IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received January 9, 2020, accepted January 21, 2020, date of publication January 30, 2020, date of current version February 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2970506

Checking Heteroscedasticity in Partially Linear Single-Index Models Using Pairwise Distance

JIAN-QIANG ZHAO[1], JIANQUAN LI[2], YAN-YONG ZHAO[3], JUNYAN HE[1], AND WALED KHALED[4]

[1]School of Mathematics and Statistics, Xuzhou University of Technology, Xuzhou 221018, China
[2]Beijing Ultrapower Software Company Ltd., Beijing 100020, China
[3]Department of Statistics, Nanjing Audit University, Nanjing 211815, China
[4]School of Mathematics, Southeast University, Nanjing 210096, China

Corresponding author: Yan-Yong Zhao (zhaoyanyong1987@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 11701286 and Grant 11961038, in part by the Natural Science Foundation of Jiangsu Province of China under Grant BK20171073, in part by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China under Grant 17KJB110006, and in part by the Xuzhou University of Technology Research Funds under Grant XKY2018120.

ABSTRACT In this article, a new test is proposed for partially linear single-index models (PLSIM) based on the pairwise distances of the sample points, to test heteroscedasticity. The statistic can be formulated as a U statistic and does not have to estimate the conditional variance function by using nonparametric methods, such as kernel, local polynomial, or spline. We derive a computationally feasible approximation to deal with the complexity of the limit zero distribution under the null hypothesis. We prove that the proposed bootstrap procedure is valid approximation to the null distribution of the test. It shows that this statistic has an asymptotically normal distribution. The algorithmic program of this test method is easy to implement and has faster convergence than some existing methods. In addition, convergence rate of the statistic does not depend on the dimensions of the covariates, which greatly reduces the impact of the dimensional curse. Finally, we give the numerical simulations and a real data example.

INDEX TERMS Dimension reduction, heteroscedasticity, partially linear single-index models.

I. INTRODUCTION

Due to the disadvantages of non-parametric models, such as curse of dimension, difficulty in interpretation and insufficient extrapolation ability, etc, some semi-parametric regression models are used to overcome these shortcomings. Based on this, we study the following partially linear single-index model (PLSIM), which is a very important semi-parametric regression model:

$$Y = g(X^T\theta) + Z^T\beta + \varepsilon, \tag{1}$$

where $X = (X_1, \cdots, X_p)^T \in R^p$ and $Z = (Z_1, \cdots, Z_q)^T \in R^q$ are covariates, $g(\cdot)$ is an unknown smooth link functions, ε is an independent random error with mean zero $E(\varepsilon|X, Z) = 0$. The parameter $\theta = (\theta_1, \cdots, \theta_p)^T \in R^p$ with $\|\theta\| = 1$ and $\theta_1 > 0$, $\beta = (\beta_1, \cdots, \beta_q)^T \in R^q$ are all unknown parameter. At present, there are many literatures to study the estimation of the parameters and the link function, see [1], [6], [9], [12]–[14], [16], [24], [27], [31].

We usually assume that the error terms in the PLSIM have a common variance. However, actual statistics often have heteroscedastic phenomena. Therefore, testing the statistical data for heteroscedasticity is an important issue. Our objective is to detect variance heterogeneity in aforementioned model (1) by testing the following hypothesis

$$H_0 : \exists\, \sigma^2 > 0,\ E(\varepsilon^2|X, Z) = \sigma^2(X, Z) = \sigma^2,$$
$$H_1 : \forall\, \sigma^2 > 0,\ E(\varepsilon^2|X, Z) \neq \sigma^2. \tag{2}$$

Under H_0, the constant σ^2 is an unconditional variance $E(\varepsilon^2)$. Consequently, the heteroscedasticity test in (2) is equivalent to determining whether the conditional variance function $E(\varepsilon^2|X, Z)$ is equal to the unconditional variance $E(\varepsilon^2)$.

Many authors have studied the heteroscedasticity test of common regression model, such as the literature [4], [7], [8], [10], [17], [18], [20], [22], [23], [28], [29]. However, there is not much literature on the heteroscedasticity test of PLSIM. Reference [30] studied the heteroscedasticity

The associate editor coordinating the review of this manuscript and approving it for publication was Amjad Gawanmeh.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

checks for single index models and [26] studied the detection of marginal heteroscedasticity for PLSIM. A drawback of some existing methods is the dimensionality problem due to the estimation inefficiency for the multivariate nonparametric function. Under the respective null hypotheses in these papers, the test statistics converge to their weak limits. When the dimension increases, the convergence speed for most of them is generally slower. Therefore, the significance level frequently cannot be well maintained when the limiting null distribution is used in moderate sample size scenarios. Asymptotically, these test statistics are less powerful for detecting alternative models.

In this paper, we formulate the proposed test statistic as a simple U-statistic for PLSIM based on the paired distances of the sample points, to test heteroscedasticity. This statistic is based on the weighted integral of the residual marked characteristic function. The weight function plays an important role in the proposed test statistic. The density function of a spherical stable law is used as the weight function. Given this particular choice, the weighted integral is transformed into an unconditional expectation with a simple form. The proposed statistic is based merely on pairwise distances between points in a sample. To the best of our knowledge, however, this study is the first to use characteristic function to detect heteroscedasticity for PLSIM. For theoretical investigations, the U-statistic theory can be applied instead of empirical process theory, and we investigate its asymptotic properties under the null, fixed alternative, and local alternative hypotheses. The asymptotic null distribution has a non-trivial form as the same as most cases for U-statistics.

The rest of this article is organized as follows. In Section II, the test procedure is presented and its asymptotic property is established. In Section III, we proposes a simple bootstrap algorithm to detect heteroscedasticity for the PLSIM. In Section IV, numerical studies to evaluate the performance of the tests are reported. In Section V, we carry out a real data example for illustrating the proposed methodology. Conclusion and discussion are given in Section VI. Technical assumptions and proofs are provided in Appendix.

II. THE TESTING PROCEDURE AND ASYMPTOTIC PROPERTIES

A. THE TEST STATISTIC

First, we let $r = \varepsilon^2 - \sigma^2$ with $\sigma^2 = E(\varepsilon^2)$, $W = (X, Z) \in R^{p+q}$. So, we can easily get $E(r|X, Z) = 0$, under the hypothesis H_0 in (2). According to the uniqueness of the Fourier transform of a function, we can do the following equivalent substitution for H_0:

$$H_0 : \phi(t) = E[re^{it^T W}] = 0, \quad \forall t \in R^{p+q}. \tag{3}$$

Because $\phi(t)$ is not a a statistic by itself, we can construct the following quantity:

$$D_\omega = \int_{R^{p+q}} |\phi(t)|^2 \omega(t)dt, \tag{4}$$

where $\omega(t) \geq 0$ is a suitable weight function. According to the definition of complex modulus, we have

$$|\phi(t)|^2 = E[cos(t^T(W - W'))rr'], \tag{5}$$

where (W', r') is an independent copy of (W, r). By the reference [15], we can get the following characteristic function of a spherical stable law in [25]:

$$\phi_z(t) = \int_{R^{p+q}} cos(t^T z)f_{a,p+q}(z)dz = e^{-\|t\|^a}, \tag{6}$$

where $\|\cdot\|$ is the Euclidean norm, and $f_{a,p+q}(\cdot)$ denotes the density of a spherical stable law in R^{p+q} with characteristic exponent $a \in (0, 2]$. The spherical stable family includes the multivariate Gaussian and Cauchy distributions as special cases, for $a = 2$ and $a = 1$, respectively. See reference [15] for the details.

We can choose the weight function $f_{a,p+q}(t)$, and get

$$D_\omega = E[e^{\|W-W'\|^a} rr']. \tag{7}$$

If the dimension $p+q$ of W is high, we will have some difficulties in dealing with the integral problem. But we can get a simple and closed form without involving high-dimensional integral, by the aforementioned weight function. So the hypothesis (2) is true if and only if $D_\omega = 0$, which can be used as a criterion for this hypothesis testing problem. When the $i-th$ sample r_i of r is available, we can estimate D_ω by its sample analogue.

Assume $(X_i, Z_i, Y_i) = (W_i, Y_i)$ are independently identical distributed (i.i.d) samples from $(X, Z, Y) = (W, Y)$, and ε_i are a independent random error with mean zero. We establish the following test statistic:

$$U_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \hat{r}_i \hat{r}_j d_{ij}, \tag{8}$$

where $\hat{r}_i = \hat{\varepsilon}_i^2 - \hat{\sigma}^2$, $\hat{\varepsilon}_i = Y_i - (\hat{g}(X_i^T\hat{\theta}) + Z_i^T\hat{\beta})$, $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n} \hat{\varepsilon}_i^2$, and $d_{ij} = e^{-\|W_i - W_j\|^a}$. $\hat{g}(\cdot)$, $\hat{\theta}$ and $\hat{\beta}$ are the estimators of $g(\cdot)$, θ and β, respectively, using a two-stage estimation method with a bandwidth parameter h (More details can be seen in [13]). With the distance measure D_w and d_{ij}, one advantage of this statistic is that we do not have to estimate the conditional variance function $E(\varepsilon^2|X, Z)$. Obviously, U_n is a moment-based test statistic and is easy to implement.

B. ASYMPTOTIC PROPERTIES

In this section, the asymptotic properties of U_n will be presented.

First, under the null hypothesis, we give the following theorem to state the limit distribution of U_n.

Theorem 1: Assume conditions in the Appendix hold and under the null hypothesis in (2), as $n \to \infty$ and $h \to 0$,

(i) if $nh^8 \to 0$, we have

$$nU_n \to \sum_{k=1}^{\infty} \lambda_k S_k^2 - E(r^2), \tag{9}$$

where S_k's are independent standard normal random variables, and λ_k's are the eigenvalues of the following integral equation:

$$\int r_j^2 \tilde{d}_{ij}\phi_k(A_j)dF(A_j) = \lambda_k \phi_k(A_i), \qquad (10)$$

where $\phi_k(A_j)$ are the associated orthonormal eigenfunctions; $A \equiv (W, r) \sim F(A)$ and $A_i \equiv (W_i, r_i)$ are independent copies of A, $\tilde{d}_{ij} = d_{ij} - E(d_{il} + d_{jl}|W_i, W_j) + E(d_{12})$ with $l \neq i, j$;

(ii) if $nh^8 \to \infty$, we have

$$\sqrt{n}h^{-4}(U_n - a(n)) \to N(0, \sigma_*^2).$$

where $a(n) = Q_{10}$ and $\sigma_*^2 > 0$ are defined in the Appendix.

If $nh^8 \to 0$, define $h(A_1, A_2) = r_1 r_2 \tilde{d}_{ij}$, due to $E(r^2) < \infty$, similar to the U-Statistics in reference [19], we get $E(h^2(A_1, A_2)) < \infty$ and $\sum_{k=1}^{\infty} \lambda_k^2 = E(h^2(A_1, A_2))$. So, $\sum_{k=1}^{\infty} \lambda_k S_k$ converges in L_2. As most cases for U-Statistics, the above limit distribution of U_n can not be applied directly for computing critical values because λ_k' s are not easy to obtain. If $nh^8 \to \infty$, the convergence rate is $h^4/\sqrt{n}$. Since it is difficult for us to directly calculate the critical value of U_n, a bootstrap approximation algorithm (in Section *III*) is designed to get the critical values.

Next, the sequence local substitution of the sensitivity test statistic we studied was $c_n\Delta(W)$, which have the following form:

$$H_{1n}: \ E(\varepsilon^2|W) = \sigma^2 + c_n\Delta(W), \qquad (11)$$

where $\sigma^2 = E(\varepsilon^2)$, c_n is a sequence of numbers converging to zero; $E(\Delta^2(W)) < \infty$, and $\Delta(W)$ is a function about W. Under H_{1n}, $r_i = \varepsilon_i^2 - \sigma^2$ can be rewritten as $r_i = u_i + c_n\Delta(W_i)$, where $E(u_i|W_i) = 0$ and $E(\Delta(W)) = 0$.

Then, we can get the following theorem based on the above hypothesis.

Theorem 2: Assume conditions in the Appendix hold and under the local alternative hypothesis in (11), as $n \to \infty$ and $h \to 0$,

(i) if $c_n = n^{-1/2}$ and $nh^8 \to 0$, we have

$$nU_n \to \sum_{k=1}^{\infty} \lambda_k(S_k + a_k)^2 - E(u^2), \qquad (12)$$

where $a_k = E(\Delta(W)\phi_k(A))$;

(ii) if $c_n = n^{-1/2}$ and $nh^8 \to \infty$, we have

$$\sqrt{n}h^{-4}(U_n - b(n)) \to N(0, \tilde{\sigma}_*^2),$$

where $b(n) = Q_{3n} + Q_{4n} + Q_{7n} + Q_{10n}$ and $\tilde{\sigma}_*^2 > 0$ are defined in the Appendix;

(iii) if $c_n = n^{-b}$ and $0 < b < 1/2$, we have $nU_n \to \infty$.

From this theorem, we can know that the test is still valid if the local alternative converges to the null hypothesis at a rate of $n^{-1/2}$. However, the asymptotic distribution of U_n when $nh^8 \to 0$ is different from that when $nh^8 \to \infty$. If we take a slower rate of $c_n = n^{-b}$, $0 < b < \frac{1}{2}$, the asymptotic power will tend to 1, which shows that the test is consistent.

If we set c_n be a fixed value $c \neq 0$, then H_{1n} turns to be the following hypothesis H_1, that is, from the local alternative hypothesis to the fixed alternative hypothesis:

$$H_1: \ E(\varepsilon^2|W) = \sigma^2 + c\Delta(W) \neq \sigma^2. \qquad (13)$$

Then, we can obtain the following theorem under H_1.

Theorem 3: Assume that conditions in the Appendix hold and under the fixed alternative hypothesis in (13), as $n \to \infty$ and $h \to 0$,

(i) if $nh^8 \to 0$, we have

$$\sqrt{n}(U_n - E(r_1 r_2 d_{12})) \to N(0, \tilde{\sigma}^2), \qquad (14)$$

where $\tilde{\sigma}^2 = var(r_1 E(r_2\, d_{12}\,|W_1\,) - 2E(r_1\, d_{12})r_1)$.

(ii) if $nh^8 \to \infty$, we have $nU_n \to \infty$.

From this theorem, we know that, if $nh^8 \to 0$, the divergence rate of U_n is also $n^{-1/2}$, which has a non-zero mean asymptote. Furthermore, we can show that the convergence rate of the statistic U_n is significantly different under hypothesis H_0 and hypothesis H_1, which does not depend on the dimension of W under different hypothesis. If $nh^8 \to \infty$, the asymptotic power of U_n tends to 1.

III. PRACTICAL IMPLEMENTATION

Because it is difficult for us to get the estimates of λ_k in (10) which involves a complex integral, we cannot directly calculate the critical values of U_n. For this, the following bootstrap approximation algorithm is designed to get the critical values. The algorithm is divided into five steps, as shown below:

(1) For a given random sample $\{X_i, Z_i, Y_i\}_{i=1}^n$, use the two-stage estimation procedure introduced in [13] to obtain estimators $\hat{\beta}, \hat{\theta}$ and $\hat{g}(\cdot)$. Here, the bandwidth h can be selected by the generalized cross validation (GCV) method proposed in [3]. More details can be seen in [13].

(2) Obtain the residuals $\hat{\varepsilon}_i = Y_i - \hat{g}(X_i^T\hat{\theta}) - Z_i^T\hat{\beta}$, $i = 1, \cdots, n$, and then calculate the test statistic U_n

$$U_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \hat{r}_i \hat{r}_j d_{ij}, \qquad (15)$$

where $\hat{r}_i = \{Y_i - [\hat{g}(X_i^T\hat{\theta}) + Z_i^T\hat{\beta}]\}^2 - \hat{\sigma}^2$, $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_i^2$, and $d_{ij} = e^{-\|W_i - W_j\|^a}$.

(3) Obtain the bootstrap error ε_i^* by randomly resampling with replacement from the set $\{\hat{\varepsilon}_i - \bar{\hat{\varepsilon}}, i = 1, \cdots, n\}$ with $\bar{\hat{\varepsilon}} = n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_i$. Then, $Y_i^* = \hat{g}(X_i^T\hat{\theta}) + Z_i^T\hat{\beta} + \varepsilon_i^*$, $i = 1, \cdots, n$

(4) Recalculate and obtain the new estimators $\hat{g}^*(\cdot), \hat{\theta}^*, \hat{\beta}^*$ using the two-stage estimation procedure. The bootstrap test statistic U_n^* is

$$U_n^* = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \hat{r}_i^* \hat{r}_j^* d_{ij}. \qquad (16)$$

where $\hat{r}_i^* = (Y_i - \hat{g}^*(X_i^T\hat{\theta}^*) - Z_i^T\hat{\beta}^*)^2 - \hat{\sigma}^{*2}$, $\hat{\sigma}^{*2} = n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_i^{*2}$, and $d_{ij} = e^{-\|W_i - W_j\|^a}$.

(5) Repeat steps 3-4 N times, get the test statistic $U_{n,1}^*, \cdots, U_{n,N}^*$. For a specified nominal level of the test, the bootstrap p-value is calculated by $p^* = N^{-1} \sum_{m=1}^{N} I(U_{n,m}^* > U_n)$, where $I(\cdot)$ is an indicator function.

The above algorithm is valid approximation to the null distribution of the U_n test. Specific details are as described in the following theorem.

Theorem 4: Assume conditions in the Appendix hold, we have

(1) Under the hypothesis H_0 or the hypothesis H_{1n} with $c_n \to 0$, the limiting conditional distribution of $nU_n^*|F_n$ is the same as the limiting null distributions of the test statistic nU_n, where $F_n = \{X_i, Z_i, Y_i\}_{i=1}^n$.

(2) Under the assumption of H_1, the limiting conditional distribution of $nU_n^*|F_n$ is a finite limit, which may be different from the limiting null distributions of the test statistic nU_n.

The above theorem shows that the proposed algorithm can control the size of the test statistic U_n well. Next, we study the power performance of this test. From Theorem 2.2, under the hypotheses H_{1n} with $c_n = n^{-b}, 0 < b < 1/2$, we can get that $nU_n \to \infty$. It shows that the proposed algorithm can have asymptotic power 1 in this case. Under the hypotheses H_{1n} with $c_n = n^{-1/2}$, the proposed algorithm can still detect the alternative hypotheses. From Theorem 2.3, under the hypothesis H_1, if $nh^8 \to \infty$, we can also get that $nU_n \to \infty$. It shows that the proposed algorithm can also have asymptotic power 1 under the hypothesis H_1. In summary, the proposed algorithm is valid.

IV. NUMERICAL STUDIES

In this section, we investigate the performance of the proposed test statistic with a finite sample size by numerical studies. To assess the power performance, the following two examples are designed and 1000 replications of the experiment are taken to calculate empirical significance level and powers at the significance level $\alpha = 0.05$. The sample sizes is $n = 300, 800$, and the number of bootstrap sample is set to be $B = 500$. For comparison, the test statistic in [29], [30] designed is also used and denoted as T_n, which is of the form

$$T_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \hat{r}_i \hat{r}_j h_0^{-p} K\left(\frac{W_i - W_j}{h_0}\right),$$

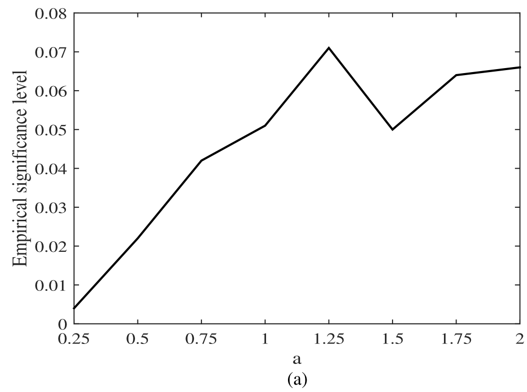
with $K(\cdot)$ being a kernel function and h_0 being the bandwidth.

Example 1: We generate the data from the following model:

$$Y_i = Z_i^T\beta + 2\exp\left\{-3\left(X_i^T\theta\right)^2\right\} + \left(\delta\left|Z_i^T\beta + X_i^T\theta\right| + 0.5\right)\varepsilon_i, \quad (17)$$

where $X_i = (X_{i1}, \cdots, X_{ip})^T$ are *i.i.d* with the common uniform distribution on the p-dimensional cube $[-1, 1]^p$; $Z_i = (Z_{i1}, \cdots, Z_{iq})^T$ is set to the standard normal distribution; $p+q$ is set to be 4 ($p = 3$, $q = 1$) and 8 ($p = 4$, $q = 4$); $\varepsilon_i \sim N(0, 1)$. The true parameters are $\beta = (1, \cdots, 1)^T$ and $\theta = (2/3, 1/3, -2/3, 0, \cdots, 0)^T$. The null hypothesis corresponds to $\delta = 0$, whereas the alternative hypothesis corresponds to $\delta \neq 0$.

For the test statistics U_n, Figure 1 shows the empirical significance level and empirical powers of the proposed test statistic for $n = 300$, $p + q = 4$ when a varies in $(0, 2]$. It is easy to see that (i) as long as a is not too small, the proposed test statistics can control empirical sizes well; (ii) when a is larger than 1.5, empirical significance level are slightly bigger than 0.05; (iii) the empirical powers becomes bigger as a increases. Thus, we suggest choosing $a = 1.5$.



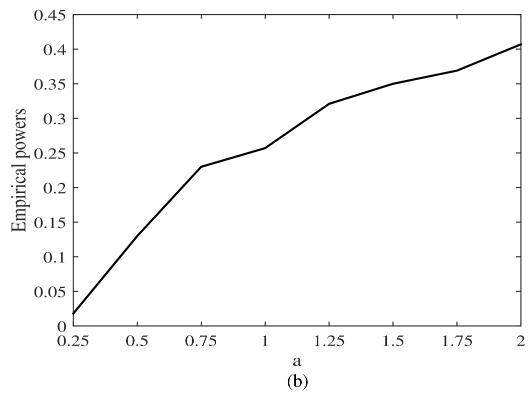


FIGURE 1. Simulation results: (a) empirical significance level, (b) empirical powers.

The power performance of the test statistics with $n = 300$, 800, $p + q = 4$, 8 and $\delta = 0, 0.5, 1, 1.5, 2$ is shown in Table 1. For the test statistic T_n, we use the Epanechnikov kernel $K(t) = 3/4(1 - t^2)_+$ and use the leave one-out cross validation to choose the proper bandwidth. From Table 1, we can find that both U_n and T_n can effectively control the sizes when $p + q = 4$ or 8. When the sample size increases, both U_n and T_n have higher power and the deviation from the hypothetical model is larger. Furthermore, the powers of U_n are larger than T_n, which is reasonable since T_n converges to its weak limit at a very slow rate due to the impact of bandwidth h_0. In addition, when dimension $p + q$ increases from 4 to 8, the power of U_n decreases. This result implies that although the convergence rate of U_n does not depend on the dimension of covariates and the dimension does affect the power performances in practice. However, we also notice that

TABLE 1. Empirical significance level and empirical powers of U_n, T_n for H_0 in example 1.

	δ	$n=300$	$n=800$
$p+q=4, U_n$	0	0.0450	0.0520
	0.5	0.1320	0.1210
	1.0	0.1930	0.3530
	1.5	0.2850	0.3840
	2.0	0.4530	0.5760
$p+q=4, T_n$	0	0.0380	0.0470
	0.5	0.0790	0.1350
	1.0	0.1480	0.2270
	1.5	0.3110	0.4760
	2.0	0.4150	0.6760
$p+q=8, U_n$	0	0.0530	0.0590
	0.5	0.0830	0.1280
	1.0	0.1530	0.1920
	1.5	0.2450	0.2830
	2.0	0.2920	0.3830
$p+q=8, T_n$	0	0.0680	0.0710
	0.5	0.0620	0.0660
	1.0	0.0560	0.0920
	1.5	0.1920	0.1340
	2.0	0.2730	0.2650

TABLE 2. Empirical significance level and empirical powers of U_n, T_n for H_0 in example 2.

	δ	$n=300$	$n=800$
$p+q=4,U_n$	0	0.0420	0.0510
	0.5	0.1650	0.1980
	1.0	0.2270	0.3890
	1.5	0.2930	0.4120
	2.0	0.4010	0.5860
$p+q=4,T_n$	0	0.0420	0.0570
	0.5	0.0810	0.1980
	1.0	0.1770	0.2870
	1.5	0.3120	0.5230
	2.0	0.3950	0.5650
$p+q=8,U_n$	0	0.0410	0.0530
	0.5	0.0850	0.1200
	1.0	0.1870	0.1690
	1.5	0.2350	0.3070
	2.0	0.2870	0.3390
$p+q=8,T_n$	0	0.0420	0.0430
	0.5	0.0790	0.0710
	1.0	0.0910	0.0950
	1.5	0.1890	0.2320
	2.0	0.2140	0.3240

even when $p+q=8$, the proposed test is still sensitive to the alternatives hypotheses. Nevertheless when the dimension $p+q$ is 8, T_n does not perform well.

Example 2: We generate the data from the following model:

$$Y_i=\sin\left\{\frac{\pi((X_i^T\theta)/\sqrt{3}-A)}{(B-A)}\right\}+Z_i^T\beta+\left(\delta\left|X_i^T\theta+Z_i^T\beta\right|+0.1\right)\varepsilon_i, \tag{18}$$

where $\theta=(1,\cdots,1)^T/\sqrt{p}$ and other settings are the same as those in Example 1.

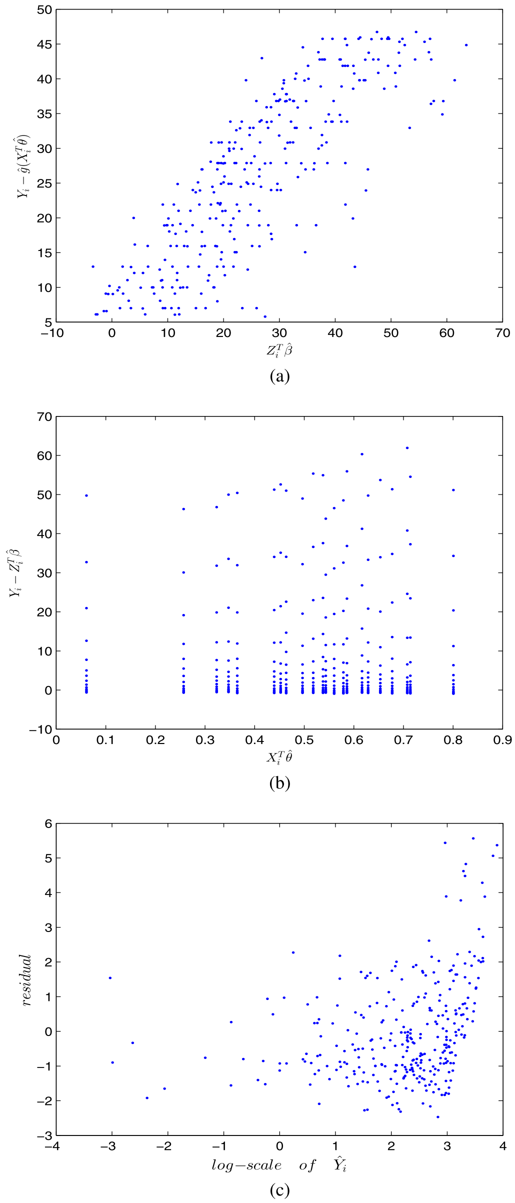


FIGURE 2. (a) Scatter plots for $Z_i^T\hat{\beta}$ and $Y_i-\hat{g}(X_i^T\hat{\theta})$, (b) Scatter plots for $X_i^T\hat{\theta}$ and $Y_i-Z_i^T\hat{\beta}$, (c) Y and residuals for the Delft data.

Table 2 lists the power performance of U_n and T_n under the null hypotheses and alternative hypothesis. The conclusions are similar to those of the model in Example 2. From the two examples, we can conclude that a dimensionality effect exists in the proposed test U_n, but the effect is not serious as in Zheng's test T_n [29].

In summary, the proposed test is a good alternative for testing heteroscedasticity.

V. REAL DATA EXAMPLE

In this section, a real data example is analyzed for illustration. We consider the Delft dataset which comprises 308 full-scale experiments. This dataset was performed at the Delft Ship Hydromechanics Laboratory, which can be obtained from the website *https://archive.ics.uci.edu/ml/machine-learning-databases/00243/*. These experiments derived from a parent form closely related to the 'Standfast 43' designed by Frans Maas. The covariates are X_{i1}-longitudinal position of the center of buoyancy, X_{i2}-prismatic coefficient, X_{i3}-length-displacement ratio, X_{i4}-beam-draught ratio, and Z_{i1}-length-beam ratio, Z_{i2}-the Froude number ranging from 0.125 to 0.450 and the response variable Y_i is the residuary resistance per unit weight of displacement.

All the variables X_i, Z_i and Y_i are centered and standardized, corresponding to covariates $X_i = (X_{i1}, X_{i2}, X_{i3}, X_{i4})^T$ and $Z_i = (Z_{i1}, Z_{i2})^T$. For this Delft dataset, we consider the following PLSIM:

$$Y_i = g(X_i^T\theta) + Z_i^T\beta + \varepsilon_i, \quad i = 1, \cdots, 308. \tag{19}$$

The scatter plots for $Z_i^T\hat{\beta}$ and $Y_i - \hat{g}(X_i^T\hat{\theta})$, $X_i^T\hat{\theta}$ and $Y_i - Z_i^T\hat{\beta}$, as well as residuals and log-scale of $\hat{Y}_i$ are given in Figure 2. These scatter plots exhibit seemingly linear and nonlinear relationships, respectively. This shows that the PLSIM is plausible. To formally check goodness of fit, we modify the test statistic U_n. We can derive a test statistic to detect possible misspecifications in the mean regression function by replacing $\hat{r}_i$ in U_n with $\hat{\varepsilon}_i = Y_i - \hat{g}(X_i^T\hat{\theta}) - Z_i^T\hat{\beta}$. With 500 bootstrap samples, the p-value is $0.216 > 0.05$, and thus, the PLSIM should be adequate for this dataset. Further, we investigate whether the heteroscedastic structure is present in the model. The p-value is now 0.001. Thus, homoscedasticity should be rejected, which is agreement with the third subplot in Figure 2. In general, we conclude that a heteroscedastic PLSIM is appropriate to this dataset.

VI. CONCLUSION

In this article, a new test is proposed for the heteroscedasticity of a PLSIM. The statistic is based on the pairwise distance between sample points. The results show that the statistic has asymptotic normal distribution of non-zero mean and the same asymptotic variance. This test method does not require assuming a distribution of random errors. The algorithmic program of this test method is easy to implement and has faster convergence than some existing methods. In addition, convergence rate of the statistic does not depend on the dimensions of the covariates, which greatly reduces the impact of the dimensional curse. The numerical simulations and a real data example verified the feasibility of the method.

For the PLSIM

$$Y = g(X^T\theta) + Z^T\beta + \varepsilon, \tag{20}$$

if X is scalar and $\theta = 1$, the model reduces to the partially linear model [21], i.e.,

$$Y = g(X) + Z^T\beta + \varepsilon, \tag{21}$$

if $\beta = 0$, the model becomes the single-index model [30], i.e.,

$$Y = g(X^T\theta) + \varepsilon. \tag{22}$$

In the following studies, we study the above two special cases using the statistics in this paper. It is just to test the variance structure separately, under the null hypothesis:

$$\begin{aligned} H_0 &: E(\varepsilon^2|X, Z) = E(\varepsilon^2|X), \\ H_0 &: E(\varepsilon^2|X) = E(\varepsilon^2|X^T\theta). \end{aligned} \tag{23}$$

How to combining these two problems together is our further work.

DATA AVAILABILITY

All the data used to support the findings of this study are included in our manuscript and can be accessed freely from the references and the URL https://archive.ics.uci.edu/ml/machine-learning-databases/00243/.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

APPENDIX

First, we introduce some regularity conditions for the asymptotic results in Section *II* and Section *III*.

(1) Suppose that the parameter space of θ, β is respectively compact subsets of R^p, R^q.

(2) $g(X^T\theta) = g$ has bounded, continuous third-order derivative; the conditional expectations $E(Z|X = x)$, $E(ZZ^T|X = x)$, $E(Z|X^T\theta = v)$ and $E(ZZ^T|X^T\theta = v)$ have bounded derivatives.

(3) With probability 1, X lies in a compact set D; the marginal density functions of X has bounded derivatives.

(4) $E(r^4) < \infty$.

Proof of Theorem 2.1: Let $f = f(X, Z, \theta, \beta) = f(W, \theta, \beta) = g(X^T\theta) + Z^T\beta$, $\hat{f} = f(X, Z, \hat{\theta}, \hat{\beta}) = f(W, \hat{\theta}, \hat{\beta}) = \hat{g}(X^T\hat{\theta}) + Z^T\hat{\beta}$, and $\varepsilon = Y - f$. Then, $\hat{\varepsilon} = Y - \hat{f}$, $\hat{r} = (Y - \hat{f})^2 - \hat{\sigma}^2$. For $\hat{r}$, we have

$$\begin{aligned} \hat{r} &= (Y - \hat{f})^2 - \hat{\sigma}^2 = [(Y - f) - (\hat{f} - f)]^2 - \sigma^2 - (\hat{\sigma}^2 - \sigma^2) \\ &= r - 2\varepsilon(\hat{f} - f) + (\hat{f} - f)^2 - (\hat{\sigma}^2 - \sigma^2). \end{aligned}$$

Under the above representation, we can decompose U_n into the following 10 parts, which are respectively recorded as Q_{in}:

$$\begin{aligned} U_n = &\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} d_{ij} r_i r_j \\ &+ 4\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} d_{ij} \varepsilon_i \varepsilon_j (\hat{f}_i - f_i)(\hat{f}_j - f_j) \\ &+ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} d_{ij} (\hat{f}_i - f_i)^2 (\hat{f}_j - f_j)^2 \\ &+ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} d_{ij} (\hat{\sigma} - \sigma)^2 \\ &- 4\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} d_{ij} r_i \varepsilon_j (\hat{f}_j - f_j) \end{aligned}$$

$$
\begin{aligned}
&+2\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} d_{ij}r_i(\hat{f}_j - f_j)^2\\
&-2\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} d_{ij}r_i(\hat{\sigma}^2 - \sigma^2)\\
&-4\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} d_{ij}\varepsilon_i(\hat{f}_i - f_i)(\hat{f}_j - f_j)^2\\
&+4\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} d_{ij}\varepsilon_i(\hat{f}_i - f_i)(\hat{\sigma}^2 - \sigma^2)\\
&-2\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} d_{ij}(\hat{f}_i - f_i)^2(\hat{\sigma}^2 - \sigma^2)\\
=:& \sum_{i=1}^{10} Q_{in}. \qquad (24)
\end{aligned}
$$

Since $\|\hat{\theta} - \theta\| = O_p(n^{-1/2})$, $\|\hat{\beta} - \beta\| = O_p(n^{-1/2})$ (see [13]) and $E(\varepsilon|W) = 0, E(\varepsilon|X) = 0, E(\varepsilon|Z) = 0$, and $E(r|W) = 0$, we can get $Q_{2n} = O_p(n^{-2} + n^{-3/2}h^2 + n^{-1}h^2)$. The proof of Q_{2n} is given as follows. Note that

$$
\begin{aligned}
\hat{f}_i - f_i &= \hat{g}(X_i^T\hat{\theta}) + Z_i^T\hat{\beta} - g(X_i^T\theta) - Z_i^T\beta\\
&= (\hat{g}(X_i^T\hat{\theta}) - g(X_i^T\hat{\theta})) + (g(X_i^T\hat{\theta})\\
&\quad - g(X_i^T\theta)) + Z_i^T(\hat{\beta} - \beta)\\
&= (\hat{g}(X_i^T\hat{\theta}) - g(X_i^T\hat{\theta})) + (\hat{\theta} - \theta)^T\nabla g(X_i^T\tilde{\theta})\\
&\quad + \frac{1}{2}(\hat{\theta} - \theta)^T\nabla g^2(X_i^T\tilde{\theta})(\hat{\theta} - \theta) + Z_i^T(\hat{\beta} - \beta),
\end{aligned}
$$

where $\tilde{\theta}$ is between $\hat{\theta}$ and θ. Let $v_i = X_i^T\theta^*$ and $u_i = X_i^T\tilde{\theta}$, where θ^* is in a δ-neighbourhood of $\hat{\theta}$ with δ is a small positive number, we have

$$
\begin{aligned}
\hat{f}_i - f_i = (\hat{g}(v_i) - g(v_i)) + (\hat{\theta} - \theta)^T\nabla g(u_i)\\
+ \frac{1}{2}(\hat{\theta} - \theta)^T\nabla g^2(u_i)(\hat{\theta} - \theta) + Z_i^T(\hat{\beta} - \beta).
\end{aligned}
$$

Notice that

$$
\begin{aligned}
&(\hat{f}_i - f_i)(\hat{f}_j - f_j)\\
&= [(\hat{g}(v_i) - g(v_i)) + (\hat{\theta} - \theta)^T\nabla g(u_i)\\
&\quad + \frac{1}{2}(\hat{\theta} - \theta)^T\nabla g^2(u_i)(\hat{\theta} - \theta) + Z_i^T(\hat{\beta} - \beta)]\\
&\quad \times[(\hat{g}(v_j) - g(v_j)) + (\hat{\theta} - \theta)^T\nabla g(u_j)\\
&\quad + \frac{1}{2}(\hat{\theta} - \theta)^T g^2(u_j)(\hat{\theta} - \theta) + Z_j^T(\hat{\beta} - \beta)]\\
&=: \sum_{i=1}^{10} M_i.
\end{aligned}
$$

where

$$
\begin{aligned}
M_1 &= [\hat{g}(v_i) - g(v_i)][\hat{g}(v_j) - g(v_j],\\
M_2 &= (\hat{\theta} - \theta)^T\nabla g(u_i)\nabla g^T(u_j)(\hat{\theta} - \theta),\\
M_3 &= \frac{1}{4}(\hat{\theta} - \theta)^T\nabla g^2(u_i)(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T\nabla g^2(u_j)(\hat{\theta} - \theta),\\
M_4 &= (\hat{\beta} - \beta)^T Z_i Z_j^T(\hat{\beta} - \beta),\\
M_5 &= (\hat{\theta} - \theta)^T[(\hat{g}(v_i) - g(v_i))\nabla g(u_j)\\
&\quad + (\hat{g}(v_j) - g(v_j))\nabla g(u_i)],\\
M_6 &= \frac{1}{2}(\hat{\theta} - \theta)^T[(\hat{g}(v_i) - g(v_i))\nabla g^2(u_j)\\
&\quad + (\hat{g}(v_j) - g(v_j))\nabla g^2(u_i)](\hat{\theta} - \theta),\\
M_7 &= [(\hat{g}(v_i) - g(v_i))Z_j^T + (\hat{g}(v_j) - g(v_j))Z_i^T](\hat{\beta} - \beta),\\
M_8 &= \frac{1}{2}(\hat{\theta} - \theta)^T[\nabla g(u_i)(\hat{\theta} - \theta)^T\nabla g^2(u_j)\\
&\quad + \nabla g^2(u_i)^T(\hat{\theta} - \theta)\nabla g^T(u_j)](\hat{\theta} - \theta),\\
M_9 &= (\hat{\beta} - \beta)^T[Z_i\nabla g^T(u_j) + Z_j\nabla g^T(u_i)](\hat{\theta} - \theta),\\
M_{10} &= \frac{1}{2}(\hat{\beta} - \beta)^T[Z_i(\hat{\theta} - \theta)^T\nabla g^2(u_j)\\
&\quad + Z_j(\hat{\theta} - \theta)^T\nabla g^2(u_i)](\hat{\theta} - \theta).
\end{aligned}
$$

We rewrite Q_{2n} as follow

$$
\begin{aligned}
Q_{2n} &= \frac{4}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} d_{ij}\varepsilon_i\varepsilon_j(M_1 + \cdots + M_{10})\\
&= \sum_{k=1}^{10} Q_{2n,i}.
\end{aligned}
$$

For the term $Q_{2n,1}$, we first note that unifoemly

$$
\begin{aligned}
\hat{g}(v_i) - g(v_i) &= \frac{\Sigma_{l\neq i}K_{il}(g(v_i) - g(v_l))}{\Sigma_{l\neq i}K_{il}}\\
&= h^2k_2\frac{g(v_i)p_i^{(2)} - (gp)_i^{(2)}}{2p_i} + o(h^2)\\
&=: h^2G_i + o(h^2),
\end{aligned}
$$

where $k_2 = \int s^2K(s)ds$ and $p_i = p(v_i)$, which is the density function of v_i. As a result, we can get

$$
\begin{aligned}
Q_{2n,1} &= \frac{4}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} d_{ij}\varepsilon_i\varepsilon_j(h^2G_i + o(h^2))(h^2G_j + o(h^2))\\
&= 4h^4E(d_{12}\varepsilon_1\varepsilon_2G_1G_2) + o(\frac{h^4}{n})\\
&= \frac{4h^4}{n}E(d_{12}G_1G_2) + o(\frac{h^4}{n})\\
&= O_p(\frac{h^4}{n}).
\end{aligned}
$$

For the term $Q_{2n,1}$, we have

$$
\begin{aligned}
&Q_{2n,2}\\
&= \frac{4}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} d_{ij}\varepsilon_i\varepsilon_j(\hat{\theta} - \theta)^T\nabla g(u_i)\nabla g^T(u_j)(\hat{\theta} - \theta)\\
&= 4(\hat{\theta} - \theta)^T\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} d_{ij}\varepsilon_i\varepsilon_j\nabla g(u_i)\nabla g^T(u_j)(\hat{\theta} - \theta)\\
&=: 4(\hat{\theta} - \theta)^T\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} L(R_i, R_j)(\hat{\theta} - \theta),
\end{aligned}
$$

where $R_i = (W_i, \varepsilon_i)$. Since $L(R_i, R_j) = L(R_j, R_i)$, $\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} L(R_i, R_j)$ is a U-statistic. Furthermore, we get $E(L(R_i, R_j)|R_i) = 0$. In term of $E(\varepsilon|X) = 0$, we have

$$\begin{aligned}E(L(R_i, R_j)|R_i) &= E(d_{ij}\varepsilon_i\varepsilon_j|X_i, \varepsilon_i)\\ &= E[E(d_{ij}\varepsilon_i\varepsilon_j|X_i, X_j, \varepsilon_i)|X_i, \varepsilon_i]\\ &= E[d_{ij}\varepsilon_i\varepsilon_j E(\varepsilon_j|X_i, X_j, \varepsilon_i)|X_i, \varepsilon_i]\\ &= 0.\end{aligned}$$

It shows that $\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} L(R_i, R_j)$ is a degenerate U-statistic of order 1. Hence, $\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} L(R_i, R_j) = O_p(n^{-1})$. According to $\|\hat\theta - \theta\| = O_p(n^{-1/2})$, we have $Q_{2n,2} = O_p(n^{-2})$. Similarly, we have $Q_{2n,3} = O_p(n^{-3})$, $Q_{2n,4} = O_p(n^{-2})$, $Q_{2n,5} = O_p(n^{-3/2}h^4)$, $Q_{2n,6} = O_p(n^{-2}h^4)$, $Q_{2n,7} = O_p(n^{-3/2}h^4)$, $Q_{2n,8} = O_p(n^{-5/2})$, $Q_{2n,9} = O_p(n^{-2})$, and $Q_{2n,10} = O_p(n^{-5/2})$. Therefore, $Q_{2n} = O_p(n^{-2} + n^{-3/2}h^2 + n^{-1}h^2)$.

Similarly, we can easily obtain $Q_{3n} = O_p(h^8 + n^{-1/2}h^6 + n^{-1}h^4 + n^{-3/2}h^2 + n^{-2})$, $Q_{5n} = O_p(n^{-1}h^2 + n^{-3/2})$, $Q_{6n} = O_p(n^{-1/2}h^4 + n^{-1}h^2 + n^{-3/2})$, $Q_{8n} = O_p(n^{-1/2}h^6 + n^{-1}h^4 + n^{-3/2}h^2 + n^{-2})$, $Q_{9n} = O_p(n^{-1/2}h^6 + n^{-1}h^2 + n^{-3/2})$, and $Q_{10n} = O_p(h^8 + n^{-1/2}h^4 + n^{-1}h^2 + n^{-3/2})$.

Consequently, we can get

$$\begin{aligned}&Q_{2n} + Q_{3n} + Q_{5n} + Q_{6n} + Q_{8n} + Q_{9n} + Q_{10n}\\ &\qquad = O_p(h^8 + n^{-1/2}h^4 + n^{-1}h^2 + n^{-3/2}).\end{aligned}$$

If $nh^8 \to 0$, we have $Q_{2n} + Q_{3n} + Q_{5n} + Q_{6n} + Q_{8n} + Q_{9n} + Q_{10n} = O_p(n^{-1/2}h^4) = o_p(\frac{1}{n})$. If $nh^8 \to \infty$, we have $Q_{2n} + Q_{3n} + Q_{5n} + Q_{6n} + Q_{8n} + Q_{9n} + Q_{10n} = O_p(h^8)$. Next, we discuss the asymptotic property of U_n in the case of $nh^8 \to 0$ and $nh^8 \to \infty$, respectively.

Note that

$$\begin{aligned}\hat\sigma^2 - \sigma^2 &= \frac{1}{n}\sum_{i=1}^{n}[(Y_i - \hat f_i) - \sigma^2]\\ &= \frac{1}{n}\sum_{i=1}^{n}[r_i - 2\varepsilon_i(\hat f_i - f_i) + (\hat f_i - f_i)^2]\\ &= \frac{1}{n}\sum_{i=1}^{n} r_i + O_p(h^4 + n^{-1/2}h^2 + n^{-1}).\end{aligned} \tag{25}$$

Then, we have

$$\begin{aligned}(\hat\sigma^2 - \sigma^2)^2 &= (\frac{1}{n}\sum_{i=1}^{n} r_i)^2 + O_p(h^8 + n^{-1/2}h^4 + n^{-1})\\ &= \frac{n-1}{n}\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} r_ir_j + \frac{1}{n^2}\sum_{i=1}^{n} r_i^2\\ &\quad + O_p(h^8 + n^{-1/2}h^4 + n^{-1})\end{aligned}$$

(i). When $nh^8 \to 0$, we have

$$\begin{aligned}U_n &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} d_{ij}\left[r_ir_j + (\hat\sigma^2 - \sigma^2)^2 - 2r_i(\hat\sigma^2 - \sigma^2)\right]\\ &\quad + o_p(\frac{1}{n})\\ &= Q_{1n} + Q_{4n} + Q_{7n} + o_p(\frac{1}{n}).\end{aligned}$$

Thus, we have

$$Q_{4n} = E(d_{12})\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} r_ir_j + \frac{1}{n}E(d_{12})E(r^2) + o_p(\frac{1}{n}). \tag{26}$$

Next, we calculate the term Q_{7n}. Similarly, we have

$$\begin{aligned}Q_{7n} &= \frac{1}{n^2(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{l=1}^{n} d_{ij}(r_i + r_j)r_l + o_p(\frac{1}{n})\\ &= \frac{1}{n^2(n-1)}\sum_{i\neq j\neq l}^{n} d_{ij}(r_i + r_j)r_l\\ &\quad + \frac{1}{n^2(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} d_{ij}(r_i + r_j)^2 + o_p(\frac{1}{n})\\ &=: Q_{7n,1} + Q_{7n,2} + o_p(\frac{1}{n}).\end{aligned}$$

Note that $E(d_{12}r_1r_2) = 0$, and we have

$$Q_{7n,2} = \frac{E[d_{12}(r_1 + r_2)^2]}{n} + o_p(\frac{1}{n}) = \frac{2E[d_{12}r_1^2]}{n} + o_p(\frac{1}{n}).$$

We rewrite $Q_{7n,1} = \frac{n-2}{2}J_n$, where

$$J_n = \binom{n}{3}^{-1}\sum\sum\sum_{1\le i<j<l\le n} H^s(A_i, A_j, A_l),$$

with $A_i = (W_i, r_i)$, $H^s(A_i, A_j, A_l) = (H_{ijl}, H_{ilj}, H_{jli})/3$ is the kernel, and $H_{ijl} = d_{ij}(r_i + r_j)r_l$. Thus, J_n is a U-statistic of order 3.

Notice that $H^s(A_i, A_j, A_l)$ is symmetric,

$$E[H^s(A_i, A_j, A_l)|A_i] = 0.$$

Because

$$\begin{aligned}E(H_{ijl}|A_i) &= E(d_{ij}(r_i + r_j)r_l|A_i)\\ &= E[E(d_{ij}(r_i + r_j)r_l|A_i, A_j)|A_i]\\ &= E[d_{ij}(r_i + r_j)E(r_l|A_i, A_j)|A_i] = 0.\end{aligned}$$

Similarly, $E(H_{ilj}|A_i) = 0$ and $E(H_{jli}|A_i) = 0$. Therefore, we have $E[H^s(A_i, A_j, A_l)|A_i] = 0$.

However, $E[H^s(A_i, A_j, A_l)|A_i, A_j] \neq 0$. Specifically,

$$\begin{aligned}E(H_{ilj}|A_i, A_j) &= E(d_{il}(r_i + r_l)r_j|A_i, A_j)\\ &= E[E(d_{il}(r_i + r_l)r_j|A_i, A_j, A_l)|A_i, A_j]\\ &= E[d_{il}r_ir_j|A_i, A_j] = r_ir_jE[d_{il}|W_i, W_j].\end{aligned}$$

Similarly, $E(H_{jli}|A_i,A_j) = r_i r_j E[d_{jl}|W_i, W_j]$. Thus, we have

$$E[H^s(A_i, A_j, A_l)|A_i, A_j] = \frac{1}{3} r_i r_j E[d_{il} + d_{jl}|W_i, W_j].$$

In term of Serfling (1980), we can get

$$\begin{aligned} J_n &= \frac{3\cdot 2}{n(n-1)} \sum\sum_{1\le i<j\le n} \frac{1}{3} r_i r_j E(d_{il} + d_{jl}|W_i, W_j) + o_p(\frac{1}{n}) \\ &= \frac{1}{n(n-1)} \sum_{i=1}\sum_{j\neq i} r_i r_j E(d_{il} + d_{jl}|W_i, W_j) + o_p(\frac{1}{n}), \end{aligned}$$

which shows that

$$Q_{7n} = \frac{1}{n(n-1)} \sum_{i=1}^{n}\sum_{j\neq i}^{n} r_i r_j E(d_{il} + d_{jl}|W_i, W_j) + \frac{2E(d_{12}r_1^2)}{n} + o_p(\frac{1}{n}). \quad (27)$$

According to (24), (26) and (27), we have

$$\begin{aligned} U_n &= \frac{1}{n(n-1)} \sum_{i=1}^{n}\sum_{j\neq i}^{n} r_i r_j [d_{ij} - E(d_{il}+d_{jl}|X_i, X_j)+E(d_{12})] \\ &\quad + \frac{E(d_{12})E(r^2) - 2E(d_{12}r_1^2)}{n} + o_p(\frac{1}{n}) \\ &=: \frac{1}{n(n-1)} \sum_{i=1}^{n}\sum_{j\neq i}^{n} h(A_i, A_j) + \frac{\mu}{n} + o_p(\frac{1}{n}), \quad (28) \end{aligned}$$

where $h(A_i, A_j) = r_i r_j \tilde{d}_{ij}$, $\tilde{d}_{ij} = d_{ij} - E(d_{il} + d_{jl}|W_i, W_j) + E(d_{12})$, and $\mu = E(d_{12})E(r^2) - 2E(d_{12}r_1^2)$. Due to $E(r|W) = 0$, we have

$$\begin{aligned} E(h(A_i, A_j)|A_i) &= E[E(r_i r_j \tilde{d}_{ij}|A_i, A_j)|A_i] \\ &= E[r_i \tilde{d}_{ij} E(r_j|A_i, A_j)|A_i] = 0. \end{aligned}$$

Because of $d_{ij} = e^{-\|X_i - X_j\|^a} \le 1$, we have $|\tilde{d}_{ij}| \le 4$. According to condition (4), we can get

$$E(h^2(A_1, A_2)) = E(r_1^2 r_2^2 \tilde{d}_{ij}^2) \le 16E(r_1^2 r_2^2) = 16E^2(r^2) < \infty.$$

Thus, via the theory of U-statistic, we have

$$n\frac{1}{n(n-1)} \sum_{i=1}^{n}\sum_{j\neq i}^{n} h(A_i, A_j) \to \sum_{k=1}^{\infty} \lambda_k (S_k^2 - 1),$$

where λ_k's are the eigenvalue of the following integral equation

$$\int h(A_i, A_j)\tilde{\phi}_k(A_j)dF(A_j) = \lambda_k \tilde{\phi}_k(A_i).$$

with F denoting the probability distribution function of A, and Z_k's are the independent standard normal random variables. Since

$$\begin{aligned} \lambda_k \tilde{\phi}_k(A_i) &= \int h(A_i, A_j)\tilde{\phi}_k(A_j)dF(A_j) \\ &= \int r_i r_j \tilde{d}_{ij}\tilde{\phi}_k(A_j)dF(A_j) \\ &= r_i \int r_j \tilde{d}_{ij}\tilde{\phi}_k(A_j)dF(A_j). \end{aligned}$$

We rewrite $\tilde{\phi}_k(A_i) = r_i\phi_k(A_i)$ and $\tilde{\phi}_k(A_j) = r_j\phi_k(A_j)$ by choosing proper $\phi_k(A_i)$ and $\phi_k(A_j)$, respectively. The integration equation can be rewritten as

$$\int r_j^2 \tilde{d}_{ij}\tilde{\phi}_k(A_j)dF(A_j) = \lambda_k \phi_k(A_i).$$

Therefore, after some simple calculation, we have

$$nU_n \to \sum_{k=1}^{\infty} \lambda_k (S_k^2 - 1) + \mu = \sum_{k=1}^{\infty} \lambda_k (S_k^2) - E(r^2).$$

(ii). When $nh^8 \to \infty$, we first compute the term Q_{6n}. From (25), we have

$$\begin{aligned} Q_{6n} &= 2\frac{1}{n(n-1)} \sum_{i=1}^{n}\sum_{j\neq i} d_{ij} r_i (\hat{f}_j - f_j)^2 \\ &= 2\frac{1}{n(n-1)} \sum_{i=1}^{n}\sum_{j\neq i} d_{ij} r_i [(\hat{g}(v_j) - g(v_j))^2 + o_p(h^4)] \\ &= 2\frac{1}{n(n-1)} \sum_{i=1}^{n}\sum_{j\neq i} d_{ij} r_i (\hat{g}(v_j) - g(v_j))^2 \\ &\quad + o_p(h^4 n^{-1/2}) \\ &= 2h^4 \frac{1}{n(n-1)} \sum_{i=1}^{n}\sum_{j\neq i} d_{ij} r_i G_j^2 + o_p(n^{-1/2}h^4) \\ &= 2h^4 \frac{1}{n(n-1)} \sum\sum_{1\le i<j\le n} d_{ij}(r_i G_j^2 + r_j G_i^2) \\ &\quad + o_p(n^{-1/2}h^4). \end{aligned}$$

It can be shown that $E[(r_i G_j^2 + r_j G_i^2)d_{ij}] = 0$ and

$$\begin{aligned} &E[(r_i G_j^2 + r_j G_i^2)d_{ij}|W_i, r_i] \\ &= E[E\{(r_i G_j^2 + r_j G_i^2)d_{ij}|W_i, r_i, W_j\}|W_i, r_i] \\ &= E[r_i G_j^2 d_{ij} + d_{ij} G_i^2 E(r_j|W_i, r_i, W_j)|W_i, r_i] \\ &= r_i E(G_j^2 d_{ij}|W_i) \\ &=: \varphi_i^*. \end{aligned}$$

By the theory of U-statistic, we can then easily derive that

$$\begin{aligned} \sqrt{n}h^{-4}Q_{6n} &= \sqrt{n}(\tfrac{2}{n(n-1)} \sum\sum_{1\le i<j\le n} d_{ij}(r_i G^2{}_j + r_j G^2{}_i) \\ &\quad + o_p(h^4 n^{-1/2})) \to N(0, \sigma_*^2), \end{aligned}$$

where $\sigma_*^2 = var(\varphi_1^*)$.

Next, we can easily get $Q_{1n} = O_p(n^{-1})$, $Q_{4n} = O_p(n^{-1} + n^{-1/2}h^4 + h^8)$, and $Q_{7n} = O_p(n^{-1} + n^{-1/2}h^4)$. Then, $Q_{1n} + Q_{2n} + Q_{5n} + Q_{8n} + Q_{9n} = O_p(n^{-1})$.

Thus, we have

$$\sqrt{n}h^{-4}(U_n - a(n)) = \sqrt{n}h^{-4}Q_{6n} + o_p(1) \to N(0, \sigma_*^2).$$

where $a(n) = Q_{3n} + Q_{4n} + Q_{7n} + Q_{10n}$.

Proof of Theorem 2.2: Under the hypothesis H_{1n}, we still divide U_n into 10 parts, similar to the proof of Theorem 2.1. As follows

$$\begin{aligned} U_n &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} d_{ij}[r_i r_j \\ &\quad +4\varepsilon_i\varepsilon_j(\hat{f}_i-f_i)(\hat{f}_j-f_j)+(\hat{f}_i-f_i)^2(\hat{f}_j-f_j)^2 \\ &\quad +(\hat{\sigma}^2-\sigma^2)^2-4r_i\varepsilon_j(\hat{f}_j-f_j) \\ &\quad +2r_i(\hat{f}_j-f_j)^2-2r_i(\hat{\sigma}^2-\sigma^2) \\ &\quad -4\varepsilon_i(\hat{f}_i-f_i)(\hat{f}_j-f_j)^2+4\varepsilon_i(\hat{f}_i-f_i)(\hat{\sigma}^2-\sigma^2) \\ &\quad -2(\hat{f}_i-f_i)^2(\hat{\sigma}^2-\sigma^2)] \\ &=: \sum_{i=1}^{10} Q_{in}. \end{aligned}$$

Since $\varepsilon_i, \hat{f}_i$ and f_i are defined previously under H_{1n}, we have $Q_{2n} = O_p(n^{-2}+n^{-3/2}h^2+n^{-1}h^2)$, $Q_{3n} = O_p(h^8+n^{-1/2}h^6+n^{-1}h^4+n^{-3/2}h^2+n^{-2})$, $Q_{8n} = O_p(n^{-1/2}h^6+n^{-1}h^4+n^{-3/2}h^2+n^{-2})$. Recalling that $r_i = u_i + c_n\Delta(W_i)$, $E(u_i|W_i) = 0$, $E(\Delta(W)) = 0$, and $\hat{\sigma}^2-\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} r_i + O_p(n^{-1}+n^{-1/2}h^2+h^4) = O_p(n^{-1/2}+h^4+c_n)$, we have $Q_{5n} = O_p(n^{-1}h^2+n^{-3/2}+c_n(n^{-1/2}h^2+n^{-1}))$, $Q_{6n} = O_p(n^{-1/2}h^4+n^{-1}h^2+n^{-3/2}+c_n(h^4+n^{-1/2}h^2+n^{-1}))$, $Q_{9n} = O_p(n^{-3/2}+n^{-1}h^2+n^{-1/2}h^6+c_n(n^{-1}h^2+n^{-1}))$, $Q_{10n} = O_p(h^8+n^{-1/2}h^4+n^{-1}h^2+n^{-3/2}+c_n(h^4+n^{-1/2}h^2+n^{-1}))$. Therefore, we can easily get $Q_{2n}+Q_{3n}+Q_{5n}+Q_{6n}+Q_{8n}+Q_{9n}+Q_{10n} = O_p(h^8+n^{-1/2}h^4+n^{-1}h^2+n^{-3/2}+c_n(h^4+n^{-1/2}h^2+n^{-1}))$.

(i) If $c_n = n^{-1/2}$ and $nh^8 \to 0$, we have $Q_{2n}+Q_{3n}+Q_{5n}+Q_{6n}+Q_{8n}+Q_{9n}+Q_{10n} = O_p(n^{-1/2}h^4) = o_p(n^{-1})$.

Similar to the proof of Theorem 2.1, it shows that

$$U_n = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} r_i r_j\widetilde{d}_{ij} + \frac{E(d_{12})E(r^2)-2E(d_{12}r_1^2)}{n} + o_p(\frac{1}{n}), \qquad (29)$$

where $\widetilde{d}_{ij} = d_{ij} - E(d_{il}+d_{jl}|W_i,W_j) + E(d_{12})$. Then, in term of $E(r|X,Z) = c_n\Delta(W)$ and the Theorem 2.1 in [5], we have

$$n \times \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} r_i r_j\widetilde{d}_{ij} \to \sum_{i=1}^{\infty}\lambda_i[(S_i+a_i)^2-1],$$

where $a_i = E(\Delta(W)\phi_i(A))$. Similar to the proof of Theorem 2.1, we have

$$E(d_{12})E(r^2)-2E(d_{12}r_1^2)-\sum_{i=1}^{\infty}\lambda_i = E(r^2) = E(u^2)+o_p(1),$$

where $u = r - c_n\Delta(W)$. Thence, if $c_n = n^{-1/2}$, we can get:

$$nU_n \to \sum_{i=1}^{\infty}\lambda_i[(S_i+a_i)^2-E(u^2)].$$

(ii) If $c_n = n^{-1/2}$ and $nh^8 \to \infty$, we have $Q_{1\,n} = O_p(n^{-1})$, $Q_{2\,n} = O_p(n^{-1}h^4)$, $Q_{3\,n} = Q_{4\,n} = Q_{10n} = O_p(h^8)$, $Q_{5n} = O_p(n^{-1}h^2)$, $Q_{7\,n} = O_p(n^{-1/2}h^4)$, $Q_{8n} = Q_{9n} = O_p(n^{-1/2}h^6)$.

Next, we compute the term Q_{6n}. Since $r_i = u_i + c_n\Delta(W_i) = u_i + n^{-1/2}\Delta(W_i)$, $E(u_i|W_i) = 0$ and $E(\Delta(W)) = 0$. Similar to the proof of Theorem 2.1, in terms of (25), we have

$$\begin{aligned} Q_{6n} &= \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} d_{ij}(u_i+n^{-1/2}\Delta(W_i))(\hat{f}_j-f_j)^2 \\ &= \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} d_{ij}u_i(\hat{g}(v_j)-g(v_j))^2 + o_p(h^4) \\ &= \frac{2h^4}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} d_{ij}u_iG_j^2 + o_p(h^4) \\ &= \frac{h^4}{n(n-1)}\sum\sum_{1\le i<j\le n} d_{ij}(u_iG_j^2+u_jG_i^2) + o_p(h^4). \end{aligned}$$

It follows that $E[(u_iG_j^2+u_jG_i^2)d_{ij}] = 0$ and

$$\begin{aligned} &E[(u_iG_j^2+u_jG_i^2)d_{ij}|W_i,u_i] \\ &= E[E\{(u_iG_j^2+u_jG_i^2)d_{ij}|W_i,u_i,W_j\}|W_i,u_i] \\ &= E[u_iG_j^2d_{ij}+d_{ij}G_i^2E(u_j|W_i,u_i,W_j)|W_i,u_i] \\ &= u_iE(G_j^2d_{ij}|W_i) =: \tilde{\varphi}_i^*. \end{aligned}$$

By the theory of U-statistic, we have

$$\sqrt{n}h^{-4}Q_{6n} = \sqrt{n}\frac{1}{n(n-1)}\sum\sum_{1\le i<j\le n} d_{ij}(u_iG_j^2+u_jG_i^2) + o_p(h^4) \to N(0,\tilde{\sigma}_*^2),$$

where $\tilde{\sigma}_*^2 = var(\tilde{\varphi}_1^*)$.

Thus, we can get

$$\sqrt{n}h^{-4}(U_n - b(n)) = \sqrt{n}h^{-4}Q_{6\,n} + o_p(1) \to N(0,\tilde{\sigma}_*^2),$$

where $b(n) = Q_{3n}+Q_{4n}+Q_{7n}+Q_{10n}$.

(iii) If $c_n = n^{-b}$ and $0 < b < \frac{1}{2}$, we have $Q_{1\,n} = O_p(n^{-2b})$, $Q_{2\,n} = O_p(n^{-2}+n^{-3/2}h^2+n^{-1}h^4)$, $Q_{3\,n} = O_p(h^8+n^{-1/2}+n^{-1}h^4+n^{-1/2}h^6+n^{-3/2}h^2)$, $Q_{4\,n} = O_p(h^8+n^{-1}+n^{-1/2}h^4+n^{-2\,b}+n^{-b}h^4+n^{-b-1/2})$, $Q_{5\,n} = O_p(n^{-1}h^2+n^{-3/2}+n^{-1/2-b}h^2+n^{-1-b})$, $Q_{6\,n} = O_P(n^{-1/2}h^4+n^{-3/2}+n^{-1}h^2+n^{-b}h^4+n^{-1-b}+n^{-b-1/2}h^2)$, $Q_{7\,n} = O_P(n^{-1/2}h^4+n^{-1}+n^{-1/2-b}+n^{-b}h^4+n^{-2\,b})$, $Q_{8\,n} = O_p(n^{-3/2}+n^{-1}h^2+n^{-1/2}h^6+n^{-1/2-b}h^2+n^{-1-b})$, $Q_{10\,n} = O_p(h^8+n^{-3/2}+n^{-1}h^2+n^{-1/2}h^4+n^{-b}h^4+n^{-1-b}+n^{-1/2-b}h^2)$. Then, we can easily get $U_n = O_p(h^8+n^{-2\,b}+n^{-b}h^4)$.

Therefore, we have

$$nU_n = O_p(nh^8+n^{1-2\,b}+n^{1-b}h^4) \to \infty.$$

Proof of Theorem 2.3:

(i) When $nh^8 \to 0$, note that $\hat{\sigma}^2-\sigma^2 = O_p(n^{-1/2})$, similar to the above proof of Theorem 2.2, under H_1 and $r_i = u_{i+}c\triangle(W)$, we can get $Q_{2\,n} = O_p(n^{-1}h^4)$, $Q_{3\,n} = O_p(h^8)$, $Q_{5\,n} = O_p(n^{-1/2}h^2)$, $Q_{6\,n} = O_p(h^4)$, $Q_{8\,n} = O_p(n^{-1/2}h^6)$, $Q_{9\,n} = O_p(h^2n^{-1})$, and $Q_{10\,n} = O_p(h^4n^{-1/2}))$. Then, we can get $Q_{2n}+Q_{3n}+Q_{4n}+Q_{5n}+Q_{6n}+Q_{8n}+Q_{9n}+Q_{10n}=O_p(h^4)=o_p(n^{-1/2})$. Therefore, we have

$$\begin{aligned}U_n &= Q_{1n}+Q_{7n}+o_p(n^{-1/2})\\ &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} d_{ij}[r_ir_j-2r_i(\hat\sigma^2-\sigma^2)]\\ &\quad+o_p(\frac{1}{\sqrt{n}}).\end{aligned}$$

For the term Q_{1n}, it shows that

$$\begin{aligned}&\sqrt{n}(Q_{1n}-E(r_1r_2d_{12}))\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[r_iE(r_jd_{ij}\,|W_i)-E(r_1r_2d_{12})]\\ &\quad+o_p(\frac{1}{\sqrt{n}}).\end{aligned}$$

For the term Q_{7n}, it follows that

$$\begin{aligned}\sqrt{n}Q_{7n} &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}[(r_i+r_j)d_{ij}]\times\frac{1}{\sqrt{n}}\sum_{i=1}^{n} r_i\\ &\quad+o_p(1)\\ &= E[(r_i+r_j)d_{12}]\times\frac{1}{\sqrt{n}}\sum_{i=1}^{n} r_i+o_p(1).\end{aligned}$$

Define $c_1=E(r_1r_2d_{12}))$ and $c_2=E[(r_i+r_j)d_{12}]$, then we have

$$\begin{aligned}\sqrt{n}(U_n-c_1) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[r_iE(r_jd_{ij}\,|X_i)-c_1-c_2r_i]\\ &\quad+o_p(\frac{1}{\sqrt{n}})\to N(0,\tilde\sigma^2),\end{aligned}$$

where $\tilde\sigma^2=var(r_iE(r_jd_{ij}|W_i)-c_2r_i)$.

(ii) When $nh^8\to\infty$, it follows that $Q_{1n}=O_p(1)$, $Q_{2n}=O_p(n^{-1}h^4)$, $Q_{3n}=Q_{4n}=Q_{10n}=O_p(h^8)$, $Q_{5n}=O_p(n^{-1/2}h^2)$, $Q_{6n}=Q_{7n}=O_p(h^4)$, and $Q_{8n}=Q_{9n}=O_p(n^{-1/2}h^6)$. Then, we have $U_n=O_p(1)$. Thus, we can easily get $nU_n\to\infty$.

Proof of Theorem 3.1: The proof of the theorem is similar to the proof of Theorem 2.1. Recalling that $\mathcal{F}_n=\{X_i,Z_i,Y_i\}_{i=1}^n=\{W_i,Y_i\}_{i=1}^n$, $\hat\varepsilon_i^*=\varepsilon_i^*-(\hat f_i^*-\hat f_i)$ with $\hat f_i^*=\hat g(X_i^T\hat\theta^*)+Z_i^T\hat\beta^*$ and $\hat f_i=\hat g(X_i^T\hat\theta)+Z_i^T\hat\beta$. Define $\sigma^{*2}=E(\varepsilon^{*2}|\mathcal{F}_n)=n^{-1}\sum_{i=1}^n(\hat\varepsilon_i-\bar\varepsilon)^2$, $r_i^*=\varepsilon_i^{*2}-\sigma^{*2}$. Then, we have

$$\begin{aligned}\hat r_i^* &= \hat\varepsilon_i^{*2}-\hat\sigma^{*2}=[\varepsilon_i^*-(\hat f_i^*-\hat f_i)]^2-\hat\sigma^{*2}\\ &= r_i^*-2\varepsilon_i^*(\hat f_i^*-\hat f_i)+(\hat f_i^*-\hat f_i)^2-(\hat\sigma^{*2}-\sigma^{*2}).\end{aligned}$$

Because ε_i^* and ε_j^* with $i\neq j$ are i.i.d., we can get $E(r_i^*|r_j^*,\mathcal{F}_n)=0$ for $i\neq j$. Moreover, it shows

$$\begin{aligned}\hat\sigma^{*2}-\sigma^{*2} &= \frac{1}{n}\sum_{i=1}^{n}(\hat\varepsilon_i^{*2}-\sigma^{*2})\\ &= \frac{1}{n}\sum_{i=1}^{n}[r_i^*-2\varepsilon_i^*(\hat f_i^*-\hat f_i)+(\hat f_i^*-\hat f_i)^2]\\ &= \frac{1}{n}\sum_{i=1}^{n} r_i^*+O_p(h^4+n^{-1/2}h^2+n^{-1}).\end{aligned}$$

Therefore, if $nh^8\to 0$, we have

$$\begin{aligned}U_n^*\,|\mathcal{F}_n &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} r_i^*r_j^*\tilde d_{ij}\\ &\quad+\frac{E(d_{12})E(r^{*2}|\mathcal{F}_n)-2E(d_{12}r_1^{*2}|\mathcal{F}_n)}{n}+o_p(\frac{1}{n}),\end{aligned}$$

Denote $r_i=\varepsilon_i^2-E(\varepsilon_i^2|W_i)$. Under the assumption (A4), we can get $E(r_i^2r_j^2\tilde d_{ij}^2)<16E^2(r^2)<\infty$ and $E(|r_i^2\tilde d_{ii}|)\leq 4E(r^2)<\infty$, which indicates that conditions A1 and A3 in [11] are satisfied. Thus, to prove the conditional asymptotic distribution of $U_n^*|\mathcal{F}_n$ is the same as that of U_n, in term of Lemma 2.1 and Lemma 2.2 in [11], there are three issues need to be proven: (1) the distribution F_r^* of r^* would converge to the distribution F_r of r, (2) $E(r_i^{*2}r_j^{*2}\tilde d_{ij}^2|\mathcal{F}_n)\to E(r_i^2r_j^2\tilde d_{ij}^2)$, and (3)$E(r_i^{*2}\tilde d_{ii}|\mathcal{F}_n)\to E(r_i^2\tilde d_{ii})$.

Since

$$F_r^*(b)=\frac{1}{n}\sum_{i=1}^{n} I((\hat\varepsilon_i-\bar\varepsilon)-\sigma^{*2}\leq b).$$

Define $\Delta f_i=\hat f_i-f_i$ and recall that $\hat\varepsilon_i=\varepsilon_i-\Delta f_i$, then we have

$$\begin{aligned}F_r^*(b) &= \frac{1}{n}\sum_{i=1}^{n} I(\varepsilon_i^2-2\Delta f_i\varepsilon_i+\Delta f_i^2-2\hat\varepsilon_i\bar\varepsilon+\bar\varepsilon^2-g^{*2}\leq b)\\ &= \frac{1}{n}\sum_{i=1}^{n} I(r_i<b+2\Delta f_i\varepsilon_i-\Delta f_i^2+2\hat\varepsilon_i\bar\varepsilon-\bar\varepsilon^2\\ &\quad+g^{*2}-E(\varepsilon^2\,|X_i)).\end{aligned}$$

Let $\Delta r_i=2\Delta f_i\varepsilon_i-\Delta f_i^2+2\hat\varepsilon_i\bar\varepsilon-\bar\varepsilon^2+\sigma^{*2}-E(\varepsilon^2\,|W_i)$. Under H_0, H_{1n} with $c_n\to 0$, we can get either $E(\varepsilon^2|X_i)=E(\varepsilon^2)=\sigma^2$ or $E(\varepsilon^2|W_i)-E(\varepsilon^2)=O_p(c_n)$. Thus, under H_0 and H_{1n}, we have $\sigma^{*2}-E(\varepsilon^2|W_i)=\sigma^{*2}-\sigma^2+\sigma^2-E(\varepsilon^2|W_i)=O_p(n^{-1/2})$ or $O_p(c_n)$. Due to $\hat\beta-\beta=O_p(n^{-1/2})$, $\hat\theta-\theta=O_p(n^{-1/2})$, $\bar\varepsilon=O_p(n^{-1/2})$, $g^{*2}-E(\varepsilon^2|X_i)=O_p(n^{-1/2})$ or $O_p(c_n)$, we can obtain that $\Delta r_i=O_p(n^{-1/2})$ or $O_p(c_n)$. So we can get

$$\begin{aligned}F_r^*(b)-\frac{1}{n}\sum_{i=1}^{n} I(r\leq b) &= \frac{1}{n}\sum_{i=1}^{n} I(r_i\leq b+\Delta r_i)\\ &\quad-\frac{1}{n}\sum_{i=1}^{n} I(r_i\leq b)\\ &\leq \frac{1}{n}\sum_{i=1}^{n} I(|r_i-b|\leq|\Delta r_i|)\\ &= o_p(1).\end{aligned}$$

Because $n^{-1}\sum_{i=1}^n I((r\leq b)\to F_r(b)$, we have $F_r^*(b)\to F_r(b)$.

Notice that

$$\begin{aligned}E(r_i^{*2}r_j^{*2}\tilde d_{ij}^2|\mathcal{F}_n) &= \frac{1}{n}\sum_{i=1}^{n}[(\hat\varepsilon_i-\bar\varepsilon)^2-\sigma^{*2}]^2\\ &\quad\times[(\hat\varepsilon_j-\bar\varepsilon)^2-\sigma^{*2}]^2\tilde d_{ij}^2\end{aligned}$$

$$= \frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i^2-\sigma^2)^2(\varepsilon_j^2-\sigma^2)^2\tilde{d}_{ij}^2+o_p(1)$$
$$= E(r_i^2 r_j^2 \tilde{d}_{ij}^2)+o_p(1).$$

Similarly, we can also get the proof of $E(r_i^{*2}\tilde{d}_{ii}|\mathcal{F}_n) \to E(r_i^2\tilde{d}_{ii})$. Therefore, when $nh^8 \to 0$, we get the conclusion that the asymptotic distribution of $U_n^*|\mathcal{F}_n$ is the same as that of U_n under H_0 and H_{1n}.

If $nh^8 \to \infty$, we have

$$U_n^* - a^*(n)\,|\mathcal{F}_n = 2h^4\frac{1}{n(n-1)}\sum\sum_{1\le i<j\le n} d_{ij}(r_i^* G_j^2 + r_j^* G_i^2) + o_p(n^{-1/2}h^4).$$

By the similar line, we can obtain that the asymptotic distribution of $U_n^*|\mathcal{F}_n$ is the same as that of U_n under H_0 and H_{1n}.

While under H_1, we still can get $E^*(r^*|W_i) = 0$. Consequently, $nU_n^*|F_n$ still converges to a finite limit, which may be different from the limiting distribution of U_n under H_0. However, under H_1, as shown in Theorem 2.3, $nU_n \to \infty$. In other words, the bootstrap algorithm is valid.

REFERENCES

[1] R. Carroll, J. Fan, I. Gijbels, and M. P. Wand, "Generalized partially linear single-index models," *J. Amer. Stat. Assoc.*, vol. 92, no. 438, pp. 477–489, 1988.
[2] R. Carroll and D. Ruppert, *Transformation and Weighting in Regression*. New York, NY, USA: Chapman & Hall, 1988.
[3] P. Craven and G. Wahba, "Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation," *Numer. Math.*, vol. 31, no. 4, pp. 377–403, Dec. 1978.
[4] R. D. Cook and S. Weisberg, "Diagnostics for heteroscedasticity in regression," *Biometrika*, vol. 70, no. 1, pp. 1–10, 1983.
[5] G. G. Gregory, "Large sample theory for U-statistics and tests of fit," *Ann. Statist.*, vol. 5, no. 1, pp. 110–123, Jan. 1977.
[6] X. Guo, T. Wang, and L. Zhu, "Model checking for parametric single-index models: A dimension reduction model-adaptive approach," *J. Roy. Stat. Soc., B (Stat. Methodol.)*, vol. 78, no. 5, pp. 1013–1035, Nov. 2016.
[7] X. Guo, X. Jiang, S. M. Zhang, and L. X. Zhu, "Pairwise distance-based heteroscedasticity test for regressions," *Sci. China, Math.*, to be published.
[8] C. Hsiao and Q. Li, "A consistent test for conditional heteroskedasticity in time-series regression models," *Econom. Theory*, vol. 17, no. 1, pp. 188–221, Feb. 2001.
[9] H. Ichimura, "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models," *J. Econometrics*, vol. 58, nos. 1–2, pp. 71–120, Jul. 1993.
[10] H. L. Koul and W. Song, "Conditional variance model checking," *J. Stat. Planning Inference*, vol. 140, no. 4, pp. 1056–1072, Apr. 2010.
[11] A. Leucht and M. H. Neumann, "Consistency of general bootstrap methods for degenerate U-type and V-type statistics," *J. Multivariate Anal.*, vol. 100, pp. 1622–1633, Sep. 2009.
[12] G. Li, P. Lai, and H. Lian, "Variable selection and estimation for partially linear single-index models with longitudinal data," *Stat. Comput.*, vol. 25, no. 3, pp. 579–593, May 2015.
[13] H. Liang, X. Liu, R. Li, and C.-L. Tsai, "Estimation and testing for partially linear single-index models," *Ann. Statist.*, vol. 38, no. 6, pp. 3811–3836, Dec. 2010.
[14] Y. Ma and L. Zhu, "Doubly robust and efficient estimators for heteroscedastic partially linear single-index models allowing high dimensional covariates," *J. Roy. Stat. Soc., B (Stat. Methodol.)*, vol. 75, no. 2, pp. 305–322, Mar. 2013.
[15] J. P. Nolan, "Multivariate elliptically contoured stable distributions: Theory and estimation," *Comput. Statist.*, vol. 28, no. 5, pp. 2067–2089, Oct. 2013.
[16] H. Peng and T. Huang, "Penalized least squares for single index models," *J. Stat. Planning Inference*, vol. 141, no. 4, pp. 1362–1379, Apr. 2011.
[17] N. Samarakoon and W. Song, "Minimum distance conditional variance function checking in heteroscedastic regression models," *J. Multivariate Anal.*, vol. 102, no. 3, pp. 579–600, Mar. 2011.
[18] N. Samarakoon and W. Song, "Empirical smoothing lack-of-fit tests for variance function," *J. Stat. Planning Inference*, vol. 142, no. 5, pp. 1128–1140, May 2012.
[19] R. H. Randles, "Approximation theorems of mathematical statistics," *Technometrics*, vol. 24, no. 3, p. 254, Aug. 1982.
[20] J. S. Simonoff and C. L. Tsai, "Use of modified profile likelihood for improved tests of constancy of variance in regression," *J. Appl. Statist.*, vol. 43, no. 2, pp. 357–370, Feb. 1994.
[21] P. Speckman, "Kernel smoothing in partial linear models," *J. Roy. Stat. Soc., B (Methodol.)*, vol. 50, no. 3, pp. 413–436, Jul. 1988.
[22] L. Su and A. Ullah, "A nonparametric goodness-of-fit-based test for conditional heteroskedasticity," *Econom. Theory*, vol. 29, no. 1, pp. 187–212, Feb. 2013.
[23] L. Wang and X.-H. Zhou, "Assessing the adequacy of variance function in heteroscedastic regression models," *Biometrics*, vol. 63, no. 4, pp. 1218–1225, Dec. 2007.
[24] W. Wang and Z.-Y. Zhu, "Variable selection for the partial linear single-index model," *Acta Math. Appl. Sin. Engl. Ser.*, vol. 33, no. 2, pp. 373–388, Apr. 2017.
[25] V. M. Zolotarev, "Integral transformations of distributions and estimates of parameters of multidimensional spherically symmetric stable laws," in *Contributions to Probability*. New York, NY, USA: Academic, 1981, pp. 283–305.
[26] J. Zhang, Y. Gai, and B. Lin, "Detection of marginal heteroscedasticity for partial linear single-index models," *Commun. Statist.-Simul. Comput.*, Feb. 2019, doi: 10.1080/03610918.2019.1565585.
[27] Y. Zhang, H. Lian, and Y. Yu, "Estimation and variable selection for quantile partially linear single-index models," *J. Multivariate Anal.*, vol. 162, pp. 215–234, Nov. 2017.
[28] J. Xu Zheng, "A consistent test of functional form via nonparametric estimation techniques," *J. Econometrics*, vol. 75, no. 2, pp. 263–289, Dec. 1996.
[29] X. Zheng, "Testing heteroscedasticity in nonlinear and nonparametric regressions," *Can. J. Statist.*, vol. 37, no. 2, pp. 282–300, Jun. 2009.
[30] X. Zhu, X. Guo, L. Lin, and L. Zhu, "Heteroscedasticity checks for single index models," *J. Multivariate Anal.*, vol. 136, pp. 41–55, Apr. 2015.
[31] X. Zhu, X. Guo, and L. Zhu, "An adaptive-to-model test for partially parametric single-index models," *Stat. Comput.*, vol. 27, no. 5, pp. 1193–1204, Sep. 2017.

JIAN-QIANG ZHAO received the B.S. and M.S. degrees from Shandong University, Jinan, China, in 2004 and 2007, respectively, and the Ph.D. degree from the China University of Mining and Technology, Xuzhou, China, in 2015.

He is currently with the School of Mathematics and Statistics, Xuzhou University of Technology. His research interests include applied statistics, operations research, and intelligent algorithms. He is also a Council Member of the Environmental and Resource Statistics Branch of the China Association for on-the-spot Statistics, the Probability and Statistics Association of Jiangsu Province, China, the Jiangsu Institute of Operations Research, and the Statistics Society of Jiangsu Province, China.

JIANQUAN LI received the M.S. degree from Tianjin University. He is currently the Director of Deep Learning Lab, Beijing Ultrapower Software Company Ltd. His research interests include natural language processing, data mining, and deep learning.

YAN-YONG ZHAO received the B.S. degree from Liaocheng University, Liaocheng, China, in 2010, the M.S. degree from the Chongqing University and Technology, Chongqing, China, in 2013, and the Ph.D. degree from Southeast University, Nanjing, China, in 2016.

He is currently with the Department of Statistics, Nanjing Audit University. His research interests include nonparametric and semiparametric model analysis, longitudinal data analysis, and statistical applications. He is currently a Council Member of the Environmental and Resource Statistics Branch of the China Association for on-the-spot Statistics and the Probability and Statistics Association of Jiangsu Province, China.

JUNYAN HE is currently pursuing the M.S. degree in applied statistics with the Xuzhou University of Technology, Xuzhou, China. She is also good at algorithm programming and statistical software.

WALED KHALED received the B.S. degree in applied statistics from Damascus University, Syria, in 2010, and the M.S. degree in statistics from Hohai University, China, in 2015. He is currently pursuing the Ph.D. degree with the School of Mathematics, Southeast University, China. His research interests include checking heteroscedasticity in regression models, simulation analysis, and estimation methods.