are close to 0.01, 0.05, 0.10 when $c = 0$, which indicates that the proposed model checking method can provide proper rejection probabilities. When $\mathcal{H}_0^*$ is not true, that is, $c \neq 0$, as the value of $c \neq 0$ increases or the sample size n increases, the empirical percentages of rejecting $\mathcal{H}_0^*$ approach to 1. Clearly, as c increases, the power increases rapidly. The powers for $n = 500$ are larger than the powers for $n = 300$ for the same c value and significance level. For $n = 500$ and large c, the powers are close to 1. We also note that the performance of test statistic is better than the ICM test statistic. This simulation shows that the proposed model checking procedure is more powerful than the transformation-based method. The results demonstrate that our proposed test procedure is powerful.

7 Real data analysis

In this section, we analyze body fat data to illustrate the application of our proposed method. The body fat data are available at http://lib.stat.cmu.edu/datasets/bodyfat. Based on the studies in Chen et al. (2016), we use model (1) to re-analyze this dataset with X_1-thigh circumference, X_2-knee circumference, X_3-ankle circumference, X_4-forearm circumference, Z_1-abdomen circumference, Z_2-height^4/weight^2, Z_3-neck circumference, Z_4- chest circumference, Z_5-age, Z_6-hip circumference, Z_7-biceps circumference and wrist circumference. The response variable Y is the percent body fat. We deleted one

Table 8 Simulation results for variable selection in Example 2

$(q_0, q - q_0)$	$\omega_{u,\boldsymbol{\alpha}_0}$ (%)	$\omega_{c,\boldsymbol{\alpha}_0}$ (%)	$\omega_{o,\boldsymbol{\alpha}_0}$ (%)			No. of zeros	
			"1 (%)"	"2" (%)	"≥ 3"(%)	$C_{\boldsymbol{\alpha}_0}$	$IN_{\boldsymbol{\alpha}_0}$
$n = 100$, homoscedasticity							
(2, 8)	6.30	83.50	9.40	0.70	0.10	7.874	0.063
(2, 13)	5.20	78.30	14.60	1.70	0.20	12.797	0.052
(2, 18)	4.10	71.20	19.90	4.10	0.70	17.674	0.041
$n = 300$, homoscedasticity							
(2, 8)	0.10	99.90	0.00	0.00	0.00	8.000	0.001
(2, 13)	0.30	99.60	0.10	0.00	0.00	12.999	0.003
(2, 18)	0.00	99.80	0.10	0.10	0.00	17.997	0.000
$n = 500$, homoscedasticity							
(2, 8)	0.00	100.00	0.00	0.00	0.00	8.000	0.000
(2, 13)	0.00	100.00	0.00	0.00	0.00	13.000	0.000
(2, 18)	0.00	100.00	0.00	0.00	0.00	18.000	0.000
$n = 100$, heteroscedasticity							
(2, 8)	4.90	82.60	11.60	0.80	0.10	7.862	0.049
(2, 13)	6.30	75.30	16.50	1.80	0.10	12.772	0.063
(2, 18)	6.00	69.80	19.80	3.60	0.80	17.675	0.060
$n = 300$, heteroscedasticity							
(2, 8)	0.30	99.60	0.10	0.00	0.00	7.999	0.003
(2, 13)	0.20	99.40	0.40	0.00	0.00	12.996	0.002
(2, 18)	0.20	99.50	0.30	0.00	0.00	17.997	0.002
$n = 500$, heteroscedasticity							
(2, 8)	0.00	100.00	0.00	0.00	0.00	8.000	0.000
(2, 13)	0.00	100.00	0.00	0.00	0.00	13.000	0.000
(2, 18)	0.00	100.00	0.00	0.00	0.00	18.000	0.000
$(p_0 - 1, p - p_0)$	$\omega_{u,\boldsymbol{\phi}_0}$(%)	$\omega_{c,\boldsymbol{\phi}_0}$ (%)	$\omega_{o,\boldsymbol{\phi}_0}$ (%)			No. of zeros	
			"1 (%)"	"2" (%)	"≥ 3"(%)	$C_{\boldsymbol{\phi}_0}$	$IN_{\boldsymbol{\phi}_0}$
$n = 100$, homoscedasticity							
(2, 7)	14.50	63.90	18.10	3.00	0.50	6.694	0.145
(2, 12)	14.20	62.60	18.70	4.00	0.50	11.606	0.142
(2, 17)	19.50	50.20	19.90	7.70	2.70	16.324	0.196
$n = 300$, homoscedasticity							
(2, 7)	0.20	98.40	1.30	0.10	0.00	6.985	0.002
(2, 12)	0.1	98.70	1.20	0.00	0.00	11.988	0.001
(2, 17)	0.10	99.30	0.60	0.00	0.00	16.994	0.001
$n = 500$, homoscedasticity							
(2, 7)	0.00	99.70	0.30	0.00	0.00	6.997	0.000
(2, 12)	0.00	99.80	0.20	0.00	0.00	11.998	0.000
(2, 17)	0.00	99.90	0.10	0.00	0.00	16.999	0.000

Table 8 continued

$(p_0-1, p-p_0)$	$\omega_{u,\boldsymbol{\phi}_0}(\%)$	$\omega_{c,\boldsymbol{\phi}_0}$ (%)	$\omega_{o,\boldsymbol{\phi}_0}$ (%)			No. of zeros	
			"1 (%)"	"2" (%)	"$\geq$ 3"(%)	$C_{\boldsymbol{\phi}_0}$	$IN_{\boldsymbol{\phi}_0}$
$n=100$, heteroscedasticity							
(2, 7)	16.70	57.20	20.40	5.10	0.60	6.605	0.167
(2, 12)	19.20	49.80	22.90	6.60	1.50	11.411	0.192
(2, 17)	21.80	41.80	23.50	9.30	3.60	16.115	0.220
$n=300$, heteroscedasticity							
(2, 7)	0.30	96.80	2.90	0.00	0.00	6.971	0.003
(2, 12)	0.40	97.80	1.80	0.00	0.00	11.980	0.004
(2, 17)	0.10	97.90	2.00	0.00	0.00	16.980	0.001
$n=500$, heteroscedasticity							
(2, 7)	0.10	99.00	0.90	0.00	0.00	6.991	0.001
(2, 12)	0.00	99.50	0.50	0.00	0.00	11.995	0.000
(2, 17)	0.10	99.80	0.10	0.00	0.00	16.999	0.000

observation with $Y=0$, giving a sample size $n=251$. In the following analysis, covariates $\boldsymbol{X}$ and $\boldsymbol{Z}$ are all standardized.

Corresponding to covariates $\boldsymbol{X}$ and $\boldsymbol{Z}$, the estimator $(\hat{\boldsymbol{\alpha}}^{\mathrm{T}}, \hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}^{\mathrm{T}})^{\mathrm{T}}$ and the penalized estimate $(\hat{\boldsymbol{\alpha}}_P^{\mathrm{T}}, \hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}_P}^{\mathrm{T}})$ are obtained in Table 10. Using Theorem 2 and Theorem 4, we conduct the tests $H_0: \boldsymbol{\alpha}_{0,s}=0$ against $H_1, \boldsymbol{\alpha}_{0,s}\neq 0$, and also $H_0: \boldsymbol{\beta}_{0,j}=0$ against $H_1, \boldsymbol{\beta}_{0,j}\neq 0$ by choosing restriction condition $\boldsymbol{A}_s=(0,\ldots,0,1,0,\ldots,0)^{\mathrm{T}}$ with that the element 1 is the s-th position in the vector $\boldsymbol{A}_s$, $s=1,2,3,4$, and $\boldsymbol{A}_j=(0,\ldots,0,1,0,\ldots,0)^{\mathrm{T}}$ with that the element 1 is the j-th position in the vector $\boldsymbol{A}_j$, $j=1,\ldots,8$. The p values $(p_{\hat{\boldsymbol{\alpha}}}, p_{\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}})$ of test statistic $\mathcal{T}_n$ and the penalized estimator $(\hat{\boldsymbol{\alpha}}_P, \hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}_P})$ show that the covariates X_3-ankle circumference, Z_1-abdomen circumference and Z_6-hip circumference should be included in the final model. In Fig. 1, we consider two kinds of estimators of $g(u)$ based on (5), along with their 95% pointwise confidence bands. The left panel in Fig. 1 is based on "synthesis" data $\{Y_i, \boldsymbol{X}_i^{\mathrm{T}}\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}^{\mathrm{T}}\boldsymbol{Z}_i\}_{i=1}^n$, and the right panel in Fig. 1 is based on "synthesis" data $\{Y_i, \boldsymbol{X}_i^{\mathrm{T}}\hat{\boldsymbol{\alpha}}_P, \hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}_P}^{\mathrm{T}}\boldsymbol{Z}_i\}_{i=1}^n$. Figure 1 shows that the local linear smoothing of the link function $g(u)$ shows a nonlinear pattern; it decreases in the beginning when the estimated single index is less around -1 and then increases. The mean and standard errors of the product relative errors $\{(Y_i-\hat{Y}_i)^2/(Y_i\hat{Y}_i)\}_{i=1}^n$, and the squared errors $\{(Y_i-\hat{Y}_i)^2\}_{i=1}^n$ are obtained as (0.2677, 0.9883), (52.5506, 75.6770), respectively. If we treat $g(u)$ as a linear model (i.e., a multiplicative linear regression model is adopted) and denote the fitted values as $\hat{Y}_{Li}$, $i=1,\ldots,n$. Then mean and standard errors of the product relative errors $\{(Y_i-\hat{Y}_{Li})^2/(Y_i\hat{Y}_{Li})\}_{i=1}^n$, and the squared errors $\{(Y_i-\hat{Y}_{Li})^2\}_{i=1}^n$ are obtained as (0.2648, 1.0534), (47.8662, 91.3414), respectively. It is seen that the partial linear single-index multiplicative model increases slightly larger mean values of the product relative errors and the squared errors, but the standard errors are both smaller with

Table 9 The simulation results for power calculations in Example 3

Significant level	0.01	0.05	0.10	0.01	0.05	0.10
	$\mathcal{T}_n^*$			ICM		
$n = 300$, homoscedasticity						
$c = 0.00$	0.012	0.044	0.096	0.010	0.040	0.089
$c = 0.25$	0.432	0.652	0.717	0.356	0.584	0.634
$c = 0.50$	0.560	0.773	0.873	0.488	0.701	0.810
$c = 0.75$	0.581	0.791	0.893	0.532	0.733	0.864
$c = 1.00$	0.617	0.816	0.915	0.582	0.769	0.882
$c = 1.25$	0.645	0.839	0.944	0.611	0.810	0.921
$n = 500$, homoscedasticity						
$c = 0.00$	0.011	0.051	0.101	0.011	0.044	0.096
$c = 0.25$	0.780	0.890	0.946	0.643	0.804	0.872
$c = 0.50$	0.823	0.913	0.973	0.796	0.842	0.910
$c = 0.75$	0.875	0.922	0.986	0.829	0.888	0.935
$c = 1.00$	0.914	0.945	0.990	0.889	0.915	0.962
$c = 1.25$	0.971	1.000	1.000	0.956	0.997	1.000
$n = 300$, heteroscedasticity						
$c = 0.00$	0.011	0.042	0.091	0.008	0.042	0.090
$c = 0.25$	0.368	0.541	0.622	0.300	0.489	0.552
$c = 0.50$	0.471	0.691	0.800	0.421	0.634	0.740
$c = 0.75$	0.522	0.739	0.831	0.474	0.679	0.806
$c = 1.00$	0.589	0.777	0.882	0.541	0.737	0.854
$c = 1.25$	0.620	0.809	0.921	0.596	0.788	0.907
$n = 500$, heteroscedasticity						
$c = 0.00$	0.012	0.048	0.097	0.011	0.046	0.095
$c = 0.25$	0.714	0.822	0.870	0.588	0.707	0.803
$c = 0.50$	0.769	0.874	0.921	0.648	0.783	0.869
$c = 0.75$	0.815	0.900	0.943	0.779	0.836	0.901
$c = 1.00$	0.888	0.921	0.970	0.854	0.889	0.942
$c = 1.25$	0.941	0.989	1.000	0.926	0.957	0.989

stability. Together with Fig. 1, we consider the partial linear single-index multiplicative model is more appropriate than the linear multiplicative model for this dataset. At last, we conducted 1000 bootstraps with the test statistic $\mathcal{T}_n^*$ to check whether the partial linear single-index multiplicative model is good enough for this dataset. The associated p value is 0.65, which indicates that the partial linear single-index multiplicative model is tenable for fitting this dataset.

Table 10 Estimation results for the bodyfat dataset

$(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_{\boldsymbol{\phi}_0})$	$\boldsymbol{\alpha}_{0,1}$	$\boldsymbol{\alpha}_{0,2}$	$\boldsymbol{\alpha}_{0,3}$	$\boldsymbol{\alpha}_{0,4}$	$\boldsymbol{\beta}_{0,1}$	$\boldsymbol{\beta}_{0,2}$	$\boldsymbol{\beta}_{0,3}$	$\boldsymbol{\beta}_{0,4}$	$\boldsymbol{\beta}_{0,5}$	$\boldsymbol{\beta}_{0,6}$	$\boldsymbol{\beta}_{0,7}$	$\boldsymbol{\beta}_{0,8}$
$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}})$	−0.0556	−0.0313	−0.0456	−0.0187	0.7736	0.1566	−0.5264	−0.0838	−0.1624	−0.0576	0.1808	−0.1740
$(p_{\hat{\boldsymbol{\alpha}}}, p_{\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}}})$	0.2189	0.3287	0.0265	0.5005	0.0413	0.1423	0.3337	0.0014	0.0875	0.0037	0.1206	0.1432
$(\hat{\boldsymbol{\alpha}}_P, \hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\phi}}_P})$	0	0	−0.1211	0	0.8919	0	−0.4433	0	0	−0.0888	0	0

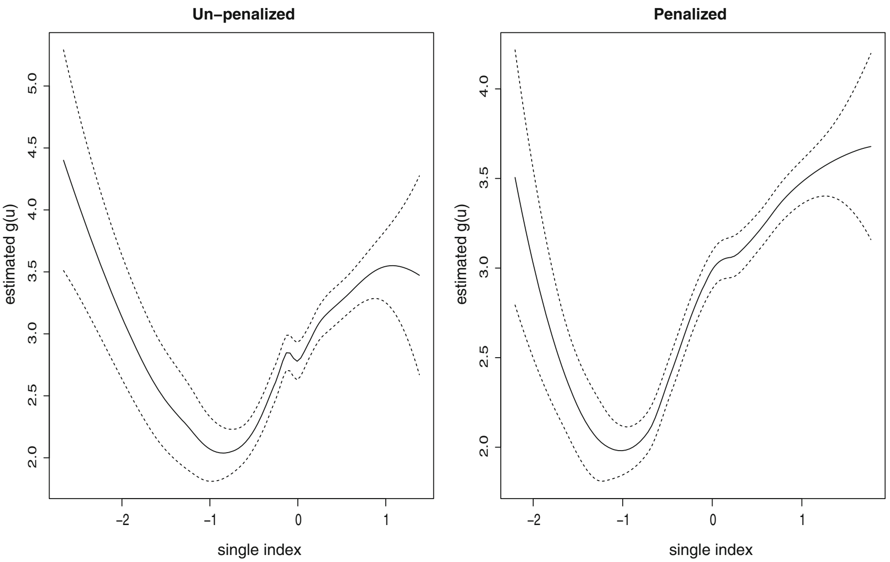

Fig. 1 The estimator $\hat{g}_L(u, \hat{\boldsymbol{\eta}})$ (solid line) against estimated single index, along with the associated 95% pointwise confidence intervals (dotted lines)

Acknowledgements The authors thank the editor, the associate editor and two referees for their constructive suggestions that helped us to improve the early manuscript. Xia Cui is a College Talent Cultivated by Thousand-Hundred-Ten Program of Guangdong Province, and her research was supported by grants from National Natural Science Foundation of China (NSFC) (No. 11471086), Humans and Social Science Research Team of Guangzhou University (No. 201503XSTD) and the Training Program for Excellent Young College Teachers of Guangdong Province (No. Yq201404). Heng Peng's research was supported in part by CEGR grant of the Research Grants Council of Hong Kong (Nos. HKBU202012 and HKBU 12302615), FRG grants from Hong Kong Baptist University (Nos. FRG2 14-15/064 and FRG2 /16-17/042).

References

Bierens, H. J. (1982). Consistent model specification tests. *Journal of Econometrics*, *20*, 105–134.

Bierens, H. J., Ploberger, W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica*, *65*, 1129–1151.

Bindele, H. F., Abebe, A., Meyer, K. N. (2018). General rank-based estimation for regression single index models. *Annals of the Institute of Statistical Mathematics*, *70*(5), 1115–1146.

Boente, G., Rodriguez, D. (2012). Robust estimates in generalized partially linear single-index models. *Test*, *21*(2), 386–411.

Chen, K., Guo, S., Lin, Y., Ying, Z. (2010). Least absolute relative error estimation. *Journal of the American Statistical Association*, *105*(491), 1104–1112.

Chen, K., Lin, Y., Wang, Z., Ying, Z. (2016). Least product relative error estimation. *Journal of Multivariate Analysis*, *144*, 91–98.

Cui, X., Härdle, W. K., Zhu, L. (2011). The EFM approach for single-index models. *The Annals of Statistics*, *39*(3), 1658–1688.

Escanciano, J. C. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory*, *22*, 1030–1051.

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

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics*, *58*, 71–120.

Jones, L. K. (1987). On a conjecture of Huber concerning the convergence of projection pursuit regression. *The Annals of Statistics*, *15*, 880–882.

Lai, P., Li, G., Lian, H. (2013). Quadratic inference functions for partially linear single-index models with longitudinal data. *Journal of Multivariate Analysis*, *118*, 115–127.

Lee, T. H., White, H., Granger, C. W. J. (2001). Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests. *Journal of Econometrics*, *56*, 208–229.

Li, G., Peng, H., Dong, K., Tong, T. (2014). Simultaneous confidence bands and hypothesis testing for single-index models. *Statistica Sinica*, *24*(2), 937–955.

Li, G., Lai, P., Lian, H. (2015). Variable selection and estimation for partially linear single-index models with longitudinal data. *Statistics and Computing*, *25*(3), 579–593.

Li, T., Yang, H., Wang, J. L., Xue, L., Zhu, L. (2011). Correction on estimation for a partial-linear single-index model. *The Annals of Statistics*, *39*(6), 3441–3443.

Lian, H., Liang, H. (2016). Separation of linear and index covariates in partially linear single-index models. *Journal of Multivariate Analysis*, *143*, 56–70.

Lian, H., Liang, H., Carroll, R. J. (2015). Variance function partially linear single-index models. *Journal of The Royal Statistical Society Series B-statistical Methodology*, *77*(1), 171–194.

Liang, H., Wang, N. (2005). Partially linear single-index measurement error models. *Statistica Sinica*, *15*(1), 99–116.

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

Lin, D. Y., Wei, L. J., Ying, Z. (2002). Model-checking techniques based on cumulative residuals. *Biometrics*, *58*, 1–12.

Liu, H., Xia, X. (2018). Estimation and empirical likelihood for single-index multiplicative models. *Journal of Statistical Planning & Inference*, *193*, 70–88.

Ma, S., Zhang, J., Sun, Z., Liang, H. (2014). Integrated conditional moment test for partially linear single index models incorporating dimension-reduction. *Electronic Journal of Statistics*, *8*(1), 523–542.

Peng, H., Huang, T. (2011). Penalized least squares for single index models. *Journal of Statistical Planning and Inference*, *141*(4), 1362–1379.

Stute, W. (1997). Nonparametric model checks for regression. *The Annals of Statistics*, *25*, 613–641.

Stute, W., Zhu, L. X. (2002). Model checks for generalized linear models. *Scandinavian Journal of Statistics Theory and Applications*, *29*, 535–545.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B Methodological*, *58*(1), 267–288.

Wang, Z., Chen, Z., Wu, Y. (2017). A relative error estimation approach for multiplicative single index model. *Journal of Systems Science & Complexity*, *30*(5), 1160–1172.

Wei, C., Wang, Q. (2012). Statistical inference on restricted partially linear additive errors-in-variables models. *Test*, *21*(4), 757–774.

Xia, Y., Härdle, W. (2006). Semi-parametric estimation of partially linear single-index models. *Journal of Multivariate Analysis*, *97*, 1162–1184.

Xia, Y., Tong, H., Li, W. K., Zhu, L. X. (2002). An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society, Series B*, *64*, 363–410.

Xia, Y., Li, W. K., Tong, H., Zhang, D. (2004). A goodness-of-fit test for single-index models. *Statistica Sinica*, *14*, 1–39.

Zhang, J., Zhu, J., Feng, Z. (2018). Estimation and hypothesis test for single-index multiplicative models. Test https://doi.org/10.1007/s11749-018-0586-2.