Ann Inst Stat Math (2018) 70:1077–1114
https://doi.org/10.1007/s10463-017-0616-y

A constructive hypothesis test for the single-index models with two groups

Jun Zhang[1] · Zhenghui Feng[2] · Xiaoguang Wang[3]

Received: 25 May 2016 / Revised: 19 August 2017 / Published online: 13 September 2017

Abstract Comparison of two-sample heteroscedastic single-index models, where both the scale and location functions are modeled as single-index models, is studied in this paper. We propose a test for checking the equality of single-index parameters when dimensions of covariates of the two samples are equal. Further, we propose two test statistics based on Kolmogorov–Smirnov and Cramér–von Mises type functionals. These statistics evaluate the difference of the empirical residual processes to test the equality of mean functions of two single-index models. Asymptotic distributions of estimators and test statistics are derived. The Kolmogorov–Smirnov and Cramér–von Mises test statistics can detect local alternatives that converge to the null hypothesis at a parametric convergence rate. To calculate the critical values of Kolmogorov–Smirnov and Cramér–von Mises test statistics, a bootstrap procedure is proposed. Simulation studies and an empirical study demonstrate the performance of the proposed procedures.

✉ Zhenghui Feng
zhfengwise@gmail.com

Jun Zhang
zhangjunstat@gmail.com

Xiaoguang Wang
wangxg@dlut.edu.cn

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

[2] School of Economics and the Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen 361005, China

[3] School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China

Keywords Empirical residual process · Single-index models · Local linear smoothing · Model checking

1 Introduction

Single-index models (SIMs) have attracted a large amount of attention. A SIM is a generalization of multivariate linear regression models with an unknown link function, and it relaxes the restrictive assumptions imposed on the parametric models of conditional mean functions. When the regression function should be estimated in a nonparametric context, the dimensionality of covariates plays a crucial role. Among the many existing dimension reduction techniques, SIMs assume the link function to be a univariate function applied to the projection of an explanatory covariate vector on to some direction, and they avoid some drawbacks of fully nonparametric methods such as the curse of dimensionality, difficulty of interpretation, and lack of extrapolation ability. Hence, SIMs have a wide range of applications in areas such as economics and finance. There have been many papers that consider the estimation of the link function, the single-index parameter and related issues. See for example, Xia et al. (2002), Wang et al. (2015), Härdle et al. (1993), Xu and Zhu (2012), Feng et al. (2013), Feng and Zhu (2012), Li et al. (2014), and Peng and Huang (2011).

How to compare two regression functions is a common topic. For example, in medical research, this problem arises when comparing two functions of the mean reaction time for drug use and the control group. In this paper, suppose we have two independent samples following the single-index heteroscedastic regression model (with two groups):

$$\begin{cases} Y_1 = g_1\left(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_1\right) + \sigma_1\left(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_1\right)\epsilon_1, \\ Y_2 = g_2\left(\boldsymbol{\gamma}_0^{\tau}\boldsymbol{X}_2\right) + \sigma_2\left(\boldsymbol{\gamma}_0^{\tau}\boldsymbol{X}_2\right)\epsilon_2, \end{cases} \tag{1}$$

where τ denotes transposition throughout this paper. In model (1), $\boldsymbol{X}_1$ is a p-dimensional covariate vector and $\boldsymbol{X}_2$ is a q-dimensional covariate vector. For $s = 1, 2$, Y_s are the response variables, and $g_s(u)$ and $\sigma_s(u)$ are unknown univariate smooth functions. Throughout this paper, we assume that functions $\sigma_s(u), s = 1, 2$ are positive. The error terms ϵ_s, $s = 1, 2$ satisfy $E(\epsilon_s) = 0$ and $E(\epsilon_s^2) = 1$. The condition $E(\epsilon_s^2) = 1$ used here is assumed for identifiability. Parameter $\boldsymbol{\beta}_0$ is an unknown index vector that belongs to the parameter space $\mathcal{B}_{\boldsymbol{\beta}} = \left\{\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^{\tau} \in \mathbb{R}^p, \|\boldsymbol{\beta}\| = 1, \beta_1 > 0\right\}$, and similarly, we define that $\boldsymbol{\gamma}_0 \in \mathcal{B}_{\boldsymbol{\gamma}} = \left\{\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_q)^{\tau} \in \mathbb{R}^q, \|\boldsymbol{\gamma}\| = 1, \gamma_1 > 0,\right\}$.

Among the various methods of estimation, there is little literature that considers the comparison of SIMs between two groups. Classical methods use parametric regression models for two groups and to compare the two regression functions g_1 and g_2, they compare the resulting parameters of the models. A disadvantage of this approach is that it requires the parametric models to be specified, which is often difficult. As we indicated above, the single-index structure is more appropriate in terms of its easier interpretation and ability to relax the restrictive assumptions. Single-index model (1) has two groups, and the comparison of the regression functions of two groups has been extensively investigated in the literature if the dimensions of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ equal each other and are both equal to one, i.e., $p = q = 1$. See for example, Kulasekera (1995), Gørgens (2002), Kulasekera and Wang (1997), Koul and Schick (1997), and Neumeyer and Dette (2003). Feng et al. (2015) proposed a Wilcoxon-type generalized likelihood ratio test to deal with the data that is possibly affected by outlying observations and heavy-tailed distributions. Zhang et al. (2010) considered test methods to compare the mean functions of two samples drawn from functional data sets by proposing L_2-norm-based and bootstrap-based test statistics for this purpose. For the comparison of SIMs, Lin and Kulasekera (2010) proposed an ANOVA-type approach to compare two or more SIMs by assuming that the dimensions of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are equal, i.e., $p = q$. Under this scenario, Lin and Kulasekera (2010) proposed an F-statistic to check $\boldsymbol{\beta}_0 = \boldsymbol{\gamma}_0$ and $g_1 \equiv g_2$ simultaneously. Suppose that $p = q$ and $g_1 \equiv g_2$, then the null hypothesis in Lin and Kulasekera (2010) becomes $\mathcal{H}_0 : \boldsymbol{\beta}_0 = \boldsymbol{\gamma}_0$, and the convergence rate of the F-statistic is $\frac{1}{(n_1+n_2)\sqrt{h}}$, where n_s is the sample size of group s, $s = 1, 2$, and h is the bandwidth used for nonparametric kernel smoothing. It is known that the convergence rates of the estimates of $(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$ are faster than the nonparametric estimates of $g_1(u)$ and $g_2(u)$. Hence, the F-statistic proposed by Lin and Kulasekera (2010) is not optimal for testing $\mathcal{H}_0 : \boldsymbol{\beta}_0 = \boldsymbol{\gamma}_0$. Let us consider another special setting. If we know $\boldsymbol{\beta}_0 = \boldsymbol{\gamma}_0$, and we further want to test $\mathcal{H}'_0 : g_1 \equiv g_2$. Neumeyer and Dette (2003) proposed using F-statistic to test this, which can detect an alternative hypothesis converging to the null hypothesis with a rate that is slower than the parametric rate. Because the test proposed by Lin and Kulasekera (2010) is not optimal for testing either the equality of single-index parameters or the equality of the mean function, this motivates us to propose test procedures with optimal convergence rates.

The first goal of this paper is to propose an estimation procedure and a hypothesis test for the unknown single-index parameters. The profile estimation equations are adopted to estimate the single-index parameters that are associated with large sample properties of the estimators. If the dimensions of covariates $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are equal, a Wald-type statistic is proposed to test $\mathcal{H}_0 : \boldsymbol{\beta}_0 = \boldsymbol{\gamma}_0$. Note that the restrictions $\|\boldsymbol{\beta}_0\| = 1$ and $\|\boldsymbol{\gamma}_0\| = 1$ mean that the single-index parameters $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}_0$ are on the boundary of a unit ball. Therefore, we use the popular "leave-one-component out" method for estimating the single-index parameter (Cui et al. 2011; Li et al. 2010; Yu and Ruppert 2002), and we show that the test statistic asymptotically converges to a standard χ^2 distribution with $p - 1$, rather than p the degree of freedom.

The second goal is to check whether the mean functions $g_1(u)$ and $g_2(u)$ are equal or not, i.e., test $\widetilde{\mathcal{H}}_0 : g_1 \equiv g_2$. The idea of constructing a test statistic for $\widetilde{\mathcal{H}}_0$ is implemented by comparing the estimated error distributions $\hat{F}_{\epsilon_s}(t)$, $s = 1, 2$ obtained under the full model (1) with the estimated error distribution functions $\hat{F}_{\widetilde{\mathcal{H}}_0,\epsilon_s}(t)$, $s = 1, 2$ obtained under the null hypothesis $\widetilde{\mathcal{H}}_0$. Under $\widetilde{\mathcal{H}}_0$, the estimator of distribution function $F_{\epsilon_s}(t)$ for ϵ_s can be obtained from the residuals based on errors $\epsilon_{\widetilde{\mathcal{H}}_0,1} = \frac{Y_1 - g_2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_1)}{\sigma_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_1)}$ and $\epsilon_{\widetilde{\mathcal{H}}_0,2} = \frac{Y_2 - g_1(\boldsymbol{\gamma}_0^\tau \boldsymbol{X}_2)}{\sigma_2(\boldsymbol{\gamma}_0^\tau \boldsymbol{X}_2)}$. It is easily seen that if $\widetilde{\mathcal{H}}_0$ is true, then $\epsilon_{\widetilde{\mathcal{H}}_0,1} = \epsilon_1$ and $\epsilon_{\widetilde{\mathcal{H}}_0,2} = \epsilon_2$. Hence, the test procedure for $\widetilde{\mathcal{H}}_0$ is completed by first estimating the distribution functions $F_{\epsilon_s}(t)$ under model (1) and $\widetilde{\mathcal{H}}_0$. We then use these estimated distribution functions to propose the Kolmogorov–Smirnov test statistic and

Cramér–von Mises test statistic. We obtain the asymptotic expressions for the estimators of $F_{\epsilon_s}(t)$ under model (1) and under $\widetilde{\mathcal{H}}_0$. The weak convergence properties for the estimators of the error distribution functions are revealed. Finally, the limiting distributions of the Kolmogorov–Smirnov and Cramér–von Mises test statistics are also derived. To mimic the null distributions of the test statistics, a bootstrap procedure is proposed to define the p values. We conducted Monte Carlo simulation experiments to examine the performance of the proposed procedures. Our simulation results show that the proposed methods perform well both for estimation and hypothesis testing.

This paper is organized as follows. In Sect. 2, we propose the estimation procedure for $\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, g_s(u)$ and $\sigma_s(u)$, $s = 1, 2$. A hypothesis for testing $\boldsymbol{\beta}_0 = \boldsymbol{\gamma}_0$ is also considered. In Sect. 3, we provide the estimators of the error distribution functions and propose test statistics to check the equality of the mean functions. A bootstrap procedure is also proposed to mimic the null distributions of test statistics. In Sect. 4, we report the results of simulation studies. All the technical proofs of the asymptotic results are given in the Appendix.

2 Estimation of $\boldsymbol{\beta}_0$, $\boldsymbol{\gamma}_0$, $g_1(u)$, $g_2(u)$, $\sigma_1^2(u)$ and $\sigma_2^2(u)$

For notational simplicity, we define $\boldsymbol{\omega}_{[1]} = \boldsymbol{\beta}$, $\boldsymbol{\omega}_{[1]}^{(1)} = \boldsymbol{\beta}^{(1)} = (\beta_2, \ldots, \beta_p)^\tau$ and $\boldsymbol{\omega}_{[2]} = \boldsymbol{\gamma}$, $\boldsymbol{\omega}_{[2]}^{(1)} = \boldsymbol{\gamma}^{(1)} = (\gamma_2, \ldots, \gamma_q)^\tau$. Similarly, we define $\boldsymbol{\omega}_{[1],0} = \boldsymbol{\beta}_0$ and $\boldsymbol{\omega}_{[1],0}^{(1)} = \boldsymbol{\beta}_0^{(1)}$ and $\boldsymbol{\omega}_{[2],0} = \boldsymbol{\gamma}_0$, $\boldsymbol{\omega}_{[2],0}^{(1)} = \boldsymbol{\gamma}_0^{(1)}$ for the true values.

Suppose that we have two samples $\{\boldsymbol{X}_{1i}, Y_{1i}, i = 1, \ldots, n_1\}$ and $\{\boldsymbol{X}_{2i}, Y_{2i}, i = 1, \ldots, n_2\}$ from model (1), where n_s and $s = 1, 2$ are the sample sizes of two groups, respectively. In the following, we propose estimation procedures for the parameters $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}_0$. The profile least squares estimation procedure used in Liang et al. (2010) is employed here. Our procedure has three steps:

1. Given $\boldsymbol{\omega}_{[s]}$, we can approximate $g_s(u)$ by $g_s(u_*)+g_s'(u_*)(u-u_*)$ in a neighborhood of u_* for $s = 1, 2$. Minimizing (2) with respect to a_{s0} and a_{s1}

$$\sum_{i=1}^{n_s} \left\{Y_{si} - a_{s0} - a_{s1}\left(\boldsymbol{\omega}_{[s]}^\tau \boldsymbol{X}_{si} - u\right)\right\}^2 K_{h_s}\left(\boldsymbol{\omega}_{[s]}^\tau \boldsymbol{X}_{si} - u\right). \tag{2}$$

   Here, $K_{h_s}(\boldsymbol{\omega}_{[s]}^\tau \boldsymbol{X}_{si} - u) = h_s^{-1} K\left(\dfrac{\boldsymbol{\omega}_{[s]}^\tau \boldsymbol{X}_{si} - u}{h_s}\right)$ where $K(u)$ is a kernel function and h_s is a bandwidth. Let $(\hat{a}_{s0}, \hat{a}_{s1})$ be the minimizer of (2). Then, the estimator of $g_s(u)$ is obtained as

$$\begin{aligned}\hat{g}_s(u, \boldsymbol{\omega}_{[s]}) &= \hat{a}_{s0} \\ &= \frac{T_{n_s,20}(u, \boldsymbol{\omega}_{[s]}) T_{n_s,01}(u, \boldsymbol{\omega}_{[s]}) - T_{n_s,10}(u, \boldsymbol{\omega}_{[s]}) T_{n_s,11}(u, \boldsymbol{\omega}_{[s]})}{T_{n_s,00}(u, \boldsymbol{\omega}_{[s]}) T_{n_s,20}(u, \boldsymbol{\omega}_{[s]}) - T_{n_s,10}^2(u, \boldsymbol{\omega}_{[s]})}.\end{aligned} \tag{3}$$

   where $T_{n_s,l_1 l_2}(u, \boldsymbol{\omega}_{[s]}) = \frac{1}{n_s} \sum_{i=1}^{n_s} K_{h_s}(\boldsymbol{\omega}_{[s]}^\tau \boldsymbol{X}_{si} - u)(\boldsymbol{\omega}_{[s]}^\tau \boldsymbol{X}_{si} - u)^{l_1} Y_{si}^{l_2}$ for $l_1 = 0, 1, 2$ and $l_2 = 0, 1$.

2. Similar to (2) and (3), the local linear smoothing technique is used to estimate variance function $\sigma_s^2(u)$, and we obtain that

$$\hat{\sigma}_s^2(u, \boldsymbol{\omega}_{[s]}) = \frac{S_{n_s,20}(u, \boldsymbol{\omega}_{[s]})S_{n_s,01}(u, \boldsymbol{\omega}_{[s]}) - S_{n_s,10}(u, \boldsymbol{\omega}_{[s]})S_{n_s,11}(u, \boldsymbol{\omega}_{[s]})}{S_{n_s,00}(u, \boldsymbol{\omega}_{[s]})S_{n_s,20}(u, \boldsymbol{\omega}_{[s]}) - S_{n_s,10}^2(u, \boldsymbol{\omega}_{[s]})}, \quad (4)$$

where

$$S_{n_s,l_1 l_2}(u, \boldsymbol{\omega}_{[s]}) = \frac{1}{n_s}\sum_{i=1}^{n_s}\left[\left(Y_{si} - \hat{g}_s(\boldsymbol{\omega}_{[s]}^\tau \boldsymbol{X}_{si}, \boldsymbol{\omega}_{[s]})\right)^2\right]^{l_2} K_{h_s}(\boldsymbol{\omega}_{[s]}^\tau \boldsymbol{X}_{si} - u)(\boldsymbol{\omega}_{[s]}^\tau \boldsymbol{X}_{si} - u)^{l_1},$$

for $l_1 = 0, 1, 2$ and $l_2 = 0, 1$.

3. We now proceed to estimate $\boldsymbol{\omega}_{[s],0}$ using the profile estimation function (Cui et al. 2011; Liang et al. 2010) and the "leave-one-component out" procedure in the following estimation equation

$$\mathcal{W}_{n_s}\left(\boldsymbol{\omega}_{[s]}^{(1)}\right) \overset{\text{def}}{=} \sum_{i=1}^{n_s} J_{\boldsymbol{\omega}_{[s]}}^\tau \hat{g}_s'(\boldsymbol{\omega}_{[s]}^\tau \boldsymbol{X}_{si}, \boldsymbol{\omega}_{[s]})\left[\boldsymbol{X}_{si} - \hat{V}_s(\boldsymbol{\omega}_{[s]}^\tau \boldsymbol{X}_{si}, \boldsymbol{\omega}_{[s]})\right]\hat{\sigma}_s^{-2}(\boldsymbol{\omega}_{[s]}^\tau \boldsymbol{X}_{si}, \boldsymbol{\omega}_{[s]}) \times \left[Y_{si} - \hat{g}_s(\boldsymbol{\omega}_{[s]}^\tau \boldsymbol{X}_{si}, \boldsymbol{\omega}_{[s]})\right], \quad (5)$$

where, $\hat{g}_s'(u, \boldsymbol{\omega}_{[s]}) = \frac{\partial \hat{g}_s(u,\boldsymbol{\omega})}{\partial u}$, $J_{\boldsymbol{\omega}_{[s]}} = \partial \boldsymbol{\omega}_{[s]}/\partial \boldsymbol{\omega}_{[s]}^{(1)}$ is the Jacobian matrix with

$$J_{\boldsymbol{\omega}_{[s]}} = \begin{pmatrix} -\boldsymbol{\omega}_{[s]}^{(1)\tau}/\sqrt{1 - \left\|\boldsymbol{\omega}_{[s]}^{(1)}\right\|^2} \\ I_{[s]-1} \end{pmatrix},$$

where $I_{[s]-1} = \text{diag}(1, \ldots, 1)$, an identity matrix of size $p-1$ for $s = 1$ and $q-1$ for $s = 2$, respectively. Moreover, $\hat{V}_s(u, \boldsymbol{\omega}_{[s]})$ is the local linear estimator of $V_{s,\boldsymbol{\omega}_{[s]}}(u)$, where $V_{s,\boldsymbol{\omega}_{[s]}}(u) = \left(E(X_{s,1}|\boldsymbol{\omega}_{[s]}^\tau \boldsymbol{X}_s = u), \ldots, E(X_{s,v_s}|\boldsymbol{\omega}_{[s]}^\tau \boldsymbol{X}_s = u)\right)^\tau$ for $s = 1, 2$, $v_1 = p$ and $v_2 = q$. The estimator $\hat{V}_s(u, \boldsymbol{\omega}_{[s]})$ is defined as $\hat{V}_s(u, \boldsymbol{\omega}_{[s]}) = \left(\hat{V}_{s,1}(u, \boldsymbol{\omega}_{[s]}), \ldots, \hat{V}_{s,v_s}(u, \boldsymbol{\omega}_{[s]})\right)^\tau$, where $\hat{V}_{s,l}(u, \boldsymbol{\omega}_{[s]}) = \frac{\sum_{i=1}^{n_s} b_{n_s,i}(u,\boldsymbol{\omega}_{[s]})\boldsymbol{X}_{s,li}}{\sum_{i=1}^{n_s} b_{n_s,i}(u,\boldsymbol{\omega}_{[s]})}$, for $l = 1, \ldots, v_s$, where $b_{n_s,i}(u, \boldsymbol{\omega}_{[s]}) = K_{h_s}(\boldsymbol{\omega}_{[s]}^\tau \boldsymbol{X}_{si} - u)\left[T_{n_s,20}(u, \boldsymbol{\omega}_{[s]}) - (\boldsymbol{\omega}_{[s]}^\tau \boldsymbol{X}_{si} - t)T_{n_s,10}(u, \boldsymbol{\omega}_{[s]})\right]$.

Let $\hat{\boldsymbol{\omega}}_{[s],0}^{(1)}$ denote the solution of the estimation equation $\mathcal{W}_{n_s}\left(\hat{\boldsymbol{\omega}}_{[s],0}^{(1)}\right) = 0$. Then, $\hat{\boldsymbol{\omega}}_{[s],0,1}$ is obtained by $\hat{\boldsymbol{\omega}}_{[s],0,1} = \sqrt{1 - \left\|\hat{\boldsymbol{\omega}}_{[s],0}^{(1)}\right\|^2}$, and the estimator of $\boldsymbol{\omega}_{[s],0}$ is defined as $\hat{\boldsymbol{\omega}}_{[s],0} = \left(\hat{\boldsymbol{\omega}}_{[s],0,1}, \hat{\boldsymbol{\omega}}_{[s],0}^{(1)\tau}\right)^{\tau}$. Finally, the estimators of $g_s(u)$ and $\sigma_s^2(u)$ are obtained by substituting $\boldsymbol{\omega}$ with $\hat{\boldsymbol{\omega}}_{[s],0}$ in (3) and (4), respectively.

In what follows, $A^{\otimes 2} = AA^{\tau}$ for any matrix or vector A. We list the conditions needed in our asymptotic results.

(C1) $E[X_{1,r}^4] < \infty$, $E[X_{2,l}^4] < \infty$ for $r = 1, \ldots, p$ and $l = 1, \ldots, q$, and the covariance matrices $\boldsymbol{\Omega}_s$, $s = 1, 2$ defined in Theorem 1 are both finite.

(C2) For $s = 1, 2$, functions $g_s(u)$, $\sigma_s(u)$, $E(\boldsymbol{X}_{s,l}|\boldsymbol{\omega}_{[s]}^{\tau}\boldsymbol{X}_s = u)$, $l = 1, \ldots, v_s$, $v_1 = p$ and $v_2 = q$, and the density function $f_{\boldsymbol{\omega}_{[s]}^{\tau}\boldsymbol{X}_s}(u)$ of random variable $\boldsymbol{\omega}_{[s]}^{\tau}\boldsymbol{X}_s$ are twice continuously differentiable with respect to u. Their second derivatives are uniformly Lipschitz continuous on $\mathcal{C}_s = \{u = \boldsymbol{\omega}_{[s]}^{\tau}x_s : x_1 \in \mathcal{X}_1 \subset \mathbb{R}^p, x_2 \in \mathcal{X}_2 \subset \mathbb{R}^q, \boldsymbol{\beta} \in \mathcal{B}_{\boldsymbol{\beta}}, \boldsymbol{\gamma} \in \mathcal{B}_{\boldsymbol{\gamma}}\}$, where $\mathcal{X}_s$ is a compact support set. Furthermore, $\inf_{u\in\mathcal{C}_s} f_{\boldsymbol{\omega}_{[s]}^{\tau}\boldsymbol{X}_s}(u) > 0$ and $\inf_{u\in\mathcal{C}_s} \sigma_s(u) \geq c_0 > 0$ for some positive constant c_0, and $\int \sigma_s^2(u) f_{\boldsymbol{\omega}_{[s],0}^{\tau}\boldsymbol{X}_s}(u)du < \infty$.

(C3) Kernel function $K(u)$ is a symmetric bounded density function supported on $[-A, A]$, satisfying a Lipschitz condition, and has twice continuous bounded derivative, satisfying $K^{(j)}(\pm A) = 0$ for $j = 0, 1, 2, 3$, and $\int s^2 K(s)ds \neq 0$.

(C4) As $n \to \infty$, the bandwidths h_s, $s = 1, 2$ satisfy $h_s(\log n_s)^{1+s_0} \to 0$, $n_s h_s^4 \to 0$ and $\frac{(\log n_s)^{2+2s_0}}{n_s h_s^2} \to 0$ for some constant $s_0 > 0$.

(C5) For $s = 1, 2$, model error ϵ_s satisfies $E[\epsilon_s^4] < \infty$, the distribution function $F_{\epsilon_s}(t)$ of ϵ_s is twice continuously differentiable. Further, the density function $f_{\epsilon_s}(t)$ of ϵ_s satisfies $\int f_{\epsilon_s}^2(t)dF_{\epsilon_s}(t) < \infty$, $\sup_{t\in\mathbb{R}} f_{\epsilon_s}(t) < \infty$, $\sup_{t\in\mathbb{R}} |t| f_{\epsilon_s}(t) < \infty$ and $\sup_{t\in\mathbb{R}} t^2 |f_{\epsilon_s}'(t)| < \infty$.

Theorem 1 *Under the conditions* (C1)–(C4)*, we have*

$$\sqrt{n_s}\left(\hat{\boldsymbol{\omega}}_{[s],0}^{(1)} - \boldsymbol{\omega}_{[s],0}^{(1)}\right) \xrightarrow{L} N\left(\mathbf{0}, \boldsymbol{\Omega}_s^{-1}\right),$$

where

$$\boldsymbol{\Omega}_s = J_{\boldsymbol{\omega}_{[s],0}}^{\tau} E\left[g_s^{\prime 2}(\boldsymbol{\omega}_{[s],0}^{\tau}\boldsymbol{X}_s)\sigma_s^{-2}(\boldsymbol{\omega}_{[s],0}^{\tau}\boldsymbol{X}_s)\left[\boldsymbol{X}_s - V_{s,\boldsymbol{\omega}_{[s],0}}(\boldsymbol{\omega}_{[s]}^{\tau}\boldsymbol{X}_1)\right]^{\otimes 2}\right] J_{\boldsymbol{\omega}_{[s],0}},$$

and $V_{s,\boldsymbol{\omega}_{[s],0}}(u) = E\left(\boldsymbol{X}_s|\boldsymbol{\omega}_{[s],0}^{\tau}\boldsymbol{X}_s = u\right)$. *Furthermore, by a simple application of the multivariate delta-method, we also have*

$$\sqrt{n_s}\left(\hat{\boldsymbol{\omega}}_{[s],0} - \boldsymbol{\omega}_0\right) \xrightarrow{L} N\left(\mathbf{0}, J_{\boldsymbol{\omega}_{[s],0}}\boldsymbol{\Omega}_s^{-1}J_{\boldsymbol{\omega}_{[s],0}}^{\tau}\right).$$

Remark 1 The population version of (5) when $\boldsymbol{\omega}_{[s]}^{(1)} = \boldsymbol{\omega}_{[s],0}^{(1)}$ is defined as

$$\mathcal{W}_{n_s}^{*}\left(\boldsymbol{\omega}_{[s],0}^{(1)}\right) = \sum_{i=1}^{n_s} J_{\boldsymbol{\omega}_{[s],0}}^{\tau} g_s'(\boldsymbol{\omega}_{[s]}^{\tau}\boldsymbol{X}_{si}) \times \left[\boldsymbol{X}_{si} - V_{s,\boldsymbol{\omega}_{[s],0}}(\boldsymbol{\omega}_{[s]}^{\tau}\boldsymbol{X}_{si})\right]\sigma_s^{-2}(\boldsymbol{\omega}_{[s]}^{\tau}\boldsymbol{X}_{si})\left[Y_{si} - g_s(\boldsymbol{\omega}_{[s]}^{\tau}\boldsymbol{X}_{si})\right].$$

The function $\mathcal{W}^*_{n_s}\left(\boldsymbol{\omega}^{(1)}_{[s],0}\right)$ satisfies the second Bartlett identity as Cui et al. (2011) claimed, that is,

$$E\left[\mathcal{W}^*_{n_s}\left(\boldsymbol{\omega}^{(1)}_{[s],0}\right)\mathcal{W}^{*\tau}_{n_s}\left(\boldsymbol{\omega}^{(1)}_{[s],0}\right)\right] = -E\left[\frac{\partial \mathcal{W}^*_{n_s}\left(\boldsymbol{\omega}^{(1)}_{[s],0}\right)}{\partial \boldsymbol{\omega}^{(1)}_{[s],0}}\right] = n_s \boldsymbol{\Omega}_s. \tag{6}$$

This makes the estimators $\hat{\boldsymbol{\omega}}^{(1)}_{[s],0}$ semi-parametric efficiency (Cui et al. 2011).

3 Test statistics and their asymptotic properties

3.1 Testing the equality of single-index parameters

When the dimension of $\boldsymbol{X}_1$ is equal to the dimension of $\boldsymbol{X}_2$, i.e., $p = q$, we are interested in

$$\mathcal{H}_0 : \boldsymbol{\beta}_0 = \boldsymbol{\gamma}_0 \ \text{ against } \ \mathcal{H}_1 : \boldsymbol{\beta}_0 \neq \boldsymbol{\gamma}_0. \tag{7}$$

By the identifiability condition and the fact that first component of $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}_0$ are all positive, (7) is equivalent to

$$\mathcal{H}^*_0 : \boldsymbol{\beta}^{(1)}_0 = \boldsymbol{\gamma}^{(1)}_0 \ \text{ against } \ \mathcal{H}^*_1 : \boldsymbol{\beta}^{(1)}_0 \neq \boldsymbol{\gamma}^{(1)}_0. \tag{8}$$

We propose the test statistic

$$\mathcal{T}_{n_1 n_2} = \left(\hat{\boldsymbol{\beta}}^{(1)}_0 - \hat{\boldsymbol{\gamma}}^{(1)}_0\right)^\tau \widehat{\boldsymbol{A}}^{-1}\left(\hat{\boldsymbol{\beta}}^{(1)}_0 - \hat{\boldsymbol{\gamma}}^{(1)}_0\right),$$

where

$$\widehat{\boldsymbol{A}} = \left[J^\tau_{\hat{\boldsymbol{\beta}}_0}\sum_{i=1}^{n_1}\frac{\hat{g}_1'^2(\hat{\boldsymbol{\beta}}^\tau_0\boldsymbol{X}_{1i},\hat{\boldsymbol{\beta}}_0)}{\hat{\sigma}^2_1(\hat{\boldsymbol{\beta}}^\tau_0\boldsymbol{X}_{1i},\hat{\boldsymbol{\beta}}_0),}\left[\boldsymbol{X}_{1i} - \hat{V}_1(\hat{\boldsymbol{\beta}}^\tau_0\boldsymbol{X}_{1i},\hat{\boldsymbol{\beta}}_0)\right]^{\otimes 2} J_{\hat{\boldsymbol{\beta}}_0}\right]^{-1}$$
$$+\left[J^\tau_{\hat{\boldsymbol{\gamma}}_0}\sum_{i=1}^{n_2}\frac{\hat{g}_2'^2(\hat{\boldsymbol{\gamma}}^\tau_0\boldsymbol{X}_{2i},\hat{\boldsymbol{\gamma}}_0)}{\hat{\sigma}^2_2(\hat{\boldsymbol{\gamma}}^\tau_0\boldsymbol{X}_{2i},\hat{\boldsymbol{\gamma}}_0),}\left[\boldsymbol{X}_{2i} - \hat{V}_2(\hat{\boldsymbol{\gamma}}^\tau_0\boldsymbol{X}_{2i},\hat{\boldsymbol{\gamma}}_0)\right]^{\otimes 2} J_{\hat{\boldsymbol{\gamma}}_0}\right]^{-1}.$$

Theorem 2 *Under the conditions of* Theorem 1, *if* $\frac{n_1}{n_1+n_2} \to \lambda \in (0, 1)$, *we have*

$$\mathcal{T}_{n_1 n_2} \xrightarrow[\mathcal{H}^*_0]{\mathcal{L}} \chi^2_{p-1}.$$

If the null hypothesis $\mathcal{H}_0$ (or $\mathcal{H}^*_0$) is true, the pooled-sample $\{\boldsymbol{X}_{1i}, Y_{1i}, \boldsymbol{X}_{2j}, Y_{2j}\}$ can be used to re-estimate $\boldsymbol{\beta}_0$ (=$\boldsymbol{\gamma}_0$). Analogous to (5), we estimate $\boldsymbol{\beta}_0$ by using the "pooled" estimation equation

$$\mathcal{W}_{n_1n_2}(\boldsymbol{\beta}^{(1)}) \stackrel{\text{def}}{=} \sum_{s=1}^{2}\sum_{i=1}^{n_s} J_{\boldsymbol{\beta}}^{\tau}\hat{g}_s'(\boldsymbol{\beta}^{\tau}\boldsymbol{X}_{si},\boldsymbol{\beta})\left[\boldsymbol{X}_{si}-\hat{V}_s(\boldsymbol{\beta}^{\tau}\boldsymbol{X}_{si},\boldsymbol{\beta})\right]\hat{\sigma}_s^{-2}(\boldsymbol{\beta}^{\tau}\boldsymbol{X}_{si},\boldsymbol{\beta}) \times\left[Y_{si}-\hat{g}_s(\boldsymbol{\beta}^{\tau}\boldsymbol{X}_{si},\boldsymbol{\beta})\right]. \quad (9)$$

Here, $\hat{g}_2(u,\boldsymbol{\beta})$, $\hat{\sigma}_2(u,\boldsymbol{\beta})$ and $\hat{V}_2(u,\boldsymbol{\beta})$ are defined according to $\hat{g}_2(u,\boldsymbol{\gamma})$, $\hat{\sigma}_2(u,\boldsymbol{\gamma})$ and $\hat{V}_2(u,\boldsymbol{\gamma})$ by substituting $\boldsymbol{\gamma}$ with $\boldsymbol{\beta}$, respectively. Moreover, $\hat{g}_2'(u,\boldsymbol{\beta})=\frac{\partial\hat{g}_2(u,\boldsymbol{\beta})}{\partial u}$. Denote the estimator from (9) as $\hat{\boldsymbol{\beta}}_{\mathcal{H}_0}^{(1)}$, and define the final estimator of $\boldsymbol{\beta}_0$ under the null hypothesis $\mathcal{H}_0$ as $\hat{\boldsymbol{\beta}}_{\mathcal{H}_0}=\left(\sqrt{1-\left\|\hat{\boldsymbol{\beta}}_{\mathcal{H}_0}^{(1)}\right\|_2^2},\hat{\boldsymbol{\beta}}_{\mathcal{H}_0}^{(1)\tau}\right)^{\tau}$. Similar to Theorem 1, we obtain the following asymptotic result.

Theorem 3 *Under the conditions of* Theorem 2, *we have*

$$\sqrt{n_1+n_2}\left(\hat{\boldsymbol{\beta}}_{\mathcal{H}_0}^{(1)}-\boldsymbol{\beta}_0^{(1)}\right)\stackrel{L}{\longrightarrow}N\left(\mathbf{0}_{p-1},\boldsymbol{\Omega}_{\mathcal{H}_0}^{-1}\right),$$

where

$$\begin{aligned}\boldsymbol{\Omega}_{\mathcal{H}_0} &= \lambda J_{\boldsymbol{\beta}_0}^{\tau}E\left[g_1'^2(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_1)\sigma_1^{-2}(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_1)\left[\boldsymbol{X}_1-V_{1,\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_1)\right]^{\otimes 2}\right]J_{\boldsymbol{\beta}_0}\\ &\quad+(1-\lambda)J_{\boldsymbol{\beta}_0}^{\tau}E\left[g_2'^2(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_2)\sigma_2^{-2}(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_2)\left[\boldsymbol{X}_2-V_{2,\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_2)\right]^{\otimes 2}\right]J_{\boldsymbol{\beta}_0}. \quad (10)\end{aligned}$$

Also, a simple application of the multivariate delta-method makes that

$$\sqrt{n_1+n_2}\left(\hat{\boldsymbol{\beta}}_{\mathcal{H}_0}-\boldsymbol{\beta}_0\right)\stackrel{L}{\longrightarrow}N\left(\mathbf{0}_p,J_{\boldsymbol{\beta}_0}\boldsymbol{\Omega}_{\mathcal{H}_0}^{-1}J_{\boldsymbol{\beta}_0}^{\tau}\right).$$

3.2 Testing the equality of the mean functions

The idea for testing the equality of the mean functions, i.e.,

$$\widetilde{\mathcal{H}}_0: g_1(u)=g_2(u)\ \text{ for each }\ u, \quad (11)$$

against

$$\widetilde{\mathcal{H}}_1: g_1(u)\neq g_2(u)\ \text{ for some }\ u,$$

is based on a comparison between the estimated error distribution $\hat{F}_{\widetilde{\mathcal{H}}_0,\epsilon_s}(t)$ obtained under $\widetilde{\mathcal{H}}_0$ and the estimated error distribution $\hat{F}_{\epsilon_s}(t)$ obtained under the alternative hypothesis (Neumeyer and Van Keilegom 2010; Van Keilegom et al. 2008; Dette et al. 2007). That is, we adopt the Kolmogorov–Smirnov or Cramér–von Mises test statistics for the processes

$$\hat{F}_{\widetilde{\mathcal{H}}_{0,\epsilon_1}}(t) - \hat{F}_{\epsilon_1}(t) \text{ and } \hat{F}_{\widetilde{\mathcal{H}}_{0,\epsilon_2}}(t) - \hat{F}_{\epsilon_2}(t).$$

We first introduce the method for obtaining estimators $\hat{F}_{\widetilde{\mathcal{H}}_{0,\epsilon_s}}(t)$ and $\hat{F}_{\epsilon_s}(t)$. Under the null hypothesis $\widetilde{\mathcal{H}}_0$, the estimators of $F_{\epsilon_1}(t)$ and $F_{\epsilon_2}(t)$ are obtained as

$$\hat{F}_{\widetilde{\mathcal{H}}_{0,\epsilon_1}}(t) = \frac{1}{n_1}\sum_{i=1}^{n_1} I\left\{\hat{\epsilon}_{\widetilde{\mathcal{H}}_{0,1i}} \le t\right\}, \quad \text{where } \hat{\epsilon}_{\widetilde{\mathcal{H}}_{0,1i}} = \frac{Y_{1i} - \hat{g}_2(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}_{1i}, \hat{\boldsymbol{\gamma}}_0)}{\hat{\sigma}_1(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}_{1i}, \hat{\boldsymbol{\beta}}_0)},$$

$$\hat{F}_{\widetilde{\mathcal{H}}_{0,\epsilon_2}}(t) = \frac{1}{n_2}\sum_{i=1}^{n_2} I\left\{\hat{\epsilon}_{\widetilde{\mathcal{H}}_{0,2i}} \le t\right\}, \quad \text{where } \hat{\epsilon}_{\widetilde{\mathcal{H}}_{0,2i}} = \frac{Y_{2i} - \hat{g}_1(\hat{\boldsymbol{\gamma}}_0^\tau \boldsymbol{X}_{2i}, \hat{\boldsymbol{\beta}}_0)}{\hat{\sigma}_2(\hat{\boldsymbol{\gamma}}_0^\tau \boldsymbol{X}_{2i}, \hat{\boldsymbol{\gamma}}_0)},$$

and under the alternative hypothesis $\widetilde{\mathcal{H}}_1$,

$$\hat{F}_{\epsilon_1}(t) = \frac{1}{n_1}\sum_{i=1}^{n_1} I\left\{\hat{\epsilon}_{1i} \le t\right\}, \quad \text{where } \hat{\epsilon}_{1i} = \frac{Y_{1i} - \hat{g}_1(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}_{1i}, \hat{\boldsymbol{\beta}}_0)}{\hat{\sigma}_1(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}_{1i}, \hat{\boldsymbol{\beta}}_0)},$$

$$\hat{F}_{\epsilon_2}(t) = \frac{1}{n_2}\sum_{i=1}^{n_2} I\left\{\hat{\epsilon}_{2i} \le t\right\}, \quad \text{where } \hat{\epsilon}_{2i} = \frac{Y_{2i} - \hat{g}_2(\hat{\boldsymbol{\gamma}}_0^\tau \boldsymbol{X}_{2i}, \hat{\boldsymbol{\gamma}}_0)}{\hat{\sigma}_2(\hat{\boldsymbol{\gamma}}_0^\tau \boldsymbol{X}_{2i}, \hat{\boldsymbol{\gamma}}_0)}.$$

Under $\widetilde{\mathcal{H}}_0$, the difference between two functions $\hat{F}_{\widetilde{\mathcal{H}}_{0,\epsilon_s}}(t)$ and $\hat{F}_{\epsilon_s}(t)$, for $s = 1, 2$, will be small. In other words, if $\widetilde{\mathcal{H}}_0$ holds, the processes are not distinguishable. Whereas under the alternative hypothesis, the differences should be obvious. To test $\widetilde{\mathcal{H}}_0$, we propose Kolmogorov–Smirnov and Cramér–von Mises type functionals based on the following test statistics:

$$\mathfrak{T}_{n_1 n_2}^{\mathrm{KS}} = \sup_{t\in\mathbb{R}} n_1^{1/2}\left|\hat{F}_{\widetilde{\mathcal{H}}_{0,\epsilon_1}}(t) - \hat{F}_{\epsilon_1}(t)\right| + \sup_{t\in\mathbb{R}} n_2^{1/2}\left|\hat{F}_{\widetilde{\mathcal{H}}_{0,\epsilon_2}}(t) - \hat{F}_{\epsilon_2}(t)\right|,$$

and

$$\mathfrak{T}_{n_1 n_2}^{\mathrm{CM}} = n_1\int\left|\hat{F}_{\widetilde{\mathcal{H}}_{0,\epsilon_1}}(t) - \hat{F}_{\epsilon_1}(t)\right|^2 d\hat{F}_{\epsilon_1}(t) + n_2\int\left|\hat{F}_{\widetilde{\mathcal{H}}_{0,\epsilon_2}}(t) - \hat{F}_{\epsilon_2}(t)\right|^2 d\hat{F}_{\epsilon_2}(t).$$

Theorem 4 shows that $\hat{F}_{\epsilon_s}(t)$ consistently estimates $F_{\epsilon_s}(t)$, and Theorems 5 and 6 show that $\hat{F}_{\widetilde{\mathcal{H}}_{0,\epsilon_s}}(t)$ can consistently estimate the distribution of error $\epsilon_{\widetilde{\mathcal{H}}_{0,s}} = \frac{Y_1 - g_2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_1)}{\sigma_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_1)} I\{s=1\} + \frac{Y_2 - g_1(\boldsymbol{\gamma}_0^\tau \boldsymbol{X}_2)}{\sigma_2(\boldsymbol{\gamma}_0^\tau \boldsymbol{X}_2)} I\{s=2\}$.

Theorem 4 *Suppose the conditions of* Theorem 3 *and condition (C5) are satisfied, we have the following asymptotic expression:*

$$\hat{F}_{\epsilon_s}(t) - F_{\epsilon_s}(t) = \frac{1}{n_s}\sum_{i=1}^{n_s}\left[I\{\epsilon_{si} \le t\} - F_{\epsilon_s}(t) + f_{\epsilon_s}(t)\left(\epsilon_{si} + \frac{t}{2}(\epsilon_{si}^2 - 1)\right)\right] + o_P\left(n_s^{-1/2}\right), \quad (12)$$

uniformly in $t \in \mathbb{R}^1$, $s = 1, 2$.

Remark 2 Note that the asymptotic result of Theorem 4 is the same as Theorem 2.1 in Neumeyer and Van Keilegom (2010). In fact, the process $\sqrt{n_s}\left(\hat{F}_{\epsilon_s}(t) - F_{\epsilon_s}(t)\right)$, $t \in \mathbb{R}$, converges weakly to a zero-mean Gaussian process $\mathcal{Z}_s(t)$ with a covariance function

$$E\Bigg[\left(I\{\epsilon_s \le t_1\} - F_{\epsilon_s}(t_1) + f_{\epsilon_s}(t_1)\left(\epsilon_s + \frac{t_1}{2}(\epsilon_s^2 - 1)\right)\right) \times \left(I\{\epsilon_s \le t_2\} - F_{\epsilon_s}(t_2) + f_{\epsilon_s}(t_2)\left(\epsilon_s + \frac{t_2}{2}(\epsilon_s^2 - 1)\right)\right)\Bigg],$$

for any $t_1 \le t_2$.

Next, we present the asymptotic expressions for $\hat{F}_{\widetilde{\mathcal{H}}_{0,\epsilon_1}}(t)$ and $\hat{F}_{\widetilde{\mathcal{H}}_{0,\epsilon_2}}(t)$. In the following, we define $D_1(u) = \frac{g_2(u)-g_1(u)}{\sigma_1(u)}$, $D_2(u) = \frac{g_1(u)-g_2(u)}{\sigma_2(u)}$, $\rho_{f,\sigma}(u) = \frac{f_{\beta_0^\tau X_1}(u)}{f_{\gamma_0^\tau X_2}(u)}\frac{\sigma_2(u)}{\sigma_1(u)}$, and

$$F^*_{\widetilde{\mathcal{H}}_{0,\epsilon_s}}(t) = E\left[F_{\epsilon_s}\left(t + D_s(\boldsymbol{\omega}_{[s],0}^\tau \boldsymbol{X}_s)\right)\right],$$

$$\mathcal{N}_s = E\left[\frac{g_s'(\boldsymbol{\omega}_{[s],0}^\tau \boldsymbol{X}_s)}{\sigma_s(\boldsymbol{\omega}_{[s],0}^\tau \boldsymbol{X}_s)} V_{s,\boldsymbol{\omega}_{[s],0}}(\boldsymbol{\omega}_{[s],0}^\tau \boldsymbol{X}_s)\right],$$

$$M_1(t) = E\left[f_{\epsilon_1}\left(t - D_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_1)\right)\frac{g_2'(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_1)}{\sigma_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_1)} V_{1,\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_1)\right],$$

$$M_2(t) = E\left[f_{\epsilon_2}\left(t - D_2(\boldsymbol{\gamma}_0^\tau \boldsymbol{X}_2)\right)\frac{g_1'(\boldsymbol{\gamma}_0^\tau \boldsymbol{X}_2)}{\sigma_2(\boldsymbol{\gamma}_0^\tau \boldsymbol{X}_2)} V_{2,\boldsymbol{\gamma}_0}(\boldsymbol{\gamma}_0^\tau \boldsymbol{X}_2)\right].$$

Theorem 5 *Under the conditions of* Theorem 4, *we have*

$$\begin{aligned}
\hat{F}_{\widetilde{\mathcal{H}}_{0,\epsilon_s}}(t) - F^*_{\widetilde{\mathcal{H}}_{0,\epsilon_s}}(t) &= \frac{1}{n_s}\sum_{i=1}^{n_s} I\left\{\epsilon_{si} - D_s(\boldsymbol{\omega}_{[s],0}^\tau \boldsymbol{X}_{si}) \le t\right\} - F^*_{\widetilde{\mathcal{H}}_{0,\epsilon_s}}(t)\\
&+ \frac{t}{2n_s}\sum_{i=1}^{n_s} f_{\epsilon_s}\left(t - D_s(\boldsymbol{\omega}_{[s],0}^\tau \boldsymbol{X}_{si})\right)(\epsilon_{si}^2 - 1)\\
&+ \left[\frac{1}{n_2}\sum_{i=1}^{n_2} f_{\epsilon_1}\left(t - D_1(\boldsymbol{\gamma}_0^\tau \boldsymbol{X}_{2i})\right)\rho_{f,\sigma}(\boldsymbol{\gamma}_0^\tau \boldsymbol{X}_{2i})\epsilon_{2i}\right] I\{s=1\}\\
&+ \left[\frac{1}{n_1}\sum_{i=1}^{n_1} f_{\epsilon_2}\left(t - D_2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i})\right)\rho_{f,\sigma}^{-1}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i})\epsilon_{1i}\right] I\{s=2\}\\
&+ M_s^\tau(t) J_{\boldsymbol{\omega}_{[s],0}} \boldsymbol{\Omega}_s^{-1}\frac{1}{n_s}\sum_{i=1}^{n_s} J_{\boldsymbol{\omega}_{[s],0}}^\tau \frac{g_s'(\boldsymbol{\omega}_{[s],0}^\tau \boldsymbol{X}_{si})}{\sigma_s(\boldsymbol{\omega}_{[s],0}^\tau \boldsymbol{X}_{si})}\left[\boldsymbol{X}_{si} - V_{s,\boldsymbol{\omega}_{[s],0}}(\boldsymbol{\omega}_{[s],0}^\tau \boldsymbol{X}_{si})\right]\epsilon_{si}\\
&+ o_P\left(n_1^{-1/2} + n_2^{-1/2}\right).
\end{aligned}$$

Under the null hypothesis $\widetilde{\mathcal{H}}_0$, we have the following asymptotic results.

Theorem 6 *Under the conditions of* Theorem 4, *if* $\widetilde{\mathcal{H}}_0$ *holds, we have*

$$\begin{aligned}\hat{F}_{\widetilde{\mathcal{H}}_0,\epsilon_s}(t) - \hat{F}_{\epsilon_s}(t) &= f_{\epsilon_s}(t)\frac{1}{n_s}\sum_{i=1}^{n_s}\epsilon_{si} + f_{\epsilon_s}(t)\mathcal{N}_s^\tau J_{\boldsymbol{\omega}_{[s],0}}\boldsymbol{\Omega}_s^{-1} \\ &\times \frac{1}{n_s}\sum_{i=1}^{n_s} J^\tau_{\boldsymbol{\omega}_{[s],0}}\frac{g_s'(\boldsymbol{\omega}^\tau_{[s],0}\boldsymbol{X}_{si})}{\sigma_s(\boldsymbol{\omega}^\tau_{[s],0}\boldsymbol{X}_{si})}\left[\boldsymbol{X}_{si} - V_{s,\boldsymbol{\omega}_{[s],0}}(\boldsymbol{\omega}^\tau_{[s],0}\boldsymbol{X}_{si})\right]\epsilon_{si} \\ &+ \left[f_{\epsilon_s}(t)\frac{1}{n_2}\sum_{i=1}^{n_2}\rho_{f,\sigma}(\boldsymbol{\gamma}_0^\tau\boldsymbol{X}_{2i})\epsilon_{2i}\right]I\{s=1\} \\ &+ \left[f_{\epsilon_s}(t)\frac{1}{n_1}\sum_{i=1}^{n_1}\rho^{-1}_{f,\sigma}(\boldsymbol{\beta}_0^\tau\boldsymbol{X}_{1i})\epsilon_{1i}\right]I\{s=2\} \\ &+ o_P\left(n_1^{-1/2}+n_2^{-1/2}\right).\end{aligned}$$

Note that the limiting process of Theorem 6 is a product of two objects: one is a deterministic density function $f_{\epsilon_s}(t)$ only depending on t, and the other is a summation of random variables with mean zero that are independent of t. According to the asymptotic results of Theorem 6, the continuous mapping theorem entails the weak convergence properties of test statistics $\mathfrak{T}^{KS}_{n_1n_2}$ and $\mathfrak{T}^{CM}_{n_1n_2}$.

Theorem 7 *Under the conditions of* Theorem 4, *if* $\widetilde{\mathcal{H}}_0$ *holds, we have*

$$\begin{aligned}\mathfrak{T}^{KS}_{n_1n_2} &\xrightarrow{\mathcal{L}} \left(\sup_{t\in\mathbb{R}} f_{\epsilon_1}(t)\right)|\boldsymbol{\xi}| + \left(\sup_{t\in\mathbb{R}} f_{\epsilon_2}(t)\right)|\boldsymbol{\eta}|, \\ \mathfrak{T}^{CM}_{n_1n_2} &\xrightarrow{\mathcal{L}} \left(\int f^2_{\epsilon_1}(t)dF_{\epsilon_1}(t)\right)\boldsymbol{\xi}^2 + \left(\int f^2_{\epsilon_2}(t)dF_{\epsilon_2}(t)\right)\boldsymbol{\eta}^2,\end{aligned}$$

where $\boldsymbol{\xi}$ *and* $\boldsymbol{\eta}$ *are two zero-mean normal random variables with the covariance and variances as*

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{\xi},\boldsymbol{\eta}) &= \sqrt{\frac{1-\lambda}{\lambda}}E\left[\rho^{-1}_{f,\sigma}(\boldsymbol{\beta}_0^\tau\boldsymbol{X}_1)\right] + \sqrt{\frac{\lambda}{1-\lambda}}E\left[\rho_{f,\sigma}(\boldsymbol{\gamma}_0^\tau\boldsymbol{X}_2)\right], \\ \mathrm{Var}(\boldsymbol{\xi}) &= \mathcal{N}_1^\tau J_{\boldsymbol{\beta}_0}\boldsymbol{\Omega}_1^{-1}J^\tau_{\boldsymbol{\beta}_0}\mathcal{N}_1 + \frac{\lambda}{1-\lambda}E\left[\rho^2_{f,\sigma}(\boldsymbol{\gamma}_0^\tau\boldsymbol{X}_2)\right]+1, \\ \mathrm{Var}(\boldsymbol{\eta}) &= \mathcal{N}_2^\tau J_{\boldsymbol{\gamma}_0}\boldsymbol{\Omega}_2^{-1}J^\tau_{\boldsymbol{\gamma}_0}\mathcal{N}_2 + \frac{1-\lambda}{\lambda}E\left[\rho^{-2}_{f,\sigma}(\boldsymbol{\beta}_0^\tau\boldsymbol{X}_1)\right]+1.\end{aligned}$$

Theorem 8 reveals that the limiting behavior of the two test statistics $\mathfrak{T}^{KS}_{n_1n_2}$ and $\mathfrak{T}^{CM}_{n_1n_2}$ can detect the local alternative $\widetilde{\mathcal{H}}_{1,n_1n_2}$ with order $O((n_1+n_2)^{-1/2})$ converging to the null hypothesis $\widetilde{\mathcal{H}}_0$.

Considering the local alternative hypothesis

$$\widetilde{\mathcal{H}}_{1,n_1n_2} : g_1(u) = g_2(u) + \frac{1}{\sqrt{n_1+n_2}}\mu(u), \text{ for every } u. \tag{13}$$

We can have the following asymptotic result.

Theorem 8 *Under the conditions of* Theorem 4, *if the local alternative hypothesis* (13) *holds, we have*

$$\mathfrak{T}^{\mathrm{KS}}_{n_1n_2} \xrightarrow{\mathcal{L}} \left(\sup_{t\in\mathbb{R}} f_{\epsilon_1}(t)\right)|\boldsymbol{\xi}+\boldsymbol{b}_1| + \left(\sup_{t\in\mathbb{R}} f_{\epsilon_2}(t)\right)|\boldsymbol{\eta}+\boldsymbol{b}_2|,$$

$$\mathfrak{T}^{\mathrm{CM}}_{n_1n_2} \xrightarrow{\mathcal{L}} \left(\int f^2_{\epsilon_1}(t)dF_{\epsilon_1}(t)\right)(\boldsymbol{\xi}+\boldsymbol{b}_1)^2 + \left(\int f^2_{\epsilon_2}(t)dF_{\epsilon_2}(t)\right)(\boldsymbol{\eta}+\boldsymbol{b}_2)^2,$$

where $\boldsymbol{b}_1 = -\sqrt{\lambda}E\left[\frac{\mu(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_1)}{\sigma_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_1)}\right]$ *and* $\boldsymbol{b}_2 = \sqrt{1-\lambda}E\left[\frac{\mu(\boldsymbol{\gamma}_0^\tau \boldsymbol{X}_2)}{\sigma_2(\boldsymbol{\gamma}_0^\tau \boldsymbol{X}_2)}\right]$.

3.3 A wild bootstrap procedure

In this subsection, we use the smooth residual bootstrap method proposed by Neumeyer and Van Keilegom (2010) and Neumeyer (2009) to mimic the distributions of the test statistics $\mathfrak{T}^{\mathrm{KS}}_{n_1n_2}$ and $\mathfrak{T}^{\mathrm{CM}}_{n_1n_2}$. The procedure is summarized as follows:

Step 1 Compute $\mathfrak{T}^{\mathrm{KS}}_{n_1n_2}$ and $\mathfrak{T}^{\mathrm{CM}}_{n_1n_2}$.

Step 2 Generate B times $i.i.d.$ variables ς_{ib} for $i = 1, \ldots, n_s$, $b = 1, \ldots, B$, and $s = 1, 2$ from a standard normal distribution $\mathcal{N}(0, 1)$. They are independent of the original sample $\{Y_{si}, \boldsymbol{X}_{si}, i = 1, \ldots, n_s, s = 1, 2\}$. Let $\hat{\epsilon}_{1i} = \frac{Y_{1i} - \hat{g}_1(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}_{1i}, \hat{\boldsymbol{\beta}}_0)}{\hat{\sigma}_1(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}_{1i}, \hat{\boldsymbol{\beta}}_0)}$, $\hat{\epsilon}_{2i} = \frac{Y_{2i} - \hat{g}_2(\hat{\boldsymbol{\gamma}}_0^\tau \boldsymbol{X}_{2i}, \hat{\boldsymbol{\gamma}}_0)}{\hat{\sigma}_2(\hat{\boldsymbol{\gamma}}_0^\tau \boldsymbol{X}_{2i}, \hat{\boldsymbol{\gamma}}_0)}$ be the estimators of ϵ_{1i} and ϵ_{2i}, respectively, and standard them by

$$\tilde{\epsilon}_{1i} = \frac{\hat{\epsilon}_{1i} - \frac{1}{n_1}\sum_{i=1}^{n_1}\hat{\epsilon}_{1i}}{\left(\frac{1}{n_1}\sum_{i=1}^{n_1}\left[\hat{\epsilon}_{1i} - \frac{1}{n_1}\sum_{i=1}^{n_1}\hat{\epsilon}_{1i}\right]^2\right)^{1/2}}, \quad i = 1, \ldots, n_1,$$

$$\tilde{\epsilon}_{2i} = \frac{\hat{\epsilon}_{2i} - \frac{1}{n_2}\sum_{i=1}^{n_2}\hat{\epsilon}_{2i}}{\left(\frac{1}{n_2}\sum_{i=1}^{n_2}\left[\hat{\epsilon}_{2i} - \frac{1}{n_2}\sum_{i=1}^{n_2}\hat{\epsilon}_{2i}\right]^2\right)^{1/2}}, \quad i = 1, \ldots, n_2.$$

Let

$$\hat{\epsilon}^*_{sib} = \tilde{\epsilon}_{si} + a_{n_s}\varsigma_{ib} \quad i = 1, \ldots, n_s, \quad b = 1, \ldots, B, \tag{14}$$

where $a_{n_s} = c_{s,1}n_s^{-1/4}$ for some positive constants $c_{s,1}$, $s = 1, 2$ (Neumeyer 2009; Neumeyer and Van Keilegom 2010). Then, define the "bootstrap"-response Y^*_{1ib} and Y^*_{2ib} as

$$Y^*_{1ib} = \hat{g}_1(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}_{1i}, \hat{\boldsymbol{\beta}}_0) + \hat{\sigma}_1(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}_{1i}, \hat{\boldsymbol{\beta}}_0)\hat{\epsilon}^*_{1ib},\ i = 1, \ldots, n_1,$$
$$Y^*_{2ib} = \hat{g}_2(\hat{\boldsymbol{\gamma}}_0^\tau \boldsymbol{X}_{2i}, \hat{\boldsymbol{\gamma}}_0) + \hat{\sigma}_2(\hat{\boldsymbol{\gamma}}_0^\tau \boldsymbol{X}_{2i}, \hat{\boldsymbol{\gamma}}_0)\hat{\epsilon}^*_{2ib},\ i = 1, \ldots, n_2.$$

Step 3 For each b, we use bootstraps $\left\{Y^*_{sib}, \boldsymbol{X}_{si}, i = 1, \ldots, n_s, s = 1, 2\right\}$ and recalculate the bootstrap estimators $\hat{\boldsymbol{\beta}}_0^{(b)}$, $\hat{\boldsymbol{\gamma}}_0^{(b)}$ $\hat{g}_1^{(b)}\left(u, \hat{\boldsymbol{\beta}}_0^{(b)}\right)$, $\hat{g}_2^{(b)}\left(u, \hat{\boldsymbol{\gamma}}_0^{(b)}\right)$, $\hat{\sigma}_1^{(b)}\left(u, \hat{\boldsymbol{\beta}}_0^{(b)}\right)$, and $\hat{\sigma}_2^{(b)}\left(u, \hat{\boldsymbol{\gamma}}_0^{(b)}\right)$. We then obtain the bootstrap test statistics $\mathfrak{B}^{\mathrm{KS}(b)}_{n_1 n_2}$ and $\mathfrak{B}^{\mathrm{CM}(b)}_{n_1 n_2}$.

Step 4 We calculate the $1-\kappa$ quantile of the bootstrap test statistics $\mathfrak{B}^{\mathrm{KS}(b)}_{n_1 n_2}$, $\mathfrak{B}^{\mathrm{CM}(b)}_{n_1 n_2}$, $b = 1, \ldots B$ as the κ-level critical value.

3.4 Extension to *k* groups

Suppose we have k ($k \geq 3$) independent samples following the single-index heteroscedastic regression model:

$$Y_s = g_s\left(\boldsymbol{\omega}^\tau_{[s],0}\boldsymbol{X}_s\right) + \sigma_s\left(\boldsymbol{\omega}^\tau_{[s],0}\boldsymbol{X}_s\right)\epsilon_s, \quad s = 1, \ldots, k. \tag{15}$$

For $s = 1, \ldots, k$, Y_s is the response variable, $g_s(u)$ and $\sigma_s(u)$, are unknown univariate smooth functions, where function $\sigma_s(u)$ is assumed to be positive. Further, $\boldsymbol{X}_s$ is the q_s-dimensional covariate vector. Error term ϵ_s satisfies $E(\epsilon_s) = 0$ and $E(\epsilon_s^2) = 1$. Parameter $\boldsymbol{\omega}_{[s],0}$ is an unknown index vector that belongs to the parameter space $\mathcal{B}_s = \left\{\boldsymbol{\omega}_{[s]} = (\omega_{[s],1}, \ldots, \omega_{[s],q_s})^\tau \in \mathbb{R}^{q_s}, \|\boldsymbol{\omega}_{[s]}\| = 1, \omega_{[s],1} > 0\right\}$.

We consider testing the equality of the mean functions, i.e.,

$$\widetilde{\mathcal{H}}_{k,0} : g_1(u) = g_2(u) = \cdots = g_k(u) \ \text{ for each } \ u. \tag{16}$$

Under null hypothesis $\widetilde{\mathcal{H}}_{k,0}$, the estimators of $F_{\epsilon_s}(t)$, $s = 1, \ldots, k$ are obtained as

$$\hat{F}_{\widetilde{\mathcal{H}}_{k,0},\epsilon_s}(t) = \frac{1}{n_s}\sum_{i=1}^{n_s} I\left\{\hat{\epsilon}_{\widetilde{\mathcal{H}}_{k,0},si} \leq t\right\},$$
$$\hat{\epsilon}_{\widetilde{\mathcal{H}}_{k,0},si} = \frac{Y_{si} - \hat{g}_s(\hat{\boldsymbol{\omega}}^\tau_{[d_s],0}\boldsymbol{X}_{d_s i}, \hat{\boldsymbol{\omega}}_{[s],0})}{\hat{\sigma}_s(\hat{\boldsymbol{\omega}}^\tau_{[s],0}\boldsymbol{X}_{si}, \hat{\boldsymbol{\omega}}_{[s],0})}, \ \text{for } d_s \neq s,$$

and set $\{1, 2, \ldots, k\}$ is equal to $\{d_1, d_2, \ldots, d_k\}$. If under the null hypothesis $\widetilde{\mathcal{H}}_{k,0}$ is not true, we define

$$\hat{F}_{\epsilon_s}(t) = \frac{1}{n_s}\sum_{i=1}^{n_s} I\left\{\hat{\epsilon}_{si} \leq t\right\}, \text{where } \hat{\epsilon}_{si} = \frac{Y_{si} - \hat{g}_s(\hat{\boldsymbol{\omega}}^\tau_{[s],0}\boldsymbol{X}_{si}, \hat{\boldsymbol{\omega}}_{[s],0})}{\hat{\sigma}_s(\hat{\boldsymbol{\omega}}^\tau_{[s],0}\boldsymbol{X}_{si}, \hat{\boldsymbol{\omega}}_{[s],0})}.$$

To test $\widetilde{\mathcal{H}}_{k,0}$, we propose using the Kolmogorov–Smirnov and Cramér–von Mises type functional based test statistics:

$$\mathfrak{T}^{\mathrm{KS}}_{k,n_1n_2} = \sum_{s=1}^{k} \sup_{t\in\mathbb{R}} n_s^{1/2} \left| \hat{F}_{\widetilde{\mathcal{H}}_{k,0,\epsilon_s}}(t) - \hat{F}_{\epsilon_s}(t) \right|,$$

and

$$\mathfrak{T}^{\mathrm{CM}}_{k,n_1n_2} = \sum_{s=1}^{k} n_s \int \left| \hat{F}_{\widetilde{\mathcal{H}}_{k,0,\epsilon_s}}(t) - \hat{F}_{\epsilon_s}(t) \right|^2 d\hat{F}_{\epsilon_s}(t).$$

Similarly, we can also use the smooth residual bootstrap method introduced in Sect. 3.3 to mimic the distributions of test statistics $\mathfrak{T}^{\mathrm{KS}}_{k,n_1n_2}$ and $\mathfrak{T}^{\mathrm{CM}}_{k,n_1n_2}$.

4 Implementation

In this section, we report simulation results to evaluate the performance of the proposed estimators and test statistics. Here, the Epanechnikov kernel $K(t) = 0.75(1-t^2)^+$ is used. Note that under-smoothing is necessary as Condition (C4) requires that $n_s h_s^4 \to 0$. To meet this requirement, we follow the suggestion in Carroll et al. (1997) by choosing the order of $O(n_s^{-1/5}) \times n_s^{-2/15} = O(n_s^{-1/3})$ for bandwidth h_s.

The selection procedure for h_s is implemented as follows. First, we minimize cross-validation score $\mathrm{CV}_s(h)$ to obtain bandwidth $h_{s,1}$; then, we use bandwidth h_s as $h_s = n_s^{-2/15} * h_{s,1}$. The cross-validation score is defined as

$$\mathrm{CV}_s(h) = n_s^{-1} \sum_{i=1}^{n_s} \left\{ Y_{si} - \hat{g}_{[s,-i]}\left(\hat{V}_{[s,-i]}, \hat{\boldsymbol{\omega}}_{[s,-i],0}\right)\right\}^2,$$

where $\hat{V}_{s,-i} = \hat{\boldsymbol{\omega}}^{\tau}_{[s,-i],0}\boldsymbol{X}_{si}$, and $\hat{\boldsymbol{\omega}}_{[s,-i],0}$ and $\hat{g}_{[s,-i]}\left(\hat{V}_{[-i]}, \hat{\boldsymbol{\omega}}_{[s,-i],0}\right)$ are computed similarly to (3) with the ith observation deleted. For the choices of a_{n_s} in (14), Neumeyer (2009); Neumeyer and Van Keilegom (2010) suggested using $a_{n_s} = c_{s,1} n_s^{-1/4}$ for some positive constants $c_{s,1}$, $s = 1, 2$. In this section, we use $a_{n_s} = n_s^{-1/4}$, and the numerical results are stable when we use a shift around a_{n_s}.

Example 1 In this example, we generate 500 realizations from models (1) and choose sample sizes $n_1 = n_2 = 50, 100, 300$ and 500:

$$g_1(u) = g_2(u) = 2\exp(u), \sigma_1(u) = \exp(u), \sigma_2(u) = (1+u)^2.$$

1. Parameter estimation. We consider $\boldsymbol{\beta}_0 = (2, 1, 0, -2, 1)^{\tau}/\sqrt{10}$, $\boldsymbol{X}_1 \sim N_5(\mathbf{0}, \boldsymbol{\Sigma}_1)$ with $\boldsymbol{\Sigma}_1 = (\sigma_{1,ij})$, $\sigma_{1,ij} = 0.5^{|i-j|}$, and $\boldsymbol{\gamma}_0 = (1, 2, 3, 1)^{\tau}/\sqrt{15}$, $\boldsymbol{X}_2 \sim N_4(\mathbf{0}, \boldsymbol{\Sigma}_2)$ with $\boldsymbol{\Sigma}_2 = (\sigma_{2,ij})$, $\sigma_{2,ij} = (-0.5)^{|i-j|}$. The model errors ϵ_1 and ϵ_2 independently follow a standard normal distribution $N(0, 1)$.
   The simulation results for $\hat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\gamma}}_0$ are reported in Tables 1 and 2, respectively.

Table 1 The mean (M), standard error (SD) and means squared error (MSE) of $\hat{\boldsymbol{\beta}}_0$ and $\arccos(\hat{\boldsymbol{\beta}}_0, \boldsymbol{\beta}_0)$

	$\hat{\boldsymbol{\beta}}_0$					$\arccos(\hat{\boldsymbol{\beta}}_0, \boldsymbol{\beta}_0)$
$n = 50$						
M	0.6012	0.3287	0.0028	−0.6083	0.3229	0.2060
SD	0.1238	0.1133	0.1201	0.0715	0.0956	0.1296
MSE	50.5148	43.6036	41.8572	19.4543	30.8682	170.6115
$n = 100$						
M	0.6298	0.3292	0.0127	−0.6134	0.3234	0.1113
SD	0.0533	0.0535	0.0550	0.0448	0.0590	0.0501
MSE	9.1718	10.9649	11.5210	8.5556	12.0104	44.7091
$n = 300$						
M	0.6323	0.3229	0.0066	−0.6249	0.3182	0.0603
SD	0.0264	0.0320	0.0311	0.0253	0.0277	0.0239
MSE	6.9070	10.5784	10.0215	6.9104	7.6257	14.2168
$n = 500$						
M	0.6349	0.3180	0.0064	−0.6267	0.3171	0.0456
SD	0.0189	0.0255	0.0248	0.0178	0.0208	0.0188
MSE	3.5930	6.4643	6.4848	3.4546	4.3016	8.3059

MSE is in the scale of $\times 10^{-4}$

The values of $\hat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\gamma}}_0$ are close to the true values of $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}_0$, respectively, and the values of $\text{MSE}(\hat{\boldsymbol{\beta}}_0, \boldsymbol{\beta}_0)$ and $\text{MSE}(\hat{\boldsymbol{\gamma}}_0, \boldsymbol{\gamma}_0)$ decrease. Moreover, the angles (in radians) of $\arccos(\hat{\boldsymbol{\beta}}_0, \boldsymbol{\beta}_0)$ and $\arccos(\hat{\boldsymbol{\gamma}}_0, \boldsymbol{\gamma}_0)$ become closer to zero when sample size n increases to 500. These simulation results show that the estimation procedure proposed in Sect. 2 works well.

2. Test procedure for $\boldsymbol{\beta}_0 = \boldsymbol{\gamma}_0$. We investigate the performance of test statistic $\mathcal{T}_{n_1 n_2}$. Let $\boldsymbol{\beta}_0 = (1, 2, 3, 4)^\tau/\sqrt{30}$, $\boldsymbol{\gamma}^{(1)}_{0,[C_o]} = (2, 3, 4)^\tau/\sqrt{30} + C_o$ and the first element of $\boldsymbol{\gamma}_{0,[C_o]}$ be $\boldsymbol{\gamma}_{0,1,[C_o]} = \sqrt{1 - \|\boldsymbol{\gamma}^{(1)}_{0,[C_o]}\|^2}$. Both $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ follow normal distribution $\sim N_4(\mathbf{0}, \boldsymbol{\Sigma}_3)$ with $\boldsymbol{\Sigma}_3 = (\sigma_{3,ij})$ and $\sigma_{3,ij} = 0.5^{|i-j|}$. The model errors ϵ_1 and ϵ_2 independently follow from a standard normal distribution $N(0, 1)$. Note that the null hypothesis $\mathcal{H}_0^*$ considered in (8) is true if and only if $C_o = 0$. The simulation results are reported in Fig. 1. Typically, when $C_0 = 0$, then $\boldsymbol{\beta}_0 = \boldsymbol{\gamma}_0$ holds. The rejection probabilities are 0.0091, 0.0277, 0.0548, 0.1123 for $n = 300$ and 0.0095, 0.0261, 0.0524, 0.1077 for $n = 500$. These simulation results of $\mathcal{T}_{n_1 n_2}$ for $C_0 = 0$ are close to 0.01, 0.025, 0.05, 0.10 when the null hypothesis $\mathcal{H}_0$ (or $\mathcal{H}_0^*$) is true. This indicates that $\mathcal{T}_{n_1 n_2}$ can provide proper rejection probabilities under the null hypothesis $\mathcal{H}_0$ (or $\mathcal{H}_0^*$). We compare these results with those of the Kullback–Leibler (KL) statistic proposed by Lin and Kulasekera (2010). In Figs. 1 and 2, we plot the empirical powers obtained by test statistic $\mathcal{T}_{n_1 n_2}$ and the KL statistic. The performances of test statistic $\mathcal{T}_{n_1 n_2}$ are more powerful than the KL statistic in both figures. This is not surprising as we indicated in Sect. 1. The

Table 2 The mean (M), standard error (SD) and means squared error (MSE) of $\hat{\boldsymbol{\gamma}}_0$ and $\arccos(\hat{\boldsymbol{\gamma}}_0, \boldsymbol{\gamma}_0)$

	$\hat{\boldsymbol{\gamma}}_0$				$\arccos(\hat{\boldsymbol{\gamma}}_0, \boldsymbol{\gamma}_0)$
$n = 50$					
M	0.2036	0.5125	0.7679	0.2581	0.1676
SD	0.1324	0.1039	0.0632	0.0878	0.1236
MSE	63.3573	34.0945	15.0437	28.3204	142.2371
$n = 100$					
M	0.2584	0.5169	0.7699	0.2660	0.0422
SD	0.0285	0.0223	0.0198	0.0288	0.0286
MSE	2.0215	1.9285	1.1102	2.8322	8.9105
$n = 300$					
M	0.2585	0.5164	0.7743	0.2581	0.0142
SD	0.0109	0.0080	0.0058	0.0083	0.0089
MSE	1.1678	0.6292	0.3373	0.6724	2.8069
$n = 500$					
M	0.2578	0.5157	0.7751	0.2582	0.0122
SD	0.0079	0.0075	0.0052	0.0078	0.0072
MSE	0.7214	0.5652	0.2709	0.4691	2.2267

MSE is in the scale of $\times 10^{-4}$

convergence rate of the KL statistic for testing the null hypothesis $\mathcal{H}_0$ (or $\mathcal{H}_0^*$) is slower than $\mathcal{T}_{n_1 n_2}$, which has a parametric convergence rate of $O\left(\frac{1}{n_1+n_2}\right)$. In Fig. 2, the power functions of both statistics increase rapidly and approach one when the value of C_o increases and sample size n increases to 500.

Example 2 In this example, we investigate the performances of the estimators $\hat{F}_{\epsilon_s}(t)$ and $\hat{F}_{\tilde{\mathcal{H}}_{0,\epsilon_s}}(t)$ for $s = 1, 2$, and test statistics $\mathfrak{T}^{\mathrm{KS}}_{n_1 n_2}$ and $\mathfrak{T}^{\mathrm{CM}}_{n_1 n_2}$. We generate 1000 realizations and choose sample sizes of $n_1 = n_2 = 50, 100, 300, 500$. The data generating process is considered as follows:

$$g_1(u) = 2\exp(0.5u), \quad g_2(u) = g_1(u) + D_o u^2, \sigma_1(u) = \sigma_1(u) = \exp(0.5u). \tag{17}$$

In this example, we set $\boldsymbol{\beta}_0 = \boldsymbol{\gamma}_0 = \frac{1}{3}(2, 1, 0, -2)^\tau$. The covariates $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independently generated from normal distribution $N_4(\mathbf{0}, \boldsymbol{\Sigma}_4)$ with $\boldsymbol{\Sigma}_4 = (\sigma_{4,ij})_{1\le i,j\le 4}$, $\sigma_{4,ij} = 0.5^{|i-j|}$. The model errors ϵ_1 and ϵ_2 independently follow a standard normal distribution $N(0, 1)$.

1. Estimation for $\hat{F}_{\epsilon_s}(t)$ and $\hat{F}_{\tilde{\mathcal{H}}_{0,\epsilon_s}}(t)$ under null hypothesis $\mathcal{H}_0$. The performance of estimator $\hat{F}_{\epsilon_s}(t)$ and its true distribution $F_{\epsilon_s}(t)$ is evaluated using the average squared error (ASE) and the average absolute error (AAE)

$$\mathrm{ASE} = n_0^{-1}\sum_{v=1}^{n_0}\left[\hat{F}_{\epsilon_s}(l_v) - F_{\epsilon_s}(l_v)\right]^2, \quad \mathrm{AAE} = n_0^{-1}\sum_{s=1}^{n_0}\left|\hat{F}_{\epsilon_s}(l_v) - F_{\epsilon_s}(l_v)\right|,$$

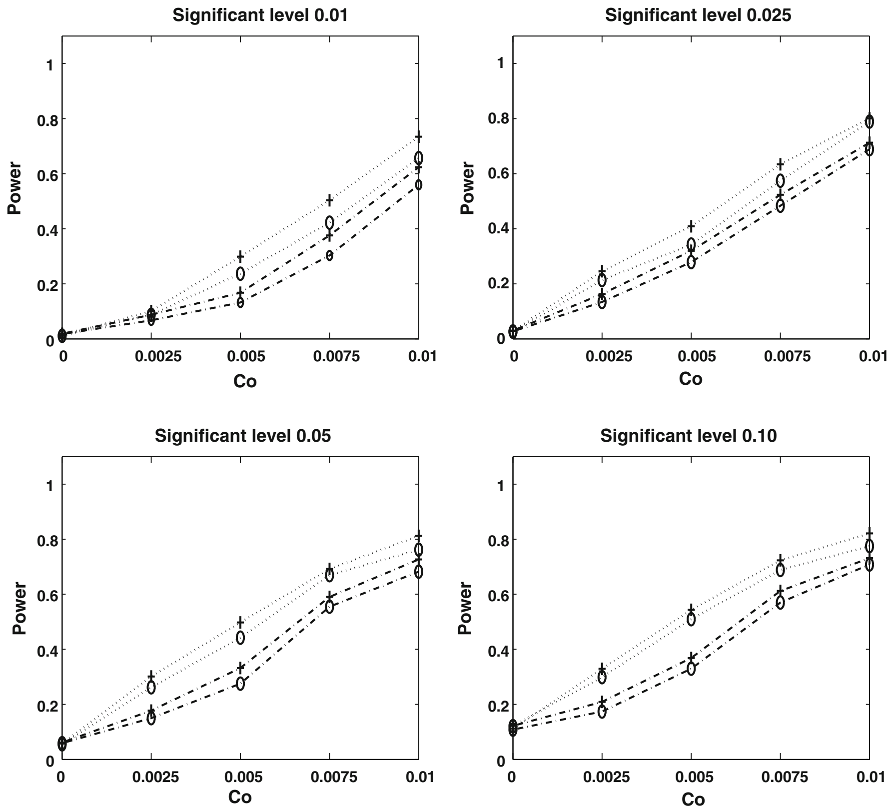

Fig. 1 Power calculations of hypothesis test $\boldsymbol{\beta}_0 = \boldsymbol{\gamma}_0$, $n = 50$ (*dashed lines*) and $n = 100$ (*dotted line*). "+" for $\mathcal{T}_{n_1 n_2}$ and "∘" for KL statistic

where $\{l_1, \ldots, l_{n_0}\}$ are the given grid points in the interval $[-2, 2]$, and $n_0 = 400$ is the number of grid points.

In Table 3, we report the numerical results of ASE and AAE for the estimators $\hat{F}_{\epsilon_s}(t)$ and $\hat{F}_{\tilde{\mathcal{H}}_{0,\epsilon_s}}(t)$ when $D_o = 0$ and $s = 1, 2$. The four estimators perform better as sample size n increases. The performance of $\hat{F}_{\epsilon_s}(t)$ is better than $\hat{F}_{\tilde{\mathcal{H}}_{0,\epsilon_s}}(t)$ in this simulation study. Figure 3a, b show that there is little difference between estimators $\hat{F}_{\epsilon_s}(t)$ and $\hat{F}_{\tilde{\mathcal{H}}_{0,\epsilon_s}}(t)$ for sample size $n = 500$ when $D_o = 0$, which indicates small values for Kolmogorov–Smirnov test statistic $\mathfrak{T}^{\mathrm{KS}}_{n_1 n_2}$ and Cramér–von Mises test statistic $\mathfrak{T}^{\mathrm{CM}}_{n_1 n_2}$. In Fig. 3, we also present the plots of $\hat{F}_{\epsilon_s}(t)$ and $\hat{F}_{\tilde{\mathcal{H}}_{0,\epsilon_s}}(t)$ when $D_o = 0.5$ and $D_o = 1$. It is easily seen that larger values of D_o lead to a larger deviation of $\hat{F}_{\tilde{\mathcal{H}}_{0,\epsilon_s}}(t)$ from $F_{\epsilon_s}(t)$, which indicates large values of Kolmogorov–Smirnov test statistic $\mathfrak{T}^{\mathrm{KS}}_{n_1 n_2}$ and Cramér–von Mises test statistic $\mathfrak{T}^{\mathrm{CM}}_{n_1 n_2}$.

2. Test statistics for $g_1(u) = g_2(u)$. In each simulation for the power calculation, 200 and 1000 bootstrap samples were generated. We also compared our results with those of the KL statistic proposed by Lin and Kulasekera (2010). The simulation

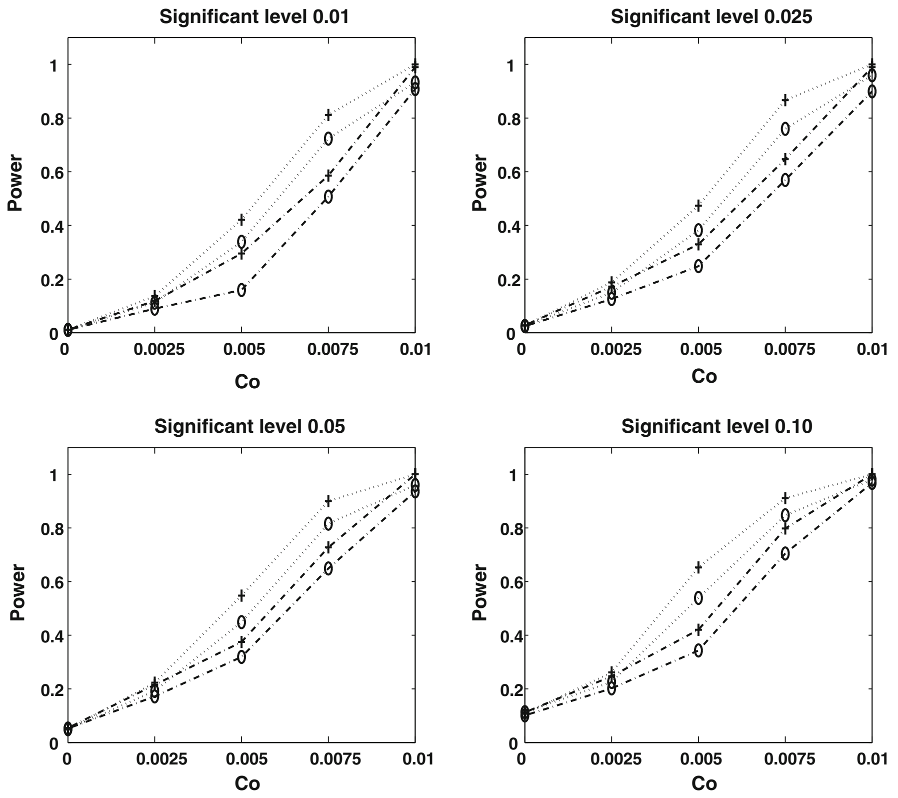

Fig. 2 Power calculations of hypothesis test $\boldsymbol{\beta}_0 = \boldsymbol{\gamma}_0$, $n = 300$ (*dashed lines*) and $n = 500$ (*dotted line*). "+" for $\mathfrak{T}_{n_1n_2}$ and "$\circ$" for KL statistic

results for the test statistics KL, $\mathfrak{T}^{\mathrm{KS}}_{n_1n_2}$ and $\mathfrak{T}^{\mathrm{CM}}_{n_1n_2}$ are reported in Tables 4 and 5. It is clear that all empirical levels obtained by the bootstrap test statistics proposed in Sect. 3.3 for $\mathfrak{T}^{\mathrm{KS}}_{n_1n_2}$ and $\mathfrak{T}^{\mathrm{CM}}_{n_1n_2}$ are close to 0.01, 0.025, 0.05, 0.10 when $D_o = 0$ and the sample size $n \geq 300$, which indicates that the bootstrap method can provide proper rejection probabilities. When the value of D_o increases, the power functions increase rapidly and approach one as sample size n increases. Moreover, in Table 4, the KL statistics are more powerful than $\mathfrak{T}^{\mathrm{KS}}_{n_1n_2}$ and $\mathfrak{T}^{\mathrm{CM}}_{n_1n_2}$ when the bootstrap sample is 200. If the bootstrap sample increases, for instance, to 1000 as shown in Table 5, the performance of statistic $\mathfrak{T}^{\mathrm{CM}}_{n_1n_2}$ becomes better than the KL statistics when the sample size $n \geq 300$, and statistic $\mathfrak{T}^{\mathrm{KS}}_{n_1n_2}$ and the KL statistic perform similarly.

5 Real data analysis

In this example, we analyze the Boston housing price dataset (available from the Machine Learning Repository at the University of California-Irvine) to illustrate our proposed method. In the Boston Housing Dataset, there are 506 instances and vari-

Table 3 The mean (M) and standard error (SD) for ASE and AAE

	ASE				AAE			
	$\hat{F}_{\epsilon_1}(t)$	$\hat{F}_{\mathcal{H}_0,\epsilon_1}(t)$	$\hat{F}_{\epsilon_2}(t)$	$\hat{F}_{\mathcal{H}_0,\epsilon_2}(t)$	$\hat{F}_{\epsilon_1}(t)$	$\hat{F}_{\mathcal{H}_0,\epsilon_1}(t)$	$\hat{F}_{\epsilon_2}(t)$	$\hat{F}_{\mathcal{H}_0,\epsilon_2}(t)$
$n = 50$								
M	14.2245	19.6098	13.7245	22.1191	0.2025	0.2198	0.1987	0.1898
SD	14.0023	18.9933	14.1012	21.9934	0.1384	0.1180	0.0997	0.1109
$n = 100$								
M	6.7123	11.8856	6.2897	14.7765	0.0991	0.1011	0.0982	0.0998
SD	6.3329	10.9987	6.3098	15.2109	0.0595	0.0559	0.0497	0.0264
$n = 300$								
M	2.5913	6.7190	2.3426	5.8062	0.0391	0.0569	0.0378	0.0543
SD	2.5775	5.8823	1.2911	5.2894	0.0125	0.0216	0.0092	0.0253
$n = 500$								
M	1.5453	3.1023	1.5139	3.2483	0.0295	0.0487	0.0294	0.0387
SD	1.3138	3.6113	1.2007	3.3413	0.0100	0.0196	0.0081	0.0228

ASE is in the scale of $\times 10^{-4}$

ables indicating the size, location and environment of the property as well as its selling price and other relevant variables that measure the socioeconomic status of neighborhood. We focus our analysis on eight attributes: MEDV (Y): the median value of owner-occupied homes in USD 1000's, RM (X_1): the average number of rooms per dwelling, AGE (X_2): the proportion of owner-occupied units built prior to 1940, DIS (X_3): the weighted distances to five Boston employment centers, RAD (X_4): an index of accessibility to radial highways, TAX (X_5): the full-value property-tax rate per USD 10,000, PTRATIO (X_6): the pupil-teacher ratio by town, BLACKS (X_7): the transformed proportion of Blacks which is calculated by $1000(\mathrm{Bk} - 0.63)^2$, where Bk is the proportion of blacks by town. For this dataset, we use the covariate NOX (the nitric oxide concentration per 10 million) to split the dataset into two groups. Group 1 is defined by the values of NOX that are less than its median, and Group 2 was defined by the values of NOX that are greater or equal to its median. We use single-index model (1) to analyze the eight attributes.

Corresponding to covariates $(X_1, X_2, \ldots, X_7)^\tau$, parameters $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}_0$ and the associated p values ($p_{\hat{\boldsymbol{\beta}}_0}$ and $p_{\hat{\boldsymbol{\gamma}}_0}$) were obtained as follows:

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_0 \\ p_{\hat{\boldsymbol{\beta}}_0} \end{pmatrix} = \begin{pmatrix} 0.2099, & 0.0996, & -0.9637, & 0.0274, & -0.0006, & -0.1029, & -0.0771 \\ 0.2708, & 0.0000, & 0.0000, & 0.5446, & 0.8177, & 0.0989, & 0.0000 \end{pmatrix},$$

and

$$\begin{pmatrix} \hat{\boldsymbol{\gamma}}_0 \\ p_{\hat{\boldsymbol{\gamma}}_0} \end{pmatrix} = \begin{pmatrix} 0.8517, & -0.5103, & -0.0852, & 0.0662, & -0.0032, & -0.0155, & 0.0486 \\ 0.0000, & 0.0000, & 0.2464, & 0.2047, & 0.1217, & 0.8205, & 0.0000 \end{pmatrix}.$$

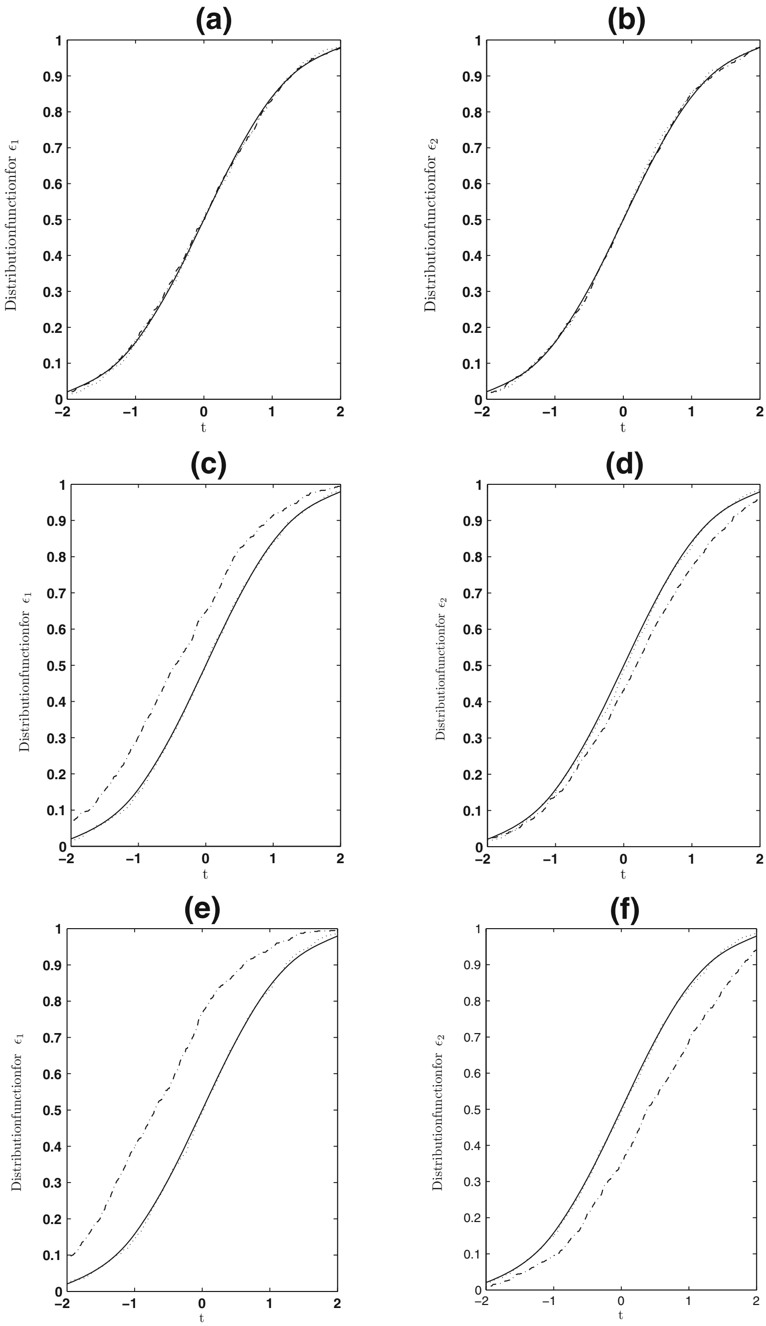

Fig. 3 The plots of true values of $F_{\epsilon_s}(t)$ (*solid line*), the plots of estimator $\hat{F}_{\epsilon_s}(t)$ (*dotted line*) and the plots of estimator $\hat{F}_{\widetilde{\mathcal{H}}_0,\epsilon_s}(t)$ (*dashed line*). **a**, **b** is the case of $D_o = 0$ (under the null hypothesis $\widetilde{\mathcal{H}}_0$), **c**, **d** is the case of $D_o = 0.5$ (under the alternative hypothesis $\widetilde{\mathcal{H}}_1$), and **e**, **f** is the case of $D_o = 1.0$ (under the alternative hypothesis $\widetilde{\mathcal{H}}_1$)

Table 4 The simulation results for power calculations in Example 2 when the bootstrap sample is 200

Significant level	KL				$\mathfrak{T}_{n_1 n_2}^{KS}$				$\mathfrak{T}_{n_1 n_2}^{CM}$			
	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10
$n = 50$												
$D_0 =$ 0.00	0.014	0.027	0.043	0.083	0.004	0.016	0.027	0.068	0.004	0.017	0.028	0.073
$D_0 =$ 0.25	0.094	0.127	0.132	0.169	0.082	0.094	0.109	0.140	0.084	0.099	0.111	0.149
$D_0 =$ 0.50	0.162	0.169	0.185	0.198	0.146	0.151	0.162	0.175	0.147	0.152	0.165	0.179
$D_0 =$ 0.75	0.245	0.262	0.295	0.332	0.212	0.222	0.247	0.283	0.220	0.229	0.250	0.289
$D_0 =$ 1.00	0.279	0.298	0.331	0.449	0.252	0.269	0.317	0.401	0.260	0.274	0.329	0.427
$n = 100$												
$D_0 =$ 0.00	0.006	0.022	0.039	0.091	0.005	0.018	0.032	0.082	0.006	0.019	0.033	0.090
$D_0 =$ 0.25	0.129	0.148	0.181	0.254	0.109	0.119	0.152	0.189	0.112	0.120	0.159	0.204
$D_0 =$ 0.50	0.215	0.226	0.238	0.258	0.176	0.182	0.207	0.212	0.180	0.188	0.211	0.222
$D_0 =$ 0.75	0.323	0.361	0.391	0.448	0.257	0.261	0.282	0.319	0.260	0.270	0.310	0.330
$D_0 =$ 1.00	0.609	0.677	0.698	0.740	0.497	0.513	0.545	0.623	0.511	0.521	0.588	0.641
$n = 300$												
$D_0 =$ 0.00	0.008	0.024	0.045	0.095	0.008	0.020	0.041	0.088	0.010	0.022	0.043	0.092
$D_0 =$ 0.25	0.315	0.357	0.442	0.519	0.256	0.298	0.367	0.448	0.260	0.310	0.378	0.452
$D_0 =$ 0.50	0.667	0.745	0.818	0.844	0.512	0.549	0.623	0.717	0.520	0.551	0.634	0.729
$D_0 =$ 0.75	0.872	0.909	0.960	0.980	0.653	0.707	0.723	0.757	0.666	0.711	0.734	0.761
$D_0 =$ 1.00	0.949	0.966	0.993	1.000	0.773	0.781	0.801	0.811	0.781	0.792	0.811	0.822

Table 4 continued

Significant level	KL				$\Upsilon^{\mathrm{KS}}_{n_1 n_2}$				$\Upsilon^{\mathrm{CM}}_{n_1 n_2}$			
	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10
$n = 500$												
$D_0 =$ 0.00	0.009	0.023	0.050	0.098	0.008	0.021	0.042	0.089	0.009	0.022	0.046	0.094
$D_0 =$ 0.25	0.392	0.491	0.537	0.628	0.313	0.369	0.427	0.509	0.321	0.371	0.435	0.510
$D_0 =$ 0.50	0.737	0.840	0.920	0.930	0.539	0.607	0.678	0.732	0.541	0.611	0.681	0.741
$D_0 =$ 0.75	0.901	0.950	0.982	0.999	0.587	0.719	0.744	0.766	0.599	0.725	0.798	0.814
$D_0 =$ 1.00	0.966	0.997	1.000	1.000	0.812	0.823	0.845	0.856	0.823	0.831	0.856	0.889

Table 5 The simulation results for power calculations in Example 2 when the bootstrap sample is 1000

Significant level	KL				$\mathfrak{T}_{n_1n_2}^{KS}$				$\mathfrak{T}_{n_1n_2}^{CM}$			
	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10
$n = 50$												
$D_0 =$ 0.00	0.014	0.029	0.043	0.085	0.005	0.017	0.029	0.074	0.006	0.018	0.031	0.081
$D_0 =$ 0.25	0.101	0.128	0.136	0.174	0.091	0.101	0.124	0.167	0.101	0.109	0.131	0.172
$D_0 =$ 0.50	0.163	0.175	0.184	0.199	0.156	0.161	0.172	0.181	0.157	0.170	0.179	0.190
$D_0 =$ 0.75	0.245	0.262	0.295	0.332	0.232	0.249	0.286	0.309	0.242	0.256	0.297	0.313
$D_0 =$ 1.00	0.282	0.301	0.368	0.454	0.272	0.294	0.353	0.429	0.280	0.308	0.369	0.439
$n = 100$												
$D_0 =$ 0.00	0.007	0.021	0.041	0.091	0.006	0.019	0.037	0.087	0.007	0.019	0.039	0.091
$D_0 =$ 0.25	0.132	0.151	0.182	0.255	0.129	0.141	0.171	0.235	0.131	0.147	0.181	0.250
$D_0 =$ 0.50	0.219	0.223	0.242	0.260	0.208	0.219	0.226	0.243	0.210	0.223	0.235	0.259
$D_0 =$ 0.75	0.322	0.368	0.389	0.451	0.313	0.353	0.379	0.429	0.320	0.369	0.387	0.441
$D_0 =$ 1.00	0.616	0.688	0.700	0.739	0.605	0.643	0.687	0.723	0.613	0.678	0.695	0.740
$n = 300$												
$D_0 =$ 0.00	0.008	0.024	0.045	0.098	0.009	0.023	0.049	0.097	0.011	0.024	0.049	0.102
$D_0 =$ 0.25	0.315	0.358	0.443	0.521	0.315	0.358	0.442	0.521	0.412	0.482	0.557	0.611
$D_0 =$ 0.50	0.668	0.744	0.820	0.843	0.667	0.747	0.819	0.843	0.684	0.787	0.865	0.917
$D_0 =$ 0.75	0.873	0.910	0.970	0.982	0.872	0.925	0.973	0.983	0.928	0.944	0.989	0.995
$D_0 =$ 1.00	0.950	0.971	0.991	1.000	0.958	0.971	0.991	1.000	0.975	0.991	1.000	1.000

Table 5 continued

Significant level	KL				$\Upsilon_{n_1n_2}^{\mathrm{KS}}$				$\Upsilon_{n_1n_2}^{\mathrm{CM}}$			
	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10
$n = 500$												
$D_0 =$ 0.00	0.009	0.024	0.051	0.099	0.009	0.024	0.051	0.099	0.010	0.025	0.049	0.101
$D_0 =$ 0.25	0.394	0.490	0.535	0.630	0.395	0.485	0.549	0.629	0.513	0.576	0.642	0.711
$D_0 =$ 0.50	0.738	0.842	0.919	0.939	0.741	0.840	0.920	0.943	0.787	0.862	0.943	0.972
$D_0 =$ 0.75	0.902	0.952	0.981	1.000	0.903	0.949	0.980	1.000	0.958	0.986	1.000	1.000
$D_0 =$ 1.00	0.972	0.996	1.000	1.000	0.972	0.998	1.000	1.000	0.995	1.000	1.000	1.000

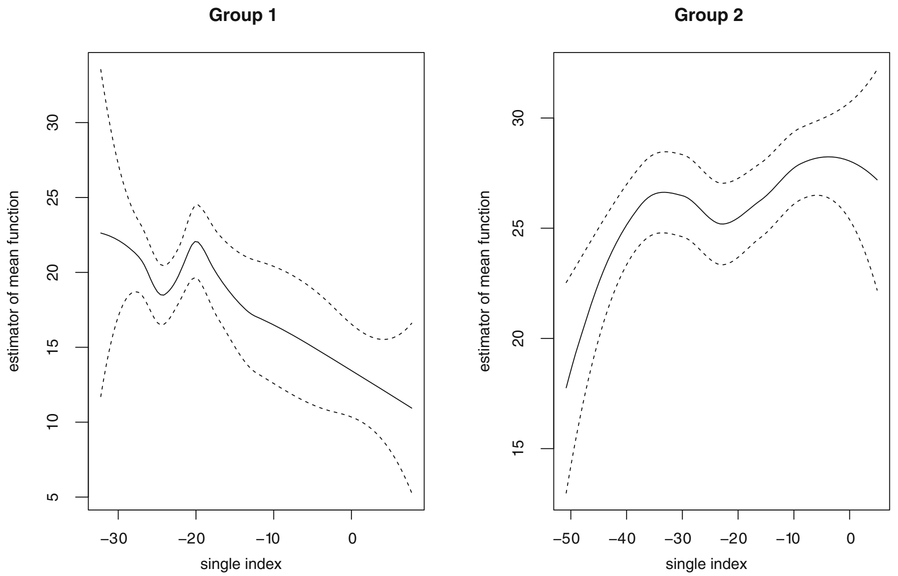

Fig. 4 Group 1: the plot for the estimator $\hat{g}_1(u)$ (*solid line*) against estimated single-index $\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}_1$ in the *left panel*, along with the associated 95% pointwise confidence intervals (*dotted lines*); Group 2: the plot for estimator $\hat{g}_2(u)$ (*solid line*) against estimated single-index $\hat{\boldsymbol{\gamma}}_0^\tau \boldsymbol{X}_2$ in the *right panel*, along with the associated 95% pointwise confidence intervals (*dotted lines*)

The p values are calculated by estimating the asymptotic variances of $\hat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\gamma}}_0$ obtained in Theorem 1. The value of test statistic $\mathcal{T}_{n_1 n_2}$ for this dataset is 301.8392, which is substantially larger than the 99% quantile of χ_6^2. The values of $\hat{\boldsymbol{\beta}}_0$, $\hat{\boldsymbol{\gamma}}_0$, and $\mathcal{T}_{n_1 n_2}$ indicate that the true values $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}_0$ are not equal for the two groups. Next, we used the test statistic proposed by Stute and Zhu (2005) to check whether the single-index models are appropriate for these two groups. The associated value of the test statistics is 1.4902 with a p value of 0.1361 for Group 1 and 1.4873 with a p value of 0.1401 for Group 2. This indicates that the single-index models are appropriate for these two groups. The estimators $\hat{g}_1(u)$ and $\hat{g}_2(u)$ along with their 95% pointwise confidence bands are presented in Fig. 4. The figure for Group 1(high NOX concentration) shows that the values of MEDV decrease with index $\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}_1$, while the figure for Group 2 (lower NOX concentration) shows that the values of MEDV increase with index $\hat{\boldsymbol{\gamma}}_0^\tau \boldsymbol{X}_2$. This is not surprising, as the air pollution index NOX is fairly strong related to life quality and hence house price. We conducted 1000 bootstraps to test $g_1(u) = g_2(u)$, and the corresponding $\mathfrak{T}_{n_1 n_2}^{\mathrm{KS}}$ and $\mathfrak{T}_{n_1 n_2}^{\mathrm{CM}}$ are both larger than the 97.5% quantile of 1000 bootstraps. This suggests a rejection of the null hypothesis $\tilde{\mathcal{H}}_0$. Lastly, we present the estimated figures for the variance functions $\sigma_1^2(u)$ and $\sigma_2^2(u)$ along with their associated 95% pointwise confidence intervals in Fig. 5. In Fig. 5, the heteroscedastic single-index regression model is appropriate for Group 1, and a homoscedastic single-index regression model is more appropriate for Group 2, as constant function $\hat{\sigma}_2^2(u) = \frac{1}{n_1} \sum_{i=1}^{n_2} \hat{\sigma}_2^2(\hat{\boldsymbol{\gamma}}_0^\tau \boldsymbol{X}_{2i})$ is encapsulated in the 95% pointwise confidence bands.

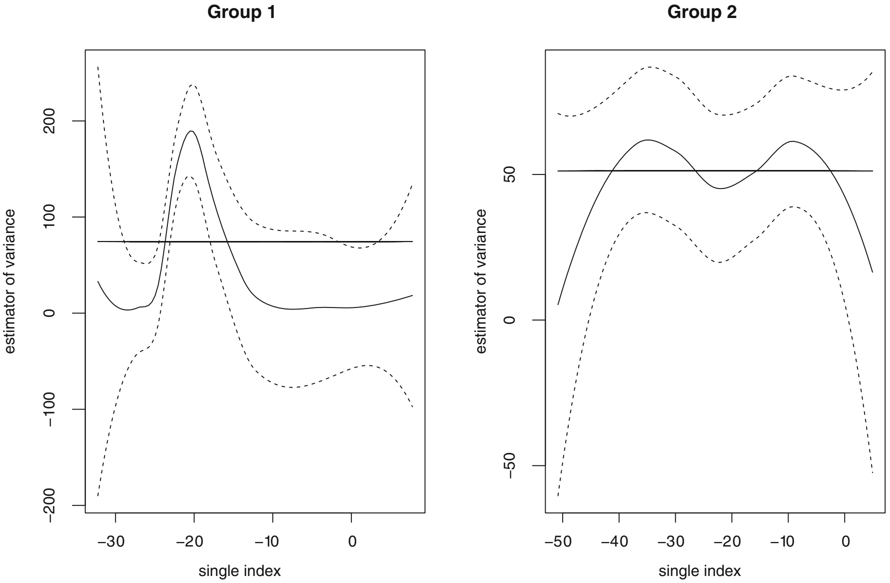

Fig. 5 Group 1: the plot for the estimator $\hat{\sigma}_1^2(u)$ (*solid line*) against estimated single-index $\hat{\boldsymbol{\beta}}_0^{\tau}\boldsymbol{X}_1$ in the *left panel*, along with the associated 95% pointwise confidence intervals (*dotted lines*) and a *horizontal line* for $\hat{\sigma}_1^2(u) = \frac{1}{n_1}\sum_{i=1}^{n_1}\hat{\sigma}_1^2(\hat{\boldsymbol{\beta}}_0^{\tau}\boldsymbol{X}_{1i})$; Group 2: the plot for estimator $\hat{\sigma}_2^2(u)$ (*solid line*) against estimated single-index $\hat{\boldsymbol{\gamma}}_0^{\tau}\boldsymbol{X}_2$ in the *right panel*, along with the associated 95% pointwise confidence intervals (*dotted lines*) a *horizontal line* for $\hat{\sigma}_2^2(u) = \frac{1}{n_2}\sum_{i=1}^{n_2}\hat{\sigma}_1^2(\hat{\boldsymbol{\gamma}}_0^{\tau}\boldsymbol{X}_{2i})$

Acknowledgements The authors thank the editor, the associate editor and two referees for their constructive suggestions that helped them to improve the early manuscript. Jun Zhang's research is supported by the National Natural Science Foundation of China (NSFC) (Grant No. 11401391). Zhenghui Feng's research is supported by the Fundamental Research Funds for the Central Universities in China (Grant No. 20720171025). Xiaoguang Wang's research is supported by the NSFC (Grant Nos. 11471065 and 11371077), and the Fundamental Research Funds for the Central Universities in China (Grant No. DUT15LK28).

Appendix

5.1 Proofs of Theorems 1 and 3

Lemma 1 *Suppose that $\boldsymbol{X}_i$, $i = 1, \ldots, n$ are i.i.d. random vectors. Let $m(\boldsymbol{x})$ be a continuous function and its derivatives up to second order are bounded, satisfying $E[m^2(\boldsymbol{X})] < \infty$. $E[m(\boldsymbol{X})|\boldsymbol{\beta}^{\tau}\boldsymbol{X} = u]$ has a continuous bounded second derivative on u. Let $K(u)$ be a bounded positive function with a bounded support satisfying the Lipschitz condition: there exists a neighborhood of the origin, say Υ, and a constant $c > 0$ such that for any $\epsilon \in \Upsilon$: $|K(u+\epsilon) - K(u)| < c|\epsilon|$. Given that $h = n^{-d}$ for some $d < 1$, we have, for $s_0 > 0$, and $j = 0, 1, 2$,*

$$\sup_{(x,\boldsymbol{\beta})\in\mathcal{X}\times\Delta}\left|\frac{1}{n}\sum_{i=1}^{n}K_h(\boldsymbol{\beta}^{\tau}\boldsymbol{X}_i - \boldsymbol{\beta}^{\tau}x)\left(\frac{\boldsymbol{\beta}^{\tau}\boldsymbol{X}_i - \boldsymbol{\beta}^{\tau}x}{h}\right)^j m(\boldsymbol{X}_i)\right.$$

$$- f_{\boldsymbol{\beta}_0^\tau \boldsymbol{X}}(\boldsymbol{\beta}_0^\tau \boldsymbol{x}) E\big[m(\boldsymbol{X})|\boldsymbol{\beta}_0^\tau \boldsymbol{X} = \boldsymbol{\beta}_0^\tau x\big]\mu_{K,j} - hS(\boldsymbol{\beta}_0^\tau \boldsymbol{x})\mu_{K,j+1}\Bigg|$$
$$= O_P(c_n),$$

where $\Delta = \{\boldsymbol{\beta} \in \Theta, \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \le Cn^{-1/2}\}$ *for some positive constant* C, $\Theta = \{\boldsymbol{\beta}, \|\boldsymbol{\beta}\| = 1, \beta_1 > 0\}$, $\mu_{K,l} = \int t^l K(t)dt$, $S(\boldsymbol{\beta}_0^\tau \boldsymbol{x}) = \frac{d}{du}\Big\{f_{\boldsymbol{\beta}_0^\tau \boldsymbol{X}}(u) E\big[m(\boldsymbol{X})|\boldsymbol{\beta}_0^\tau \boldsymbol{X} = u\big]\Big\}\,|_{u=\boldsymbol{\beta}_0^\tau \boldsymbol{x}}$, *and* $c_n = \left\{\frac{(\log n)^{1+s_0}}{nh}\right\}^{1/2} + h^2$.

Proof This proof can be completed by a similar argument of Lemma A.4 in Wang et al. (2010). See also the Lemma A6.1 in Xia (2006). □

Proofs of Theorems 1 and 3 We present the proof of Theorem 3. The proof of Theorem 1 is similar and we omit the details. We define $c_{n_s} = \left\{\frac{(\log n_s)^{1+s_0}}{n_s h_s}\right\}^{1/2} + h_s^2$ for $s = 1, 2$ for simplicity in the following.

Proof Note that $\mathcal{W}_{n_1 n_2}\left(\hat{\boldsymbol{\beta}}_{\mathcal{H}_0}^{(1)}\right) = \mathbf{0}$. Taylor expansion entails that

$$-\frac{1}{\sqrt{n_1+n_2}}\mathcal{W}_{n_1 n_2}\left(\boldsymbol{\beta}_0^{(1)}\right) = \left[\frac{1}{n_1+n_2}\frac{\partial \mathcal{W}_{n_1 n_2}\left(\boldsymbol{\beta}^{(1)}\right)}{\partial \boldsymbol{\beta}^{(1)}}\Big|_{\boldsymbol{\beta}^{(1)}=\tilde{\boldsymbol{\beta}}_0^{(1)}}\right]\left[\sqrt{n_1+n_2}\left(\hat{\boldsymbol{\beta}}_{\mathcal{H}_0}^{(1)} - \boldsymbol{\beta}_0^{(1)}\right)\right], \quad (18)$$

where $\tilde{\boldsymbol{\beta}}_0^{(1)}$ is between $\hat{\boldsymbol{\beta}}_{\mathcal{H}_0}^{(1)}$ and $\boldsymbol{\beta}_0^{(1)}$.

Step 1 In the following, we define $N = n_1 + n_2$ for simplicity. In this step, we deal with $N^{-1/2}\mathcal{W}_{n_1 n_2}\left(\boldsymbol{\beta}_0^{(1)}\right)$. Using Lemma 1 and the detailed proofs of Lemma A.4 in Zhang et al. (2014), we have $\hat{g}_s(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{si}, \boldsymbol{\beta}_0) = g_s(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{si}) + O_P(c_{ns})$, $\hat{V}_s(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{si}, \boldsymbol{\beta}_0) = V_{s,\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{si}) + O_P(c_{ns})$, for $s = 1, 2$. Moreover,

$$S_{n_1,l_1 1}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i}, \boldsymbol{\beta}_0)$$
$$= \frac{1}{n_1}\sum_{j=1}^{n_1} K_{h_1}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1j} - \boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i})(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1j} - \boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i})^{l_1}\sigma_1^2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1j})\epsilon_{1j}^2$$
$$+ \frac{2}{n_1}\sum_{j=1}^{n_1} K_{h_1}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1j} - \boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i})(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1j} - \boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i})^{l_1}$$
$$\times \big[g_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1j}) - \hat{g}_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1j}, \boldsymbol{\beta}_0)\big]\sigma_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1j})\epsilon_{1j}$$
$$+ \frac{1}{n_1}\sum_{j=1}^{n_1} K_{h_1}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1j} - \boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i})(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1j} - \boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i})^{l_1}$$

$$\times \left[g_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1j}) - \hat{g}_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1j}, \boldsymbol{\beta}_0)\right]^2$$
$$= h_1^{l_1} f_{\boldsymbol{\beta}_0^\tau \boldsymbol{X}_1}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i})\sigma_1^2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i})\mu_{Kl_1} + O_P(h_1^{l_1} c_{n1} + h_1^{l_1} c_{n1}^2), \quad (19)$$

for $l_1 = 0, 1, 2$. Using (19), we obtain $\hat{\sigma}_1^2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i}, \boldsymbol{\beta}_0) = \sigma_1^2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i}) + O_P(c_{n1})$. Similarly, $\hat{\sigma}_2^2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{2i}, \boldsymbol{\beta}_0) = \sigma_2^2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{2i}) + O_P(c_{n2})$.

Let $G_{1,w}^{\boldsymbol{x}}(u, \boldsymbol{\beta}) = E\left[Y_1^w\{\boldsymbol{X}_1 - \boldsymbol{x}\}|\boldsymbol{\beta}^\tau \boldsymbol{X}_1 = u\right] f_{\boldsymbol{\beta}^\tau \boldsymbol{X}}(u)$, $K'_{h_1}(u) = \frac{1}{h_1}K'(u/h_1)$. Using condition (C3), we have

$$\begin{aligned}
&E\left[\frac{\partial}{\partial \boldsymbol{\beta}} T_{n_1,l_1l_2}(\boldsymbol{\beta}^\tau \boldsymbol{x}, \boldsymbol{\beta})\right] \\
&= \frac{1}{n_1}\sum_{i=1}^{n_1} E\left[K'_{h_1}(\boldsymbol{\beta}^\tau \boldsymbol{X}_{1i} - \boldsymbol{\beta}^\tau \boldsymbol{x}) J_{\boldsymbol{\beta}}^\tau \left(\frac{\boldsymbol{X}_{1i} - \boldsymbol{x}}{h_1}\right)(\boldsymbol{\beta}^\tau \boldsymbol{X}_{1i} - \boldsymbol{\beta}^\tau \boldsymbol{x})^{l_1} Y_{1i}^{l_2}\right] \\
&\quad + \frac{1}{n_1}\sum_{i=1}^{n_1} E\left[K_{h_1}(\boldsymbol{\beta}^\tau \boldsymbol{X}_{1i} - \boldsymbol{\beta}^\tau \boldsymbol{x}) J_{\boldsymbol{\beta}}^\tau (\boldsymbol{X}_{1i} - \boldsymbol{x})\, l_1(\boldsymbol{\beta}^\tau \boldsymbol{X}_{1i} - \boldsymbol{\beta}^\tau \boldsymbol{x})^{l_1-1} I\{l_1 \ge 1\} Y_{1i}^{l_2}\right] \\
&= -\sum_{v=0}^{2} \frac{l_1 + v}{v!} J_{\boldsymbol{\beta}}^\tau G_{1,l_2}^{\boldsymbol{x}(v)}(\boldsymbol{\beta}^\tau \boldsymbol{x}, \boldsymbol{\beta}) h_1^{l_1-1+v} \mu_{K,l_1-1+v} I\{l_1 + v \ge 1\} \\
&\quad + \sum_{v=0}^{3} \frac{l_1}{v!} J_{\boldsymbol{\beta}}^\tau G_{1,l_2}^{\boldsymbol{x}(v)}(\boldsymbol{\beta}^\tau \boldsymbol{x}, \boldsymbol{\beta}) h_1^{l_1-1+v} \mu_{K,l_1-1+v} I\{l_1 \ge 1\} + O(h_1^{l_1+2}),
\end{aligned} \quad (20)$$

where $G_{1,l_2}^{\boldsymbol{x}(v)}(u, \boldsymbol{\beta}) = \frac{\partial^v}{\partial u^v} G_{1,l_2}^{\boldsymbol{x}}(u, \boldsymbol{\beta})$, and $I\{u\}$ is the indicator function. Similar to the proof of Theorem 3.1 in Fan and Gijbels (1996) and Lemma A.5 in Zhang et al. (2014), together with (20) and Lemma 1, we can have

$$\begin{aligned}
&\frac{\partial \hat{g}_1(\boldsymbol{\beta}^\tau \boldsymbol{X}_{1i}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^{(1)}}\bigg|_{\boldsymbol{\beta}^{(1)} = \tilde{\boldsymbol{\beta}}_0^{(1)}} \\
&= J_{\boldsymbol{\beta}_0}^\tau \left[\boldsymbol{X}_{1i} - V_{1,\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i})\right] g'_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i}) + O_P\left(h_1^2 + \sqrt{\frac{(\log n_1)^{1+s_0}}{n_1 h_1^3}}\right).
\end{aligned} \quad (21)$$

Under the null hypothesis $\mathcal{H}_0$,

$$\begin{aligned}
&\frac{\partial \hat{g}_2(\boldsymbol{\beta}^\tau \boldsymbol{X}_{2i}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^{(1)}}\bigg|_{\boldsymbol{\beta}^{(1)} = \tilde{\boldsymbol{\beta}}_0^{(1)}} \\
&= J_{\boldsymbol{\beta}_0}^\tau \left[\boldsymbol{X}_{2i} - V_{2,\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{2i})\right] g'_2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{2i}) + O_P\left(h_2^2 + \sqrt{\frac{(\log n_2)^{1+s_0}}{n_2 h_2^3}}\right).
\end{aligned} \quad (22)$$

Define that $\mathcal{Q}_{n_1}(u, \boldsymbol{\beta}_0) = \frac{1}{n_1 h_1^2} T_{n_1,20}(u, \boldsymbol{\beta}_0) \frac{1}{n_1} T_{n_1,00}(u, \boldsymbol{\beta}_0) - \frac{1}{n_1^2 h_1^2} T_{n_1,10}^2(u, \boldsymbol{\beta}_0)$ and $\mathcal{L}_{n_1}(u, \boldsymbol{\beta}_0) = \frac{1}{n_1^2 h_1^2} T_{n_1,20}(u, \boldsymbol{\beta}_0) T_{n_1,01}(u, \boldsymbol{\beta}_0) - \frac{1}{n_1^2 h_1^2} T_{n_1,10}(u, \boldsymbol{\beta}_0) T_{n_1,11}$

$(u, \boldsymbol{\beta}_0)$. Then, $\hat{g}_1(u, \boldsymbol{\beta}_0) = \frac{\mathcal{L}_{n_1}(u,\boldsymbol{\beta}_0)}{\mathcal{Q}_{n_1}(u,\boldsymbol{\beta}_0)}$ and $\hat{g}'_1(u, \boldsymbol{\beta}_0) = \frac{\partial \mathcal{L}_{n_1}(u,\boldsymbol{\beta}_0)/\partial u}{\mathcal{Q}_{n_1}(u,\boldsymbol{\beta}_0)} - \frac{\mathcal{L}_{n_1}(u,\boldsymbol{\beta}_0)\partial \mathcal{Q}_{n_1}(u,\boldsymbol{\beta}_0)/\partial u}{\mathcal{Q}^2_{n_1}(u,\boldsymbol{\beta}_0)}$. Following the proof of Lemma A.5 in Zhang et al. (2014), together with Lemma 1 and (20), we have $\hat{g}'_1(u, \boldsymbol{\beta}_0) = g'_1(u) + O_P\left(h_1^2 + \sqrt{\frac{(\log n_1)^{1+s_0}}{n_1 h_1^3}}\right)$ and $\hat{g}'_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i}, \boldsymbol{\beta}_0) = g'_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i}) + O_P\left(h_1^2 + \sqrt{\frac{(\log n_1)^{1+s_0}}{n_1 h_1^3}}\right)$. Similarly, $\hat{g}'_2(u, \boldsymbol{\beta}_0) = g'_2(u) + O_P\left(h_2^2 + \sqrt{\frac{(\log n_2)^{1+s_0}}{n_2 h_2^3}}\right)$, $\hat{g}'_2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{2i}, \boldsymbol{\beta}_0) = g'_2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{2i}) + O_P\left(h_2^2 + \sqrt{\frac{(\log n_2)^{1+s_0}}{n_2 h_2^3}}\right)$.

Using the asymptotic results (21) and (22) and the condition of that $\frac{n_1}{n_1+n_2} = \frac{n_1}{N} \to \lambda \in (0, 1)$, as $\max\left\{\frac{(\log n_1)^{2+2s_0}}{n_1 h_1^2}, \frac{(\log n_2)^{2+2s_0}}{n_2 h_2^2}\right\} \to 0$, and also $\max\{n_1 h_1^8, n_2 h_2^8\} \to 0$, we have

$$\begin{aligned}
&(n_1+n_2)^{-1/2}\mathcal{W}_{n_1 n_2}\left(\boldsymbol{\beta}_0^{(1)}\right)\\
&= \sqrt{\frac{n_1}{n_1+n_2}} n_1^{-1/2} \sum_{i=1}^{n_1} J_{\boldsymbol{\beta}_0}^\tau \frac{\hat{g}'_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i}, \boldsymbol{\beta}_0)}{\hat{\sigma}_1^2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i}, \boldsymbol{\beta}_0)} \left[\boldsymbol{X}_{1i} - \hat{V}_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i}, \boldsymbol{\beta}_0)\right]\\
&\quad \times \left[Y_{1i} - \hat{g}_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i}, \boldsymbol{\beta}_0)\right]\\
&\quad + \sqrt{\frac{n_2}{n_1+n_2}} n_2^{-1/2} \sum_{i=1}^{n_2} J_{\boldsymbol{\beta}_0}^\tau \frac{\hat{g}'_2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{2i}, \boldsymbol{\beta}_0)}{\hat{\sigma}_2^2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{2i}, \boldsymbol{\beta}_0)} \left[\boldsymbol{X}_{2i} - \hat{V}_2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{2i}, \boldsymbol{\beta}_0)\right]\\
&\quad \times \left[Y_{2i} - \hat{g}_2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{2i}, \boldsymbol{\beta}_0)\right]\\
&= \sqrt{\frac{n_1}{n_1+n_2}} n_1^{-1/2} \sum_{i=1}^{n_1} J_{\boldsymbol{\beta}_0}^\tau g'_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i}) \left[\boldsymbol{X}_{1i} - V_{1,\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i})\right] \sigma_1^{-1}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i}) \epsilon_{1i}\\
&\quad + \sqrt{\frac{n_2}{n_1+n_2}} n_2^{-1/2} \sum_{i=1}^{n_2} J_{\boldsymbol{\beta}_0}^\tau g'_2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{2i}) \left[\boldsymbol{X}_{2i} - V_{2,\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{2i})\right] \sigma_2^{-1}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{2i}) \epsilon_{2i}\\
&\quad + o_P(1),
\end{aligned} \tag{23}$$

where $V_{2,\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_2) = E[\boldsymbol{X}_2 | \boldsymbol{\beta}_0^\tau \boldsymbol{X}_2]$.

Step 2 In this sub-step, we deal with $\frac{1}{n_1+n_2} \frac{\partial \mathcal{W}_{n_1 n_2}\left(\boldsymbol{\beta}^{(1)}\right)}{\partial \boldsymbol{\beta}^{(1)}}\Big|_{\boldsymbol{\beta}^{(1)} = \tilde{\boldsymbol{\beta}}_0^{(1)}}$. Define

$$\begin{aligned}
\mathcal{S}_{n_1 n_2}(\tilde{\boldsymbol{\beta}}_0^{(1)}) &\overset{\text{def}}{=} \frac{1}{n_1+n_2} \sum_{s=1}^{2} \sum_{i=1}^{n_s} \left[Y_{si} - \hat{g}_s(\tilde{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}_{si}, \tilde{\boldsymbol{\beta}}_0)\right]\\
&\times \frac{\partial}{\partial \boldsymbol{\beta}^{(1)}} \left\{J_{\boldsymbol{\beta}}^\tau \hat{g}'_s(\boldsymbol{\beta}^\tau \boldsymbol{X}_{si}, \boldsymbol{\beta}) \left[\boldsymbol{X}_{si} - \hat{V}_s(\boldsymbol{\beta}^\tau \boldsymbol{X}_{si}, \boldsymbol{\beta})\right] \hat{\sigma}_s^{-2}(\boldsymbol{\beta}^\tau \boldsymbol{X}_{si}, \boldsymbol{\beta})\right\}\Big|_{\boldsymbol{\beta}^{(1)} = \tilde{\boldsymbol{\beta}}_0^{(1)}},
\end{aligned}$$

and

$$\mathcal{L}_{n_1n_2}(\tilde{\boldsymbol{\beta}}_0^{(1)}) \stackrel{\text{def}}{=} \frac{1}{n_1+n_2}\sum_{s=1}^{2}\sum_{i=1}^{n_s}\left\{J_{\tilde{\boldsymbol{\beta}}_0}^{\tau}\hat{g}_s'(\tilde{\boldsymbol{\beta}}_0^{\tau}\boldsymbol{X}_{si},\tilde{\boldsymbol{\beta}}_0)\left[\boldsymbol{X}_{si}-\hat{V}_s(\tilde{\boldsymbol{\beta}}_0^{\tau}\boldsymbol{X}_{si},\tilde{\boldsymbol{\beta}}_0)\right]\right. \left.\times\hat{\sigma}_s^{-2}(\tilde{\boldsymbol{\beta}}_0^{\tau}\boldsymbol{X}_{si},\tilde{\boldsymbol{\beta}}_0)\right\}\frac{\partial\hat{g}_s(\boldsymbol{\beta}^{\tau}\boldsymbol{X}_{si},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^{(1)}}\bigg|_{\boldsymbol{\beta}^{(1)}=\tilde{\boldsymbol{\beta}}_0^{(1)}}.$$

Then,

$$\frac{1}{n_1+n_2}\frac{\partial\mathcal{W}_{n_1n_2}\left(\boldsymbol{\beta}^{(1)}\right)}{\partial\boldsymbol{\beta}^{(1)}}\bigg|_{\boldsymbol{\beta}^{(1)}=\tilde{\boldsymbol{\beta}}_0^{(1)}} = \mathcal{S}_{n_1n_2}(\tilde{\boldsymbol{\beta}}_0^{(1)})+\mathcal{L}_{n_1n_2}(\tilde{\boldsymbol{\beta}}_0^{(1)}), \tag{24}$$

where $\tilde{\boldsymbol{\beta}}_0=\left(\sqrt{1-\tilde{\boldsymbol{\beta}}_0^{(1)\tau}\tilde{\boldsymbol{\beta}}_0^{(1)}},\tilde{\boldsymbol{\beta}}_0^{(1)\tau}\right)^{\tau}$. Note that $\tilde{\boldsymbol{\beta}}_0^{(1)}$ is between $\hat{\boldsymbol{\beta}}_{\mathcal{H}_0}^{(1)}$ and $\boldsymbol{\beta}_0^{(1)}$. By using (18), we have $\hat{\boldsymbol{\beta}}_{\mathcal{H}_0}^{(1)}=\boldsymbol{\beta}_0^{(1)}+O_P((n_1+n_2)^{-1/2})$. Note that $\tilde{\boldsymbol{\beta}}_0^{(1)}\xrightarrow{P}\boldsymbol{\beta}_0^{(1)}$, $\tilde{\boldsymbol{\gamma}}_0^{(1)}\xrightarrow{P}\boldsymbol{\beta}_0^{(1)}$ and $\tilde{\boldsymbol{\beta}}_0\xrightarrow{P}\boldsymbol{\gamma}_0$, $\tilde{\boldsymbol{\gamma}}_0\xrightarrow{P}\boldsymbol{\gamma}_0$. Together with (21)–(22) and condition of that $\frac{n_1}{n_1+n_2}\to\lambda\in(0,1)$, we have

$$\mathcal{L}_{n_1n_2}(\tilde{\boldsymbol{\beta}}_0^{(1)})\xrightarrow{P}\lambda J_{\boldsymbol{\beta}_0}^{\tau}E\left[\frac{g_1^{\prime 2}(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_1)}{\sigma_1^2(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_1)}\left[\boldsymbol{X}_1-V_{1,\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_1)\right]^{\otimes 2}\right]J_{\boldsymbol{\beta}_0} +(1-\lambda)J_{\boldsymbol{\beta}_0}^{\tau}E\left[\frac{g_2^{\prime 2}(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_2)}{\sigma_2^2(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_2)}\left[\boldsymbol{X}_2-V_{2,\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_2)\right]^{\otimes 2}\right]J_{\boldsymbol{\beta}_0}. \tag{25}$$

Moreover, a direct calculation for $\mathcal{S}_{n_1n_2}(\tilde{\boldsymbol{\beta}}_0^{(1)})$ and Lemma 1 entail that $\mathcal{S}_{n_1n_2}(\tilde{\boldsymbol{\beta}}_0^{(1)})$ $=o_P(1)$. Together with (23) and (25), we complete the proof of Theorem 2. □

5.2 Proof of Theorem 3

Proof From the proof of Theorem 3, we can have that

$$\sqrt{n_1}\left(\hat{\boldsymbol{\beta}}_0^{(1)}-\boldsymbol{\beta}_0^{(1)}\right) = \boldsymbol{\Omega}_1^{-1}n_1^{-1/2}\sum_{i=1}^{n_1}J_{\boldsymbol{\beta}_0}^{\tau}g_1'(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_{1i})\left[\boldsymbol{X}_{1i}-V_{1,\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_{1i})\right] \sigma_1^{-1}(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_{1i})\epsilon_{1i}+o_P(1), \tag{26}$$

$$\sqrt{n_2}\left(\hat{\boldsymbol{\gamma}}_0^{(1)}-\boldsymbol{\gamma}_0^{(1)}\right) = \boldsymbol{\Omega}_2^{-1} n_2^{-1/2} \sum_{i=1}^{n_2} J_{\boldsymbol{\gamma}_0}^{\tau} g_2'(\boldsymbol{\gamma}_0^{\tau}\boldsymbol{X}_{2i})\left[\boldsymbol{X}_{2i}-V_{2,\boldsymbol{\gamma}_0}(\boldsymbol{\gamma}_0^{\tau}\boldsymbol{X}_{2i})\right]\sigma_2^{-1}(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_{2i})\epsilon_{2i}+o_P(1). \tag{27}$$

Under the null hypothesis $\mathcal{H}_0 : \boldsymbol{\beta}_0 = \boldsymbol{\gamma}_0$, we can have

$$\begin{aligned}
&\sqrt{n_1+n_2}\left(\hat{\boldsymbol{\beta}}_0^{(1)}-\hat{\boldsymbol{\gamma}}_0^{(1)}\right)\\
&=\left(\frac{\sqrt{n_1}}{\sqrt{\lambda}}\left(\hat{\boldsymbol{\beta}}_0^{(1)}-\boldsymbol{\beta}_0^{(1)}\right)-\frac{\sqrt{n_2}}{\sqrt{1-\lambda}}\left(\hat{\boldsymbol{\gamma}}_0^{(1)}-\boldsymbol{\beta}_0^{(1)}\right)\right)+o_P(1)\\
&\xrightarrow[\mathcal{H}_0^*]{\mathcal{L}} N\left(\mathbf{0}_{p-1}, \frac{1}{\lambda}\boldsymbol{\Omega}_1^{-1}+\frac{1}{1-\lambda}\boldsymbol{\Omega}_2^{-1}\right).
\end{aligned}$$

Moreover,

$$\begin{aligned}
&(n_1+n_2)\widehat{\boldsymbol{A}}\\
&=\frac{n_1+n_2}{n_1}\left[J_{\hat{\boldsymbol{\beta}}_0}^{\tau}\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{\hat{g}_1'^2(\hat{\boldsymbol{\beta}}_0^{\tau}\boldsymbol{X}_{1i},\hat{\boldsymbol{\beta}}_0)}{\hat{\sigma}_1^2(\hat{\boldsymbol{\beta}}_0^{\tau}\boldsymbol{X}_{1i},\hat{\boldsymbol{\beta}}_0),}\left[\boldsymbol{X}_{1i}-\hat{V}_1(\hat{\boldsymbol{\beta}}_0^{\tau}\boldsymbol{X}_{1i},\hat{\boldsymbol{\beta}}_0)\right]^{\otimes 2}J_{\hat{\boldsymbol{\beta}}_0}\right]^{-1}\\
&\quad+\frac{n_1+n_2}{n_2}\left[J_{\hat{\boldsymbol{\gamma}}_0}^{\tau}\frac{1}{n_2}\sum_{i=1}^{n_2}\frac{\hat{g}_2'^2(\hat{\boldsymbol{\gamma}}_0^{\tau}\boldsymbol{X}_{2i},\hat{\boldsymbol{\gamma}}_0)}{\hat{\sigma}_2^2(\hat{\boldsymbol{\gamma}}_0^{\tau}\boldsymbol{X}_{2i},\hat{\boldsymbol{\gamma}}_0),}\left[\boldsymbol{X}_{2i}-\hat{V}_2(\hat{\boldsymbol{\gamma}}_0^{\tau}\boldsymbol{X}_{2i},\hat{\boldsymbol{\gamma}}_0)\right]^{\otimes 2}J_{\hat{\boldsymbol{\gamma}}_0}\right]^{-1}\\
&\xrightarrow{P}\frac{1}{\lambda}\boldsymbol{\Omega}_1^{-1}+\frac{1}{1-\lambda}\boldsymbol{\Omega}_2^{-1}.
\end{aligned}$$

Then, the Slutsky Theorem and continuous mapping theorem entail that

$$\begin{aligned}
&\mathcal{T}_{n_1n_2}\\
&=\left[\sqrt{n_1+n_2}\left(\hat{\boldsymbol{\beta}}_0^{(1)}-\hat{\boldsymbol{\gamma}}_0^{(1)}\right)\right]^{\tau}\left((n_1+n_2)\widehat{\boldsymbol{A}}\right)^{-1}\left[\sqrt{n_1+n_2}\left(\hat{\boldsymbol{\beta}}_0^{(1)}-\hat{\boldsymbol{\gamma}}_0^{(1)}\right)\right]\\
&\times\xrightarrow[\mathcal{H}_0^*]{\mathcal{L}}\chi_{p-1}^2.
\end{aligned}$$

We complete the proof of Theorem 3. □

5.3 Proof of Theorem 4

Lemma 2 *Suppose that conditions* (C1)–(C5) *hold. Let* $F_{\hat{\epsilon}_s}(t|\mathcal{Q}_{n_s})$ *be the distribution function of* $\hat{\epsilon}_s=\frac{Y_s-\hat{g}_s(\hat{\omega}_{s,0}^{\tau}\boldsymbol{X}_s,\hat{\omega}_{s,0})}{\hat{\sigma}_s(\hat{\omega}_{s,0}^{\tau}\boldsymbol{X}_s,\hat{\omega}_{s,0})}$ *conditional on the data* $\mathcal{Q}_{n_s}=\{\boldsymbol{X}_{si},Y_{si}\}_{i=1}^{n_s}$ *(i.e., considering* $\hat{g}_s\left(\hat{\omega}_{s,0}^{\tau}\boldsymbol{x}_s,\hat{\omega}_{s,0}\right)$, $\hat{\sigma}_s\left(\hat{\omega}_{s,0}^{\tau}\boldsymbol{x}_s,\hat{\omega}_{s,0}\right)$ *as fixed functions on* $\boldsymbol{x}_s$*) for* $s=1,2$

respectively. Here, $\hat{\omega}_{1,0} = \hat{\boldsymbol{\beta}}_0$ *and* $\hat{\omega}_{2,0} = \hat{\boldsymbol{\gamma}}_0$*. Then, we have*

$$\sup_{t\in\mathbb{R}} \left| n_1^{-1} \sum_{i=1}^{n_1} \left[I\left\{\hat{\epsilon}_{1i} \le t\right\} - I\{\epsilon_{1i} \le t\} - F_{\hat{\epsilon}_1}(t|\mathcal{Q}_{n_1}) + F_{\epsilon_1}(t)\right] \right| = o_P\left(n_1^{-1/2}\right), \tag{28}$$

$$\sup_{t\in\mathbb{R}} \left| n_2^{-1} \sum_{i=1}^{n_2} \left[I\left\{\hat{\epsilon}_{2i} \le t\right\} - I\{\epsilon_{2i} \le t\} - F_{\hat{\epsilon}_2}(t|\mathcal{Q}_{n_2}) + F_{\epsilon_2}(t)\right] \right| = o_P\left(n_2^{-1/2}\right). \tag{29}$$

Proof In the following, we only prove (28), the proof of (29) is similar and we omit the details. Let

$$\mathscr{O} = \Big\{ I\{\epsilon_1 \le t f_2(\boldsymbol{X}_1) + f_1(\boldsymbol{X}_1)\} - I\{\epsilon_1 \le t\} - P(\epsilon_1 \le t f_2(\boldsymbol{X}_1) + f_1(\boldsymbol{X}_1)) + P(\epsilon_1 \le t); t \in \mathbb{R}, f_1, f_2 \in M_1^{1+\delta}(\mathfrak{R}_c^p) \Big\},$$

where $M_1^{1+\delta}(\mathfrak{R}_c^p)$ is the class of all differential functions $f(u)$ defined on the domain $\mathfrak{R}_c^p$ of $\boldsymbol{x}_1$ and $\|f\|_{1+\delta} \le 1$. Here $\mathfrak{R}_c^p$ is a compact set of $\mathbb{R}^p$ and

$$\|f\|_{1+\delta} = \sup_{\boldsymbol{x}_1\in\mathfrak{R}_c^p} |f(\boldsymbol{x}_1)| + \sum_{l=1}^{p} \sup_{\boldsymbol{x}_1\in\mathfrak{R}_c^p} \left|\frac{\partial f(\boldsymbol{x}_1)}{\partial x_{1l}}\right| + \sup_{\boldsymbol{x}_{1,1},\boldsymbol{x}_{1,2}\in\mathfrak{R}_c^p} \frac{|\partial f(\boldsymbol{x}_{1,1}) - \partial f(\boldsymbol{x}_{1,2})|}{\|\boldsymbol{x}_{1,1} - \boldsymbol{x}_{1,2}\|^\delta}.$$

Using Lemma 1 and $\|\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0\| = O_P(n_1^{-1/2})$, and similar to the proofs of (21) and (22), we have that

$$\hat{g}_1(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{x}_1, \hat{\boldsymbol{\beta}}_0) = g_1(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1) + O_P\left(n_1^{-1/2} + c_{n_1}\right), \tag{30}$$

$$\hat{\sigma}_1(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{x}_1, \hat{\boldsymbol{\beta}}_0) = \sigma_1(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1) = O_P\left(n_1^{-1/2} + c_{n_1}\right), \tag{31}$$

uniformly in $\boldsymbol{x}_1 \in \mathfrak{R}_c^p$. Let $A_{n_1}(\boldsymbol{x}_1) = \frac{\hat{g}_1(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{x}_1, \hat{\boldsymbol{\beta}}_0) - g_1(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1)}{\sigma_1(\boldsymbol{\beta}_0^\tau \boldsymbol{x})}$, $B_{n_1}(\boldsymbol{x}) = \frac{\hat{\sigma}_1(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{x}_1, \hat{\boldsymbol{\beta}}_0)}{\sigma_1(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1)}$. So, (30) and (31) entail $P\left(A_{n_1} \in M_1^{1+\delta}(\mathfrak{R}_c^p)\right) \to 1$, $P\left(B_{n_1} \in M_1^{1+\delta}(\mathfrak{R}_c^p)\right) \to 1$ as $n_1 \to \infty$, $h_1 \to 0$ and $\frac{n_1 h_1}{(\log n_1)^{1+s}} \to \infty$.

By directly using the Corollary 2.7.2 of van der Vaart and Wellner (1996), the bracketing number $N_{[\,]}\left(\upsilon^2, M_1^{1+\delta}(\mathfrak{R}_c^p), L_2(P)\right)$ can be at most $\exp\left(c_0 \upsilon^{-\frac{2p}{1+\delta}}\right)$ for some positive constant c_0, According to the proof of Lemma 1 in Appendix B of Akritas and Van Keilegom (2001), and then the class $\mathscr{O}$ defined above is a Donsker class, i.e., we have that $\int_0^\infty \sqrt{N_{[\,]}(\bar{\upsilon}, \mathscr{O}, L_2(P))} d\bar{\upsilon} < \infty$. Then, the proof of (28) is complete. □

Proof of Theorem 4 We can have that

$$
\begin{aligned}
&\hat{F}_{\epsilon_1}(t) - F_{\epsilon_1}(t) \\
&= \frac{1}{n_1}\sum_{i=1}^{n_1} I\{\epsilon_{1i} \le t\} - F_{\epsilon_1}(t) + \left(F_{\hat{\epsilon}_1}(t|\mathscr{Q}_{n_1}) - F_{\epsilon_1}(t)\right) + R_{n_1,1}(t), \quad (32)
\end{aligned}
$$

where $R_{n_1,1}(t) = o_P(n_1^{-1/2})$ uniformly in $t \in \mathbb{R}$ by using Lemma 2. Taylor expansion entails that

$$
\begin{aligned}
&F_{\hat{\epsilon}_1}(t|\mathscr{Q}_{n_1}) - F_{\epsilon_1}(t) \\
&= \int \left[F_{\epsilon_1}(t + t[B_{n_1}(\boldsymbol{x}_1) - 1] + A_{n_1}(\boldsymbol{x}_1)) - F_{\epsilon_1}(t)\right] dF_{\boldsymbol{X}_1}(\boldsymbol{x}_1) \\
&= f_{\epsilon_1}(t)t\int [B_{n_1}(\boldsymbol{x}_1) - 1]dF_{\boldsymbol{X}_1}(\boldsymbol{x}_1) + f_{\epsilon_1}(t)\int A_{n_1}(\boldsymbol{x}_1)dF_{\boldsymbol{X}_1}(\boldsymbol{x}_1) \\
&\quad + \int f'_{\epsilon_1}(t + v^*_{n_1}(t, \boldsymbol{x}_1))\left\{t[B_{n_1}(\boldsymbol{x}_1) - 1] + A_{n_1}(\boldsymbol{x}_1)\right\}^2 dF_{\boldsymbol{X}_1}(\boldsymbol{x}_1) \\
&= R_{n_1,2}(t) + R_{n_1,3}(t) + R_{n_1,4}(t), \quad (33)
\end{aligned}
$$

where $v^*_{n_1}(t, \boldsymbol{x}_1)$ is between 0 and $t[B_{n_1}(\boldsymbol{x}_1) - 1] + A_{n_1}(\boldsymbol{x}_1)$. Note that

$$
\begin{aligned}
&A_{n_1}(\boldsymbol{x}_1) \\
&= \frac{\hat{g}_1(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{x}_1, \boldsymbol{\beta}_0) - \hat{g}_1(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1, \boldsymbol{\beta}_0)}{\sigma_1(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1)} + \frac{\hat{g}_1(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{x}_1, \boldsymbol{\beta}_0) - g(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1)}{\sigma_1(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1)}. \quad (34)
\end{aligned}
$$

Recall the definition of $\hat{g}_1(u, \boldsymbol{\beta}_0)$ and using Lemma 1,

$$
\begin{aligned}
&\hat{g}_1(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1, \boldsymbol{\beta}_0) - g_1(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1) \\
&= \frac{1}{n_1 f_{\boldsymbol{\beta}_0^\tau \boldsymbol{X}_1}(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1)}\sum_{i=1}^{n_1} K_{h_1}\left(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i} - \boldsymbol{\beta}_0^\tau \boldsymbol{x}_1\right)\sigma_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i})\epsilon_{1i} + O_P(c_{n_1}). \quad (35)
\end{aligned}
$$

Similar to (21), we can also have

$$
\begin{aligned}
&\hat{g}_1(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{x}_1, \boldsymbol{\beta}_0) - \hat{g}_1(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1, \boldsymbol{\beta}_0) \\
&= \left[\boldsymbol{x}_1 - V_{1,\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1)\right]^\tau g_1'(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1)\left(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0\right) + O_P\left(h_1^2 + \sqrt{\frac{(\log n_1)^{1+s}}{n_1 h_1^3}}\right). \quad (36)
\end{aligned}
$$

Together with (34), (35) and (36), we have

$$
R_{n_1,3}(t) = f_{\epsilon_1}(t)\int A_{n_1}(\boldsymbol{x}_1)dF_{\boldsymbol{X}_1}(\boldsymbol{x}_1) = \frac{f_{\epsilon_1}(t)}{n_1}\sum_{i=1}^{n_1}\epsilon_{1i} + o_P(n_1^{-1/2}). \quad (37)
$$

Similarly,

$$\hat{\sigma}_1^2\left(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{x}_1, \hat{\boldsymbol{\beta}}_0\right) - \sigma_1^2(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1)$$
$$= \frac{1}{f_{\boldsymbol{\beta}_0^\tau \boldsymbol{X}_1}(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1) n_1} \sum_{i=1}^{n_1} K_{h_1}\left(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i} - \boldsymbol{\beta}_0^\tau \boldsymbol{x}_1\right) \sigma_1^2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i})(\epsilon_{1i}^2 - 1)$$
$$+ 2\left[\boldsymbol{x}_1 - V_{1,\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1)\right]^\tau \sigma_1(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1)\sigma_1'(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1)\left(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0\right) + o_P(n_1^{-1/2}). \quad (38)$$

Then, Taylor expansion for $\sqrt{\hat{\sigma}_1^2\left(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{x}, \hat{\boldsymbol{\beta}}_0\right)} - \sqrt{\sigma_1^2(\boldsymbol{\beta}_0^\tau \boldsymbol{x})}$ and asymptotic expression (38) entail that

$$R_{n_1,2}(t) = t f_{\epsilon_1}(t) \int [B_{n_1}(\boldsymbol{x}_1) - 1] dF_{\boldsymbol{X}_1}(\boldsymbol{x}_1)$$
$$= t f_{\epsilon_1}(t) \frac{1}{2n_1} \sum_{i=1}^{n_1} (\epsilon_{1i}^2 - 1) + o_P(n_1^{-1/2}). \quad (39)$$

Moreover, (34), (39) and Condition (C5) entail that $R_{n_1,4}(t) = o_P(n_1^{-1/2})$ uniformly in t. Together with (32), (26) and (37)–(39), we complete the proof of Theorem 2. □

5.4 Proof of Theorems 5 and 6

Proof Recalling the definition of $F^*_{\widetilde{\mathcal{H}}_{0,\epsilon_1}}(t) = E\left[F_{\epsilon_1}\left(t + \frac{g_2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_1) - g_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_1)}{\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})}\right)\right]$.

$$\hat{F}_{\widetilde{\mathcal{H}}_{0,\epsilon_1}}(t) - F^*_{\widetilde{\mathcal{H}}_{0,\epsilon}}(t) = \frac{1}{n_1} \sum_{i=1}^{n_1} I\{\hat{\epsilon}_{\widetilde{\mathcal{H}}_0,i} \le t\} - F^*_{\widetilde{\mathcal{H}}_{0,\epsilon_1}}(t)$$
$$= \frac{1}{n_1} \sum_{i=1}^{n_1} I\left\{\epsilon_{1i} + \frac{g_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i}) - g_2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i})}{\sigma_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i})} \le t\right\} - F^*_{\widetilde{\mathcal{H}}_{0,\epsilon_1}}(t)$$
$$+ \left[F_{\widetilde{\mathcal{H}}_{0,\hat{\epsilon}_1}}(t|\mathscr{V}_{n_1 n_2}) - F^*_{\widetilde{\mathcal{H}}_{0,\epsilon_1}}(t)\right] + S_{n_1,1}(t), \quad (40)$$

where $F_{\widetilde{\mathcal{H}}_{0,\hat{\epsilon}_1}}(t|\mathscr{V}_{n_1 n_2})$ be the distribution function of $\hat{\epsilon}_{\widetilde{\mathcal{H}}_0,1} = \frac{Y - \hat{g}_2(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}, \hat{\boldsymbol{\gamma}}_0)}{\hat{\sigma}_1(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}, \hat{\boldsymbol{\beta}}_0)}$ conditional on the data $\mathscr{V}_{n_1 n_2} = \{\boldsymbol{X}_{1i}, Y_{1i}, \boldsymbol{X}_{2j}, Y_{2j}, 1 \le i \le n_1, 1 \le j \le n_2\}$, and similar to the analysis of Lemma 2, we have $\sup_{t \in \mathbb{R}} |S_{n_1,1}(t)| = o_P(n_1^{-1/2})$. Taylor expansion entails that

$$F_{\widetilde{\mathcal{H}}_{0,\hat{\epsilon}_1}}(t|\mathscr{V}_{n_1 n_2}) - F^*_{\widetilde{\mathcal{H}}_{0,\epsilon_1}}(t)$$
$$= \int F_{\epsilon_1}\Bigg(t + \frac{g_1(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1) - g_2(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1)}{\sigma_1(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1)} + [B_{n_1}(\boldsymbol{x}_1) - 1]t$$

$$
\begin{aligned}
&+\frac{\hat{g}_2(\hat{\boldsymbol{\beta}}_0^{\tau}\boldsymbol{x}_1,\hat{\boldsymbol{\gamma}}_0)-g_2(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x}_1)}{\sigma_1(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x}_1)}\Bigg)dF_{X_1}(\boldsymbol{x}_1)-F^{*}_{\widetilde{\mathcal{H}}_{0,\epsilon_1}}(t)\\
&=t\int f_{\epsilon_1}\left(t+\frac{g_1(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x}_1)-g_2(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x}_1)}{\sigma_1(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x}_1)}\right)[B_{n_1}(\boldsymbol{x}_1)-1]dF_{X_1}(\boldsymbol{x}_1)\\
&+\int f_{\epsilon_1}\left(t+\frac{g_1(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x}_1)-g_2(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x}_1)}{\sigma_1(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x}_1)}\right)\frac{\hat{g}_2(\hat{\boldsymbol{\beta}}_0^{\tau}\boldsymbol{x}_1,\hat{\boldsymbol{\gamma}}_0)-g_2(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x}_1)}{\sigma_1(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x}_1)}dF_{X_1}(\boldsymbol{x}_1)\\
&+R_{n_1n_2}(t). \qquad (41)
\end{aligned}
$$

Similar to the analysis of (21), we can have that

$$
\begin{aligned}
&\hat{g}_2(\hat{\boldsymbol{\beta}}_0^{\tau}\boldsymbol{x}_1,\boldsymbol{\gamma}_0)-\hat{g}_2(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x}_1,\boldsymbol{\gamma}_0)\\
&\quad=g_2'(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x}_1)\boldsymbol{x}_1^{\tau}\left(\hat{\boldsymbol{\beta}}_0-\boldsymbol{\beta}_0\right)+O_P\left(h_2^2+\sqrt{\frac{(\log n_2)^{1+s_0}}{n_2h_2^3}}\right). \qquad (42)
\end{aligned}
$$

Recall the definition of $\hat{g}_2(u,\boldsymbol{\gamma}_0)$ and using Lemma 1,

$$
\begin{aligned}
&\hat{g}_2(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x}_1,\boldsymbol{\gamma}_0)-g_2(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x}_1)\\
&\quad=\frac{1}{n_2 f_{\boldsymbol{\gamma}_0^{\tau}X_2}(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x}_1)}\sum_{i=1}^{n_2}K_{h_2}\left(\boldsymbol{\gamma}_0^{\tau}\boldsymbol{X}_{2i}-\boldsymbol{\beta}_0^{\tau}\boldsymbol{x}_1\right)\sigma_2(\boldsymbol{\gamma}_0^{\tau}\boldsymbol{X}_{2i})\epsilon_{2i}+O_P(c_{n_2}). \qquad (43)
\end{aligned}
$$

We can also show that $R_{n_1n_2}(t)$ defined in (41) is $o_P(n_1^{-1/2}+n_2^{-1/2})$ uniformly in $t\in\mathbb{R}$. Together with (39), (42) and (43), we have

$$
\begin{aligned}
&F_{\widetilde{\mathcal{H}}_{0,\hat{\epsilon}_1}}(t|\mathscr{V}_{n_1n_2})-F^{*}_{\widetilde{\mathcal{H}}_{0,\epsilon_1}}(t)\\
&=\frac{t}{2n_1}\sum_{i=1}^{n_1}(\epsilon_{1i}^2-1)f_{\epsilon_1}\left(t+\frac{g_1(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_{1i})-g_2(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_{1i})}{\sigma_1(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_{1i})}\right)\\
&+E\left[f_{\epsilon_1}\left(t+\frac{g_1(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_1)-g_2(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_1)}{\sigma_1(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_1)}\right)\frac{g_2'(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})}{\sigma_1(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})}V_{1,\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})\right]^{\tau}\left(\hat{\boldsymbol{\beta}}_0-\boldsymbol{\beta}_0\right)\\
&+\frac{1}{n_2}\sum_{i=1}^{n_2}f_{\epsilon_1}\left(t+\frac{g_1(\boldsymbol{\gamma}_0^{\tau}\boldsymbol{X}_{2i})-g_2(\boldsymbol{\gamma}_0^{\tau}\boldsymbol{X}_{2i})}{\sigma_1(\boldsymbol{\gamma}_0^{\tau}\boldsymbol{X}_{2i})}\right)\frac{f_{\boldsymbol{\beta}_0^{\tau}X_1}(\boldsymbol{\gamma}_0^{\tau}\boldsymbol{X}_{2i})}{f_{\boldsymbol{\gamma}_0^{\tau}X_2}(\boldsymbol{\gamma}_0^{\tau}\boldsymbol{X}_{2i})}\frac{\sigma_2(\boldsymbol{\gamma}_0^{\tau}\boldsymbol{X}_{2i})}{\sigma_1(\boldsymbol{\gamma}_0^{\tau}\boldsymbol{X}_{2i})}\epsilon_{2i}\\
&+o_P(n^{-1/2}).
\end{aligned}
$$

Recalling the definitions of $D_1(u)$ and $\rho_{f,\sigma}(u)$, we complete the proof of Theorem 5. Moreover, the proof of Theorem 6 is completed by following the asymptotic result of Theorem 5 and recalling that $D_1(u)\equiv D_2(u)\equiv 0$ under the null hypothesis, we omit the details. □

5.5 Proof of Theorem 7

Proof By using the detailed proof of Theorem 1 in Stute et al. (2008), the class of functions

$$\ell_t(\epsilon_1, \boldsymbol{x}_1) = I\left\{\epsilon_1 \le t - \frac{1}{\sqrt{n_1+n_2}}\frac{\mu(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1)}{\sigma_1(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1)}\right\} - I\left\{\epsilon_1 \le t\right\} \\ - F_{\epsilon_1}\left(t - \frac{1}{\sqrt{n_1+n_2}}\frac{\mu(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1)}{\sigma_1(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_1)}\right) + F_{\epsilon_1}(t)$$

is a Vapnik-Chervonenkis class with envelop function 4 (Pollard 1984, Ch. 2). Then, we can have that

$$n_1^{-1/2}\left|\sum_{i=1}^{n_1}\left[I\left\{\epsilon_{1i} \le t - \frac{1}{\sqrt{n_1+n_2}}\frac{\mu(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i})}{\sigma_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i})}\right\} - I\left\{\epsilon_{1i} \le t\right\}\right]\right. \\ \left. - E\left[F_{\epsilon_1}\left(t - \frac{1}{\sqrt{n_1+n_2}}\frac{\mu(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_1)}{\sigma_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_1)}\right)\right] + F_{\epsilon_1}(t)\right| = o_P\big(n_1^{-1/2}\big). \quad (44)$$

Moreover, Taylor expansion entails that

$$E\left[F_{\epsilon_1}\left(t - \frac{1}{\sqrt{n_1+n_2}}\frac{\mu(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_1)}{\sigma_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_1)}\right)\right] - F_{\epsilon_1}(t) \\ = -\frac{1}{\sqrt{n_1+n_2}} f_{\epsilon_1}(t) E\left[\frac{\mu(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_1)}{\sigma_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_1)}\right] + o\big(n_1^{-1/2} + n_2^{-1/2}\big). \quad (45)$$

If the local alternative hypothesis $\mathcal{H}_{1n_1n_2}$ is true, together with (44) and (45), we have

$$\hat{F}_{\widetilde{\mathcal{H}}_{0,\epsilon_1}}(t) - \hat{F}_{\epsilon_1}(t) = -\frac{1}{\sqrt{n_1+n_2}} f_{\epsilon_1}(t) E\left[\frac{\mu(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_1)}{\sigma_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_1)}\right] \\ + f_{\epsilon_1}(t)\mathcal{N}_1^\tau J_{\boldsymbol{\beta}_0}\boldsymbol{\Omega}_1^{-1}\frac{1}{n_1}\sum_{i=1}^{n_1} J_{\boldsymbol{\beta}_0}^\tau \frac{g_1'(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i})}{\sigma_1(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i})}\left[\boldsymbol{X}_{1i} - V_{1,\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_{1i})\right]\epsilon_{1i} \\ + f_{\epsilon_1}(t)\frac{1}{n_2}\sum_{i=1}^{n_2}\rho_{f,\sigma}(Y\boldsymbol{\gamma}_0^\tau \boldsymbol{X}_{2i})\epsilon_{2i} + f_{\epsilon_1}(t)\frac{1}{n_1}\sum_{i=1}^{n_1}\epsilon_{1i} \\ + o_P(n_1^{-1/2} + n_2^{-1/2}).$$

We can also obtain a similar expression for $\hat{F}_{\widetilde{\mathcal{H}}_{0,\epsilon_2}}(t) - \hat{F}_{\epsilon_2}(t)$ and we omit the details. Using the continuous mapping theorem, we complete the proof of Theorem 7. □

References

Akritas, M. G., Van Keilegom, I. (2001). Non-parametric estimation of the residual distribution. *Scandinavian Journal of Statistics*, *28*(3), 549–567.

Carroll, R. J., Fan, J., Gijbels, I., Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, *92*(438), 477–489.

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

Dette, H., Neumeyer, N., Keilegom, I. V. (2007). A new test for the parametric form of the variance function in non-parametric regression. *Journal of the Royal Statistical Society*: *Series B (Statistical Methodology)*, *69*(5), 903–917.

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

Feng, L., Zou, C., Wang, Z., Zhu, L. (2015). Robust comparison of regression curves. *TEST*, *24*(1), 185–204.

Feng, Z., Zhu, L. (2012). An alternating determination-optimization approach for an additive multi-index model. *Computational Statistics & Data Analysis*, *56*(6), 1981–1993.

Feng, Z., Wen, X. M., Yu, Z., Zhu, L. (2013). On partial sufficient dimension reduction with applications to partially linear multi-index models. *Journal of the American Statistical Association*, *108*(501), 237–246.

Gørgens, T. (2002). Nonparametric comparison of regression curves by local linear fitting. *Statistics & Probability Letters*, *60*(1), 81–89.

Härdle, W., Hall, P., Ichimura, H. (1993). Optimal smoothing in single-index models. *The Annals of Statistics*, *21*, 157–178.

Koul, H. L., Schick, A. (1997). Testing for the equality of two nonparametric regression curves. *Journal of Statistical Planning and Inference*, *65*(2), 293–314.

Kulasekera, K. (1995). Comparison of regression curves using quasi-residuals. *Journal of the American Statistical Association*, *90*(431), 1085–1093.

Kulasekera, K., Wang, J. (1997). Smoothing parameter selection for power optimality in testing of regression curves. *Journal of the American Statistical Association*, *92*(438), 500–511.

Li, G., Zhu, L., Xue, L., Feng, S. (2010). Empirical likelihood inference in partially linear single-index models for longitudinal data. *Journal of Multivariate Analysis*, *101*(5), 718–732.

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.

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, W., Kulasekera, K. (2010). Testing the equality of linear single-index models. *Journal of Multivariate Analysis*, *101*(5), 1156–1167.

Neumeyer, N. (2009). Smooth residual bootstrap for empirical processes of nonparametric regression residuals. *Scandinavian Journal of Statistics*, *36*(2), 204–228.

Neumeyer, N., Dette, H. (2003). Nonparametric comparison of regression curves: An empirical process approach. *The Annals of Statistics*, *31*(3), 880–920.

Neumeyer, N., Van Keilegom, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *Journal of Multivariate Analysis*, *101*(5), 1067–1078.

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

Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer.

Stute, W., Zhu, L. X. (2005). Nonparametric checks for single-index models. *The Annals of Statistics*, *33*, 1048–1083.

Stute, W., Xu, W. L., Zhu, L. X. (2008). Model diagnosis for parametric regression in high-dimensional spaces. *Biometrika*, *95*(2), 451–467.

van der Vaart, A.W., Wellner, J.A. (1996). *Weak convergence and empirical processes*, *Springer series in statistics*. New York: Springer. (**with applications to statistics**).

Van Keilegom, I., Manteiga, W. G., Sellero, C. S. (2008). Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *TEST*, *17*(2), 401–415.

Wang, J. L., Xue, L., Zhu, L., Chong, Y. S. (2010). Estimation for a partial-linear single-index model. *The Annals of Statistics*, *38*, 246–274.

Wang, T., Xu, P., Zhu, L. (2015). Variable selection and estimation for semi-parametric multiple-index models. *Bernoulli*, *21*(1), 242–275.

Xia, Y. (2006). Asymptotic distributions for two estimators of the single-index model. *Econometric Theory*, *22*, 1112–1137.

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 (Statistical Methodology)*, *64*, 363–410.

Xu, P., Zhu, L. (2012). Estimation for a marginal generalized single-index longitudinal model. *Journal of Multivariate Analysis*, *105*, 285–299.
Yu, Y., Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association*, *97*, 1042–1054.
Zhang, C., Peng, H., Zhang, J. T. (2010). Two samples tests for functional data. *Communications in Statistics—Theory and Methods*, *39*(4), 559–578.
Zhang, J., Wang, X., Yu, Y., Gai, Y. (2014). Estimation and variable selection in partial linear single index models with error-prone linear covariates. *Statistics*, *48*, 1048–1070.