Brazilian Journal of Probability and Statistics
2019, Vol. 33, No. 1, 57–86
https://doi.org/10.1214/17-BJPS377

Exploring the constant coefficient of a single-index variation

Jun Zhang[a], Cuizhen Niu[b] and Gaorong Li[c]

[a]*Shenzhen University*
[b]*Beijing Normal University*
[c]*Beijing University of Technology*

Abstract. We consider a problem of checking whether the coefficient of the scale and location function is a constant. Both the scale and location functions are modeled as single-index models. Two test statistics based on Kolmogorov–Smirnov and Cramér–von Mises type functionals of the difference of the empirical residual processes are proposed. The asymptotic distribution of the estimator for single-index parameter is derived, and the empirical distribution function of residuals is shown to converge to a Gaussian process. Moreover, the proposed test statistics can be able to detect local alternatives that converge to zero at a parametric convergence rate. A bootstrap procedure is further proposed to calculate critical values. Simulation studies and a real data analysis are conducted to demonstrate the performance of the proposed methods.

1 Introduction

Single-index models, a generalization of multivariate linear regression models with an unknown link function, have been paid great attention because they gain more flexibility and relax restrictive assumptions imposed on parametric models of conditional mean functions. There have been many papers to consider the consistency estimation of the single-index parameter and the nonparametric link function. See, for example, Wang, Xu and Zhu (2012), Xia et al. (2002), Ichimura (1993), Wang and Zhu (2015), Feng et al. (2013), Guo, Wang and Zhu (2016), Härdle, Hall and Ichimura (1993), Peng and Huang (2011), Li et al. (2014), Wang, Xu and Zhu (2015). In this paper, we consider a single-index heteroscedasticity regression model:

$$Y = g(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}) + \sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})\epsilon, \tag{1.1}$$

where "τ" denotes the transpose operator on a vector or a matrix throughout this paper. In model (1.1), Y is the response variable and $\boldsymbol{X}$ is a p-dimensional covariate vector. $g(\cdot)$ and $\sigma(\cdot)$ are two unknown univariate smooth functions, and throughout this paper we assume that the function $\sigma(\cdot)$ in the model (1.1) is positive. The error term ϵ satisfies $E(\epsilon) = 0$ and $E(\epsilon^2) = 1$. The last condition for ϵ is assumed for identifiability of the model. The parameter $\boldsymbol{\beta}_0$ is an unknown

Key words and phrases. Empirical residual process, single-index models, local linear smoothing.
Received April 2017; accepted September 2017.

index vector which belongs to the parameter space $\mathcal{B} = \{\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^\tau, \|\boldsymbol{\beta}\| = 1, \beta_1 > 0, \boldsymbol{\beta} \in \mathbb{R}^p\}$.

Amongst the various methods of estimation, we are interested in testing whether a constant coefficient of variation exists in a dataset, that is,

$$\mathcal{H}_0 : g(\cdot) = c\sigma(\cdot), \tag{1.2}$$

for some nonzero constant c. Carroll and Ruppert (1988) investigated a parametric model with a constant coefficient of variation as a special case of the hypothesis (1.2). Eagleson and Müller (1997) considered the problem of nonparametric estimation of the mean regression function where the standard deviation function is proportional to the mean regression function. Model (1.1) generalizes the models considered by Dette, Marchlewski and Wagener (2012), Dette and Wieczorek (2009) who focused on the one dimensional case of $\boldsymbol{X}$. Dette and Wieczorek (2009) proposed an estimate of the L_2-distance between the variance and squared regression function. As claimed in Dette, Marchlewski and Wagener (2012), the test statistic proposed by Dette and Wieczorek (2009) is not able to detect alternatives converging to the null hypothesis at the rate of $n^{-1/2}$, while the empirical process statistic proposed by Dette, Marchlewski and Wagener (2012) succeeds. Moreover, the Kolmogorov–Smirnov and Cramér–von Mises statistics by using the empirical process constructed from the residuals-based empirical distribution functions to test model assumptions also have the the advantage of detecting local alternatives that converge to zero at the rate of $n^{-1/2}$, independent of the dimension of $\boldsymbol{X}$ (Neumeyer and Van Keilegom (2010)). However, the estimation and test procedures for $g(\cdot)$ and $\sigma(\cdot)$ proposed in Dette and Wieczorek (2009), Dette, Marchlewski and Wagener (2012), Dette, Pardo-Fernández and Van Keilegom (2009) can not be directly extended to the case of multivariate $\boldsymbol{X}$ due to the "curse of dimensionality". This motivates us to investigate the estimation method for model (1.1) and test procedure for the constant coefficient of variation hypothesis (1.2) with a single-index structure.

In this article, the first goal is to estimate the unknown single-index parameter $\boldsymbol{\beta}_0$ and unknown $g(\cdot)$ and $\sigma(\cdot)$. The profile estimation equation is proposed to estimate $\boldsymbol{\beta}_0$, and the large sample properties of the estimator is obtained. The second goal is to check whether the constant coefficient of variation hypothesis (1.2) is true or not. Under the null hypothesis (1.2), the estimator of distribution function $F_\epsilon(s)$ of model error ϵ is obtained from the residuals based on the error $\epsilon = \frac{Y}{\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})} - \frac{E[Y]}{E[\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})]}$. At the same time, the estimator of $F_\epsilon(s)$ under the full model (1.1) can be also obtained by using the residuals based on $\frac{Y - g(\boldsymbol{\beta}_0^\tau \boldsymbol{X})}{\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})}$. Then, we obtain the asymptotic expressions for the estimators of $F_\epsilon(s)$ under the full model (1.1) and under the null hypothesis (1.2). Two test statistics, namely, Kolmogorov–Smirnov test statistic and Cramér–von Mises test statistic are used to check whether the null hypothesis (1.2) is true or not. The limiting distributions of these two test statistics are also derived. To mimic the null distributions of the test statistics, a bootstrap procedure is proposed to define p-values. We conduct Monte Carlo simulation experiments to examine the performance of the proposed procedures. Our simulation results show that the proposed methods perform well both in estimation and hypothesis testing.

This paper is organized as follows. In Section 2, we propose the estimation procedure for $\boldsymbol{\beta}_0$, $g(\cdot)$ and $\sigma(\cdot)$. In Section 3, we provide the estimators of error distribution function, and two test statistics for the testing problem. A bootstrap procedure is also proposed to mimic the null distribution of test statistics. In Section 4, we report the results of simulation studies. In Section 5, a real data is analyzed as an illustration. All the technical proofs of the asymptotic results are given in Appendix.

2 Estimation method for $\boldsymbol{\beta}_0$, $g(\cdot)$ and $\sigma(\cdot)$

Suppose that we have an *i.i.d.* sample $\{\boldsymbol{X}_i, Y_i\}_{i=1}^n$, $\boldsymbol{X}_i = (X_{1i}, \dots, X_{pi})^\tau$ from the model (1.1). We now employ the profile least squares estimation procedure proposed in Cui, Härdle and Zhu (2011), Liang et al. (2010), Liang and Wang (2005).

(1) A local linear smoothing technique is used to estimate $g(\cdot)$. We approximate $g(u)$ by $g(u_*) + g'(u_*)(u - u_*)$ in a neighborhood of u_*. Given $\boldsymbol{\beta}$, the local linear estimators of $g(u)$ and its derivative $g'(u)$ are obtained by minimizing (2.1) with respect to a_0 and a_1,

$$\sum_{i=1}^{n}\{Y_i - a_0 - a_1(\boldsymbol{\beta}^\tau \boldsymbol{X}_i - u)\}^2 K_h(\boldsymbol{\beta}^\tau \boldsymbol{X}_i - u), \tag{2.1}$$

where $K_h(\boldsymbol{\beta}^\tau \boldsymbol{X}_i - u) = h^{-1}K(\frac{\boldsymbol{\beta}^\tau \boldsymbol{X}_i - u}{h})$ with $K(\cdot)$ being a kernel function and h being a bandwidth. Let $(\hat{a}_0, \hat{a}_1)$ be the minimizer of (2.1), denoted as $(\hat{g}(u,\boldsymbol{\beta}), \hat{g}'(u,\boldsymbol{\beta}))$. Then, the estimator of $g(u)$ is obtained as

$$\hat{g}(u,\boldsymbol{\beta}) = \hat{a}_0 = \frac{T_{n,20}(u,\boldsymbol{\beta})T_{n,01}(u,\boldsymbol{\beta}) - T_{n,10}(u,\boldsymbol{\beta})T_{n,11}(u,\boldsymbol{\beta})}{T_{n,00}(u,\boldsymbol{\beta})T_{n,20}(u,\boldsymbol{\beta}) - T_{n,10}^2(u,\boldsymbol{\beta})}, \tag{2.2}$$

where $T_{n,l_1l_2}(u,\boldsymbol{\beta}) = \sum_{i=1}^n K_h(\boldsymbol{\beta}^\tau \boldsymbol{X}_i - u)(\boldsymbol{\beta}^\tau \boldsymbol{X}_i - u)^{l_1} Y_i^{l_2}$ for $l_1 = 0, 1, 2$, $l_2 = 0, 1$.

(2) The local linear smoothing technique is used to estimate the variance function $\sigma^2(\cdot)$. Similar to (2.1) and (2.2), we estimate $\sigma^2(u)$ as:

$$\hat{\sigma}^2(u,\boldsymbol{\beta}) = \frac{S_{n,20}(u,\boldsymbol{\beta})S_{n,01}(u,\boldsymbol{\beta}) - S_{n,10}(u,\boldsymbol{\beta})S_{n,11}(u,\boldsymbol{\beta})}{S_{n,00}(u,\boldsymbol{\beta})S_{n,20}(u,\boldsymbol{\beta}) - S_{n,10}^2(u,\boldsymbol{\beta})}, \tag{2.3}$$

$S_{n,l_1l_2}(u,\boldsymbol{\beta}) = \sum_{i=1}^n K_h(\boldsymbol{\beta}^\tau \boldsymbol{X}_i - u)(\boldsymbol{\beta}^\tau \boldsymbol{X}_i - u)^{l_1}[(Y_i - \hat{g}(\boldsymbol{\beta}^\tau \boldsymbol{X}_i, \boldsymbol{\beta}))^2]^{l_2}$ for $l_1 = 0, 1, 2$, $l_2 = 0, 1$.

(3) A noted in Zhu et al. (2010), the restriction of $\|\boldsymbol{\beta}\| = 1$ leads to a non-differential problem at the point $\boldsymbol{\beta}$ lying on the boundary of a unit ball. To solve it, we transform the boundary of a unit ball in $\mathbb{R}^p$ to the interior of a unit ball in $\mathbb{R}^{p-1}$. We now proceed to estimate $\boldsymbol{\beta}_0$ by using the profile estimation function (Liang et al. (2010), Liang and Wang (2005)) and the "leave-one-component" procedure (Cui, Härdle and Zhu (2011), Zhu et al. (2010)),

$$\begin{aligned} \mathcal{W}_n(\boldsymbol{\beta}^{(1)}) \stackrel{\text{def}}{=} \sum_{i=1}^{n} & J_{\boldsymbol{\beta}}^{\tau} \hat{g}'(\boldsymbol{\beta}^{\tau}\boldsymbol{X}_i, \boldsymbol{\beta})[\boldsymbol{X}_i - \hat{V}(\boldsymbol{\beta}^{\tau}\boldsymbol{X}_i)]\hat{\sigma}^{-2}(\boldsymbol{\beta}^{\tau}\boldsymbol{X}_i, \boldsymbol{\beta}) \\ & \times [Y_i - \hat{g}(\boldsymbol{\beta}^{\tau}\boldsymbol{X}_i, \boldsymbol{\beta})], \end{aligned} \tag{2.4}$$

in which, $\hat{g}'(u, \boldsymbol{\beta}) = \frac{\partial \hat{g}(u,\boldsymbol{\beta})}{\partial u}$, $J_{\boldsymbol{\beta}} = \partial\boldsymbol{\beta}/\partial\boldsymbol{\beta}^{(1)}$ is the Jacobian matrix of size $d \times (p-1)$ with

$$J_{\boldsymbol{\beta}} = \begin{pmatrix} -\boldsymbol{\beta}^{(1)\tau}/\sqrt{1 - \|\boldsymbol{\beta}^{(1)}\|^2} \\ I_{p-1} \end{pmatrix}, \tag{2.5}$$

where $I_{p-1} = \operatorname{diag}(1, \dots, 1)$, an identity matrix of size $p-1$. Moreover, $\hat{V}(t)$ is the local linear estimator of $V(u) = E(\boldsymbol{X}|\boldsymbol{\beta}^{\tau}\boldsymbol{X} = u) = (V_1(u), \dots, V_p(u))^{\tau}$, defined as $\hat{V}(u) = \frac{\sum_{i=1}^{n} b_{n,i}(u)\boldsymbol{X}_i}{\sum_{i=1}^{n} b_{n,i}(u)}$, where $b_{n,i}(u) = K_h(\boldsymbol{\beta}^{\tau}\boldsymbol{X}_i - u)[T_{n,20}(u, \boldsymbol{\beta}) - (\boldsymbol{\beta}^{\tau}\boldsymbol{X}_i - u)T_{n,10}(u, \boldsymbol{\beta})]$. To solve (2.4), an consistent initial estimate of $\boldsymbol{\beta}_0^{(1)}$ will speed up to obtain its final estimator. We suggest to use more stable, more robust and widely used dimension reduction methods for this initial estimator in practice. For example, Xia et al. (2002), Xia and Härdle (2006).

Denote that $\hat{\boldsymbol{\beta}}_0^{(1)}$ is the solution of the estimation equation $\mathcal{W}_n(\hat{\boldsymbol{\beta}}_0^{(1)}) = \mathbf{0}_{p-1}$. Then, we apply the equation $\boldsymbol{\beta}_{0,1} = \sqrt{1 - \|\boldsymbol{\beta}_0^{(1)}\|^2}$ to estimate $\boldsymbol{\beta}_{0,1}$ by $\hat{\boldsymbol{\beta}}_{0,1} = \sqrt{1 - \|\hat{\boldsymbol{\beta}}_0^{(1)}\|^2}$, and the estimator of $\boldsymbol{\beta}_0$ is obtained as $\hat{\boldsymbol{\beta}}_0 = (\hat{\boldsymbol{\beta}}_{0,1}, \hat{\boldsymbol{\beta}}_0^{(1)\tau})^{\tau}$. Finally, the estimators of $g(u)$ and $\sigma^2(u)$ are obtained by substituting $\boldsymbol{\beta}$ with $\hat{\boldsymbol{\beta}}_0$ in (2.1) and (2.3), respectively.

It is noted that the model (1.1) is different from the generalized single-index proposed in Cui, Härdle and Zhu (2011). Cui, Härdle and Zhu (2011) assumed that $E(Y|\boldsymbol{X}) = \mu(g(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}))$ and $\operatorname{Var}(Y|\boldsymbol{X}) = V(g(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}))\sigma^2$, where $\mu(\cdot)$ is a known monotonic function, $V(\cdot)$ is a known covariance function. Cui, Härdle and Zhu (2011) assumed that the variance function $\operatorname{Var}(Y|\boldsymbol{X})$ is linked with the mean function $E(Y|\boldsymbol{X})$. Our model (1.1) is different with Cui, Härdle and Zhu (2011) because we assume that the variance function $\operatorname{Var}(Y|\boldsymbol{X}) = \sigma^2(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})$ does not need to involve the mean function $E(Y|\boldsymbol{X}) = g(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})$. Next, when the null hypothesis $\mathcal{H}_0$ holds, the model (1.1) becomes to $Y = c\sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}) + \sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})\epsilon$, and also $E(Y|\boldsymbol{X}) = c\sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})$, $\operatorname{Var}(Y|\boldsymbol{X}) = \sigma^2(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})$. Thus, our model (1.1) is a special one proposed in Cui, Härdle and Zhu (2011) under the null hypothesis $\mathcal{H}_0$. However, if the null hypothesis $\mathcal{H}_0$ fails, our model (1.1) is different from Cui, Härdle and Zhu (2011). It is also noted that the profile estimation function (2.5) is different from EFM method proposed in Cui, Härdle and Zhu (2011). It is seen that if the null hypothesis $\mathcal{H}_0$ holds, we can use both the estimation equation (2.4) and also the EFM approach proposed in Cui, Härdle and Zhu (2011) to estimate the single-index parameter $\boldsymbol{\beta}_0$.

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

(C1) $E[|X_r|^3] < \infty$ for $r = 1, \dots, p$, and the covariance matrix $\boldsymbol{\Omega}_0$ defined in Theorem 2.1 is a positive definite matrix and finite.

(C2) The functions $g(u)$, $\sigma(u)$, $V(u) = E(\boldsymbol{X}|\boldsymbol{\beta}^\tau \boldsymbol{X} = u)$ and the density function $f_{\boldsymbol{\beta}^\tau \boldsymbol{X}}(u)$ of the random variable $\boldsymbol{\beta}^\tau \boldsymbol{X}$ are twice continuously differentiable with respect u. Their second derivatives are uniformly Lipschitz continuous on $\mathcal{C} = \{u = \boldsymbol{\beta}^\tau x : x \in \mathcal{X} \subset \mathbb{R}^p, \boldsymbol{\beta} \in \mathcal{B}\}$, where $\mathcal{X}$ is a compact support set. Furthermore, $\inf_{u \in \mathcal{C}} f_{\boldsymbol{\beta}^\tau \boldsymbol{X}}(u) \ge c_0 > 0$, $\inf_{u \in \mathcal{C}} \sigma(u) \ge c_0 > 0$ for some positive constant c_0, and $\int \sigma^2(u) f_{\boldsymbol{\beta}^\tau \boldsymbol{X}}(u)\, du < \infty$.

(C3) The kernel function $K(\cdot)$ is a symmetric bounded density function supported on $[-A, A]$ and satisfies a Lipschitz condition. Moreover, the kernel function $K(\cdot)$ has twice continuous bounded derivatives, satisfying $K^{(j)}(\pm A) = 0$ for $j = 0, 1$ and $\int s^2 K(s)\, ds \ne 0$.

(C4) As $n \to \infty$, the bandwidth h satisfies $nh^4 \to 0$ and $\frac{(\log n)^{1+s}}{nh^2} \to 0$ for some $s > 0$.

(C5) The model error ϵ satisfies $E[\epsilon^4] < \infty$. The distribution function $F_\epsilon(s)$ of ϵ is twice continuously differentiable, and the density function $f_\epsilon(s)$ of ϵ satisfies $\int f_\epsilon^2(s)\, dF_\epsilon(s) < \infty$, $\sup_{-\infty<s<\infty} f_\epsilon(s) < \infty$, $\sup_{-\infty<s<\infty} |s| f_\epsilon(s) < \infty$ and $\sup_{-\infty<s<\infty} s^2 |f'_\epsilon(s)| < \infty$.

We now present the asymptotic properties of $\hat{\boldsymbol{\beta}}_0^{(1)}$ and $\hat{\boldsymbol{\beta}}_0$.

Theorem 2.1. *Under the conditions* (C1)–(C4), *we have*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_0^{(1)} - \boldsymbol{\beta}_0^{(1)}) \xrightarrow{L} N(\mathbf{0}_{p-1}, \boldsymbol{\Omega}_0^{-1}),$$

where $\boldsymbol{\Omega}_0 = J_{\boldsymbol{\beta}_0}^\tau E[g'^2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}) \sigma^{-2}(\boldsymbol{\beta}_0^\tau \boldsymbol{X})[\boldsymbol{X} - V(\boldsymbol{\beta}_0^\tau \boldsymbol{X})]^{\otimes 2}] J_{\boldsymbol{\beta}_0}$. *Moreover, by a simple application of the multivariate delta-method, we also have*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0) \xrightarrow{L} N(\mathbf{0}_p, J_{\boldsymbol{\beta}_0} \boldsymbol{\Omega}_0^{-1} J_{\boldsymbol{\beta}_0}^\tau).$$

Remark 1. A population version of (2.4) when $\boldsymbol{\beta}^{(1)} = \boldsymbol{\beta}_0^{(1)}$ is defined as

$$\mathcal{W}_n^*(\boldsymbol{\beta}_0^{(1)}) = \sum_{i=1}^n J_{\boldsymbol{\beta}_0}^\tau g'(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)[\boldsymbol{X}_i - V(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)] \sigma^{-2}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)[Y_i - g(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)].$$

The function $\mathcal{W}_n^*(\boldsymbol{\beta}_0^{(1)})$ entails the second Bartlett identity as Cui, Härdle and Zhu (2011) claimed, that is,

$$E\big[\mathcal{W}_n^*(\boldsymbol{\beta}_0^{(1)})\mathcal{W}_n^{*\tau}(\boldsymbol{\beta}_0^{(1)})\big] = -E\left[\frac{\partial \mathcal{W}_n^*(\boldsymbol{\beta}_0^{(1)})}{\partial \boldsymbol{\beta}_0^{(1)}}\right] = \boldsymbol{\Omega}_0. \tag{2.6}$$

The second Bartlett identity (2.6) makes the estimator $\hat{\boldsymbol{\beta}}_0^{(1)}$ obtained from (2.4) is possible semiparametric efficiency (Cui, Härdle and Zhu (2011)).

3 The test statistics and their asymptotic properties

The idea for testing the hypothesis (1.2) is to compare the estimated error distribution $\hat{F}_\epsilon(s)$ obtained under the full model (1.1) with the estimated error distribution function $\hat{F}_{0\epsilon}(s)$ obtained under the null hypothesis (1.2). That is, we adopt Kolmogorov–Smirnov or Cramer-von Mises test statistics based on the difference between $\hat{F}_\epsilon(s)$ and $\hat{F}_{0\epsilon}(s)$ by using the process

$$\sqrt{n}\big(\hat{F}_{0\epsilon}(s) - \hat{F}_\epsilon(s)\big). \tag{3.1}$$

In the following, we introduce the estimators $\hat{F}_\epsilon(s)$, $\hat{F}_{0\epsilon}(s)$ and the test statistics, and present the associated theoretical results.

3.1 Test statistics

After obtaining $\hat{\boldsymbol{\beta}}_0$, we define the estimator of the error distribution $F_\epsilon(s)$ under the full model (1.1) as

$$\hat{F}_\epsilon(s) = \frac{1}{n}\sum_{i=1}^n I\{\hat{\epsilon}_i \le s\}, \qquad \text{where } \hat{\epsilon}_i = \frac{Y_i - \hat{g}(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}_i, \hat{\boldsymbol{\beta}}_0)}{\hat{\sigma}(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}_i, \hat{\boldsymbol{\beta}}_0)}, \tag{3.2}$$

where, $\hat{g}(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}_i, \hat{\boldsymbol{\beta}}_0)$, $\hat{\sigma}(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}_i, \hat{\boldsymbol{\beta}}_0)$ are obtained from (2.2) and (2.3) respectively.

If the null hypothesis $\mathcal{H}_0$ is true, it is easily seen that $E[Y] = cE[\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})]$, $c = \frac{E[Y]}{E[\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})]}$, and $\epsilon_i = \frac{Y_i}{\sigma(\boldsymbol{\beta}^\tau \boldsymbol{X}_i)} - \frac{E[Y]}{E[\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})]}$. This motivates us to estimate the error distribution $F_\epsilon(s)$ as

$$\hat{F}_{0\epsilon}(s) = \frac{1}{n}\sum_{i=1}^n I\{\hat{\epsilon}_{0i} \le s\}, \tag{3.3}$$

$$\hat{\epsilon}_{0i} = \frac{Y_i}{\hat{\sigma}(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}_i, \hat{\boldsymbol{\beta}}_0)} - \frac{\bar{Y}}{\frac{1}{n}\sum_{i=1}^n \hat{\sigma}(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}_i, \hat{\boldsymbol{\beta}}_0)}, \tag{3.4}$$

and $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$. Then the test statistics based on Kolmogorov–Smirnov and Cramér–von Mises type functionals of the process $\sqrt{n}(\hat{F}_{0\epsilon}(s) - \hat{F}_\epsilon(s))$ are defined as

$$\mathfrak{T}_{n,\mathrm{KS}} = \sup_{-\infty<s<+\infty} n^{1/2}\big|\hat{F}_{0\epsilon}(s) - \hat{F}_{\epsilon}(s)\big|,$$

$$\mathfrak{T}_{n,\mathrm{CM}} = n\int \big(\hat{F}_{0\epsilon}(s) - \hat{F}_{\epsilon}(s)\big)^2\, d\hat{F}_{\epsilon}(s).$$

3.2 Theoretical results

We now present some asymptotic properties of the proposed estimators and test statistics. In this subsection, Theorems 3.1–3.2 present the asymptotic normalities of $\hat{F}_{\epsilon}(s)$ and $\hat{F}_{0\epsilon}(s)$, respectively. Theorem 3.3 is the asymptotic result for the difference of process $\hat{F}_{\epsilon}(s) - \hat{F}_{0\epsilon}(s)$ and the test statistics $\mathfrak{T}_{n,\mathrm{KS}}$ and $\mathfrak{T}_{n,\mathrm{CM}}$. Theorems 3.4–3.5 reveal the properties when the local alternatives converge to zero at an $n^{-1/2}$ rate.

Theorem 3.1. *Assumed that the conditions of Theorem* 2.1 *and condition* (C5) *are satisfied*, *we have*

$$\hat{F}_{\epsilon}(s) - F_{\epsilon}(s) = \frac{1}{n}\sum_{i=1}^{n}\bigg[I\{\epsilon_i \le s\} - F_{\epsilon}(s) + f_{\epsilon}(s)\bigg(\epsilon_i + \frac{s}{2}(\epsilon_i^2 - 1)\bigg)\bigg] + o_P\big(n^{-1/2}\big),$$

uniformly in $s \in \mathbb{R}^1$.

Remark 2. The process $\sqrt{n}(\hat{F}_{\epsilon}(s) - F_{\epsilon}(s))$ $(-\infty < s < \infty)$ converges weakly to a zero-mean Gaussian process $\mathcal{N}(s)$ with covariance function

$$\begin{aligned} &F_{\epsilon}\big(\min\{s_1, s_2\}\big) - F_{\epsilon}(s_1)F_{\epsilon}(s_2) + f_{\epsilon}(s_1)f_{\epsilon}(s_2) \\ &\quad + f_{\epsilon}(s_1)E\big[\epsilon I\{\epsilon \le s_2\}\big] + f_{\epsilon}(s_2)E\big[\epsilon I\{\epsilon \le s_1\}\big] \\ &\quad + \frac{1}{2}\big\{s_1 f_{\epsilon}(s_1)E\big[\epsilon^2 I\{\epsilon \le s_2\}\big] + s_2 f_{\epsilon}(s_2)E\big[\epsilon^2 I\{\epsilon \le s_1\}\big]\big\} \\ &\quad + \frac{1}{4} f_{\epsilon}(s_1)f_{\epsilon}(s_2)\big\{2(s_1+s_2)E\big[\epsilon^3\big] + s_1 s_2 E\big[\epsilon^4\big] - s_1 s_2\big\} \\ &\quad - \frac{1}{2}\big\{s_1 f_{\epsilon}(s_1)F_{\epsilon}(s_2) + s_2 f_{\epsilon}(s_2)F_{\epsilon}(s_1)\big\}. \end{aligned}$$

Next, we present the asymptotic expansion for the estimator $\hat{F}_{0\epsilon}(s)$. Define $m_c = \frac{E[g(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})]}{E[\sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})]}$ and

$$F_{\epsilon}^{*}(s) = E\bigg[F_{\epsilon}\bigg(s + m_c - \frac{g(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})}{\sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})}\bigg)\bigg],$$

$$m_{f_{\epsilon},c}(s) = E\bigg[f_{\epsilon}\bigg(s + m_c - \frac{g(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})}{\sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})}\bigg)\bigg].$$

Theorem 3.2. *Under the conditions of Theorem 3.1, we have*

$$
\begin{aligned}
\hat{F}_{0\epsilon}(s) - F_\epsilon^*(s) = {} & \frac{1}{n}\sum_{i=1}^{n} I\left\{\epsilon_i + \frac{g(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)}{\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)} - m_c \le s\right\} - F_\epsilon^*(s) \\
& + m_{f_\epsilon,c}(s)\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i - m_c\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)}{E[\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})]} \\
& + \frac{1}{n}\sum_{i=1}^{n}\frac{s+m_c}{2} f_\epsilon\left(s + m_c - \frac{g(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)}{\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)}\right)(\epsilon_i^2 - 1) \\
& - \frac{1}{n}\sum_{i=1}^{n}\frac{m_{f_\epsilon,c}(s) m_c}{2E[\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})]}\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)(\epsilon_i^2 - 1) + o_P(n^{-1/2})
\end{aligned}
\tag{3.5}
$$

uniformly in $s \in \mathbb{R}^1$.

Under $\mathcal{H}_0$, $\frac{g(\boldsymbol{\beta}_0^\tau \boldsymbol{x})}{\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{x})} = c$, $m_c = c$, then $f_\epsilon(s + m_c - \frac{g(\boldsymbol{\beta}_0^\tau \boldsymbol{x})}{\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{x})}) = f_\epsilon(s)$, $F_\epsilon^*(s) = F_\epsilon(s)$, $m_{f_\epsilon,c}(s) = f_\epsilon(s)$. From Theorem 3.1 and Theorem 3.2, we have the following theorem.

Theorem 3.3. *Under the conditions of Theorem 3.1, if the null hypothesis* $\mathcal{H}_0$ *holds, we have*

$$
\hat{F}_{0\epsilon}(s) - \hat{F}_\epsilon(s) = \frac{f_\epsilon(s)}{n}\sum_{i=1}^{n}\left(\frac{\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)}{E[\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})]} - 1\right)\left[\epsilon_i - \frac{c}{2}(\epsilon_i^2 - 1)\right] + o_P(n^{-1/2}).
$$

Moreover, $n^{1/2}(\hat{F}_{0\epsilon}(s) - \hat{F}_\epsilon(s))$, $s \in \mathbb{R}^1$, *converges weakly to a zero-mean Gaussian process* $f_\epsilon(s)\mathfrak{N}$, *where* $\mathfrak{N}$ *is a zero-mean normal random variable with the variance*

$$
\mathrm{Var}(\mathfrak{N}) = \left(1 + \frac{c^2}{4}(E[\epsilon^4] - 1) - cE[\epsilon^3]\right)\frac{\mathrm{Var}(\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X}))}{(E[\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})])^2}.
$$

The continuous mapping theorem further entails that

$$
\mathfrak{T}_{n,\mathrm{KS}} \xrightarrow{L} \sup_{-\infty < s < +\infty} f_\epsilon(s)|\mathfrak{N}|, \qquad \mathfrak{T}_{n,\mathrm{CM}} \xrightarrow{L} \int f_\epsilon^2(s)\, dF_\epsilon(s)\mathfrak{N}^2.
$$

We now investigate the asymptotic properties of the test statistics by considering the local alternative hypothesis:

$$
\mathcal{H}_{1n} : g(\boldsymbol{\beta}_0^\tau \boldsymbol{x}) = c\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{x}) + n^{-1/2}\gamma(\boldsymbol{x}),
\tag{3.6}
$$

where $\gamma(\cdot) \not\equiv 0$. Under $\mathcal{H}_{1n}$, the estimators $\hat{\boldsymbol{\beta}}_0^{(1)}$ and $\hat{\boldsymbol{\beta}}_0$ are still $\sqrt{n}$-consistent.

Theorem 3.4. *Under the conditions of Theorem 3.1, if the local alternative hypothesis $\mathcal{H}_{1n}$ holds, we have*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_0^{(1)} - \boldsymbol{\beta}_0^{(1)}) \xrightarrow{L} N(\boldsymbol{\Omega}_{c,0}^{-1}\boldsymbol{B}_{c,0}, \boldsymbol{\Omega}_{c,0}^{-1}),$$

where,

$$\boldsymbol{B}_{c,0} = cJ_{\boldsymbol{\beta}_0}^{\tau} E\Big[\frac{\sigma'(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})}{\sigma^2(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})}(\boldsymbol{X} - V(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}))\gamma(\boldsymbol{X})\Big],$$

$$\boldsymbol{\Omega}_{c,0} = c^2 J_{\boldsymbol{\beta}_0}^{\tau} E\Big[\frac{\sigma'^2(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})}{\sigma^2(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})}[\boldsymbol{X} - V(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})]^{\otimes 2}\Big] J_{\boldsymbol{\beta}_0}.$$

Moreover, we have $\sqrt{n}(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0) \xrightarrow{L} N(J_{\boldsymbol{\beta}_0}\boldsymbol{\Omega}_{c,0}^{-1}\boldsymbol{B}_{c,0}, J_{\boldsymbol{\beta}_0}\boldsymbol{\Omega}_{c,0}^{-1}J_{\boldsymbol{\beta}_0}^{\tau})$.

Remark 3. From the definition of $\boldsymbol{B}_{c,0}$, it is easily seen that if $\gamma(\boldsymbol{x}) = \omega(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x})$ such that $E[\omega^2(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x})] < \infty$ or $\gamma(\boldsymbol{x})$ is a constant function, the biased term $\boldsymbol{B}_{c,0}$ equals to zero by using the fact of that $E[\boldsymbol{X} - V(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})|\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}] = \mathbf{0}_p$.

Theorem 3.5. *Under the conditions of Theorem 3.1, if the local alternative hypothesis $\mathcal{H}_{1n}$ holds, we have*

$$\hat{F}_{0\epsilon}(s) - \hat{F}_{\epsilon}(s) = \frac{f_{\epsilon}(s)}{n}\sum_{i=1}^{n}\Big(\frac{\sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_i)}{E[\sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})]} - 1\Big)\Big[\epsilon_i - \frac{c}{2}(\epsilon_i^2 - 1)\Big] + \frac{f_{\epsilon}(s)}{\sqrt{n}}\delta + o_P(n^{-1/2}),$$

where $\delta = \frac{E[\gamma(\boldsymbol{X})]}{E[\sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})]} - E[\frac{\gamma(\boldsymbol{X})}{\sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})}]$. *Then, under the local hypothesis $\mathcal{H}_{1n}$, the continuous mapping theorem entails that*

$$\mathfrak{T}_{n,\mathrm{KS}} \xrightarrow{L} \sup_{-\infty<s<+\infty} f_{\epsilon}(s)|\mathfrak{N} + \delta|, \qquad \mathfrak{T}_{n,\mathrm{CM}} \xrightarrow{L} \int f_{\epsilon}^2(s)\,dF_{\epsilon}(s)(\mathfrak{N} + \delta)^2.$$

Remark 4. This theorem tells us the test statistics $\mathfrak{T}_{n,\mathrm{KS}}$ and $\mathfrak{T}_{n,\mathrm{CM}}$ detect the local alternative hypothesises converging to null hypothesis at an $n^{-1/2}$ rate when $\gamma(\boldsymbol{x})$ satisfies $\{\gamma(\boldsymbol{x}) : \frac{E[\gamma(\boldsymbol{X})]}{E[\sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})]} \neq E[\frac{\gamma(\boldsymbol{X})}{\sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})}]\}$.

3.3 A wild bootstrap procedure

We use the smooth residual bootstrap (Neumeyer and Van Keilegom (2010), Neumeyer (2009)) to mimic the distributions of the test statistics $\mathfrak{T}_{n,\mathrm{KS}}$ and $\mathfrak{T}_{n,\mathrm{CM}}$. The procedure is summarized as follows:

Step 1: Compute $\mathfrak{T}_{n,\mathrm{KS}}$, $\mathfrak{T}_{n,\mathrm{CM}}$.

Step 2: Generate N times *i.i.d.* variables ς_{ib}, $i=1,\dots,n, b=1,\dots,B$ from a standard normal distribution $N(0,1)$. They are independent from the original sample $\{Y_i, \boldsymbol{X}_i\}_{i=1}^n$. Let $\hat{\epsilon}_i = \frac{Y_i - \hat{g}(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}_i, \hat{\boldsymbol{\beta}}_0)}{\hat{\sigma}(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}_i, \hat{\boldsymbol{\beta}}_0)}$, $i=1,\dots,n$, and define

$$\begin{aligned}
\hat{\epsilon}_{ib}^* &= \tilde{\epsilon}_i + a_n \varsigma_{ib}, \qquad \tilde{\epsilon}_i = \frac{\hat{\epsilon}_i - \frac{1}{n}\sum_{i=1}^n \hat{\epsilon}_i}{(\frac{1}{n}\sum_{i=1}^n [\hat{\epsilon}_i - \frac{1}{n}\sum_{i=1}^n \hat{\epsilon}_i]^2)^{1/2}}, \\
Y_{ib}^* &= \hat{g}(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}_i, \hat{\boldsymbol{\beta}}_0) + \hat{\sigma}(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}_i, \hat{\boldsymbol{\beta}}_0)\hat{\epsilon}_{ib}^*.
\end{aligned} \tag{3.7}$$

Step 3: For each b, using bootstraps $\{Y_{ib}^*, \boldsymbol{X}_i\}_{i=1}^n$, we re-calculate the bootstrap estimators $\hat{\boldsymbol{\beta}}_0^{[b]}$ and $\hat{g}^{[b]}(u, \hat{\boldsymbol{\beta}}_0^{[b]})$, $\hat{\sigma}^{[b]}(u, \hat{\boldsymbol{\beta}}_0^{[b]})$. Let

$$\hat{\epsilon}_{0i,b} = \frac{Y_{ib}^*}{\hat{\sigma}^{[b]}(\boldsymbol{\beta}_0^{[b]\tau} \boldsymbol{X}_i, \boldsymbol{\beta}_0^{[b]})} - \frac{\frac{1}{n}\sum_{i=1}^n Y_{ib}^*}{\frac{1}{n}\sum_{i=1}^n \hat{\sigma}^{[b]}(\boldsymbol{\beta}_0^{[b]\tau} \boldsymbol{X}_i, \boldsymbol{\beta}_0^{[b]})},$$

and

$$\hat{\epsilon}_{i,b} = \frac{Y_{ib}^* - \hat{g}^{[b]}(\boldsymbol{\beta}_0^{[b]\tau} \boldsymbol{X}_i, \boldsymbol{\beta}_0^{[b]})}{\hat{\sigma}^{[b]}(\boldsymbol{\beta}_0^{[b]\tau} \boldsymbol{X}_i, \boldsymbol{\beta}_0^{[b]})}.$$

Then, define

$$\hat{F}_{0\epsilon}^{[b]}(s) = \frac{1}{n}\sum_{i=1}^n I\{\hat{\epsilon}_{0i,b} \le s\}, \qquad \hat{F}_{\epsilon}^{[b]}(s) = \frac{1}{n}\sum_{i=1}^n I\{\hat{\epsilon}_{i,b} \le s\},$$

and the bootstrap statistics

$$\begin{aligned}
\mathfrak{T}_{n,\mathrm{KS}}^{[b]} &= \sup_{-\infty < s < +\infty} n^{1/2} \big|\hat{F}_{0\epsilon}^{[b]}(s) - \hat{F}_{\epsilon}^{[b]}(s)\big|, \\
\mathfrak{T}_{n,\mathrm{CM}}^{[b]} &= n \int \big(\hat{F}_{0\epsilon}^{[b]}(s) - \hat{F}_{\epsilon}^{[b]}(s)\big)^2 \, d\hat{F}_{\epsilon}^{[b]}(s).
\end{aligned}$$

Step 4: We calculate the $1-\kappa$ quantile of the bootstrap test statistics $\mathcal{T}_{n,\mathrm{KS}}^{[b]}$, $\mathcal{T}_{n,\mathrm{CM}}^{[b]}$, $b=1,\dots N$ as the κ-level critical value.

4 Implementation

In this section, we report simulation results to evaluate the performance of the proposed estimators and test statistics. In the following simulations, the Epanechnikov kernel $K(t) = 0.75(1-t^2)^+$ is used. It is worthwhile to point out that undersmoothing is necessary as condition (C4) requires that $nh^4 \to 0$. To meet this requirement, we follow Carroll et al. (1997)'s suggestion by using the order of

$O(n^{-1/5}) \times n^{-2/15} = O(n^{-1/3})$ for the bandwidth h. Define the cross-validation score as

$$\mathrm{CV}(h) = n^{-1} \sum_{i=1}^{n} \{Y_i - \hat{g}_{-i}(\hat{\boldsymbol{\beta}}_{0,-i}^{\tau} \boldsymbol{X}_i, \hat{\boldsymbol{\beta}}_{0,-i})\}^2,$$

where $\hat{\boldsymbol{\beta}}_{0,-i}$ and $\hat{g}_{-i}(\hat{\boldsymbol{\beta}}_{0,-i}^{\tau} \boldsymbol{X}_i, \hat{\boldsymbol{\beta}}_{0,-i})$ are computed analogous to (2.2)–(2.4) from the data with the ith observation deleted. Let $h_1 = \arg\min_h \mathrm{CV}(h)$. Then the final choice for h is defined as $h = n^{-2/15} * h_1$. For the choice of a_n in (3.7), Neumeyer (2009), Neumeyer and Van Keilegom (2010) suggested to use $a_n = c_1 n^{-1/4}$ for some positive constant c_1. We used $a_n = n^{-1/4}$ in the following simulations and the numerical results were stable when we shifted several values around the selected values.

Example 1. We generate 500 realizations and choose the sample size to be $n = 300, 500, 1000$ from model:

$$Y = 2\exp(\boldsymbol{\beta}_0^{\tau} \boldsymbol{X}) + \exp(\boldsymbol{\beta}_0^{\tau} \boldsymbol{X})\epsilon.$$

Here, $\boldsymbol{\beta}_0 = (1, 0, 3, -2, 1)^{\tau}/\sqrt{15}$, $\boldsymbol{X} \sim N_5(\boldsymbol{\mu}, \Sigma)$ with $\boldsymbol{\mu} = (2, \dots, 2)^{\tau}$, $\Sigma = (\sigma_{ij})$, $\sigma_{ij} = 0.5^{|i-j|}$. For the model error ϵ, we consider two cases: $\epsilon \sim N(0, 1)$, and $\epsilon \sim 2(\mathrm{Exp}(2) - 1/2)$, where Exp(2) is an exponential distribution with expectation 1/2.

The performance of estimators $\hat{F}_{\epsilon}(s)$ and $\hat{F}_{0\epsilon}(s)$ are evaluated using the average squared error (ASE) and the average absolute error (AAE)

$$\mathrm{ASE} = n_0^{-1} \sum_{v=1}^{n_0} [\hat{F}_{\epsilon}(s_v) - F_{\epsilon}(s_v)]^2, \qquad \mathrm{AAE} = n_0^{-1} \sum_{s=1}^{n_0} |\hat{F}_{\epsilon}(s_v) - F_{\epsilon}(s_v)|,$$

where $\{s_1, \dots, s_{n_0}\}$ are the given grid points, and $n_0 = 400$ is the number of grid points.

The simulation results for $\hat{\boldsymbol{\beta}}_0$ are reported in Table 1 and Table 2. The values of $\hat{\boldsymbol{\beta}}_0$ are close to the true value of $\boldsymbol{\beta}_0$, and the values of $\mathrm{MSE}(\hat{\boldsymbol{\beta}}_0, \boldsymbol{\beta}_0)$ become smaller as the sample size n increases to 1000. Moreover, the angles (in radians) of $\arccos(\hat{\boldsymbol{\beta}}_0, \boldsymbol{\beta}_0)$ become closer to zero when the sample size n increases to 1000. In Table 3 and Table 4, we also report the numerical results of ASE and AAE for the estimators $\hat{F}_{\epsilon}(s)$ and $\hat{F}_{0\epsilon}(s)$. Both estimators perform better as the sample size n increases. The performance of $\hat{F}_{\epsilon}(s)$ is better than $\hat{F}_{0\epsilon}(s)$ in this simulation study. It is seen that $\hat{\epsilon}_{0i} = \hat{\epsilon}_i + \frac{\hat{g}(\hat{\boldsymbol{\beta}}^{\tau} \boldsymbol{X}_i)}{\hat{\sigma}(\hat{\boldsymbol{\beta}}^{\tau} \boldsymbol{X}_i)} - \frac{\bar{Y}}{\frac{1}{n}\sum_{i=1}^{n} \hat{\sigma}(\hat{\boldsymbol{\beta}}^{\tau} \boldsymbol{X}_i)}$. From this equation, the performance of estimator $\hat{F}_{0\epsilon}(s)$ involves both $\hat{\epsilon}_i$ and $\frac{\hat{g}(\hat{\boldsymbol{\beta}}^{\tau} \boldsymbol{X}_i)}{\hat{\sigma}(\hat{\boldsymbol{\beta}}^{\tau} \boldsymbol{X}_i)} - \frac{\bar{Y}}{\frac{1}{n}\sum_{i=1}^{n} \hat{\sigma}(\hat{\boldsymbol{\beta}}^{\tau} \boldsymbol{X}_i)}$. Under the null hypothesis, $\frac{\bar{Y}}{\frac{1}{n}\sum_{i=1}^{n} \hat{\sigma}(\hat{\boldsymbol{\beta}}^{\tau} \boldsymbol{X}_i)} \xrightarrow{P} c$ with root-n convergence rate, and

Table 1 *The Mean (M), Standard Errors (SD) and Mean Squared Errors (MSE) of $\hat{\boldsymbol{\beta}}_0$ when $\epsilon \sim N(0, 1)$*[a]

	$\hat{\boldsymbol{\beta}}_{0,1}$	$\hat{\boldsymbol{\beta}}_{0,2}$	$\hat{\boldsymbol{\beta}}_{0,3}$	$\hat{\boldsymbol{\beta}}_{0,4}$	$\hat{\boldsymbol{\beta}}_{0,5}$	$\arccos(\hat{\boldsymbol{\beta}}_0, \boldsymbol{\beta}_0)$
			$n = 300$			
M	0.2633	0.0015	0.7768	−0.5044	0.2590	0.0525
SD	0.0260	0.0294	0.0178	0.0230	0.0261	0.0218
MSE	6.8593	9.5309	3.1337	5.6836	7.0219	32.2438
			$n = 500$			
M	0.2582	0.0078	0.7776	−0.5077	0.2596	0.0389
SD	0.0206	0.0214	0.0131	0.0197	0.0183	0.0175
MSE	4.1818	4.5705	1.7471	4.2543	3.4318	18.1908
			$n = 1000$			
M	0.2574	0.0027	0.7759	−0.5067	0.2603	0.0301
SD	0.0149	0.0131	0.0095	0.0097	0.0110	0.0126
MSE	2.4792	3.4727	0.9085	1.5780	2.1688	10.6087

[a]Note: MSE is in the scale of $\times 10^{-4}$.

Table 2 *The Mean (M), Standard Errors (SD) and Mean Squared Errors (MSE) of $\hat{\boldsymbol{\beta}}_0$ when $\epsilon \sim 2(\mathrm{Exp}(2) - 1/2)$*[a]

	$\hat{\boldsymbol{\beta}}_{0,1}$	$\hat{\boldsymbol{\beta}}_{0,2}$	$\hat{\boldsymbol{\beta}}_{0,3}$	$\hat{\boldsymbol{\beta}}_{0,4}$	$\hat{\boldsymbol{\beta}}_{0,5}$	$\arccos(\hat{\boldsymbol{\beta}}_0, \boldsymbol{\beta}_0)$
			$n = 300$			
M	0.2588	0.0077	0.7761	−0.5078	0.2652	0.0494
SD	0.0251	0.0263	0.0146	0.0229	0.0226	0.0173
MSE	6.2339	7.4574	2.1391	5.9357	5.5536	27.3290
			$n = 500$			
M	0.2611	0.0059	0.7752	−0.5114	0.2602	0.0372
SD	0.0183	0.0191	0.0140	0.0165	0.0208	0.0166
MSE	3.4077	3.9480	1.9557	2.9528	4.3063	16.5743
			$n = 1000$			
M	0.2611	0.0040	0.7753	−0.5131	0.2577	0.0287
SD	0.0143	0.0172	0.0101	0.0113	0.0143	0.0118
MSE	2.1145	3.0960	1.0081	1.3707	2.0369	9.6275

[a]Note: MSE is in the scale of $\times 10^{-4}$.

$\frac{\hat{g}(\hat{\boldsymbol{\beta}}^{\tau} \boldsymbol{X}_i)}{\hat{\sigma}(\hat{\boldsymbol{\beta}}^{\tau} \boldsymbol{X}_i)} \xrightarrow{P} c$ with root-(nh) convergence rate, slowly then the former one. Moreover, the variance function estimator $\hat{\sigma}(\hat{\boldsymbol{\beta}}^{\tau} \boldsymbol{X}_i)$ performs not as well as the mean function estimator $\hat{g}(\hat{\boldsymbol{\beta}}^{\tau} \boldsymbol{X}_i)$ because the variance function are usually more difficult to estimate than the mean function. So the estimator $\hat{F}_{0\epsilon}(s)$ performs not better

Table 3 *The Mean (M) and Standard Errors (SD) for ASE and AAE when $\epsilon \sim N(0, 1)$*[a]

	$\hat{F}_\epsilon(s)$		$\hat{F}_{0\epsilon}(s)$	
	ASE	AAE	ASE	AAE
		$n = 300$		
M	4.8929	0.0171	21.9599	0.0363
SD	3.6795	0.0064	22.0420	0.0203
		$n = 500$		
M	6.2402	0.0142	8.2263	0.0222
SD	7.6969	0.0173	9.1560	0.0114
		$n = 1000$		
M	1.4782	0.0094	4.5914	0.0159
SD	1.1630	0.0037	6.4110	0.0110

[a]Note: ASE is in the scale of $\times 10^{-4}$.

Table 4 *The Mean (M) and Standard Errors (SD) for ASE and AAE when $\epsilon \sim 2(\text{Exp}(2) - 1/2)$*[a]

	$\hat{F}_\epsilon(s)$		$\hat{F}_{0\epsilon}(s)$	
	ASE	AAE	ASE	AAE
		$n = 300$		
M	5.7764	0.0181	18.3054	0.0303
SD	8.9399	0.0128	39.9751	0.0278
		$n = 500$		
M	10.0468	0.0145	10.0037	0.0228
SD	57.6307	0.0274	20.4832	0.0195
		$n = 1000$		
M	1.0768	0.0083	3.0179	0.0137
SD	0.9591	0.0040	3.4882	0.0087

[a]Note: ASE is in the scale of $\times 10^{-4}$.

than $\hat{F}_\epsilon(s)$. Figure 1 presents fours plots for the estimators $\hat{g}(t)$, $\hat{\sigma}(t)$, $\hat{F}_\epsilon(s)$ and $\hat{F}_{0\epsilon}(s)$ against their true values when $\epsilon \sim N(0, 1)$ with the sample size $n = 1000$. From Figure 1, we see that those estimators perform well.

Example 2. In this example, we investigate the performances of test statistics $\mathfrak{T}_{n,\mathrm{KS}}$ and $\mathfrak{T}_{n,\mathrm{CM}}$. 500 data sets consisting of $n = 300, 500, 1000$ observations are generated, and 1000 bootstrap samples are generated in each simulation for power calculation. We consider the data generating process (DGP) as follows:

$$g(\boldsymbol{X}) = 2\exp(\boldsymbol{\beta}_0^\tau \boldsymbol{X}) + C_o(\boldsymbol{\beta}_0^\tau \boldsymbol{x})^2 + \exp(\boldsymbol{\beta}_0^\tau \boldsymbol{X})\epsilon. \tag{4.1}$$

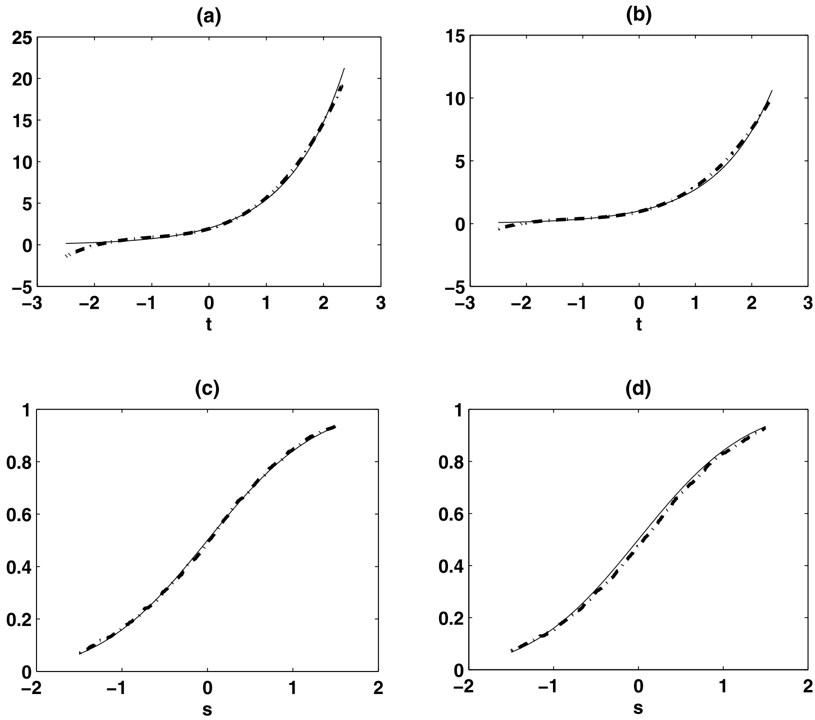

Figure 1 (a) *The plots of estimator* $\hat{g}(t)$ (*dashed line*) *against* $g(t)$ (*solid line*). (b) *The plots of estimator* $\hat{\sigma}(t)$ (*dashed line*) *against* $\sigma(t)$ (*solid line*). (c) *The plots of estimator* $\hat{F}_\epsilon(s)$ (*dashed line*) *against* $F_\epsilon(s)$ (*solid line*). (d) *The plots of estimator* $\hat{F}_{0\epsilon}(s)$ (*dashed line*) *against* $F_\epsilon(s)$ (*solid line*).

The parameter $\boldsymbol{\beta}_0$, covariate $\boldsymbol{X}$ and model error ϵ are the same as Example 1. The null hypothesis $\mathcal{H}_0$ considered in (1.2) is true if and only is $C_o = 0$. The simulation results are reported in Table 5 and Table 6. All empirical levels obtained by the bootstrap test statistics are close to four nominal levels when the null hypothesis $\mathcal{H}_0$ is true, which indicates that the bootstrap method provides proper Type I errors. That is, when $C_o = 0$, the percentages of $\mathcal{H}_0$ being rejected are close to the corresponding nominal level for all four nominal levels, and they are much closer to the significance levels when $n = 1000$. As the value of C_0 increases, the power functions increase rapidly and approach to one when the sample size n increases. In this example, the Cramér–von Mises test statistic $\mathcal{T}_{n,\mathrm{CM}}$ performs more powerful than the Kolmogorov–Smirnov test statistic $\mathcal{T}_{n,KM}$. This phenomenon is the same as the simulation results obtained in Neumeyer and Van Keilegom (2010), they also found that the performances of $\mathcal{T}_{n,\mathrm{CM}}$ show larger values of power functions than that of $\mathcal{T}_{n,KM}$ in testing additivity of a multivariate regression function.

Table 5 *The simulation results for power calculations in Example 2 when $\epsilon \sim N(0, 1)$*

	$\mathfrak{T}_{KS}$				$\mathfrak{T}_{CM}$			
Significant level	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10
				$n = 300$				
$C_0 = 0.00$	0.0091	0.0239	0.0489	0.0967	0.0113	0.0251	0.0491	0.1079
$C_0 = 0.05$	0.6423	0.6585	0.6829	0.6911	0.7723	0.7845	0.8247	0.8423
$C_0 = 0.10$	0.7154	0.7317	0.7449	0.7723	0.8011	0.8247	0.8754	0.9011
$C_0 = 0.15$	0.8537	0.9024	0.9357	0.9768	0.9277	0.9545	0.9689	0.9912
$C_0 = 0.20$	0.9423	0.9524	0.9632	0.9814	0.9872	0.9948	1.0000	1.0000
				$n = 500$				
$C_0 = 0.00$	0.0091	0.0251	0.0487	0.1108	0.0112	0.0253	0.0561	0.1011
$C_0 = 0.05$	0.7317	0.7967	0.8211	0.8780	0.8456	0.8823	0.9131	0.9277
$C_0 = 0.10$	0.8211	0.8943	0.9105	0.9512	0.8999	0.9345	0.9620	1.0000
$C_0 = 0.15$	0.9023	0.9468	0.9749	1.0000	0.9367	0.9825	0.9933	1.0000
$C_0 = 0.20$	0.9857	0.9946	1.0000	1.0000	0.9978	0.9997	1.0000	1.0000
				$n = 1000$				
$C_0 = 0.00$	0.0101	0.0247	0.0489	0.0979	0.0107	0.0251	0.0511	0.0989
$C_0 = 0.05$	0.8834	0.9011	0.9578	0.9923	0.9315	0.9559	0.9733	0.9821
$C_0 = 0.10$	0.9457	0.9722	0.9978	1.0000	0.9850	0.9979	1.0000	1.0000
$C_0 = 0.15$	0.9837	0.9921	1.0000	1.0000	0.9919	1.0000	1.0000	1.0000
$C_0 = 0.20$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 6 *The simulation results for power calculations in Example 2 when $\epsilon \sim 2(\mathrm{Exp}(2) - 1/2)$*

	$\mathfrak{T}_{KS}$				$\mathfrak{T}_{CM}$			
Significant level	0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10
				$n = 300$				
$C_0 = 0.00$	0.0082	0.0227	0.0467	0.0925	0.0101	0.0246	0.0476	0.0990
$C_0 = 0.05$	0.6122	0.6367	0.6589	0.6724	0.7319	0.7622	0.7935	0.8117
$C_0 = 0.10$	0.6926	0.7166	0.7287	0.7569	0.7878	0.8166	0.8572	0.8989
$C_0 = 0.15$	0.8269	0.8917	0.9155	0.9614	0.9159	0.9360	0.9521	0.9844
$C_0 = 0.20$	0.9115	0.9269	0.9332	0.9711	0.9799	0.9821	0.9974	1.0000
				$n = 500$				
$C_0 = 0.00$	0.0088	0.0248	0.0491	0.1024	0.0101	0.0257	0.0535	0.1100
$C_0 = 0.05$	0.7126	0.7790	0.8087	0.8692	0.8334	0.8769	0.9099	0.9197
$C_0 = 0.10$	0.8154	0.8845	0.9027	0.9316	0.8879	0.9165	0.9528	0.9897
$C_0 = 0.15$	0.8911	0.9275	0.9613	0.9829	0.9289	0.9739	0.9821	0.9997
$C_0 = 0.20$	0.9788	0.9822	0.9990	1.0000	0.9914	0.9979	1.0000	1.0000
				$n = 1000$				
$C_0 = 0.00$	0.0098	0.0263	0.0482	0.0996	0.0101	0.0249	0.0508	0.1019
$C_0 = 0.05$	0.8721	0.8934	0.9299	0.9819	0.9217	0.9431	0.9655	0.9939
$C_0 = 0.10$	0.9320	0.9658	0.9819	0.9930	0.9769	0.9894	0.9930	1.0000
$C_0 = 0.15$	0.9621	0.9857	0.9933	1.0000	0.9822	0.9929	1.0000	1.0000
$C_0 = 0.20$	0.9870	0.9940	1.0000	1.0000	0.9990	1.0000	1.0000	1.0000

5 A 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 variables about environment of the property as well as its selling price and other relevant variables. For this dataset, we consider to use the values of NOX (the nitric oxide concentration per 10 million) which are greater or equal to the median of NOX. Eight attributes are included in the model (1.1): MEDV (Y) – the median value of owner-occupied homes in \$ 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 centres, RAD (X_4) – the index of accessibility to radial highways, TAX (X_5) – the full-value property-tax rate per \$10,000, PTRATIO (X_6) – the pupil-teacher ratio by town, BLACKS (X_7) – the transformed proportion of Blacks which is calculated by $1000(\mathrm{Bk} - 0.63)^2$ and Bk is the proportion of blacks by town.

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

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_0 \\ p_{\hat{\boldsymbol{\beta}}_0} \end{pmatrix} = \begin{pmatrix} 0.852, & -0.510, & -0.085, & 0.066, & -0.003, & -0.016, & 0.049 \\ 0.000, & 0.000, & 0.246, & 0.205, & 0.122, & 0.821, & 0.000 \end{pmatrix}.$$

The p-values are calculated by estimating the asymptotic variances of $\hat{\boldsymbol{\beta}}_0$ obtained in Theorem 2.1. When the significance level is set at 0.05, we find that RM-X_1, Age-X_2 and BLACKS-X_7 are significant. Note that the high-values of NOX is reasonably highly related to life quality and then house price. The significance of these indices RM and Age are very reasonable. Next, we used the test statistic proposed by Stute and Zhu (2005) to check whether the single-index model $g(\boldsymbol{\beta}_0^\tau \boldsymbol{X})$ is appropriate for this dataset. The associated value of the test statistics is 1.8941 with a p-value of 0.2131. This indicates that the single-index model $g(\boldsymbol{\beta}_0^\tau \boldsymbol{X})$ is not a constant function, see also Figure 2. We conducted 1000 bootstraps to test $g(u) = c\sigma(u)$ for some constant c, and the corresponding $\mathcal{T}_{n,\mathrm{KS}}$, $\mathcal{T}_{n,\mathrm{CM}}$ are both larger than the 99% quantile of 1000 bootstraps, and this suggests a rejection of the null hypothesis $\mathcal{H}_0$. The estimators $\hat{g}(u)$, along with its 95% pointwise confidence bands, are presented in Figure 2. We also present the estimated figure for the error function $\hat{c}\hat{\sigma}(u)$ in Figure 2, where $\hat{c}$ is obtained as $\hat{c} = \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{g}(\hat{\boldsymbol{\beta}}^\tau \boldsymbol{X}_i)}{\hat{\sigma}(\hat{\boldsymbol{\beta}}^\tau \boldsymbol{X}_i)}$. In this figure, the estimated function $\hat{c}\hat{\sigma}(u)$ is not a constant function, which indicates that the heteroscedasticity exists in this dataset. Moreover, it is seen that $\hat{c}\hat{\sigma}(u)$ is not encapsulated in the 95% pointwise confidence bands. This indicates that the null hypothesis $\mathcal{H}_0$ is not true, although $\hat{g}(u)$ and $\hat{c}\hat{\sigma}(u)$ has a similar variation tendency in Figure 2.

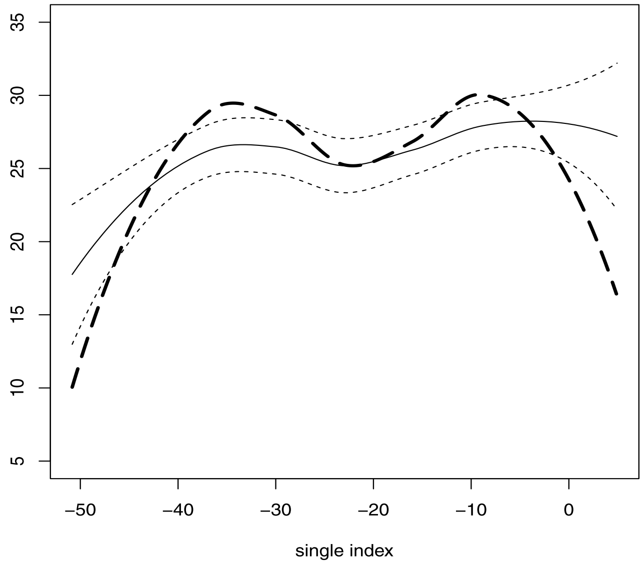

Figure 2 *The plot for the estimator* $\hat{g}(u)$ (*solid line*) *against estimated single-index* $\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}$, *along with the associated* 95% *pointwise confidence intervals* (*dotted lines*); *The plot for estimator* $\hat{c}\hat{\sigma}(u)$ (*thick longdash line*) *against estimated single-index* $\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}$.

Appendix

A.1 Proof of Theorem 2.1

Lemma A.1. *Suppose that* $\boldsymbol{X}_i$, $i = 1, \ldots, n$ *are i.i.d. random vector, and* $E[m(\boldsymbol{X})|\boldsymbol{\beta}^\tau \boldsymbol{X} = u]$ *have a continuous bounded second derivative on* u, *satisfying* $E[m^2(\boldsymbol{X})] < \infty$. *Let* $K(\cdot)$ *be a bounded positive function with a bounded support satisfying the Lipschitz condition*: *there exits a neighbourhood 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* $s > 0$, $j = 0, 1, 2$,

$$\sup_{(x,\boldsymbol{\beta})\in\mathcal{X}\times\Delta}\Bigg|\frac{1}{n}\sum_{i=1}^{n} K_h(\boldsymbol{\beta}^\tau \boldsymbol{X}_i - \boldsymbol{\beta}^\tau x)\left(\frac{\boldsymbol{\beta}^\tau \boldsymbol{X}_i - \boldsymbol{\beta}^\tau x}{h}\right)^j m(\boldsymbol{X}_i)$$
$$- f_{\boldsymbol{\beta}_0^\tau \boldsymbol{X}}(\boldsymbol{\beta}_0^\tau \boldsymbol{x}) E[m(\boldsymbol{X})|\boldsymbol{\beta}_0^\tau \boldsymbol{X} = \boldsymbol{\beta}_0^\tau x]\mu_{K,j} - hS(\boldsymbol{\beta}_0^\tau \boldsymbol{x})\mu_{K,j+1}\Bigg|$$
$$= O_P(c_{n1}),$$

where $\Delta = \{\boldsymbol{\beta} \in \Theta, \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \le Cn^{-1/2}\}$ *for some positive constant* C, $\mu_{K,l} = \int t^l K(t)\,dt$, $S(\boldsymbol{\beta}_0^\tau \boldsymbol{x}) = \frac{d}{du}\{f_{\boldsymbol{\beta}_0^\tau \boldsymbol{X}}(u) E[m(\boldsymbol{X})|\boldsymbol{\beta}_0^\tau \boldsymbol{X} = u]\}|_{u=\boldsymbol{\beta}_0^\tau \boldsymbol{x}}$, *and* $c_{n1} = \{\frac{(\log n)^{1+s}}{nh}\}^{1/2} + h^2$.

Proof. This proof is completed by the Lemma A6.1 in Xia (2006). □

Proof of Theorem 2.1. Note that $\mathcal{W}_n(\hat{\boldsymbol{\beta}}^{(1)}) = \mathbf{0}$. Taylor expansion entails that

$$-\frac{1}{\sqrt{n}}\mathcal{W}_n(\boldsymbol{\beta}_0^{(1)}) = \left[\frac{1}{n}\frac{\partial \mathcal{W}_n(\boldsymbol{\beta}^{(1)})}{\partial \boldsymbol{\beta}^{(1)}}\bigg|_{\boldsymbol{\beta}^{(1)}=\tilde{\boldsymbol{\beta}}_0^{(1)}}\right][\sqrt{n}(\hat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}_0^{(1)})], \tag{A.1}$$

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

Step 1. In this sub-step, we analyze $n^{-1/2}\mathcal{W}_n(\boldsymbol{\beta}_0^{(1)})$. Directly using Lemma A.1 we have that $\hat{g}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i, \boldsymbol{\beta}_0) = g(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i) + O_P(c_{n1})$, $\hat{V}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i, \boldsymbol{\beta}_0) = V(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i) + O_P(c_{n1})$. Moreover,

$$\begin{aligned}
&\frac{1}{n}S_{n,l_1 1}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i, \boldsymbol{\beta}_0)\\
&\quad= \frac{1}{n}\sum_{j=1}^n K_h(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_j - \boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_j - \boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)^{l_1}\sigma^2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_j)\epsilon_j^2\\
&\qquad + \frac{2}{n}\sum_{j=1}^n K_h(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_j - \boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_j - \boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)^{l_1}\\
&\qquad \times [g(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_j) - \hat{g}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_j, \boldsymbol{\beta}_0)]\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_j)\epsilon_j \qquad\qquad (A.2)\\
&\qquad + \frac{1}{n}\sum_{j=1}^n K_h(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_j - \boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_j - \boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)^{l_1}\\
&\qquad \times [g(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_j) - \hat{g}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_j, \boldsymbol{\beta}_0)]^2\\
&\quad= h^{l_1} f_{\boldsymbol{\beta}_0^\tau X}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)\sigma^2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)\mu_{Kl_1} + O_P(h^{l_1}c_{n1} + h^{l_1}c_{n1}^2),
\end{aligned}$$

for $l_1 = 0, 1, 2$. Directly using (A.2), we obtain $\hat{\sigma}^2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i, \boldsymbol{\beta}_0) = \sigma^2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i) + O_P(c_{n1})$.

Define

$$\begin{aligned}
&\mathcal{G}_{n2}(u, \boldsymbol{\beta}_0)\\
&\quad = \frac{1}{nh^2}T_{n,20}(u, \boldsymbol{\beta}_0)\frac{1}{n}T_{n,00}(u, \boldsymbol{\beta}_0) - \frac{1}{n^2h^2}T_{n,10}^2(u, \boldsymbol{\beta}_0),\\
&\mathcal{G}_{n1}(u, \boldsymbol{\beta}_0)\\
&\quad = \frac{1}{nh^2}T_{n,20}(u, \boldsymbol{\beta}_0)\frac{1}{n}T_{n,01}(u, \boldsymbol{\beta}_0) - \frac{1}{nh}T_{n,10}(u, \boldsymbol{\beta}_0)\frac{1}{nh}T_{n,11}(u, \boldsymbol{\beta}_0).
\end{aligned}$$

Then, $\hat{g}(u, \boldsymbol{\beta}_0) = \frac{\mathcal{G}_{n1}(u,\boldsymbol{\beta}_0)}{\mathcal{G}_{n2}(u,\boldsymbol{\beta}_0)}$, $\hat{g}'(u, \boldsymbol{\beta}_0) = \frac{\partial \mathcal{G}_{n1}(u,\boldsymbol{\beta}_0)/\partial u}{\mathcal{G}_{n2}(u,\boldsymbol{\beta}_0)} - \frac{\mathcal{G}_{n1}(u,\boldsymbol{\beta}_0)\partial \mathcal{G}_{n2}(u,\boldsymbol{\beta}_0)/\partial u}{\mathcal{G}_{n2}^2(u,\boldsymbol{\beta}_0)}$. Using Lemma A.1, we have that $\hat{g}'(u, \boldsymbol{\beta}_0) = g'(u) + O_P(h^2 + \sqrt{\frac{(\log n)^{1+s}}{nh^3}})$ and also $\hat{g}'(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i, \boldsymbol{\beta}_0) = g'(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i) + O_P(h^2 + \sqrt{\frac{(\log n)^{1+s}}{nh^3}})$. As a result, as $nh^8 \to 0$, $\frac{(\log n)^{2+2s}}{nh^2} \to 0$, we have that

$$\begin{aligned} n^{-1/2}&\mathcal{W}_n(\boldsymbol{\beta}_0^{(1)}) \\ &= n^{-1/2}\sum_{i=1}^{n} J_{\boldsymbol{\beta}_0}^\tau \hat{g}'(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i, \boldsymbol{\beta}_0)[\boldsymbol{X}_i - \hat{V}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i, \boldsymbol{\beta}_0)]\hat{\sigma}^{-2}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i, \boldsymbol{\beta}_0) \\ &\quad \times [Y_i - \hat{g}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i, \boldsymbol{\beta}_0)] \\ &= n^{-1/2}\sum_{i=1}^{n} J_{\boldsymbol{\beta}_0}^\tau g'(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)[\boldsymbol{X}_i - V(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)]\sigma^{-1}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)\epsilon_i + o_P(1). \end{aligned} \tag{A.3}$$

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

$$\begin{aligned} \mathcal{S}_{n1}&(\tilde{\boldsymbol{\beta}}_0^{(1)}) \\ &\stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n}\bigg\{\bigg[\frac{\partial}{\partial \boldsymbol{\beta}^{(1)}}\{J_{\boldsymbol{\beta}}^\tau \hat{g}'(\boldsymbol{\beta}^\tau \boldsymbol{X}_i, \boldsymbol{\beta})[\boldsymbol{X}_i - \hat{V}(\boldsymbol{\beta}^\tau \boldsymbol{X}_i, \boldsymbol{\beta})]\hat{\sigma}^{-2}(\boldsymbol{\beta}^\tau \boldsymbol{X}_i, \boldsymbol{\beta})\}\bigg] \\ &\quad \times [Y_i - \hat{g}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i, \boldsymbol{\beta}_0)]\bigg\}\bigg|_{\boldsymbol{\beta}^{(1)}=\tilde{\boldsymbol{\beta}}_0^{(1)}}, \end{aligned}$$

and

$$\begin{aligned} \mathcal{S}_{n2}&(\tilde{\boldsymbol{\beta}}_0^{(1)}) \\ &\stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n}\{J_{\tilde{\boldsymbol{\beta}}_0}^\tau \hat{g}'(\tilde{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}_i, \tilde{\boldsymbol{\beta}}_0)[\boldsymbol{X}_i - \hat{V}(\tilde{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}_i, \tilde{\boldsymbol{\beta}}_0)]\hat{\sigma}^{-2}(\tilde{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}_i, \tilde{\boldsymbol{\beta}}_0)\} \\ &\quad \times \left\{\frac{\partial \hat{g}(\boldsymbol{\beta}^\tau \boldsymbol{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^{(1)}}\bigg|_{\boldsymbol{\beta}^{(1)}=\tilde{\boldsymbol{\beta}}_0^{(1)}}\right\}. \end{aligned}$$

Then,

$$\frac{1}{n}\frac{\partial \mathcal{W}_n(\boldsymbol{\beta}^{(1)})}{\partial \boldsymbol{\beta}^{(1)}}\bigg|_{\boldsymbol{\beta}^{(1)}=\tilde{\boldsymbol{\beta}}_0^{(1)}} = \mathcal{S}_{n1}(\tilde{\boldsymbol{\beta}}_0^{(1)}) + \mathcal{S}_{n2}(\tilde{\boldsymbol{\beta}}_0^{(1)}), \tag{A.4}$$

where $\tilde{\boldsymbol{\beta}}_0 = (\sqrt{1 - \tilde{\boldsymbol{\beta}}_0^{(1)\tau}\tilde{\boldsymbol{\beta}}_0^{(1)}}, \tilde{\boldsymbol{\beta}}_0^{(1)\tau})^\tau$. Note that $\tilde{\boldsymbol{\beta}}_0^{(1)}$ is between $\hat{\boldsymbol{\beta}}^{(1)}$ and $\boldsymbol{\beta}_0^{(1)}$, and (A.1) entails that $\hat{\boldsymbol{\beta}}_{(1)} = \boldsymbol{\beta}_0^{(1)} + O_P(n^{-1/2})$. Then, Using Lemma A.1, we have that

$$\begin{aligned} &\frac{\partial \hat{g}(\boldsymbol{\beta}^\tau \boldsymbol{X}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^{(1)}}\bigg|_{\boldsymbol{\beta}^{(1)}=\tilde{\boldsymbol{\beta}}_0^{(1)}} \\ &\quad = J_{\boldsymbol{\beta}_0}^\tau[\boldsymbol{X}_i - V(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)]g'(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i) + O_P\bigg(h^2 + \sqrt{\frac{(\log n)^{1+s}}{nh^3}}\bigg). \end{aligned} \tag{A.5}$$

By using the fact that $\tilde{\boldsymbol{\beta}}_0^{(1)} \xrightarrow{P} \boldsymbol{\beta}_0^{(1)}$ and $\tilde{\boldsymbol{\beta}}_0 \xrightarrow{P} \boldsymbol{\beta}_0$ and (A.5), we have

$$\mathcal{S}_{n2}(\tilde{\boldsymbol{\beta}}_0^{(1)}) \xrightarrow{P} J_{\boldsymbol{\beta}_0}^\tau E[g'^2(\boldsymbol{\beta}_0^\tau \boldsymbol{X})\sigma^{-2}(\boldsymbol{\beta}_0^\tau \boldsymbol{X})[\boldsymbol{X} - V(\boldsymbol{\beta}_0^\tau \boldsymbol{X})]^{\otimes 2}]J_{\boldsymbol{\beta}_0}. \tag{A.6}$$

Moreover, a direct calculation for $\mathcal{S}_{n1}(\tilde{\boldsymbol{\beta}}_0^{(1)})$ and Lemma A.1 entail that $\mathcal{S}_{n1}(\tilde{\boldsymbol{\beta}}_0^{(1)}) = o_P(1)$. Together with (A.3) and (A.6), we have completed the proof of Theorem 2.1. □

A.2 Proof of Theorem 3.1

Lemma A.2. *Suppose that Conditions* (A1)–(A5) *hold. Let* $F_{\hat{\epsilon}}(s|\mathscr{H}_n)$ *be the distribution function of* $\hat{\epsilon} = \frac{Y - \hat{g}(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}, \hat{\boldsymbol{\beta}}_0)}{\hat{\sigma}(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}, \hat{\boldsymbol{\beta}}_0)}$ *conditional on the data* $\mathscr{H}_n = \{\boldsymbol{X}_i, Y_i\}_{i=1}^n$ (*i.e., considering* $\hat{g}(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{x}, \hat{\boldsymbol{\beta}}_0), \hat{\sigma}(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{x}, \hat{\boldsymbol{\beta}}_0)$ *as fixed functions on* $\boldsymbol{x}$). *Then, we have*

$$\begin{aligned}\sup_{-\infty<s<+\infty}\left|n^{-1}\sum_{i=1}^n [I\{\hat{\epsilon}_i \le s\} - I\{\epsilon_i \le s\} - F_{\hat{\epsilon}}(s|\mathscr{H}_n) + F_\epsilon(s)]\right| \\ = o_P(n^{-1/2}).\end{aligned} \tag{A.7}$$

Proof. Define $d_{n1}(\boldsymbol{x}) = \frac{\hat{g}(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{x}, \hat{\boldsymbol{\beta}}_0) - g(\boldsymbol{\beta}_0^\tau \boldsymbol{x})}{\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{x})}$, $d_{n2}(\boldsymbol{x}) = \frac{\hat{\sigma}(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{x}, \hat{\boldsymbol{\beta}}_0)}{\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{x})}$ and

$$\begin{aligned}\mathscr{F} = \{&I\{\epsilon \le s d_2(\boldsymbol{X}) + d_1(\boldsymbol{X})\} - I\{\epsilon \le s\} - P(\epsilon \le s d_2(\boldsymbol{X}) + d_1(\boldsymbol{X})) \\ &+ P(\epsilon \le s); -\infty < s < +\infty, d_1, d_2 \in C_1^{1+\delta}(\mathfrak{R}_c^p)\},\end{aligned}$$

where $C_1^{1+\delta}(\mathfrak{R}_c^p)$ is the class of all differential functions $d(\cdot)$ defined on the domain $\mathfrak{R}_c^p$ of $\boldsymbol{x}$ and $\|d\|_{1+\delta} \le 1$. Here $\mathfrak{R}_c^p$ is a compact set of $\mathbb{R}^p$ and

$$\begin{aligned}\|d\|_{1+\delta} = \max&\left\{\sup_{\boldsymbol{x}\in\mathfrak{R}_c^p}|d(\boldsymbol{x})| + \sum_{l=1}^p \sup_{\boldsymbol{x}\in\mathfrak{R}_c^p}\left|\frac{\partial d(\boldsymbol{x})}{\partial x_l}\right|\right\} \\ &+ \sup_{\boldsymbol{x}_1,\boldsymbol{x}_2\in\mathfrak{R}_c^p}\frac{|\partial d(\boldsymbol{x}_1) - \partial d(\boldsymbol{x}_2)|}{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^\delta}.\end{aligned}$$

Using Lemma A.1 and $\hat{\boldsymbol{\beta}}_0 = \boldsymbol{\beta}_0 + O_P(n^{-1/2})$, we have $\hat{g}(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{x}, \hat{\boldsymbol{\beta}}_0) = g(\boldsymbol{\beta}_0^\tau \boldsymbol{x}) + O_P(h^2 + \sqrt{\frac{(\log n)^{1+s}}{nh}})$ and $\hat{g}'(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{x}, \hat{\boldsymbol{\beta}}_0) = g'(\boldsymbol{\beta}_0^\tau \boldsymbol{x}) + O_P(h^2 + \sqrt{\frac{(\log n)^{1+s}}{nh^3}})$ uniformly in $\boldsymbol{x} \in \mathfrak{R}_c^p$. From Theorem 2.1, we have $|\hat{\sigma}(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{x}, \hat{\boldsymbol{\beta}}_0) - \sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{x})| = O_P(h^2 + \sqrt{\frac{(\log n)^{1+s}}{nh}})$, $|\hat{\sigma}'(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{x}, \hat{\boldsymbol{\beta}}_0) - \sigma'(\boldsymbol{\beta}_0^\tau \boldsymbol{x})| = O_P(h^2 + \sqrt{\frac{(\log n)^{1+s}}{nh^3}})$. So, $P(d_{n1} \in C_1^{1+\delta}(\mathfrak{R}_c^p)) \to 1$, $P(d_{n2} \in C_1^{1+\delta}(\mathfrak{R}_c^p)) \to 1$ as $n \to \infty$, $h \to 0$ and $\frac{nh^3}{(\log n)^{1+s}} \to \infty$.

By directly using the Corollary 2.7.2 of van der Vaart and Wellner (1996), we have that the bracketing number $N_{[\,]}(\varepsilon^2, C_1^{1+\delta}(\mathfrak{R}_c^p), L_2(P))$ will be at most $\exp(c_0\varepsilon^{-\frac{2p}{1+\delta}})$ for some positive constant c_0, Following the proof of Lemma 1 in Appendix B of Akritas and Van Keilegom (2001) and the proof of Lemma A.3 in Neumeyer and Van Keilegom (2010), the afore-defined class $\mathscr{F}$ is a Donsker class, i.e., $\int_0^\infty \sqrt{N_{[\,]}(\bar{\varepsilon}, \mathscr{F}, L_2(P))}\,d\bar{\varepsilon} < \infty$. Then, we finish the proof of (A.7). □

Proof of Theorem 3.1. Using Lemma A.2, we have that

$$\begin{aligned}
&\hat{F}_\epsilon(s) - F_\epsilon(s) \\
&\quad = \frac{1}{n}\sum_{i=1}^n I\{\hat{\epsilon}_i \le s\} - F_\epsilon(s) \\
&\quad = \frac{1}{n}\sum_{i=1}^n I\{\epsilon_i \le s\} - F_\epsilon(s) + \bigl(F_{\hat{\epsilon}}(s|\mathscr{H}_n) - F_\epsilon(s)\bigr) + R_{n,1}(s).
\end{aligned} \tag{A.8}$$

$R_{n,1}(s) = \frac{1}{n}\sum_{i=1}^n [I\{\hat{\epsilon}_i \le s\} - I\{\epsilon_i \le s\}] + F_\epsilon(s) - F_{\hat{\epsilon}}(s|\mathscr{H}_n) = o_P(n^{-1/2})$ uniformly in $s \in \mathbb{R}$. For $F_{\hat{\epsilon}}(s|\mathscr{H}_n) - F_\epsilon(s)$, Taylor expansion entails that

$$\begin{aligned}
&F_{\hat{\epsilon}}(s|\mathscr{H}_n) - F_\epsilon(s) \\
&\quad = \int \bigl[F_\epsilon\bigl(s + s[d_{n2}(\boldsymbol{x}) - 1] + d_{n1}(\boldsymbol{x})\bigr) - F_\epsilon(s)\bigr]\,dF_{\boldsymbol{X}}(\boldsymbol{x}) \\
&\quad = f_\epsilon(s)s\int [d_{n2}(\boldsymbol{x}) - 1]\,dF_{\boldsymbol{X}}(\boldsymbol{x}) + f_\epsilon(s)\int d_{n1}(\boldsymbol{x})\,dF_{\boldsymbol{X}}(\boldsymbol{x}) \\
&\qquad + \int f_\epsilon'\bigl(s + s_n^*(s,\boldsymbol{x})\bigr)\bigl\{s[d_{n2}(\boldsymbol{x}) - 1] + d_{n1}(\boldsymbol{x})\bigr\}^2\,dF_{\boldsymbol{X}}(\boldsymbol{x}) \\
&\quad = R_{n,2}(s) + R_{n,3}(s) + R_{n,4}(s),
\end{aligned} \tag{A.9}$$

where $s_n^*(s,\boldsymbol{x})$ is between 0 and $s[d_{n2}(\boldsymbol{x}) - 1] + d_{n1}(\boldsymbol{x})$.

Step 2.1. Recall the definition of $\hat{g}(u,\boldsymbol{\beta})$ and $d_{n1}(\boldsymbol{x})$, we have that

$$\begin{aligned}
&d_{n1}(\boldsymbol{x}) \\
&\quad = \frac{T_{n,01}(\hat{\boldsymbol{\beta}}_0^\tau\boldsymbol{x}, \hat{\boldsymbol{\beta}}_0) - T_{n,00}(\hat{\boldsymbol{\beta}}_0^\tau\boldsymbol{x}, \hat{\boldsymbol{\beta}}_0)g(\hat{\boldsymbol{\beta}}_0^\tau\boldsymbol{x}) + \frac{T_{n,10}^2(\hat{\boldsymbol{\beta}}_0^\tau\boldsymbol{x}, \hat{\boldsymbol{\beta}}_0)}{T_{n,20}(\hat{\boldsymbol{\beta}}_0^\tau\boldsymbol{x}, \hat{\boldsymbol{\beta}}_0)}g(\hat{\boldsymbol{\beta}}_0^\tau\boldsymbol{x})}{\sigma(\boldsymbol{\beta}_0^\tau\boldsymbol{x})\Bigl(T_{n,00}(\hat{\boldsymbol{\beta}}_0^\tau\boldsymbol{x}, \hat{\boldsymbol{\beta}}_0) - \frac{T_{n,10}^2(\hat{\boldsymbol{\beta}}_0^\tau\boldsymbol{x}, \hat{\boldsymbol{\beta}}_0)}{T_{n,20}(\hat{\boldsymbol{\beta}}_0^\tau\boldsymbol{x}, \hat{\boldsymbol{\beta}}_0)}\Bigr)} \\
&\qquad - \frac{T_{n,10}(\hat{\boldsymbol{\beta}}_0^\tau\boldsymbol{x}, \hat{\boldsymbol{\beta}}_0)T_{n,11}(\hat{\boldsymbol{\beta}}_0^\tau\boldsymbol{x}, \hat{\boldsymbol{\beta}}_0)}{T_{n,20}(\hat{\boldsymbol{\beta}}_0^\tau\boldsymbol{x}, \hat{\boldsymbol{\beta}}_0)T_{n,00}(\hat{\boldsymbol{\beta}}_0^\tau\boldsymbol{x}, \hat{\boldsymbol{\beta}}_0) - T_{n,10}^2(\hat{\boldsymbol{\beta}}_0^\tau\boldsymbol{x}, \hat{\boldsymbol{\beta}}_0)} \\
&\qquad + \frac{g(\hat{\boldsymbol{\beta}}_0^\tau\boldsymbol{x}) - g(\boldsymbol{\beta}_0^\tau\boldsymbol{x})}{\sigma(\boldsymbol{\beta}_0^\tau\boldsymbol{x})}.
\end{aligned} \tag{A.10}$$

From Theorem 2.1, we have $\hat{\boldsymbol\beta}_0 = \boldsymbol\beta_0 + O_P(n^{-1/2})$. Using Lemma A.1, we have that

$$\frac{1}{nh}T_{n,1m}(\hat{\boldsymbol\beta}_0^\tau \boldsymbol{x}, \hat{\boldsymbol\beta}_0) = O_P\Big(h + \sqrt{\frac{(\log n)^{1+s}}{nh}}\Big), \qquad m = 0, 1.$$

Moreover

$$\begin{aligned}
&\frac{1}{n}T_{n,01}(\hat{\boldsymbol\beta}_0^\tau \boldsymbol{x}, \hat{\boldsymbol\beta}_0) - \frac{1}{n}T_{n,00}(\hat{\boldsymbol\beta}_0^\tau \boldsymbol{x}, \hat{\boldsymbol\beta}_0) g(\hat{\boldsymbol\beta}_0^\tau \boldsymbol{x}) \\
&\quad = \frac{1}{n}\sum_{i=1}^n K_h(\boldsymbol\beta_0^\tau \boldsymbol{X}_i - \boldsymbol\beta_0^\tau \boldsymbol{x})\sigma(\boldsymbol\beta_0^\tau \boldsymbol{X}_i)\epsilon_i \\
&\qquad - \frac{1}{n}\sum_{i=1}^n K_h(\boldsymbol\beta_0^\tau \boldsymbol{X}_i - \boldsymbol\beta_0^\tau \boldsymbol{x})[g(\hat{\boldsymbol\beta}_0^\tau \boldsymbol{X}_i) - g(\boldsymbol\beta_0^\tau \boldsymbol{X}_i)] \\
&\qquad + \frac{1}{n}\sum_{i=1}^n [K_h(\hat{\boldsymbol\beta}_0^\tau \boldsymbol{X}_i - \hat{\boldsymbol\beta}_0^\tau \boldsymbol{x}) - K_h(\boldsymbol\beta_0^\tau \boldsymbol{X}_i - \boldsymbol\beta_0^\tau \boldsymbol{x})]\sigma(\boldsymbol\beta_0^\tau \boldsymbol{X}_i)\epsilon_i \qquad\qquad (A.11) \\
&\qquad + \frac{1}{n}\sum_{i=1}^n K_h(\hat{\boldsymbol\beta}_0^\tau \boldsymbol{X}_i - \hat{\boldsymbol\beta}_0^\tau \boldsymbol{x})[g(\hat{\boldsymbol\beta}_0^\tau \boldsymbol{X}_i) - g(\hat{\boldsymbol\beta}_0^\tau \boldsymbol{x})] \\
&\quad \stackrel{\text{def}}{=} D_{n,1} - D_{n,2} + D_{n,3} + D_{n,4}.
\end{aligned}$$

For $D_{n,2}$, as $h \to 0$, $\frac{(\log n)^{1+s}}{nh} \to 0$, Taylor expansion and Lemma A.1 entail that

$$\begin{aligned}
D_{n,2} &= \frac{1}{n}\sum_{i=1}^n K_h(\boldsymbol\beta_0^\tau \boldsymbol{X}_i - \boldsymbol\beta_0^\tau \boldsymbol{x}) g'(\boldsymbol\beta_0^\tau \boldsymbol{X}_i)\boldsymbol{X}_i^\tau(\hat{\boldsymbol\beta}_0 - \boldsymbol\beta_0) \\
&\quad + O_P(\|\hat{\boldsymbol\beta}_0 - \boldsymbol\beta_0\|_2^2) \qquad\qquad (A.12) \\
&= f_{\boldsymbol\beta_0^\tau X}(\boldsymbol\beta_0^\tau \boldsymbol{x}) g'(\boldsymbol\beta_0^\tau \boldsymbol{x}) V^\tau(\boldsymbol\beta_0^\tau \boldsymbol{x})(\hat{\boldsymbol\beta}_0 - \boldsymbol\beta_0) + o_P(n^{-1/2}).
\end{aligned}$$

Let $K_h'(\cdot) = h^{-1}K'(\cdot)$, and $\tilde{\boldsymbol\beta}_0$ is between $\hat{\boldsymbol\beta}_0$ and $\boldsymbol\beta_0$. Using lemma A.1, we have

$$\begin{aligned}
D_{n,3} &= \frac{1}{nh}\sum_{i=1}^n \sigma(\boldsymbol\beta_0^\tau \boldsymbol{X}_i)\epsilon_i K_h'(\tilde{\boldsymbol\beta}_0^\tau \boldsymbol{X}_i - \tilde{\boldsymbol\beta}_0^\tau \boldsymbol{x})(\boldsymbol{X}_i - \boldsymbol{x})^\tau(\hat{\boldsymbol\beta}_0 - \boldsymbol\beta_0) \qquad (A.13) \\
&= o_P(n^{-1/2}),
\end{aligned}$$

and as $h(\log n)^{1+s} \to 0$, $nh^4 \to 0$,

$$\begin{aligned}
D_{n,4} &= hg'(\hat{\boldsymbol\beta}_0^\tau \boldsymbol{x})\frac{1}{n}\sum_{i=1}^n K_h(\hat{\boldsymbol\beta}_0^\tau \boldsymbol{X}_i - \hat{\boldsymbol\beta}_0^\tau \boldsymbol{x})\Big(\frac{\hat{\boldsymbol\beta}_0^\tau \boldsymbol{X}_i - \hat{\boldsymbol\beta}_0^\tau \boldsymbol{x}}{h}\Big) + O_P(h^2) \qquad (A.14) \\
&= O_P\Big(h^2 + h\Big(h + \sqrt{\frac{(\log n)^{1+s}}{nh}}\Big)\Big) = o_P(n^{-1/2}).
\end{aligned}$$

Note that $\frac{1}{nh}T_{n,00}(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{x}, \hat{\boldsymbol{\beta}}_0) = f_{\boldsymbol{\beta}_0^\tau X}(\boldsymbol{\beta}_0^\tau \boldsymbol{x}) + O_P(h^2 + \sqrt{\frac{(\log n)^{1+s}}{nh}})$, together with (A.10)–(A.14), we have that

$$d_{n1}(\boldsymbol{x}) = \frac{\frac{1}{n}\sum_{i=1}^n K_h(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i - \boldsymbol{\beta}_0^\tau \boldsymbol{x})\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)\epsilon_i}{f_{\boldsymbol{\beta}_0^\tau X}(\boldsymbol{\beta}_0^\tau \boldsymbol{x})\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{x})} + \frac{g'(\boldsymbol{\beta}_0^\tau \boldsymbol{x})[\boldsymbol{x} - V(\boldsymbol{\beta}_0^\tau \boldsymbol{x})]^\tau(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0)}{\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{x})} + o_P(n^{-1/2}). \tag{A.15}$$

Step 2.2. Similar to the analysis of (A.10), we have that

$$\begin{aligned}
&\hat{\sigma}^2(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{x}, \hat{\boldsymbol{\beta}}_0) - \sigma^2(\boldsymbol{\beta}_0^\tau \boldsymbol{x}) \\
&= f_{\boldsymbol{\beta}_0^\tau X}^{-1}(\boldsymbol{\beta}_0^\tau \boldsymbol{x})\frac{1}{n}\sum_{i=1}^n K_h(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i - \boldsymbol{\beta}_0^\tau \boldsymbol{x})\sigma^2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)(\epsilon_i^2 - 1) \\
&\quad - f_{\boldsymbol{\beta}_0^\tau X}^{-1}(\boldsymbol{\beta}_0^\tau \boldsymbol{x})\frac{2}{n}\sum_{i=1}^n K_h(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i - \boldsymbol{\beta}_0^\tau \boldsymbol{x})\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)\epsilon_i \\
&\quad \times (\hat{g}(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}_i, \hat{\boldsymbol{\beta}}_0) - g(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)) \\
&\quad + o_P(n^{-1/2}) \\
&\stackrel{\text{def}}{=} D_{n,5} - D_{n,6} + o_P(n^{-1/2}).
\end{aligned}$$

Directly using (A.15), we have that

$$\begin{aligned}
D_{n,6} &= f_{\boldsymbol{\beta}_0^\tau X}^{-1}(\boldsymbol{\beta}_0^\tau \boldsymbol{x})\frac{2}{n}\sum_{i=1}^n K_h(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i - \boldsymbol{\beta}_0^\tau \boldsymbol{x})\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)g'(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)\epsilon_i \\
&\quad \times (\boldsymbol{X}_i - V(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i))^\tau(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0) \\
&\quad + f_{\boldsymbol{\beta}_0^\tau X}^{-1}(\boldsymbol{\beta}_0^\tau \boldsymbol{x})\frac{2}{n^2}\sum_{i=1}^n\sum_{s=1}^n f_{\boldsymbol{\beta}_0^\tau X}^{-1}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)K_h(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i - \boldsymbol{\beta}_0^\tau \boldsymbol{x}) \\
&\quad \times K_h(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_s - \boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_s)\epsilon_i\epsilon_s \\
&\stackrel{\text{def}}{=} D_{n,6}[1] + D_{n,6}[2].
\end{aligned}$$

Using Lemma A.1 and $\hat{\boldsymbol{\beta}}_0 = \boldsymbol{\beta}_0 + O_P(n^{-1/2})$, we have that $D_{n,6}[1] = o_P(n^{-1/2})$. The projection of U-statistics with second order (Serfling (1980), Section 5.5.2) entails that $nD_{n,6}[2]$ converges in distribution to a weighted sum of independent χ^2 random variables, i.e., $D_{n,6}[2] = O_P(n^{-1}) = o_P(n^{-1/2})$. Then,

$$\begin{aligned}
&d_{n2}^2(\boldsymbol{x}) - 1 \\
&= \frac{\hat{\sigma}^2(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{x}, \hat{\boldsymbol{\beta}}_0) - \sigma^2(\boldsymbol{\beta}_0^\tau \boldsymbol{x})}{\sigma^2(\boldsymbol{\beta}_0^\tau \boldsymbol{x})}
\end{aligned} \tag{A.16}$$

$$= \frac{1}{\sigma^2(\boldsymbol{\beta}_0^\tau \boldsymbol{x}) f_{\boldsymbol{\beta}_0^\tau X}(\boldsymbol{\beta}_0^\tau \boldsymbol{x})} \frac{1}{n} \sum_{i=1}^n K_h(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i - \boldsymbol{\beta}_0^\tau \boldsymbol{x}) \sigma^2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)(\epsilon_i^2 - 1) + o_P(n^{-1/2}).$$

Using (A.16), we have that

$$d_{n2}(\boldsymbol{x}) - 1 = \frac{1}{2\sigma^2(\boldsymbol{\beta}_0^\tau \boldsymbol{x}) f_{\boldsymbol{\beta}_0^\tau X}(\boldsymbol{\beta}_0^\tau \boldsymbol{x})} \frac{1}{n} \times \sum_{i=1}^n K_h(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i - \boldsymbol{\beta}_0^\tau \boldsymbol{x}) \sigma^2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)(\epsilon_i^2 - 1) + o_P(n^{-1/2}). \tag{A.17}$$

Step 2.3. Note that $E[\boldsymbol{X} - V(\boldsymbol{\beta}_0^\tau \boldsymbol{X})] = 0$, then $\int \frac{g'(\boldsymbol{\beta}_0^\tau \boldsymbol{x})[\boldsymbol{x} - V(\boldsymbol{\beta}_0^\tau \boldsymbol{x})]}{\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{x})} dF_{\boldsymbol{X}}(\boldsymbol{x}) = E[\frac{g'(\boldsymbol{\beta}_0^\tau \boldsymbol{X})[\boldsymbol{X} - V(\boldsymbol{\beta}_0^\tau \boldsymbol{X})]}{\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})}] = 0$. Together with (A.9), (A.15) and (A.17), we have

$$\begin{aligned} &F_{\hat{\epsilon}}(s|\mathscr{H}_n) - F_\epsilon(s) \\ &= s f_\epsilon(s) \frac{1}{2n} \sum_{i=1}^n (\epsilon_i^2 - 1)\sigma^2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i) \int \frac{K_h(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i - \boldsymbol{\beta}_0^\tau \boldsymbol{x})}{\sigma^2(\boldsymbol{\beta}_0^\tau \boldsymbol{x}) f_{\boldsymbol{\beta}_0^\tau X}(\boldsymbol{\beta}_0^\tau \boldsymbol{x})} dF_{\boldsymbol{X}}(\boldsymbol{x}) \\ &\quad + f_\epsilon(s) \frac{1}{n} \sum_{i=1}^n \sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)\epsilon_i \\ &\quad \times \int \frac{K_h(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i - \boldsymbol{\beta}_0^\tau \boldsymbol{x})}{f_{\boldsymbol{\beta}_0^\tau X}(\boldsymbol{\beta}_0^\tau \boldsymbol{x}) \sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{x})} dF_{\boldsymbol{X}}(\boldsymbol{x}) + o_P(n^{-1/2}) \\ &= s f_\epsilon(s) \frac{1}{2n} \sum_{i=1}^n (\epsilon_i^2 - 1) + f_\epsilon(s) \frac{1}{n} \sum_{i=1}^n \epsilon_i + o_P(n^{-1/2}). \end{aligned} \tag{A.18}$$

Moreover, (A.15), (A.17) and Condition (C5) entail $R_{n,4}(s) = o_P(n^{-1/2})$ uniformly in s. Together with (A.8), (A.9) and (A.18), we have completed the proof of Theorem 3.1. □

A.3 Proof of Theorem 3.2

Recalling $F_\epsilon^*(s) = E[F_\epsilon(s + m_c - \frac{g(\boldsymbol{\beta}_0^\tau \boldsymbol{X})}{\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})})]$, $m_c = \frac{E[g(\boldsymbol{\beta}_0^\tau \boldsymbol{X})]}{E[\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})]}$, and we further define $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$ and $\bar{\hat{\sigma}} = \frac{1}{n}\sum_{i=1}^n \hat{\sigma}(\hat{\boldsymbol{\beta}}_0 \boldsymbol{X}_i, \hat{\boldsymbol{\beta}}_0)$.

$$\begin{aligned} \hat{F}_{0\epsilon}(s) - F_\epsilon^*(s) &= \frac{1}{n} \sum_{i=1}^n I\{\hat{\epsilon}_{0i} \le s\} - F_\epsilon^*(s) \\ &= \frac{1}{n} \sum_{i=1}^n I\left\{\epsilon_i + \frac{g(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)}{\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)} - m_c \le s\right\} - F_\epsilon^*(s) \\ &\quad + [F_{0\hat{\epsilon}}(s|\mathscr{H}_n) - F_\epsilon^*(s)] + Q_{n,1}(s), \end{aligned} \tag{A.19}$$

where $F_{0\hat{\epsilon}}(s|\mathscr{H}_n)$ be the distribution function of $\hat{\epsilon}_0 = \frac{Y}{\hat{\sigma}(\hat{\boldsymbol{\beta}}_0^\tau \boldsymbol{X}, \hat{\boldsymbol{\beta}}_0)} - \bar{Y}/\bar{\hat{\sigma}}$ conditional on the data $\mathscr{H}_n = \{\boldsymbol{X}_i, Y_i\}_{i=1}^n$, and

$$Q_{n,1}(s) = \frac{1}{n}\sum_{i=1}^{n}\big[I\{\hat{\epsilon}_{0i} \le s\} - I\{\epsilon_i + g(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)/\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i) - m_c \le s\}\big]$$
$$+ F_\epsilon^*(s) - F_{0\hat{\epsilon}}(s|\mathscr{H}_n) = o_P(n^{-1/2}),$$

uniformly in $s \in \mathbb{R}$. Similar to Lemma A.2, we have $\sup_{-\infty<s<\infty}|Q_{n,1}(s)| = o_P(n^{-1/2})$. Directly using (A.17) and the projection of U-statistics (Serfling (1980), Section 5.3.1), we have that

$$\begin{aligned}
&\bar{\hat{\sigma}} - E[\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})] \\
&\quad = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{2\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i) f_{\boldsymbol{\beta}_0^\tau \boldsymbol{X}}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)} K_h(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i - \boldsymbol{\beta}_0^\tau \boldsymbol{X}_j)\sigma^2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_j) \\
&\qquad \times (\epsilon_j^2 - 1) \\
&\qquad + \frac{1}{n}\sum_{i=1}^{n}\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i) - E[\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})] + o_P(n^{-1/2}) \\
&\quad = \frac{1}{n}\sum_{i=1}^{n}\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i) - E[\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})] + \frac{1}{2n}\sum_{i=1}^{n}\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)(\epsilon_i^2 - 1) + o_P(n^{-1/2}).
\end{aligned} \tag{A.20}$$

From (A.20), we have that $\bar{\hat{\sigma}} - E[\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})] = O_P(n^{-1/2})$ and

$$\begin{aligned}
&\bar{Y}/\bar{\hat{\sigma}} - m_c \\
&\quad = \frac{\bar{Y} - E[Y]}{E[\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})]} - \frac{E[Y]}{(E[\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})])^2}(\bar{\hat{\sigma}} - E[\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})]) + o_P(n^{-1/2}).
\end{aligned} \tag{A.21}$$

Then, (A.20) and (A.21) entail that $\bar{Y}/\bar{\hat{\sigma}} - m_c = o_P(1)$. As a result, we use the Taylor expansion and have that

$$\begin{aligned}
&F_{0\hat{\epsilon}}(s|\mathscr{H}_n) - F_\epsilon^*(s) \\
&\quad = \int F_\epsilon\bigg(s + m_c - \frac{g(\boldsymbol{\beta}_0^\tau \boldsymbol{x})}{\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{x})} + (s + \bar{Y}/\bar{\hat{\sigma}})[d_{n2}(\boldsymbol{x}) - 1] + \frac{\bar{Y}}{\bar{\hat{\sigma}}} - m_c\bigg) dF_{\boldsymbol{X}}(\boldsymbol{x}) \\
&\qquad - F_\epsilon^*(s) \\
&\quad = \int f_\epsilon\bigg(s + m_c - \frac{g(\boldsymbol{\beta}_0^\tau \boldsymbol{x})}{\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{x})}\bigg) \\
&\qquad \times \bigg[(s + \bar{Y}/\bar{\hat{\sigma}})[d_{n2}(\boldsymbol{x}) - 1] + \frac{\bar{Y}}{\bar{\hat{\sigma}}} - m_c\bigg] dF_{\boldsymbol{X}}(\boldsymbol{x})
\end{aligned} \tag{A.22}$$

$$
\begin{aligned}
&+\int f'_\epsilon\Big(s+m_c-\frac{g(\boldsymbol{\beta}_0^\tau \boldsymbol{x})}{\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{x})}+s_n^{**}(s,\boldsymbol{x})\Big)\\
&\times\Big[\Big(s+\frac{\bar{Y}}{\bar{\hat{\sigma}}}\Big)[d_{n2}(\boldsymbol{x})-1]+\frac{\bar{Y}}{\bar{\hat{\sigma}}}-m_c\Big]^2\,dF_{\boldsymbol{X}}(\boldsymbol{x})\\
&=Q_{n,2}(s)+Q_{n,3}(s),
\end{aligned}
$$

where $s_n^{**}(s,\boldsymbol{x})$ is between 0 and $(s+\bar{Y}/\bar{\hat{\sigma}})[d_{n2}(\boldsymbol{x})-1]+\bar{Y}/\bar{\hat{\sigma}}-m_c$. Recalling the definition of $m_{f_\epsilon,c}(s)=E[f_\epsilon(s+m_c-\frac{g(\boldsymbol{\beta}_0^\tau \boldsymbol{X})}{\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})})]$, then

$$
\begin{aligned}
&Q_{n,2}(s)\\
&\quad=\frac{s+\bar{Y}/\bar{\hat{\sigma}}}{2n}\sum_{i=1}^n\int f_\epsilon\Big(s+m_c-\frac{g(\boldsymbol{\beta}_0^\tau \boldsymbol{x})}{\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{x})}\Big)\frac{K_h(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i-\boldsymbol{\beta}_0^\tau \boldsymbol{x})}{\sigma^2(\boldsymbol{\beta}_0^\tau \boldsymbol{x})f_{\boldsymbol{\beta}_0^\tau X}(\boldsymbol{\beta}_0^\tau \boldsymbol{x})}\,dF_{\boldsymbol{X}}(\boldsymbol{x})\\
&\qquad\times\sigma^2(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)(\epsilon_i^2-1)+(\bar{Y}/\bar{\hat{\sigma}}-m_c)m_{f_\epsilon,c}(s)+o_P(n^{-1/2}) \qquad\qquad (A.23)\\
&\quad=\frac{1}{2n}\sum_{i=1}^n(s+m_c)f_\epsilon\Big(s+m_c-\frac{g(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)}{\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{x}_i)}\Big)(\epsilon_i^2-1)\\
&\qquad+m_{f_\epsilon,c}(s)(\bar{Y}/\bar{\hat{\sigma}}-m_c)+o_P(n^{-1/2}).
\end{aligned}
$$

Similar to the analysis of (A.23), we have that $Q_{n,3}(s)=o_P(n^{-1/2})$ uniformly in $s\in\mathbb{R}$. From (A.19)–(A.23), we have completed the proof of Theorem 3.2.

A.4 Proofs of Theorems 3.4–3.5

Proof. Under the local alternative hypothesis $\mathcal{H}_{1n}$, we have that

$$
\begin{aligned}
&n^{-1/2}\mathcal{W}_n(\boldsymbol{\beta}_0^{(1)})\\
&\quad=n^{-1/2}\sum_{i=1}^n J_{\boldsymbol{\beta}_0}^\tau c\sigma'(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)[\boldsymbol{X}_i-V(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)]\sigma^{-1}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)\epsilon_i\\
&\qquad+n^{-1}\sum_{i=1}^n J_{\boldsymbol{\beta}_0}^\tau c\sigma'(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)[\boldsymbol{X}_i-V(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)]\sigma^{-2}(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)\gamma(\boldsymbol{X}_i)\qquad (A.24)\\
&\qquad+o_P(1).
\end{aligned}
$$

It is easily seen that the second summation of (A.24) converges in probability to

$$
cJ_{\boldsymbol{\beta}_0}^\tau E\Big[\frac{\sigma'(\boldsymbol{\beta}_0^\tau \boldsymbol{X})}{\sigma^2(\boldsymbol{\beta}_0^\tau \boldsymbol{X})}(\boldsymbol{X}-V(\boldsymbol{\beta}_0^\tau \boldsymbol{X}))\gamma(\boldsymbol{X})\Big].
$$

Moreover, using the fact that $Y_i=c\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)+n^{-1/2}\gamma(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)+\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)\epsilon_i$ under the local alternative hypothesis $\mathcal{H}_{1n}$, similar to the analysis of (A.10) and (A.17), we have that

$$
\begin{aligned}
d_{c,n1}(\boldsymbol{x}) &= \frac{\hat{g}(\hat{\boldsymbol{\beta}}_0^{\tau}\boldsymbol{x}, \hat{\boldsymbol{\beta}}_0) - c\sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x})}{\sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x})} \\
&= \frac{\frac{1}{n}\sum_{i=1}^{n} K_h(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_i - \boldsymbol{\beta}_0^{\tau}\boldsymbol{x})[\sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_i)\epsilon_i + n^{-1/2}\gamma(\boldsymbol{X}_i)]}{f_{\boldsymbol{\beta}_0^{\tau}X}(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x})\sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x})} \\
&\quad + \frac{c\sigma'(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x})[\boldsymbol{x} - V(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x})]^{\tau}(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0)}{\sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x})} + o_P\big(n^{-1/2}\big) \\
&= \frac{\frac{1}{n}\sum_{i=1}^{n} K_h(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_i - \boldsymbol{\beta}_0^{\tau}\boldsymbol{x})\sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_i)\epsilon_i}{f_{\boldsymbol{\beta}_0^{\tau}X}(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x})\sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x})} \\
&\quad + \frac{c\sigma'(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x})[\boldsymbol{x} - V(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x})]^{\tau}(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}_0)}{\sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x})} \\
&\quad + \frac{1}{\sqrt{n}}\frac{E[\gamma(\boldsymbol{X})|\boldsymbol{\beta}_0^{\tau}\boldsymbol{X} = \boldsymbol{\beta}_0^{\tau}\boldsymbol{x}]}{\sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x})} + o_P\big(n^{-1/2}\big),
\end{aligned}
\tag{A.25}
$$

and similar to the analysis of (A.25), we have that $d_{n2}(\boldsymbol{x}) - 1$ has the same asymptotic expression with (A.17) under the local alternative hypothesis $\mathcal{H}_{1n}$. Moreover, together with (A.25), (A.9) becomes to

$$
\begin{aligned}
&F_{\hat{\epsilon}}(s|\mathscr{H}_n) - F_{\epsilon}(s) \\
&\quad \stackrel{\text{def}}{=} \int F_{\epsilon}\Big(s + s\big[d_{n2}(\boldsymbol{x}) - 1\big] + d_{c,n1}(\boldsymbol{x}) - n^{-1/2}\frac{\gamma(\boldsymbol{x})}{\sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x})}\Big)\,dF_{\boldsymbol{X}}(\boldsymbol{x}) \\
&\qquad - F_{\epsilon}(s) \\
&\quad = f_{\epsilon}(s)\Big(\int d_{c,n1}(\boldsymbol{x})\,dF_{\boldsymbol{X}}(\boldsymbol{x}) - \frac{1}{\sqrt{n}}E\Big[\frac{\gamma(\boldsymbol{X})}{\sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X})}\Big]\Big) \\
&\qquad + f_{\epsilon}(s)s\int\big[d_{n2}(\boldsymbol{x}) - 1\big]\,dF_{\boldsymbol{X}}(\boldsymbol{x}) + o_P\big(n^{-1/2}\big) \\
&\quad = sf_{\epsilon}(s)\frac{1}{2n}\sum_{i=1}^{n}\big(\epsilon_i^2 - 1\big) + f_{\epsilon}(s)\frac{1}{n}\sum_{i=1}^{n}\epsilon_i + o_P\big(n^{-1/2}\big).
\end{aligned}
\tag{A.26}
$$

We see that the asymptotic expression (A.26) is also the same as the one obtained in Theorem 3.1. Moreover, similar to the analysis of (A.20) and (A.21), using (A.26) and that $\bar{Y} = \frac{c}{n}\sum_{i=1}^{n}\sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_i) + \frac{1}{\sqrt{n}}\frac{1}{n}\sum_{i=1}^{n}\gamma(\boldsymbol{X}_i) + \frac{1}{n}\sum_{i=1}^{n}\sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{X}_i)\epsilon_i$, we have that

$$
\begin{aligned}
&F_{0\hat{\epsilon}}(s|\mathscr{H}_n) - F_{\epsilon}(s) \\
&\quad = \int F_{\epsilon}\Big(s + (s + \bar{Y}/\bar{\hat{\sigma}})\big[d_{n2}(\boldsymbol{x}) - 1\big] + \bar{Y}/\bar{\hat{\sigma}} - c - \frac{1}{\sqrt{n}}\frac{\gamma(\boldsymbol{x})}{\sigma(\boldsymbol{\beta}_0^{\tau}\boldsymbol{x})}\Big)\,dF_{\boldsymbol{X}}(\boldsymbol{x})
\end{aligned}
$$

$$
\begin{aligned}
&- F_\epsilon(s) \\
&= f_\epsilon(s) \int \bigg[(s + \overline{Y/\sigma})[d_{n2}(\boldsymbol{x}) - 1] + \bar{Y}/\bar{\hat{\sigma}} \\
&\quad - c - \frac{1}{\sqrt{n}} \frac{\gamma(\boldsymbol{x})}{\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{x})} \bigg] dF_{\boldsymbol{X}}(\boldsymbol{x}) \\
&\quad + o_P(n^{-1/2}) \\
&= f_\epsilon(s) \Bigg[\frac{s}{2n} \sum_{i=1}^n (\epsilon_i^2 - 1) + \frac{1}{n} \sum_{i=1}^n \frac{\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)}{E[\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})]} \epsilon_i \Bigg] \\
&\quad - f_\epsilon(s) \Bigg[\frac{c}{2n} \sum_{i=1}^n \bigg(\frac{\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X}_i)}{E[\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})]} - 1 \bigg) (\epsilon_i^2 - 1) \Bigg] \\
&\quad + \frac{f_\epsilon(s)}{\sqrt{n}} \bigg[\frac{E[\gamma(\boldsymbol{X})]}{E[\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})]} - E\bigg[\frac{\gamma(\boldsymbol{X})}{\sigma(\boldsymbol{\beta}_0^\tau \boldsymbol{X})} \bigg] \bigg] + o_P(n^{-1/2}).
\end{aligned}
\tag{A.27}
$$

Together with (A.26) and (A.27), we have completed the proof of Theorem 3.4. □

Acknowledgments

The authors thank the Editor, the Associate Editor and a referee for their constructive comments and suggestions, which led to the substantial improvement of an early manuscript. Cuizhen Niu's research was supported by the National Natural Science Foundation of China (Grant No. 11701034), and the Fundamental Research Funds for the Central Universities, China Postdoctoral Science Foundation (Grant No. 2016M600951). Gaorong Li's research is supported by the National Natural Science Foundation of China (Grant No. 11871001 and 11471029), and Beijing Natural Science Foundation (Grant No. 1182003). Jun Zhang's research is supported by the National Natural Science Foundation of China (Grant No. 11401391).

References

Akritas, M. G. and Van Keilegom, I. (2001). Non-parametric estimation of the residual distribution. *Scandinavian Journal of Statistics* **28**, 549–567. MR1858417

Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association* **92**, 477–489. MR1467842

Carroll, R. J. and Ruppert, D. (1988). *Transformation and Weighting in Regression*. London: Chapman and Hall. MR1014890

Cui, X., Härdle, W. K. and Zhu, L. (2011). The EFM approach for single-index models. *The Annals of Statistics* **39**, 1658–1688. MR2850216

Dette, H., Marchlewski, M. and Wagener, J. (2012). Testing for a constant coefficient of variation in nonparametric regression by empirical processes. *Annals of the Institute of Statistical Mathematics* **64**, 1045–1070. MR2960957

Dette, H., Pardo-Fernández, J. C. and Van Keilegom, I. (2009). Goodness-of-fit tests for multiplicative models with dependent data. *Scandinavian Journal of Statistics* **36**, 782–799. MR2573308

Dette, H. and Wieczorek, G. (2009). Testing for a constant coefficient of variation in nonparametric regression. *Journal of Statistical Theory and Practice* **3**, 587–612. MR2750460

Eagleson, G. K. and Müller, H. G. (1997). Transformations for smooth regression models with multiplicative errors. *Journal of the Royal Statistical Society, Series B* **59**, 173–189. MR1436562

Feng, Z., Wen, X. M., Yu, Z. and Zhu, L. (2013). On partial sufficient dimension reduction with applications to partially linear multi-index models. *Journal of the American Statistical Association* **108**, 237–246. MR3174616

Guo, X., Wang, T. and Zhu, L. (2016). Model checking for parametric single-index models: A dimension reduction model-adaptive approach. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **78**, 1013–1035. MR3557187

Härdle, W. K., Hall, P. and Ichimura, H. (1993). Optimal smoothing in single-index models. *The Annals of Statistics* **21**, 157–178. MR1212171

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics* **58**, 71–120. MR1230981

Li, G., Peng, H., Dong, K. and Tong, T. (2014). Simultaneous confidence bands and hypothesis testing for single-index models. *Statistica Sinica* **24**, 937–955. MR3235406

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

Liang, H. and Wang, N. (2005). Partially linear single-index measurement error models. *Statistica Sinica* **15**, 99–116. MR2125722

Neumeyer, N. (2009). Smooth residual bootstrap for empirical processes of nonparametric regression residuals. *Scandinavian Journal of Statistics* **36**, 204–228. MR2528982

Neumeyer, N. and Van Keilegom, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *Journal of Multivariate Analysis* **101**, 1067–1078. MR2595293

Peng, H. and Huang, T. (2011). Penalized least squares for single index models. *Journal of Statistical Planning and Inference* **141**, 1362–1379. MR2747907

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. New York: Wiley. MR0595165

Stute, W. and Zhu, L. (2005). Nonparametric checks for single-index models. *The Annals of Statistics* **33**, 1048–1083. MR2195628

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. *Springer Series in Statistics*. New York: Springer. With applications to statistics. MR1385671

Wang, T., Xu, P. and Zhu, L. (2012). Non-convex penalized estimation in high-dimensional models with single-index structure. *Journal of Multivariate Analysis* **109**, 221–235. MR2922865

Wang, T., Xu, P. and Zhu, L. (2015). Variable selection and estimation for semi-parametric multiple-index models. *Bernoulli* **21**, 242–275. MR3322318

Wang, T. and Zhu, L. (2015). A distribution-based LASSO for a general single-index model. *Science China Mathematics* **58**, 109–130. MR3296334

Xia, Y. (2006). Asymptotic distributions for two estimators of the single-index model. *Econometric Theory* **22**, 1112–1137. MR2328530

Xia, Y. and Härdle, W. K. (2006). Semi-parametric estimation of partially linear single-index models. *Journal of Multivariate Analysis* **97**, 1162–1184. MR2276153

Xia, Y., Tong, H., Li, W. and Zhu, L. (2002). An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society, Series B* **64**, 363–410. MR1924297

Zhu, L., Lin, L., Cui, X. and Li, G. (2010). Bias-corrected empirical likelihood in a multi-link semiparametric model. *Journal of Multivariate Analysis* **101**, 850–868. MR2584904

J. Zhang
College of Mathematics and Statistics
Institute of Statistical Sciences
Shenzhen University
Shenzhen, 518060
China

C. Niu
School of Statistics
Beijing Normal University
Beijing, 100875
China
E-mail: nczlbc_890@126.com

G. Li
Beijing Institute for Scientific
and Engineering Computing
Beijing University of Technology
Beijing, 100124
China