SCIENCE CHINA
Mathematics

• ARTICLES •

doi: 10.1007/s11425-000-0000-0

Pairwise distance-based heteroscedasticity test for regressions

Xu Guo[1], Xuejun Jiang[2,*], Shumei Zhang[1] & Lixing Zhu[1,3]

[1]*School of Statistics, Beijing Normal University, Beijing* 100875*, China;*
[2]*Department of Mathematics, Southern University of Science and Technology of China, Shenzhen* 518055*, China;*
[3]*Department of Mathematics, Hong Kong Baptist University, Hong Kong, China*

Email: xustat12@bnu.edu.cn, jiangxj@sustc.edu.cn, zsm1963@bnu.edu.cn, lzhu@hkbu.edu.hk

Received February 15, 2018; accepted January 8, 2019

Abstract In this study, we propose nonparametric testing for heteroscedasticity in nonlinear regression models based on pairwise distances between points in a sample. The test statistic can be formulated such that U-statistic theory can be applied to it. Although the limiting null distribution of the statistic is complicated, we can derive a computationally feasible bootstrap approximation for such a distribution; the validity of the introduced bootstrap algorithm is proven. The test can detect any local alternatives that are different from the null at a nearly optimal rate in hypothesis testing. The convergence rate of this test statistic does not depend on the dimension of the covariates, which significantly alleviates the impact of dimensionality. We provide three simulation studies and a real-data example to evaluate the performance of the test and demonstrate its applications.

Keywords dimensionality; heteroscedasticity testing; pairwise distance; U-statistic theory.

MSC(2010) 62G08, 62G20

Citation: Guo, X., Jiang, X. J., Zhang, S. M. and Zhu, L. X.. Pairwise distance-based heteroscedasticity test. Sci China Math, 2017, 60, doi: 10.1007/s11425-000-0000-0

1 Introduction

In regression analysis, all error terms are frequently assumed to have common variance: under such an assumption, the variance of errors can be easily estimated, and an estimation on the regression function can be calculated; without such an assumption, however, more complicated methods are required to estimate the regression function. Therefore, it is important to detect heteroscedasticity in various regression models.

In this study, we consider the following nonlinear regression model:

$$Y = f(X, \beta) + \epsilon, \tag{1.1}$$

where ϵ is the error term with unknown distribution, $Y \in \mathbb{R}$ is the response with covariates $X \in \mathbb{R}^p$, and the function $f(\cdot, \beta)$ is known as well as twice differentiable up to a d-dimensional vector of parameters β.

* Corresponding author

Furthermore, $E(\epsilon|X) = 0$ is assumed. Model (1.1) is commonly used in this context because it is easy to interpret and presents well-developed theories. When $f(X,\beta) = X^\top\beta$, model (1.1) is transformed into the classical linear model. When compared with the classical linear model, model (1.1) is more flexible and applicable because high-order terms of the covariates can be included.

Our objective is to detect variance heterogeneity in the aforementioned model by focusing on the following hypothesis testing:

$$\begin{aligned} &H_0 : \exists\ \sigma^2 > 0,\ E(\epsilon^2|X) \equiv \sigma^2(X) = \sigma^2, \\ &H_1 : \forall\ \sigma^2 > 0,\ E(\epsilon^2|X) \neq \sigma^2. \end{aligned} \tag{1.2}$$

Under H_0, the constant σ^2 is an unconditional variance $E(\epsilon^2)$. Consequently, the heteroscedasticity test in (1.2) is equivalent to determining whether the conditional variance function $E(\epsilon^2|X)$ is equal to the unconditional variance $E(\epsilon^2)$.

A number of authors have contributed to the study of heteroscedasticity tests in nonlinear regression models. Cook and Weisberg (1983) as well as Tsai (1986) proposed score tests for a parametric structure variance function in linear regression models and first-order autoregressive models, respectively. Simonoff and Tsai (1994) further developed a modified score test in linear models. More recently, Lin and Wei (2003) developed score tests for heteroscedasticity in nonlinear regression models. Following the work of Zheng (1996) on checking lack-of-fit in the mean function, Zheng (2009) proposed a quadratic form of the conditional moment test for heteroscedasticity in nonlinear regression models. Hsiao and Li (2001) investigated the tests of heteroscedasticity for nonlinear time-series regression models, while Su and Ullah (2013) introduced a nonparametric test for conditional heteroscedasticity in nonlinear regression models based on a measure of nonparametric goodness-of-fit (R^2). In addition to heteroscedasticity tests, several authors, including Wang and Zhou (2007), Koul and Song (2010), and Samarakoon and Song (2011, 2012), have discussed testing the goodness-of-fit of a given parametric variance function in nonlinear regression models.

A drawback of some existing methods is the dimensionality problem, which arises due to estimation inefficiency for the multivariate nonparametric function. This particular problem is discussed in the work of Hsiao and Li (2001), Zheng (2009), and Su and Ullah (2013). Under the null hypotheses in these studies, the test statistics that are multiplied by $O(n^{1/2}h^{p/4})$ converge to their weak limits, with n being the sample size and h being the bandwidth used in nonparametric estimation. Under the assumption that $h \to 0$ as $n \to \infty$, the rate $O(n^{1/2}h^{p/4})$ can be extremely slow when p is large. Therefore, the significance level frequently cannot be sufficiently maintained when the limiting null distribution is used in scenarios with a moderate sample size. Furthermore, these tests can only detect alternative hypotheses that differ from the null hypothesis at the rate of order $O(n^{-1/2}h^{-p/4})$. Asymptotically, these test statistics are less powerful for detecting alternative models.

In the present study, we propose a new statistic for testing heteroscedasticity in nonlinear regression models. This statistic is based on the weighted integral of the residual-marked characteristic function. The weight function—in this case, the density function of a spherical stable law—plays an important role in the proposed test statistic. Given this choice, the weighted integral is transformed into an unconditional expectation with a simple form. The proposed statistic is based only on pairwise distances between points in a sample. The basic concept is adopted from the approach of Bierens (1982). The characteristic function has been used in many hypothesis-testing problems such as those by Alba-Fernández et al. (2008), Székely and Rizzo (2013), and Fan et al. (2017). Meintanis (2016) provided an informative review of the aforementioned use of the characteristic function. To the best of our knowledge, however, this study is the first to use the characteristic function to detect heteroscedasticity for nonlinear regression models.

For theoretical investigations, we formulate the proposed test statistic as a simple U-statistic to apply U-statistic theory instead of empirical process theory and investigate its asymptotic properties under null, fixed alternative, and local alternative hypotheses. The asymptotic null distribution exhibits a non-trivial form, similar to most cases for U-statistics. Subsequently, we propose a residual-based bootstrap algorithm to approximate the critical values of the test statistic and provide a rigorous proof for its consistency. The test statistic multiplied by n can converge to its weak limit and can detect local alternatives that are distinct from the null hypothesis at a rate that is as close as possible to $n^{-1/2}$. This rate is the fastest convergence rate in hypothesis testing. Moreover, the convergence rate of the test statistic does not depend on the dimensions of the covariates. Thus, the test is not sensitive to dimensionality. Remarkably, estimating the nonparametric conditional variance function $E(\epsilon^2|X)$ is not required in the construction of our test statistic. Therefore, typical bandwidth selection is completely avoided.

The rest of this paper is organized as follows. In Section 2, we describe the construction of test statistic and establish its asymptotic properties under all types of hypotheses. In Section 3, a bootstrap algorithm for implementing the proposed test is presented and the validity of this algorithm is justified. In Section 4, we conduct numerical studies to evaluate the performance of the test and include a real-data example to demonstrate its applications. Concluding remarks and discussions are presented in Section 5, and proofs are provided in the Appendix.

2 Test statistic and asymptotic results

2.1 Construction

Let $\eta = \epsilon^2 - \sigma^2$ with $\sigma^2 = E(\epsilon^2)$. Under the null hypothesis, we have $E(\eta|X) = 0$. Given the uniqueness of a function's Fourier transform, the null hypothesis in (1.2) is equivalent to

$$H_0 : \phi(t) = E[\eta e^{it^\top X}] = 0, \ \ \forall t \in \mathbb{R}^p.$$

However, $\phi(t)$ cannot be a statistic by itself. This limitation motivates us to consider the following quantity:

$$D_\omega = \int_{\mathbb{R}^p} |\phi(t)|^2 \omega(t) dt, \tag{2.1}$$

where $\omega(t) \geqslant 0$ is a suitable weight function.

From the definition of a complex modulus, we can easily obtain

$$|\phi(t)|^2 = E\left[\cos(t^\top (X - X'))\eta\eta'\right]$$

with (X', η') existing as an independent copy of (X, η). As Nolan (2013) presents, the characteristic function of a spherical stable law is given by

$$\phi_Z(t) = \int_{\mathbb{R}^p} \cos(t^\top z) f_{a,p}(z) dz = e^{-||t||^a},$$

where $\|\cdot\|$ is the Euclidean norm, and $f_{a,p}(\cdot)$ denotes the density of a spherical stable law in $\mathbb{R}^p$ with the 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. For more information about this family, see Nolan (2013).

Using the weight function $\omega(t) = f_{a,p}(t)$, we obtain

$$D_\omega = E\left[e^{-\|X-X'\|^a}\eta\eta'\right]. \tag{2.2}$$

When the dimensionality p of X is high, the integral $\int_{\mathbb{R}^p} |\phi(t)|^2 \omega(t) dt$ can be problematic. However, we end up with a simple and closed form without involving a high-dimensional integral with the aforementioned weight function. A significant fact is that the null hypothesis (1.2) is true if and only if $D_\omega = 0$. This fact can then be used as a criterion for this hypothesis testing problem. When the i-th sample, η_i, of η is available, we can then estimate D_ω through its sample analogue.

Suppose $(X_i, Y_i), i = 1, \cdots, n$ are independent and identically distributed (i.i.d.) samples from the population (X, Y). We consider the following test statistic

$$T_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \hat{\eta}_i \hat{\eta}_j d_{ij},$$

where $d_{ij} = e^{-\|X_i - X_j\|^a}$, $\hat{\eta}_i = \hat{\epsilon}_i^2 - \hat{\sigma}^2$, $\hat{\epsilon}_i = Y_i - f(X_i, \hat{\beta})$ with $\hat{\beta}$ being the nonlinear least squares estimation of β, and $\hat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} \hat{\epsilon}_i^2$. Through the distance measure D_ω, the conditional variance function $E(\epsilon^2|X)$ should no longer be estimated using nonparametric methods such as kernel, local polynomial, and spline. Evidently, T_n is a pairwise distance-based test statistic, which presents an alternative method.

2.2 Asymptotic results

In this subsection, we present the asymptotic properties of the test statistic T_n. To obtain the asymptotic properties of T_n, we need the following conditions.

(C1). The parameter space $\mathcal{B}$ of β is a compact subset of $\mathbb{R}^d$; $E[Y - f(X, \beta)]^2$ has a unique minimum at β_0, an interior point of $\mathcal{B}$;

(C2). The regression function $f(x, \beta)$ is continuously differentiable of order 2 in β. Let $\nabla f(x, \beta) \equiv \partial f(x, \beta)/\partial \beta$ and $\nabla^2 f(x, \beta) \equiv \partial^2 f(x, \beta)/\partial \beta \partial \beta^\top$. The terms $\nabla f(x, \cdot)$ and $\nabla^2 f(x, \cdot)$ are continuous in x and are dominated by functions $F_1(x)$ and $F_2(x)$, respectively. The functions $F_1(X)$, $F_2(X)$ have finite fourth and second moments, respectively.

(C3). $\Sigma = E(\nabla f(X, \beta) \nabla f^\top(X, \beta))$ is nonsingular.

(C4). $E(\eta^2) < \infty$.

The first three conditions are commonly assumed to derive the asymptotic normalities of $\hat{\beta}$ and $\hat{\sigma}^2$. For example, see Hsiao and Li (2001), Zheng (2009), and Su and Ullah (2013). The last condition, also extremely mild, is required for the asymptotic properties of related U-statistics and is equivalent to the requirement that $E(\epsilon^4) < \infty$.

Let $W \equiv (X, \eta) \sim F(W)$ and $W_i \equiv (X_i, \eta_i)$ be independent copies of W, denoted by $\tilde{d}_{ij} = d_{ij} - E(d_{il} + d_{jl}|X_i, X_j) + E(d_{12})$ with $l \neq i, j$, and $h(W_1, W_2) = \eta_1 \eta_2 \tilde{d}_{12}$. The following theorem states the limit distribution of T_n under the null hypothesis.

Theorem 2.1. *Under the null hypothesis in (1.2) and the conditions (C1)–(C4), we have*

$$nT_n \Rightarrow \sum_{k=1}^{\infty} \lambda_k Z_k^2 - E(\eta^2),$$

as $n \to \infty$, where Z_ks are independent standard normal random variables and λ_ks are the eigenvalues of the integral equation

$$\int \eta_j^2 \tilde{d}_{ij} \phi_k(W_j)\, dF(W_j) = \lambda_k \phi_k(W_i),$$

with $\phi_k(W_j)$ being the associated orthonormal eigenfunctions.

Under $E(\eta^2) < \infty$, we have $Eh^2(W_1, W_2) < \infty$. From P197 in Serfling (1980), we know that $\sum_{k=1}^{\infty} \lambda_k^2 = Eh^2(W_1, W_2)$; thus, the infinite sum $\sum_{k=1}^{\infty} \lambda_k Z_k^2$ actually converges in L_2. See also Leucht and Neumann (2009). Similar to most cases for U-statistics, the aforementioned limit distribution of T_n cannot be directly applied to compute critical values because λ_ks are difficult to obtain. To overcome this challenge, we propose a bootstrap approximation for critical values and demonstrate the consistency of the developed algorithm, which is described in Section 3.

To provide additional insight on the structure of the limit, we introduce the oracle test statistic as if the true error η_i were observable:

$$T_n^O = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \eta_i \eta_j d_{ij}.$$

This oracle statistic weakly converges to $\sum_{k=1}^{\infty}\tilde{\lambda}_k Z_k^2 - E(\eta^2)$, where the constant $\tilde{\lambda}_k$ is the eigenvalue of another integral equation, i.e.,

$$\int h^*(W_i, W_j)\phi_k^*(W_j)\, dF(W_j) = \tilde{\lambda}_k \phi_k^*(W_i),$$

where $h^*(W_i, W_j) = \eta_i\eta_j d_{ij}$ and Z_k denotes an independent standard normal random variable. Evidently, the limiting null distribution of the test statistic with the residual $\hat{\eta}_i$ exhibits the same structure as that of the aforementioned oracle statistic with varying eigenvalues. That is, the estimations of parameters β and σ^2 notably impact the weak limit.

Subsequently, we investigate the sensitivity of the test statistic using a sequence of local alternatives with the following form:

$$H_{1n} : E(\epsilon^2|X) = \sigma^2 + c_n V(X), \tag{2.3}$$

where $\sigma^2 = E(\epsilon^2)$, $E(V^2(X)) < \infty$ and c_n represents a sequence converging to zero. Following the above local alternative hypotheses, $\eta_i = \epsilon_i^2 - \sigma^2$ can be rewritten as $\eta_i = u_i + c_n V(X_i)$, where $E(u_i|X_i) = 0$ and $E(V(X)) = 0$.

Then, we have the following theorem under H_{1n}.

Theorem 2.2. *Under the local alternative hypotheses in (2.3) and the conditions presented in Theorem 2.1, we have*

- *with* $c_n = n^{-1/2}$,

$$nT_n \Rightarrow \sum_{k=1}^{\infty}\lambda_k(Z_k + a_k)^2 - E(u^2),$$

where $a_k = E(V(X)\phi_k(W))$ *and* $\phi_k(W)$ *is defined in Theorem 2.1;*

- *with* $c_n = n^{-r}, 0 < r < 1/2$, $nT_n \Rightarrow \infty$.

The preceding theorem implies that the test is still valid when the local alternatives converge to the null hypothesis at the rate $n^{-1/2}$. When the local alternatives are distinct from the null hypothesis at a slower rate, n^{-r} with $0 < r < 1/2$, the asymptotic power tends towards 1, thereby implying that the test is consistent.

If we set c_n as a fixed value other than 0, the local alternative hypotheses H_{1n} defined in (2.3) is transformed into the fixed alternative hypothesis, H_1. That is,

$$H_1 : E(\epsilon^2|X) = \sigma^2 + cV(X) \neq \sigma^2. \tag{2.4}$$

Here $\sigma^2 = E(\epsilon^2)$. We also consider the asymptotic property of the proposed test statistic under the fixed alternative hypothesis. The result is presented as follows.

Theorem 2.3. *Under the fixed alternative hypothesis* H_1 *in (2.4) and the conditions in Theorem 2.1, we have*

$$\sqrt{n}(T_n - E(\eta_1\eta_2 d_{12})) \Rightarrow N(0, \tilde{\sigma}^2),$$

with $\tilde{\sigma}^2 = var\Big(\eta_1 E(\eta_2 d_{12}|X_1) - 2E(\eta_1 d_{12})\eta_1\Big)$.

The convergence rates of T_n are interestingly very different under H_0 and H_1 and do not depend on the dimensions of covariates X under either the null, fixed, or local alternative hypotheses. These advantages are also observed in the following simulation studies. The proposed test can perform well in finite sample cases, particularly when the covariate X has a relatively large dimension.

3 Numerical implementation

A bootstrap approximation is adopted to determine critical values. The algorithm is residual-based and given as follows:

(1). The residuals $\hat{\epsilon}_i = Y_i - f(X_i, \hat{\beta})$, where $\hat{\beta}$ is the nonlinear least squares estimator of β, are obtained;

(2). The bootstrap error ϵ_i^* is obtained by randomly resampling with replacement from the set $\{\hat{\epsilon}_i - \bar{\hat{\epsilon}}, i = 1, \cdots, n\}$ with $\bar{\hat{\epsilon}} = n^{-1}\sum_{i=1}^n \hat{\epsilon}_i$. Then, $Y_i^* = f(X_i, \hat{\beta}) + \epsilon_i^*$;

(3). Y_i^* is regressed on X_i to obtain the estimator $\hat{\beta}^*$. The bootstrap residuals, $\hat{\epsilon}^* = Y_i^* - f(X_i, \hat{\beta}^*)$ and $\hat{\eta}_i^* = \hat{\epsilon}_i^{*2} - \hat{\sigma}^{*2}$ with $\hat{\sigma}^{*2} = n^{-1}\sum_{i=1}^n \hat{\epsilon}_i^{*2}$, are calculated. The test statistic T_n^* is constructed based on $(X_1, \hat{\eta}_1^*), \ldots, (X_n, \hat{\eta}_n^*)$;

(4). Steps 2–3 are repeated B times, and the bootstrap statistics are denoted as $\{T_{n,b}^*\}_{b=1}^B$. The bootstrap p-value is calculated using $p^* = B^{-1}\sum_{b=1}^B I(T_{n,b}^* > T_n)$, where $I(\cdot)$ is an indicator function.

The preceding algorithm has been adopted by many authors, including Hsiao and Li (2001), Wang and Zhou (2007), and Su and Ullah (2013). The number of bootstrap samples is set to $B = 500$. Heuristically, steps 2 and 3 ensure that conditional to the original sample $\mathcal{F}_n = (X_i, Y_i)_{i=1}^n$, the bootstrap sample (X_i, Y_i^*) satisfies the null hypothesis. That is, conditional to the original sample $\mathcal{F}_n$, the bootstrap replicates ϵ_i^* are i.i.d. with a mean of 0 and variance of $n^{-1}\sum_{i=1}^n(\hat{\epsilon}_i - \bar{\hat{\epsilon}})^2$. Therefore, the bootstrap distribution obtained in step 3, conditional to the random sample $\mathcal{F}_n$, approximates the null distribution of the test statistic T_n even when the null hypothesis is false.

The following theorem demonstrates that the preceding bootstrap procedure provides a valid approximation of the null distribution for the T_n test.

Theorem 3.1. *Under the conditions stated in Theorem 2.1, we have the following:*

(i) Under the null hypothesis, H_0, or the local alternative hypotheses, H_{1n}, with $c_n \to 0$, the limiting conditional distribution of $nT_n^*|\mathcal{F}_n$ is the same as the limiting null distribution of the test statistic, nT_n.

(ii) Under the fixed alternative hypothesis H_1, the limiting conditional distribution of $nT_n^*|\mathcal{F}_n$ is a finite limit, which may differ from the limiting null distribution of the test statistic nT_n.

Theorem 3.1 demonstrates that the bootstrap algorithm can effectively control the size of the test statistic T_n. We then investigate the power performance of this test. From Theorem 2.2, under the local alternative hypotheses H_{1n} with $c_n = n^{-r}(0 < r < 1/2)$, $nT_n \Rightarrow \infty$. This implies that the preceding bootstrap procedure can have the asymptotic power of 1 in this case. Under the local alternative hypotheses with $c_n = n^{-1/2}$, the bootstrap procedure can still detect the alternative hypotheses. From Theorem 2.3, under the fixed alternative hypothesis H_1 defined in equation (2.4), $nT_n \Rightarrow \infty$. This indicates the preceding bootstrap procedure can have the asymptotic power of 1 under the fixed alternative hypothesis. In summary, the bootstrap algorithm is valid.

4 Numerical studies

In this section, we first conduct simulation studies to demonstrate the performance of our proposed test statistic and then to conduct real-data analysis.

4.1 Simulations

In the simulations, $1,000$ replications of the experiment were used to compute empirical sizes and powers at significance level $\alpha = 0.05$. To examine power performance, the following three examples were designed. In the first two examples, the sample sizes $n = 100, 150$ and 200 were considered. The true parameter was $\beta = (1, 2, 3, 0, \cdots, 0)^\top/\sqrt{14}$; p was set to 4 and 8; the observations $X_i = (X_{1i}, X_{2i}, \cdots, X_{pi})$ with $i = 1, \cdots, n$, independent of the standard normal errors ε_i, were i.i.d with the common uniform distribution on the p-dimensional cube $[-1, 1]^p$. Linear regression models were used in the two examples; when $p = 4$, the two examples were the same as Examples 1 and 2 in Zhu et al. (2015) for the purpose of comparison. In the third example, the model was nonlinear. In this example, $n = 100, 200$; $p = 4, 8$; $\beta = (1, 1, \cdots, 1)^\top/\sqrt{p}$; $X_i \sim N(0, I_p)$; and $\varepsilon_i \sim N(0, 1)$. The test statistic by Zheng (2009), designed for nonlinear regression models, was also considered and denoted as T_n^{ZH}, which resulted in the following form:

$$T_n^{ZH} = \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j\neq i}^n \hat{\eta}_i\hat{\eta}_j\frac{1}{h^p}K\left(\frac{X_i - X_j}{h}\right),$$

Table 1 Empirical sizes ($\delta = 0$) and powers ($\delta = 1$) for H_0 in Example 1 with $n = 100, p = 4$, and different a values.

a	0.25	0.5	0.75	1	1.25	1.5	1.75	2
$\delta = 0$	0.0040	0.0220	0.0420	0.0510	0.0710	0.0500	0.0640	0.0660
$\delta = 1$	0.0180	0.1300	0.2300	0.2570	0.3210	0.3500	0.3690	0.4070

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

Example 1: Data are generated from the following model:

$$Y = 2 + 2(X^\top \beta) + (\delta |X^\top \beta| + 0.5) \times \varepsilon.$$

The null hypothesis corresponds to $\delta = 0$, whereas the alternative hypothesis corresponds to $\delta \neq 0$.

To use the proposed test, we need to first determine the value of the characteristic exponent $a \in (0, 2]$. We conduct a simple simulation study for this purpose. Table 1 reports the empirical sizes and powers of our proposed test statistic for $n = 100, p = 4$ when a varies. From this table, we find that as long as a is not too small ($a \geqslant 0.5$), the proposed test can efficiently control empirical sizes extremely well. However, when a is larger than 1.5, empirical sizes are slightly higher than the significance level 0.05. For empirical powers, the powers increase with increasing a. Consequently, we use $a = 1.5$ in the following studies.

For T_n^{ZH}, the multiple Gaussian kernel function $K(u_1, u_2, \cdots, u_p) = k(u_1)k(u_2)\cdots k(u_p)$ is used with $k(u) = 1/(\sqrt{2\pi})e^{-\frac{u^2}{2}}$. To select a bandwidth, we use simulations to choose an appropriate value from the grid points $j/100$ for $j = 11, 15, \cdots, 99$. From $1,000$ simulations, we plot the estimated power curves against the aforementioned bandwidth sequences with a sample size of 100, dimensions of covariates $p = 4$, and $\delta = 0, 1$, which are shown in Figure 1. This strategy has also been used by many authors, including Sun and Wang (2009) and Lopez and Patilea (2009). As shown in Figure 1, bandwidth affects the size and power performances of T_n^{ZH}. In particular, when the bandwidth h is too small or too large, T_n^{ZH} cannot efficiently control empirical size and has extremely low power. From this experiment, $h = 0.45$ is a reasonable choice.

The power performance of the proposed test statistics with $p = 4, 8$; $n = 100, 150, 200$ and $\delta = 0, 0.5, \cdots, 2$ is presented in Table 2. This table shows that even when $p = 8$, the proposed test T_n and Zheng's test T_n^{ZH} can efficiently control the sizes. Both tests have higher power when the sample size increases and the deviation from the hypothetical model is larger. Moreover, test T_n has higher powers than T_n^{ZH}. This finding is appropriate given that T_n^{ZH} converges to its weak limit at a very slow rate of $n^{-1/2}h^{-p/4}$.

When dimension p increases from 4 to 8, the power of T_n decreases. This result implies that although the convergence rate of T_n does not depend on the dimension of covariates, the dimension of covariates does affect power performances in practice. However, we also notice that even when $p = 8$, the proposed test is still sensitive to the alternatives. Nevertheless, when the dimension p is 8, T_n^{ZH} fails completely. It has extremely low power even when $n = 200$ and $\delta = 2$. Thus, when the dimension of covariates is high, T_n works better.

Example 2: The data are generated from the following model:

$$Y = 1 + 2(X^\top \beta) + (\delta |X_1 + X_2| + 1) \times \varepsilon.$$

where X, ε, β, and δ have the same settings as those in Example 1.

The simulation results are summarized in Table 3. The conclusions are similar to those of the model in Example 1. From the two examples, we can conclude that a dimensionality effect occurs in the proposed test T_n; however, the effect is not as serious as that in Zheng's test. When $p = 4$, the two examples are the same as Examples 1 and 2 in Zhu et al. (2015). Their test is denoted as T_n^Z. The power curves of T_n and T_n^Z in the two examples with $n = 100, 200$ are plotted in Figure 2. As shown in the figure, T_n is not as powerful as T_n^Z for Example 1 because an additional dimension reduction structure for the conditional variance function is applied in T_n^Z, or equivalently, a single-index structure $E(\epsilon^2|X) = E(\epsilon^2|X^\top\beta)$ is assumed for the null hypothesis of T_n^Z. By contrast, the proposed test T_n does

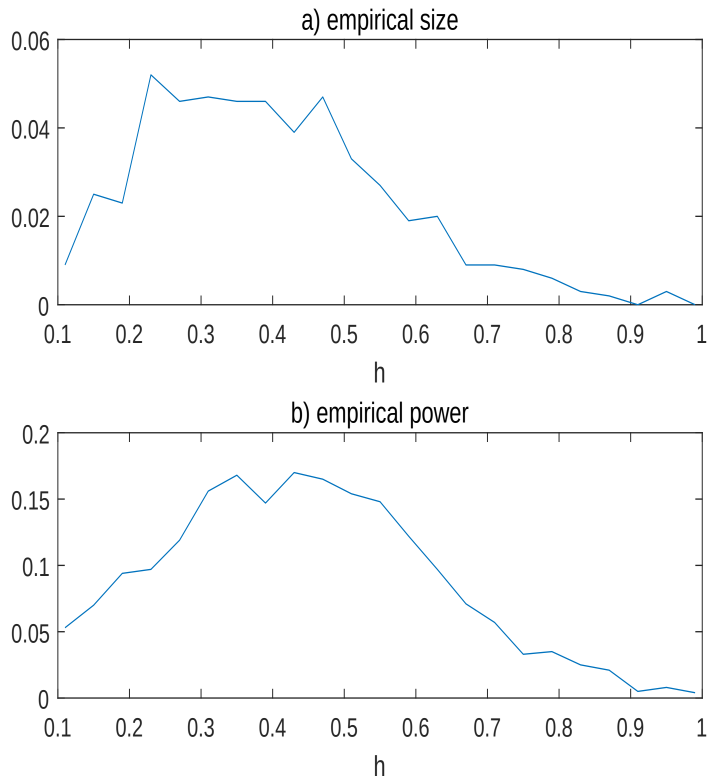

Figure 1 Estimated size and power curves of test T_n^{ZH} against bandwidth h with $n = 100, p = 4$ under $\delta = 0, 1$ in Example 1.

Table 2 Empirical sizes and powers for H_0 in Example 1 with $p = 4, 8; n = 100, 150, 200;$ and different δ values.

	δ	$n = 100$	$n = 150$	$n = 200$
$p = 4, T_n$	0	0.0510	0.0490	0.0540
	0.5	0.1930	0.2580	0.3960
	1.0	0.3500	0.5360	0.7380
	1.5	0.4320	0.6550	0.8420
	2.0	0.5240	0.7600	0.9010
$p = 4, T_n^{ZH}$	0	0.0460	0.0580	0.0470
	0.5	0.0680	0.1570	0.2260
	1.0	0.1760	0.3780	0.6280
	1.5	0.2830	0.5850	0.8230
	2.0	0.3350	0.6520	0.8790
$p = 8, T_n$	0	0.0560	0.0540	0.0520
	0.5	0.1330	0.1680	0.2070
	1.0	0.2050	0.2890	0.4100
	1.5	0.2460	0.3620	0.5010
	2.0	0.2990	0.4200	0.5450
$p = 8, T_n^{ZH}$	0	0.0470	0.0550	0.0560
	0.5	0.0290	0.0500	0.0640
	1.0	0.0650	0.0980	0.1320
	1.5	0.0610	0.1120	0.2370
	2.0	0.1020	0.1760	0.2740

Table 3 Empirical sizes and powers for H_0 in Example 2 with $p = 4, 8; n = 100, 150, 200$; and different δ values.

	δ	$n = 100$	$n = 150$	$n = 200$
$p = 4, T_n$	0	0.0610	0.0470	0.0560
	0.5	0.1440	0.1940	0.2460
	1.0	0.2670	0.4070	0.5920
	1.5	0.3750	0.5970	0.7580
	2.0	0.4670	0.6880	0.8630
$p = 4, T_n^{ZH}$	0	0.0500	0.0490	0.0470
	0.5	0.0510	0.0920	0.1310
	1.0	0.1210	0.2730	0.4640
	1.5	0.1860	0.4500	0.6910
	2.0	0.2590	0.5660	0.8210
$p = 8, T_n$	0	0.0620	0.0560	0.0510
	0.5	0.1210	0.1300	0.1560
	1.0	0.1600	0.2420	0.2800
	1.5	0.2200	0.2880	0.4260
	2.0	0.2520	0.3690	0.4980
$p = 8, T_n^{ZH}$	0	0.0390	0.0530	0.0550
	0.5	0.0540	0.0460	0.0450
	1.0	0.0420	0.0630	0.0920
	1.5	0.0690	0.1080	0.1680
	2.0	0.0730	0.1510	0.2080

Table 4 Empirical sizes and powers for H_0 in Example 3 with $p = 4, 8; n = 100, 200$; and different δ values.

	δ	$n = 100$	$n = 200$
$p = 4, T_n$	0	0.0500	0.0570
	0.5	0.3540	0.8200
	1.0	0.6640	0.9620
	1.5	0.6820	0.9860
	2.0	0.7060	0.9940
$p = 4, T_n^{ZH}$	0	0.0440	0.0500
	0.5	0.2140	0.5740
	1.0	0.4380	0.8520
	1.5	0.5260	0.9160
	2.0	0.5800	0.9300
$p = 8, T_n$	0	0.0520	0.0480
	0.5	0.1280	0.3360
	1.0	0.2060	0.5740
	1.5	0.2540	0.6420
	2.0	0.2560	0.7020
$p = 8, T_n^{ZH}$	0	0.0280	0.0240
	0.5	0.0720	0.1040
	1.0	0.1420	0.2400
	1.5	0.1260	0.2860
	2.0	0.1460	0.3500

not require this assumption. Moreover, when the sample size increases from 100 to 200, the differences between the powers of T_n and T_n^Z can be extremely small for Example 1, as shown in Figure 2. When the single index structure fails to hold, such as in Example 2, T_n^Z has very low powers, as shown in the figure. Furthermore, when the sample size increases, our proposed test statistic T_n can have more powers than T_n^Z.

Example 3: The data are generated from the following model:

$$Y = (X^\top \beta + 0.3)^2 + (\delta |X^\top \beta_1| + 0.5) \times \varepsilon.$$

In this example, $E(Y|X) = (X^\top \beta + 0.3)^2$ is a nonlinear form; $\beta_1 = (\underbrace{1, \cdots, 1}_{p/2}, 0, \cdots, 0)^\top / \sqrt{p/2}$. The simulation results are summarized in Table 4. We can obtain conclusions similar to those for Examples 1 and 2.

In Examples 1–3, the dimension of covariates, 8, is actually large relative to the sample sizes, $n = 100, 200$. For sample size $n = 200$, $8 > 200^{\frac{1}{3}} \approx 5.848$ and is slightly smaller than $200^{\frac{2}{5}} \approx 8.3255$. Furthermore, for sample size $n = 100$, $8 > 100^{\frac{2}{5}} \approx 6.3096$. Therefore, the proposed test is a satisfactory alternative for testing heteroscedasticity, especially when the dimensions of covariates are relatively high.

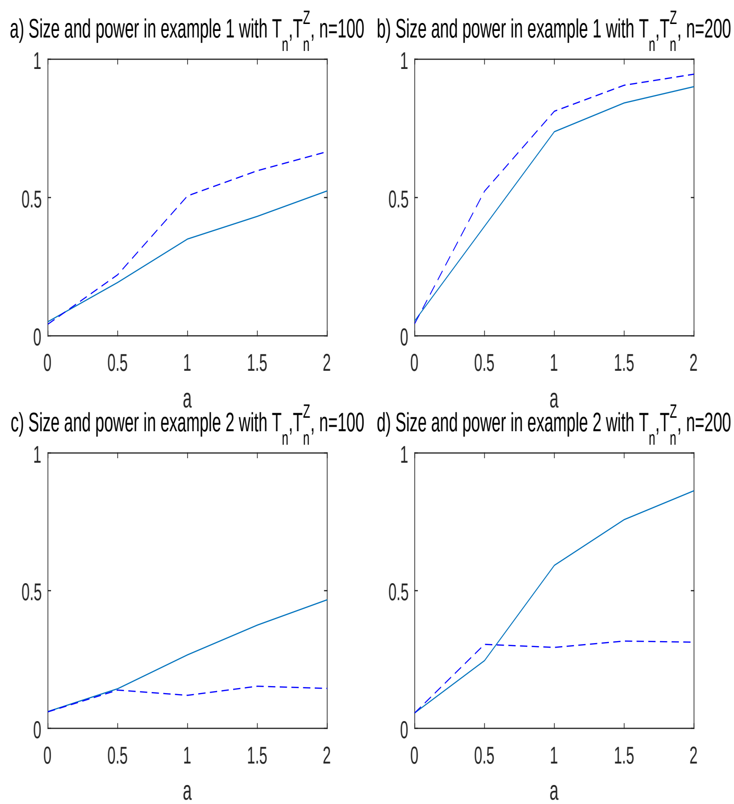

Figure 2 Empirical size and power curves in Examples 1 and 2 with $p = 4, n = 100, 200$. The dashed and solid lines represent the results for T_n^Z and T_n, respectively.

4.2 Real data analysis

A real-data example is analyzed for illustration. We consider the data set comprising ultrasonic calibration. The data is publicly available from

$$https://www.itl.nist.gov/div898/strd/nls/data/chwirut1.shtml.$$

There are 214 observations. The response, Y, represents an ultrasonic response, and the predictor variable, X, represents metal distance. From the upper scatter plot in Figure 3, it is clear that the relationship between the response and predictor is not linear: rather, it is an exponentially decaying pattern in the data. According to modeling suggestions from NIST, the following nonlinear regression model is considered:

$$Y = \frac{\exp(-\beta_1 X)}{\beta_2 + \beta_3 X} + \epsilon.$$

The three parameters in the above model are estimated as $0.1903, 0.0061$, and 0.0105, respectively. From the upper scatter plot for X and the regression residuals in Figure 3, we find that the errors have greater variance for the values of metal distance that are < 1 in comparison to those elsewhere. This implies that the assumption of homogeneous variances seems to be violated. Formally, the p-value of T_n is 0 with 500 bootstrap samples, and thus, the homoscedasticity assumption is strongly rejected.

Parameter inference will become more complex when heteroscedasticity exists within the model. In the following, we try a square root transformation of the response variable to solve this problem. The three parameters in the nonlinear regression model with the square root transformation are now estimated as $-0.0154, 0.0807$, and 0.0639, respectively. The middle plot of the predicted values with the transformed data in Figure 3 indicates a good fit. The middle scatter plot for X and the regression residuals now suggest that the errors satisfy the assumption of homogeneous variances. Formally, the p-value of T_n is 0.542 with 500 bootstrap samples, and thus, the homoscedasticity assumption cannot be rejected.

Next, we observe the three parameters. The standard errors of these regression coefficients are estimated as $0.0086, 0.0015$, and 0.0029. The p-values of the corresponding t-test statistics are 0.0733, 0, and 0. These values imply that the last two parameters are strongly significant, while the first one is significant if we set the nominal level at 0.1. We also try the regression model with β_1 set to zero for the square root transformation. At this time, β_2 and β_3 are estimated as 0.0827 and 0.0590. The standard errors of the two parameters are estimated as 0.0009 and 0.008. Again, the p-values of the corresponding t-test statistics are both zero. The bottom scatter plot in Figure 3 displays that the regression model without β_1 also fits the data very well. Heteroscedasticity is not found from the scatter plot for X and the regression residuals. Further, we employ our test statistic, T_n, to test this assumption and get p-value 0.404, which supports the homoscedasticity assumption.

In sum, for this data set, we find heteroscedasticity when using the nonlinear regression model to fit the original response. This finding then motivates us to consider the square root transformation of the response, which leads to a good fit and eliminates heteroscedasticity. Based on significance testing results, the first parameter seems to lack significance. Then, we try the regression model without β_1. However, the models with and without β_1 generate predicted values that are quite close to each other, since the estimate of β_1, -0.0154, is close to zero in magnitude. In fact, the model with the original response also fits the data very well. The R^2 values from fitting the three models are all about 0.98. However, it is important to note that statistical inference on regression parameters commonly needs the homoscedasticity assumption. Checking whether or not this assumption holds and developing a suitable model that satisfies the homoscedasticity assumption can give us more arguments to justify the conclusions and analyses obtained from the fitted models.

5 Conclusions and discussions

In this study, we consider the heteroscedasticity testing problem in regression models. We propose a pairwise distance-based test statistic with a simple and closed form. The proposed methodology exhibits

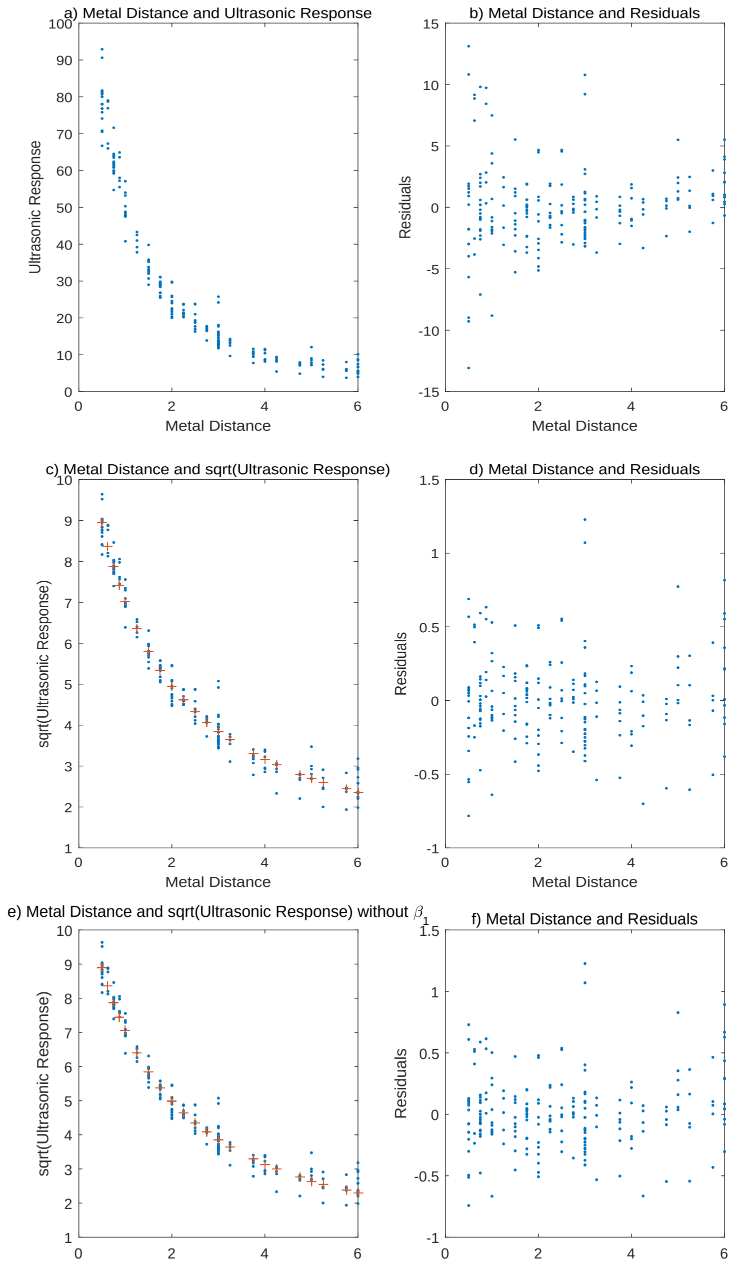

Figure 3 The upper two are the scatter plots for metal distance and ultrasonic response and the corresponding residual; the middle two are the scatter plots for metal distance and the square root of ultrasonic response and the corresponding residual; the lower two are the scatter plots for metal distance and the square root of ultrasonic response without β_1 and the corresponding residual. The red '+' points denote the predicted values.

several advantages over existing methods. For example, it is easy to implement, does not require a bandwidth, and has a faster convergence rate than several existing methods.

The concept introduced in this study is easy to extend to testing the goodness-of-fit of a given parametric variance function, as studied by Wang and Zhou (2007) and Koul and Song (2010). In particular, we can consider the following hypothesis:

$$H_0 : E(\epsilon^2|X) = g(X, \beta, \theta),$$

where $g(X, \cdot, \cdot)$ is a known parametric function and θ is an unknown parameter in $\mathbb{R}^q$. This formulation includes the popular log-linear model and the power-of-the-mean model. The former corresponds to $f(X_i, \beta) = X_i^\top \beta$ and $g(X_i, \beta, \theta) = \exp(X_i^\top \beta)$, whereas the latter corresponds to $g(X_i, \beta, \theta) = \theta_1(f(X_i, \beta))^{\theta_2}$. Let $e = \epsilon^2 - g(X, \beta, \theta)$. The following test statistic,

$$S_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \hat{e}_i \hat{e}_j d_{ij},$$

can be defined, where $\hat{e}_i = (Y_i - f(X_i, \hat{\beta}))^2 - g(X_i, \hat{\beta}, \hat{\theta})$, $(\hat{\beta}, \hat{\theta})$ is any $\sqrt{n}$-consistent estimator of (β, θ), and $d_{ij} = e^{-\|X_i - X_j\|^a}$. Similar to the decomposition of $\hat{\eta}$ shown in the Appendix, we can obtain

$$\hat{e} = e - 2\epsilon(f(X, \hat{\beta}) - f(X, \beta)) + (f(X, \hat{\beta}) - f(X, \beta))^2 - (g(X, \hat{\beta}, \hat{\theta}) - g(X, \beta, \theta)).$$

The asymptotic properties of S_n can be similarly derived. In particular, nS_n converges to a nondegenerate limiting distribution under the null hypothesis. S_n can detect local alternative hypotheses that deviate from the null hypothesis at a rate of $n^{-1/2}$; under the fixed alternative hypotheses, the divergence rate of S_n is $n^{-1/2}$, which is asymptotically normal.

In practice, nonlinear regression models may be incorrectly specified. To avoid this problem, nonparametric and semi-parametric regression models have been developed. Interests in testing heteroscedasticity in nonparametric and semiparametric regression models have been recently observed. Studies on this topic include those of Zhu et al. (2001), Dette (2002), and Dette et al. (2007) for nonparametric regression models, You and Chen (2005), Dette and Marchlewski (2008), and Lin and Qu (2012) for partial linear regression models and Zhu et al. (2015) for single index models. The proposed methodology is generic. In the future, we aim to extend the proposed methodology to more complicated models, such as partial linear regression models and single index models. Finally, extending the approach to time-series nonlinear regression models will also be interesting.

Acknowledgements The authors are grateful to the editor, the associate editor and the anonymous referee for the constructive comments and suggestions that led to significant improvement of an early manuscript. This work is supported by National Natural Science Foundation of China (11701034, 11601227, 11871263, 11671042), the Natural Science Foundation of Guangdong Province of China (2016A030313856), Shenzhen Sci-Tech Fund No. JCYJ20170307110329106, and a grant from the University Grants Council of Hong Kong. All correspondence should be addressed to Xuejun Jiang, Department of Mathematics, Southern University of Science and Technology of China; E-mail: jiangxj@sustc.edu.cn.

References

1 Alba-Fernández, V., Jiménez-Gamero, M.D., Muñoz-García, J. (2008). A test for the two-sample problem based on empirical characteristic functions. *Computational Statistics and Data Analysis*, **52**, 3730-3748.

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

3 Cook, R. D. and Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika*, **70**, 1-10.

4 Dette, H. (2002). A consistent test for heteroscedasticity in nonparametric regression based on the kernel method. *Journal of Statistical Planning and Inference*, **103**, 311-329.

5 Dette, H. and Marchlewski, M. (2008). A test for the parametric form of the variance function in a partial linear regression model. *Journal of Statistical Planning and Inference*, **138(10)**, 3005-3021.

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

7 Fan, Y. N., Lafaye de Micheaux, P., Penev, S. and Salopek, D. (2017). Multivariate nonparametric test of independence. *Journal of Multivariate Analysis*, **153**, 189-210.

8 Gregory, G. G. (1977). Large sample theory for U-statistics and tests of fit. *Annals of Statistics*, **7**, 110-123.

9 Hsiao, C. and Li, Q. (2001). A consistent test for conditional heteroskedasticity in time-series regression models. *Econometric Theory*, **17**, 188-221.

10 Koul, L. H. and Song, W. X. (2010). Conditional variance model checking. *Journal of Statistical Planning and Inference*, **140(4)**, 1056-1072.

11 Leucht, A. and Neumann, M. H. (2009). Consistency of general bootstrap methods for degenerate U-type and V-type statistics. *Journal of Multivariate Analysis*, **100**, 1622-1633.

12 Lin, J. G. and Qu, X. Y. (2012). A consistent test for heteroscedasticity in semiparametric regression with nonparametric variance function based on the kernel method. *Statistics*, **46**, 565-576.

13 Lin, J. G. and Wei, B. C. (2003). Testing for heteroscedasticity in nonlinear regression models. *Communications in Statistics-Theory and Methods*, **32**, 171-192.

14 Lopez, O. and Patilea, V. (2009). Nonparametric lack-of-fit tests for parametric mean-regression models with censored data. *Journal of Multivariate Analysis*, **100**, 210-230.

15 Meintanis, S. G. (2016). A review of testing procedures based on the empirical characteristic function. *South African Statistical Journal*, **50**, 1-14.

16 Nolan, J. P. (2013). Multivariate elliptically contoured stable distributions: theory and estimation. *Computational Statistics*, **28**, 2067-2089.

17 Samarakoon, N. and Song, W. X. (2011). Minimum distance conditional variance function checking in heteroscedastic regression models. *Journal of Multivariate Analysis*, **102(3)**, 579-600.

18 Samarakoon, N. and Song, W. X. (2012). Empirical smoothing lack-of-fit tests for variance function. *Journal of Statistical Planning and Inference* , **42(5)**, 1128-1140.

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

20 Simonoff, J. S. and Tsai, C. L. (1994). Improved tests for nonconstant variance in regression based on the modified profile likelihood. *Journal of Applied Statistics*, **43**, 357-370.

21 Su, L. and Ullah, A. (2013). A nonparametric goodness-of-fit-based test for conditional heteroskedasticity. *Econometric Theory*, **29**, 187-212.

22 Sun, Z. H. and Wang, Q. H. (2009). Checking the adequacy of a general linear model with responses missing at random. *Journal of Statistical Planning and Inference*, **139**, 3588-3604.

23 Tsai, C. L. (1986). Score test for the first-order autoregressive model with heteroscedasticity. *Biometrika*, **73**, 455-460.

24 Wang, L. and Zhou, X. H. (2007). Assessing the adequacy of variance function in heteroscedastic regression models. *Biometrics*, **63(4)**, 1218-1225.

25 You, J. H. and Chen, G. M. (2005). Testing heteroscedasticity in partially linear regression models. *Statistics and Probability Letters*, **73**, 61-70.

26 Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*, **75**, 263-289.

27 Zheng, J. X. (2009). Testing heteroscedasticity in nonlinear and nonparametric regressions. *Canadian Journal of Statistics*, **37**, 282-300.

28 Zhu, L. X., Fujikoshi, Y. and Naito, K. (2001). Heteroscedasticity test for regression models. *Science in China: Series A*, **44**, 1237-1252.

29 Zhu, X. H., Guo, X., Lin, L. and Zhu, L. X. (2015). Heteroscedasticity checks for single index models. *Journal of Multivariate Analysis*, **136**, 41-55.

Appendix A

Denote $\hat{f} = f(X, \hat{\beta}), f = f(X, \beta), \epsilon = Y - f, \eta = (Y - f)^2 - \sigma^2$ with $\sigma^2 = E(\epsilon^2)$.

Proof of Theorem 2.1. For $\hat{\eta}$, we have the following decomposition:

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

Then T_n can be decomposed into 10 parts.

$$T_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} d_{ij} \eta_i \eta_j + 4 \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} d_{ij} \epsilon_i \epsilon_j (\hat{f}_i - f_i)(\hat{f}_j - f_j)$$

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

Given that $\hat{\beta}-\beta=O_p(1/\sqrt{n})$ and $E(\epsilon|X)=E(\eta|X)=0$, we can easily obtain that: $Q_{2n}=O_p(n^{-2}), Q_{3n}=O_p(n^{-2}), Q_{5n}=O_p(n^{-3/2}), Q_{6n}=O_p(n^{-3/2}), Q_{8n}=O_p(n^{-2}), Q_{9n}=O_p(n^{-3/2}), Q_{10n}=O_p(n^{-3/2})$.

Here, we only present a sketch proof for the order of Q_{2n}. Others can be similarly obtained, and thus the details are omitted in this paper. Notably

$$
\hat{f}_i-f_i=f(X_i,\hat{\beta})-f(X_i,\beta)=(\hat{\beta}-\beta)^{\top}\nabla f(X_i,\beta)+\frac{1}{2}(\hat{\beta}-\beta)^{\top}\nabla f^2(X_i,\tilde{\beta})(\hat{\beta}-\beta).
$$

Here $\tilde{\beta}$ is a value between β and $\hat{\beta}$. Let

$$
\begin{aligned}
Q_{2n,1} &= 4(\hat{\beta}-\beta)^{\top}\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}d_{ij}\epsilon_i\epsilon_j\nabla f(X_i,\beta)\nabla f^{\top}(X_j,\beta)(\hat{\beta}-\beta)\\
&=: 4(\hat{\beta}-\beta)^{\top}\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}L(R_i,R_j)(\hat{\beta}-\beta).
\end{aligned}
$$

Here $R_i=(X_i,\epsilon_i)$. We obtain $Q_{2n}=Q_{2n,1}+o_p(Q_{2n,1})$. Furthermore, $L(R_i,R_j)=L(R_j,R_i)$. Thus $n(n-1)^{-1}\sum_{i=1}^{n}\sum_{j\neq i}L(R_i,R_j)$ is a U-statistic. In addition, we have $E(L(R_i,R_j)|R_i)=0$. Notably, $E(\epsilon|X)=0$ and also

$$
\begin{aligned}
E(L(R_i,R_j)|R_i) &= E\left[d_{ij}\epsilon_i\epsilon_j\nabla f(X_i,\beta)\nabla f^{\top}(X_j,\beta)|X_i,\epsilon_i\right]\\
&= E\left[E(d_{ij}\epsilon_i\epsilon_j\nabla f(X_i,\beta)\nabla f^{\top}(X_j,\beta)|X_i,X_j,\epsilon_i)|X_i,\epsilon_i\right]\\
&= E\left[d_{ij}\epsilon_i\nabla f(X_i,\beta)\nabla f^{\top}(X_j,\beta)E(\epsilon_j|X_i,X_j,\epsilon_i)|X_i,\epsilon_i\right]=0.
\end{aligned}
$$

This finding implies that $n(n-1)^{-1}\sum_{i=1}^{n}\sum_{j\neq i}L(R_i,R_j)$ is a degenerate U-statistic of order 1, and thus, $n(n-1)^{-1}\sum_{i=1}^{n}\sum_{j\neq i}L(R_i,R_j)=O_p(n^{-1})$. Since $\hat{\beta}-\beta=O_p(1/\sqrt{n})$, we get $Q_{2n,1}=O_p(n^{-2})$, and thus, $Q_{2n}=O_p(n^{-2})$.

Consequently, we can obtain:

$$
\begin{aligned}
T_n &= \frac{1}{n(n-1)}\sum_{i=1}\sum_{j\neq i}d_{ij}[\eta_i\eta_j+(\hat{\sigma}^2-\sigma^2)^2-2\eta_i(\hat{\sigma}^2-\sigma^2)]+o_p(\frac{1}{n})\\
&= Q_{1n}+Q_{4n}-Q_{7n}+o_p(\frac{1}{n}). \qquad (A.1)
\end{aligned}
$$

Notably,

$$
\begin{aligned}
\hat{\sigma}^2-\sigma^2 &= \frac{1}{n}\sum_{i=1}^{n}[(Y_i-\hat{f}_i)^2-\sigma^2]=\frac{1}{n}\sum_{i=1}^{n}[\eta_i-2\epsilon_i(\hat{f}_i-f_i)+(\hat{f}_i-f_i)^2]\\
&= \frac{1}{n}\sum_{i=1}^{n}\eta_i+O_p(\frac{1}{n}),
\end{aligned}
$$

which implies that:

$$
\begin{aligned}
(\hat{\sigma}^2-\sigma^2)^2 &= \left(\frac{1}{n}\sum_{i=1}^{n}\eta_i\right)^2 + o_p(\frac{1}{n}) \\
&= \frac{n-1}{n}\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\eta_i\eta_j + \frac{1}{n^2}\sum_{i=1}^{n}\eta_i^2 + o_p(\frac{1}{n}).
\end{aligned}
$$

Consequently, we have:

$$
Q_{4n} = E(d_{12})\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\eta_i\eta_j + \frac{1}{n}E(d_{12})E(\eta^2) + o_p(\frac{1}{n}). \tag{A.2}
$$

Then, we consider the term Q_{7n}. Similarly we have:

$$
\begin{aligned}
Q_{7n} &= \frac{1}{n^2(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{l=1}^{n} d_{ij}(\eta_i+\eta_j)\eta_l + o_p(\frac{1}{n}) \\
&= \frac{1}{n^2(n-1)}\sum_{i\neq j\neq l}^{n} d_{ij}(\eta_i+\eta_j)\eta_l + \frac{1}{n^2(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} d_{ij}(\eta_i+\eta_j)^2 + o_p(\frac{1}{n}) \\
&=: Q_{7n,1} + Q_{7n,2} + o_p(\frac{1}{n}).
\end{aligned} \tag{A.3}
$$

Given that $E(d_{12}\eta_1\eta_2) = 0$, the following expression can be easily shown:

$$
Q_{7n,2} = \frac{E[d_{12}(\eta_1+\eta_2)^2]}{n} + o_p(\frac{1}{n}) = \frac{2E[d_{12}\eta_1^2]}{n} + o_p(\frac{1}{n}).
$$

We can write $Q_{7n,1} = \frac{n-2}{n}U_n$, where

$$
U_n = \binom{n}{3}^{-1}\sum\sum\sum_{1\leqslant i<j<l\leqslant n} H^s(W_i, W_j, W_l).
$$

U_n is a third-order U-statistic. Here, $W_i = (X_i, \eta_i)$, $H^s(W_i, W_j, W_l) = (H_{ijl} + H_{ilj} + H_{jli})/3$ is the kernel with $H_{ijl} = d_{ij}(\eta_i+\eta_j)\eta_l$.

By construction the kernel $H^s(W_i, W_j, W_l)$ is symmetric in its three arguments, and $E(H^s(W_i, W_j, W_l)|W_i) = 0$. Notably,

$$
\begin{aligned}
E(H_{ijl}|W_i) &= E(d_{ij}(\eta_i+\eta_j)\eta_l|W_i) = E[E(d_{ij}(\eta_i+\eta_j)\eta_l|W_i, W_j)|W_i] \\
&= E[d_{ij}(\eta_i+\eta_j)E(\eta_l|W_i, W_j)|W_i] = 0.
\end{aligned}
$$

Similarly, we can obtain $E(H_{ilj}|W_i) = 0$. Moreover, we observe that

$$
\begin{aligned}
E(H_{jli}|W_i) &= E(d_{jl}(\eta_j+\eta_l)\eta_i|W_i) = \eta_i E(d_{jl}(\eta_j+\eta_l)) \\
&= \eta_i E[E(d_{jl}(\eta_j+\eta_l)|X_j, X_l)] = 0.
\end{aligned}
$$

In summary, we conclude that $E(H^s(W_i, W_j, W_l)|W_i) = 0$.

However, $E(H^s(W_i, W_j, W_l)|W_i, W_j) \neq 0$. To be precise, although $E(H_{ijl}|W_i, W_j) = E(d_{ij}(\eta_i + \eta_j)\eta_l|W_i, W_j) = 0$, we obtain

$$
\begin{aligned}
E(H_{ilj}|W_i, W_j) &= E(d_{il}(\eta_i+\eta_l)\eta_j|W_i, W_j) = E[E(d_{il}(\eta_i+\eta_l)\eta_j|W_i, W_j, X_l)|W_i, W_j] \\
&= E[d_{il}\eta_i\eta_j|W_i, W_j] = \eta_i\eta_j E(d_{il}|X_i, X_j).
\end{aligned}
$$

Similarly, we derive $E(H_{jli}|W_i, W_j) = \eta_i\eta_j E(d_{jl}|X_i, X_j)$. In summary, we have

$$
E(H^s(W_i, W_j, W_l)|W_i, W_j) = \frac{1}{3}\eta_i\eta_j E(d_{il} + d_{jl}|X_i, X_j).
$$

From Serfling (1980, section 5.3.4), we have:

$$\begin{aligned} U_n &= \frac{3\cdot 2}{n(n-1)}\sum\sum_{1\leqslant i<j\leqslant n}\frac{1}{3}\eta_i\eta_j E(d_{il}+d_{jl}|X_i,X_j)+o_p(\frac{1}{n}) \\ &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\eta_i\eta_j E(d_{il}+d_{jl}|X_i,X_j)+o_p(\frac{1}{n}), \end{aligned}$$

which implies that

$$Q_{7n} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\eta_i\eta_j E(d_{il}+d_{jl}|X_i,X_j)+\frac{2E(d_{12}\eta_1^2)}{n}+o_p(\frac{1}{n}). \tag{A.4}$$

From Equations, (A.1), (A.2) and (A.4), we conclude that

$$\begin{aligned} T_n &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\eta_i\eta_j\left[d_{ij}-E(d_{il}+d_{jl}|X_i,X_j)+E(d_{12})\right] \\ &\quad +\frac{E(d_{12})E(\eta^2)-2E(d_{12}\eta_1^2)}{n}+o_p(\frac{1}{n}) \\ &=: \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}h(W_i,W_j)+\frac{\mu}{n}+o_p(\frac{1}{n}). \end{aligned} \tag{A.5}$$

Here $\tilde{d}_{ij}=d_{ij}-E(d_{il}+d_{jl}|X_i,X_j)+E(d_{12}), h(W_i,W_j)=\eta_i\eta_j\tilde{d}_{ij}$ and $\mu=E(d_{12})E(\eta^2)-2E(d_{12}\eta_1^2)$.

Notably, $E(\eta|X)=0$ under H_0, and thus,

$$\begin{aligned} E(h(W_i,W_j)|W_i) &= E[E(\eta_i\eta_j\tilde{d}_{ij}|W_i,X_j)|W_i] \\ &= E[\eta_i\tilde{d}_{ij}E(\eta_j|W_i,X_j)|W_i]=0. \end{aligned}$$

Recall that $d_{ij}=e^{-\|X_i-X_j\|^a}\leqslant 1$. Hence, $|\tilde{d}_{ij}|\leqslant 4$ and under condition C4,

$$E(h^2(W_1,W_2)) = E(\eta_1^2\eta_2^2\tilde{d}_{ij}^2)\leqslant 16E(\eta_1^2\eta_2^2)=16E^2(\eta^2)<\infty.$$

Accordingly, we can obtain $n\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}h(W_i,W_j)\Rightarrow\sum_{i=1}^{\infty}\lambda_i(Z_i^2-1)$ based on the standard theory of U-statistics, see e.g. Serfling (1980, section 5.5). Here, Z_k's are independent standard normal random variables, and the constants λ_k's are the eigenvalues of the integral equation

$$\int h(W_i,W_j)\tilde{\phi}_k(W_j)\,dF(W_j)=\lambda_k\tilde{\phi}_k(W_i)$$

with F being the pdf of W.

We derive

$$\begin{aligned} \lambda_k\tilde{\phi}_k(W_i) &= \int h(W_i,W_j)\tilde{\phi}_k(W_j)\,dF(W_j)=\int\eta_i\eta_j\tilde{d}_{ij}\tilde{\phi}_k(W_j)\,dF(W_j) \\ &= \eta_i\int\eta_j\tilde{d}_{ij}\tilde{\phi}_k(W_j)\,dF(W_j). \end{aligned}$$

Then we can write $\tilde{\phi}_k(W_i)=\eta_i\phi_k(W_i)$ by appropriately choosing $\phi_k(W_i)$. Similarly, $\tilde{\phi}_k(W_j)=\eta_j\phi_k(W_j)$. The integration equation can be rewritten as

$$\int\eta_j^2\tilde{d}_{ij}\phi_k(W_j)\,dF(W_j)=\lambda_k\phi_k(W_i).$$

We immediately obtain $nT_n\Rightarrow\sum_{i=1}^{\infty}\lambda_i(Z_i^2-1)+\mu$. Furthermore

$$\sum_{i=1}^{\infty}\lambda_i=E[h(W_1,W_1)]=E[\eta_1^2\big(1-2E(d_{13}|X_1)+E(d_{12})\big)]$$

$$= E(\eta^2) + E(d_{12})E(\eta^2) - 2E(\eta^2 E(d_{12}|X_1))$$
$$= E(\eta^2) + E(d_{12})E(\eta^2) - 2E(\eta^2 d_{12}).$$

Consequently, we derive that $nT_n \Rightarrow \sum_{i=1}^{\infty} \lambda_i Z_i^2 - E(\eta^2)$.

Proof of Theorem 2.2. Under the local alternative hypothesis, T_n is still able to be decomposed into 10 parts.

$$\begin{aligned} T_n = & \frac{1}{n(n-1)} \sum_{i=1} \sum_{j\neq i} d_{ij}[\eta_i\eta_j + 4\epsilon_i\epsilon_j(\hat{f}_i - f_i)(\hat{f}_j - f_j) + (\hat{f}_i - f_i)^2(\hat{f}_j - f_j)^2 \\ & +(\hat{\sigma}^2 - \sigma^2)^2 - 4\eta_i\epsilon_j(\hat{f}_j - f_j) + 2\eta_i(\hat{f}_j - f_j)^2 - 2\eta_i(\hat{\sigma}^2 - \sigma^2) \\ & -4\epsilon_i(\hat{f}_i - f_i)(\hat{f}_j - f_j)^2 + 4\epsilon_i(\hat{f}_i - f_i)(\hat{\sigma}^2 - \sigma^2) - 2(\hat{f}_i - f_i)^2(\hat{\sigma}^2 - \sigma^2)] \equiv: \sum_{i=1}^{10} Q_{in}. \end{aligned}$$

Under the alternative hypothesis, $\epsilon_i, \hat{f}_i$ and f_i do not change. Thus we can also obtain that $Q_{kn} = O_p(n^{-2}), k = 2,3,8$. Notably, η_i can be rewritten as $\eta_i = u_i + c_n V(X_i)$, here $E(u_i|X_i) = 0$ and $E(V(X)) = 0$. Furthermore, recall that

$$\hat{\sigma}^2 - \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} \eta_i + O_p(\frac{1}{n}) = O_p(\frac{1}{\sqrt{n}}),$$

which results in $Q_{kn} = O_p(c_n n^{-1} + n^{-3/2}), k = 5,6$ and, $Q_{kn} = O_p(n^{-3/2}), k = 9,10$. Accordingly, $nQ_{kn} = o_p(1), k = 2,3,5,6,8,9,10$. Similar to the arguments in the proof for Theorem 2.1, we have:

$$T_n = \frac{1}{n(n-1)} \sum_{i=1}^{n}\sum_{j\neq i}^{n} \eta_i\eta_j\tilde{d}_{ij} + \frac{E(d_{12})E(\eta^2) - 2E(d_{12}\eta_1^2)}{n} + o_p(\frac{1}{n}). \tag{A.6}$$

Here $\tilde{d}_{ij} = d_{ij} - E(d_{il} + d_{jl}|X_i, X_j) + E(d_{12})$. Thus if $c_n = n^{-1/2}$, given that $E(\eta|X) = c_n V(X)$, from Theorem 2.1 in Gregory (1977), then we have:

$$n \times \frac{1}{n(n-1)} \sum_{i=1}^{n}\sum_{j\neq i}^{n} \eta_i\eta_j\tilde{d}_{ij} \Rightarrow \sum_{i=1}^{\infty} \lambda_i[(Z_i + a_i)^2 - 1].$$

Here $a_i = E(V(X)\phi_i(W))$. Similarly, we can obtain that

$$E(d_{12})E(\eta^2) - 2E(d_{12}\eta_1^2) - \sum_{i=1}^{\infty}\lambda_i = E(\eta^2) = E(u^2) + o_p(1).$$

Here $u = \eta - c_n V(X)$. In summary, if $c_n = n^{-1/2}$, then we have:

$$nT_n \Rightarrow \sum_{i=1}^{\infty} \lambda_i(Z_i + a_i)^2 - E(u^2).$$

Furthermore we also derive:

$$\begin{aligned} T_n &= \frac{1}{n(n-1)} \sum_{i=1}^{n}\sum_{j\neq i}^{n} \eta_i\eta_j\tilde{d}_{ij} + \frac{E(d_{12})E(\eta^2) - 2E(d_{12}\eta_1^2)}{n} + o_p(\frac{1}{n}) \\ &= \frac{1}{n(n-1)} \sum_{i=1}^{n}\sum_{j\neq i}^{n} \tilde{d}_{ij}[u_iu_j + 2c_n u_i V_j + c_n^2 V_i V_j] + \frac{E(d_{12})E(u^2) - 2E(d_{12}u_1^2)}{n} + o_p(\frac{1}{n}) \\ &= : T_{n1} + c_n T_{n2} + c_n^2 T_{n3} + \frac{\mu^*}{n} + o_p(\frac{1}{n}). \end{aligned}$$

By using the standard U-statistic theory, we determine that $T_{n1} = O_p(n^{-1}), T_{n2} = O_p(n^{-1/2})$ and $T_{n3} = E[\tilde{d}_{12}V_1V_2] + o_p(1)$. Thus if $c_n = n^{-r}, 0 < r < 1/2$, then we can easily obtain $nT_n \Rightarrow \infty$. □

Proof of Theorem 2.3. From the Proof of Theorem 2.2, under the fixed alternative hypothesis, we have $Q_{kn} = O_p(n^{-2}), k = 2, 3, 8$ and $Q_{kn} = O_p(n^{-1}), k = 5, 6$ for the decomposition of T_n. Since $\hat{\sigma}^2 - \sigma^2 = O_p(n^{-1/2})$, we obtain that $Q_{kn} = O_p(n^{-3/2}), k = 9, 10$ and $Q_{4n} = O_p(n^{-1})$. Then we have

$$T_n = \frac{1}{n(n-1)} \sum_{i=1} \sum_{j \neq i} d_{ij}[\eta_i \eta_j - 2\eta_i(\hat{\sigma}^2 - \sigma^2)] + o_p(\frac{1}{\sqrt{n}}).$$

For the term Q_{1n}, we have

$$\sqrt{n}(Q_{1n} - E(\eta_1 \eta_2 d_{12})) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\eta_i E(\eta_j d_{ij}|X_i) - E(\eta_1 \eta_2 d_{12})] + o_p(\frac{1}{\sqrt{n}}).$$

For the term Q_{7n}, we derive

$$\begin{aligned}\sqrt{n} Q_{7n} &= \frac{1}{n(n-1)} \sum_{i=1} \sum_{j \neq i} [(\eta_i + \eta_j) d_{ij}] \times \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \eta_i + o_p(1) \\ &= E((\eta_1 + \eta_2) d_{12}) \times \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \eta_i + o_p(1).\end{aligned}$$

Let $c_1 = E(\eta_1 \eta_2 d_{12})$ and $c_2 = E((\eta_1 + \eta_2) d_{12})$. Then, we can obtain

$$\sqrt{n}(T_n - c_1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\eta_i E(\eta_j d_{ij}|X_i) - c_1 - c_2 \eta_i] + o_p(\frac{1}{\sqrt{n}}) \Rightarrow N(0, \tilde{\sigma}^2).$$

Here, $\tilde{\sigma}^2 = var\big(\eta_i E(\eta_j d_{ij}|X_i) - c_2 \eta_i\big)$. □

Proof of Theorem 3.1. The proof of Theorem 3.1 follows steps similar to those in the proof of Theorem 2.1. Therefore, we only sketch the proof in this paper. Recall that $\mathcal{F}_n = \{X_i, Y_i\}_{i=1}^n$. Furthermore, denote $E^*(\cdot) = E(\cdot|\mathcal{F}_n)$. Note that $\hat{\epsilon}_i^* = \epsilon_i^* - [f(X_i, \hat{\beta}^*) - f(X_i, \hat{\beta})]$. Define $\sigma^{*2} = E^*(\epsilon^{*2}) = n^{-1} \sum_{i=1}^n (\hat{\epsilon}_i - \bar{\epsilon})^2$ and $\eta_i^* = \epsilon_i^{*2} - \sigma^{*2}$. Consequently, we have:

$$\begin{aligned}\hat{\eta}_i^* &= \hat{\epsilon}_i^{*2} - \hat{\sigma}^{*2} = \Big(\epsilon_i^* - [f(X_i, \hat{\beta}^*) - f(X_i, \hat{\beta})]\Big)^2 - \hat{\sigma}^{*2} \\ &= \eta_i^* - 2\epsilon_i^*[f(X_i, \hat{\beta}^*) - f(X_i, \hat{\beta})] + [f(X_i, \hat{\beta}^*) - f(X_i, \hat{\beta})]^2 - (\hat{\sigma}^{*2} - \sigma^{*2}).\end{aligned}$$

Recall that ϵ_i^* and ϵ_j^* with $i \neq j$ are independent with each other conditional on the random sample, $\mathcal{F}_n = \{X_i, Y_i\}_{i=1}^n$. Thus, we have $E^*(\eta_i^*|\eta_j^*, X_i, X_j) = 0$ for $i \neq j$. Moreover, $\hat{\beta}^* - \hat{\beta} = O_p(n^{-1/2})$. In addition, note that

$$\begin{aligned}&\hat{\sigma}^{*2} - \sigma^{*2} = \frac{1}{n} \sum_{i=1}^{n} (\hat{\epsilon}_i^{*2} - \sigma^{*2}) \\ &= \frac{1}{n} \sum_{i=1}^{n} [\eta_i^* - 2\epsilon_i^*[f(X_i, \hat{\beta}^*) - f(X_i, \hat{\beta})] + [f(X_i, \hat{\beta}^*) - f(X_i, \hat{\beta})]^2] \\ &= \frac{1}{n} \sum_{i=1}^{n} \eta_i^* + o_p(\frac{1}{\sqrt{n}}).\end{aligned}$$

Consequently, we obtain that

$$T_n^*|\mathcal{F}_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \eta_i^* \eta_j^* \tilde{d}_{ij} + \frac{E(d_{12}) E^*(\eta^{*2}) - 2E^*(d_{12} \eta_1^{*2})}{n} + o_p(\frac{1}{n}).$$

Subsequently, we define $\eta_i = \epsilon_i^2 - E(\epsilon_i^2|X_i)$. As shown in the proof of theorem 2.1, we have $E[\eta_i^2 \eta_j^2 \tilde{d}_{ij}^2] \leqslant 16E^2(\eta^2) < \infty$ and $E[|\eta_i^2 \tilde{d}_{ii}|] \leqslant 4E(\eta^2) < \infty$ under our assumed condition $E(\eta^2) < \infty$, which implies that the conditions A1 and A3 in Leucht and Neumann (2009) are satisfied. Thus, according to Lemma 2.1 and Lemma 2.2 in Leucht and Neumann (2009), to prove the conditional asymptotic distribution of T_n^* given $\mathcal{F}_n$ is the same as the asymptotic distribution of T_n, we need to prove the following three issues. First, the common distribution F_η^* of η^* will converge to the common distribution F_η of η. Second, $E^*[\eta_i^{*2}\eta_j^{*2}\tilde{d}_{ij}^2] \Rightarrow E[\eta_i^2\eta_j^2\tilde{d}_{ij}^2]$, and third, $E^*[\eta_i^{*2}\tilde{d}_{ii}] \Rightarrow E[\eta_i^2\tilde{d}_{ii}]$.

Notably,

$$F_\eta^*(r) = \frac{1}{n}\sum_{i=1}^n I((\hat{\epsilon}_i - \bar{\epsilon})^2 - \sigma^{*2} \leqslant r).$$

Denote $\Delta f_i = f(X_i, \hat{\beta}) - f(X_i, \beta)$ and recall that $\hat{\epsilon}_i = \epsilon_i - \Delta f_i$. Then, we can obtain

$$\begin{aligned} F_\eta^*(r) &= \frac{1}{n}\sum_{i=1}^n I(\epsilon_i^2 - 2\Delta f_i\epsilon_i + \Delta f_i^2 - 2\hat{\epsilon}_i\bar{\epsilon} + \bar{\epsilon}^2 - \sigma^{*2} \leqslant r) \\ &= \frac{1}{n}\sum_{i=1}^n I(\eta_i \leqslant r + 2\Delta f_i\epsilon_i - \Delta f_i^2 + 2\hat{\epsilon}_i\bar{\epsilon} - \bar{\epsilon}^2 + \sigma^{*2} - E(\epsilon^2|X_i)). \end{aligned}$$

Denote $\Delta\eta_i = 2\Delta f_i\epsilon_i - \Delta f_i^2 + 2\hat{\epsilon}_i\bar{\epsilon} - \bar{\epsilon}^2 + \sigma^{*2} - E(\epsilon^2|X_i)$. Under the null hypothesis and local alternative hypothesis with $c_n \to 0$, we have either $E(\epsilon^2|X_i) = E(\epsilon^2) = \sigma^2$ or $E(\epsilon^2|X_i) - E(\epsilon^2) = O_p(c_n)$. In summary, under H_0 and H_{1n}, we always have $\sigma^{*2} - E(\epsilon^2|X_i) = \sigma^{*2} - \sigma^2 + \sigma^2 - E(\epsilon^2|X_i) = O_p(n^{-1/2})$ or $O_p(c_n)$. Given that $\hat{\beta} - \beta = O_p(n^{-1/2}), \bar{\epsilon} = O_p(n^{-1/2})$ and $\sigma^{*2} - E(\epsilon^2|X_i) = O_p(n^{-1/2})$ or $O_p(c_n)$, we obtain $\Delta\eta_i = O_p(n^{-1/2})$ or $O_p(c_n)$. Consequently, we obtain that

$$\begin{aligned} F_\eta^*(r) - \frac{1}{n}\sum_{i=1}^n I(\eta \leqslant r) &= \frac{1}{n}\sum_{i=1}^n I(\eta_i \leqslant r + \Delta\eta_i) - \frac{1}{n}\sum_{i=1}^n I(\eta_i \leqslant r) \\ &\leqslant \frac{1}{n}\sum_{i=1}^n I(|\eta_i - r| \leqslant |\Delta\eta_i|) = o_p(1), \end{aligned}$$

given that $n^{-1}\sum_{i=1}^n I(\eta \leqslant r) \Rightarrow F_\eta(r)$, we conclude that $F_\eta^*(r) \Rightarrow F_\eta(r)$.

Next, we show that $E^*[\eta_i^{*2}\eta_j^{*2}\tilde{d}_{ij}^2] \Rightarrow E[\eta_i^2\eta_j^2\tilde{d}_{ij}^2]$ hold. Notably,

$$\begin{aligned} E^*[\eta_i^{*2}\eta_j^{*2}\tilde{d}_{ij}^2] &= \frac{1}{n}\sum_{i=1}^n [(\hat{\epsilon}_i - \bar{\epsilon})^2 - \sigma^{*2}]^2[(\hat{\epsilon}_j - \bar{\epsilon})^2 - \sigma^{*2}]^2\tilde{d}_{ij}^2 \\ &= \frac{1}{n}\sum_{i=1}^n [\epsilon_i^2 - \sigma^2]^2[\epsilon_j^2 - \sigma^2]^2\tilde{d}_{ij}^2 + o_p(1) = E[\eta_i^2\eta_j^2\tilde{d}_{ij}^2] + o_p(1). \end{aligned}$$

$E^*[\eta_i^{*2}\tilde{d}_{ii}^2] \Rightarrow E[\eta_i^2\tilde{d}_{ii}^2]$ can be similarly proven. In summary, we prove that the conditional asymptotic distribution of T_n^* given $\mathcal{F}_n$ is the same as the asymptotic distribution of T_n under H_0 and H_{1n}.

Under fixed alternative hypothesis H_1, we still have $E^*(\eta^*|X_i) = 0$. Consequently, $nT_n^*|\mathcal{F}_n$ still converges to a finite limit, which may differ from the limiting distribution of T_n under the null hypothesis. However, $nT_n \Rightarrow \infty$ under H_1, as shown in Theorem 2.3. In other words, the bootstrap algorithm is valid.