Journal of Statistical Planning and Inference 142 (2012) 1128–1140

Contents lists available at SciVerse ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Empirical smoothing lack-of-fit tests for variance function

Nithansha Samarakoon, Weixing Song *

Kansas State University, United States

A R T I C L E I N F O

Article history:
Received 23 June 2011
Received in revised form
26 October 2011
Accepted 28 November 2011
Available online 8 December 2011

Keywords:
Lack-of-fit test
Empirical L_2-distance
Consistency and local power

A B S T R A C T

This paper discusses a nonparametric empirical smoothing lack-of-fit test for the functional form of the variance in regression models. The proposed test can be treated as a nontrivial modification of Zheng's nonparametric smoothing test, Koul and Ni's minimum distance test for the mean function in the classic regression models. The paper establishes the asymptotic normality of the proposed test under the null hypothesis. Consistency at some fixed alternatives and asymptotic power under some local alternatives are also discussed. A simulation study is conducted to assess the finite sample performance of the proposed test. Simulation study also shows that the proposed test is more powerful and computationally more efficient than some existing tests.

Published by Elsevier B.V.

1. Introduction

It is a common assumption of a regression model that the random error terms are mutually independent with mean zero and equal variances. Occasionally, this homoscedasticity assumption on the variance is not satisfied, and the real data generated from the applications often exhibit certain heteroscedasticity structures. Heteroscedasticity may be caused by many things such as data pooling, different levels of determination, different measurements of error, important variables that may be omitted from the model, as well as many other factors.

A remarkable amount of statistical research that has already been carried out for assessing the heteroscedasticity in both parametric and nonparametric regression models. Early works in this area include some graphical procedures and some formal tests, most of which are based on the residuals obtained by fitting a model with a completely specified regression and variance function. See Harrison and Mccabe (1979), Breusch and Pagan (1979), White (1980), Koenker and Bassett (1981), Diblasi and Bowman (1997) and the references therein. When the covariate is one dimensional and fixed, Dette and Munk (1998), Dette (2002), Dette and Hetzler (2009a) propose some lack-of-it tests for heteroscedasticity in nonparametric regression setups using pseudo-residuals. Later, Dette and Hetzler (2009b) construct a Khamaladze transformation for the test proposed in Dette and Hetzler (2009a) and show that this new test is asymptotically distribution free. Dette and Munk (1998)'s methodology is extended by You and Chen (2005) to develop a heteroscedasticity test in the partially linear regression models. When the covariate is one dimensional and random, Liero (2003), Dette and Marchlewski (2010) propose some testing procedures to check the homoscedasticity. In multi-dimensional covariate case, a Cramèr–von Mises type test based on cumulative estimated residuals is proposed by Zhu et al. (2001). This test is able to detect the local alternatives converging to the null at the parametric rate of $1/\sqrt{n}$, regardless of the type of regression function and the variance function. The asymptotic distributions of the above test statistics are usually

* Corresponding author.
E-mail address: weixing@ksu.edu (W. Song).

0378-3758/$ - see front matter Published by Elsevier B.V.
doi:10.1016/j.jspi.2011.11.022

complicated and are not asymptotically distribution free. Some resampling methods are used to find the critical values or p-values.

Compared to the research of testing heteroscedasticity, fewer meticulous procedures for testing the adequacy of a given variance function are proposed in the literature. When the covariate is one dimensional, Dette et al. (2007) studied the problem of testing the parametric form of the conditional variance in nonparametric regression models. They proposed a Kolmogorov–Smirnov and a Cramèr–von Mises type tests, which are constructed from some stochastic processes based on the difference between the empirical processes that are obtained from the standardized nonparametric residuals under the null and alternative hypotheses. In the multi-dimensional covariate case, Wang and Zhou (2006) presented a kernel smoothing based nonparametric test for checking the adequacy of a parametric variance function. Their test can detect $1/\sqrt{nh^{d/2}}$ local alternatives, where n is the sample size, d is the dimension of the covariates, and h is the bandwidth in constructing the test statistic. Samarakoon and Song (2011) considered the same question and proposed a minimum distance test. Wang and Zhou (2006) provided a bootstrap algorithm to implement their test, while Samarakoon and Song (2011)'s test requires the calculation of the integrations in the test statistic.

This paper is organized as follows. The test statistic and the technical assumptions are stated in Section 2; the asymptotic null distribution, the consistency and local power study of the test are presented in Section 3; Section 4 contains simulation studies to show the finite sample performance of the test and a comparison of the proposed method with the minimum distance test proposed by Samarakoon and Song (2011). All the proofs of main results are presented in Section 5.

2. Test statistic and assumptions

In this section, we propose a new lack-of-fit test to check the adequacy of a parametric form of the variance function in the heteroscedastic regression models. To be specific, consider the following regression model:

$$Y = m(X;\beta)+\sqrt{v(X)}\varepsilon, \tag{2.1}$$

where Y is a one dimensional response variable, X is a d-dimensional explanatory variable, $m(x;\beta)$ is the mean function of known form characterized by the unknown p-dimensional parameter β, and $v(x)$ is the conditional variance function of Y given $X=x$. The hypothesis to be tested is

$$H_0 : v(x) = v(x;\beta_0,\theta_0) \text{ for some } (\beta_0,\theta_0) \in \Gamma \times \Theta \quad \text{v.s.} \quad H_a : H_0 \text{ is not true}, \tag{2.2}$$

where Γ and Θ are the parameter spaces of β and θ, respectively. The actual meaning of the alternative hypothesis H_a above should be understood as $P(v(X) = v(X;\beta,\theta)) < 1$ for all β and θ. Assuming that the error term ε satisfies $E(\varepsilon|X) = 0$ and $E(\varepsilon^2|X) = 1$, we have

$$E[(Y-m(X;\beta))^2|X=x] = v(x),$$

which implies that testing the variance function in model (2.1) is amount to testing the mean function in the following regression model:

$$(Y-m(X;\beta))^2 = v(X)+\xi$$

if β is known, where $(Y-m(X;\beta))^2$ is viewed as the response variable, and $\xi = (Y-m(X;\beta))^2 - v(X)$ as the error term which is uncorrelated with X. Similar to Koul and Ni (2004), Samarakoon and Song (2011) developed a lack-of-fit test for H_0 in (2.2) based on the following quantity:

$$T_n(\beta,\theta) = \int_{\mathcal{C}} \left[\frac{h^{-d}\sum_{i=1}^n K_h(x-X_i)[(Y_i-m(X_i;\beta))^2 - v(X_i;\beta,\theta)]}{w^{-d}\sum_{i=1}^n K_w(x-X_i)}\right]^2 dG(x), \tag{2.3}$$

where $\mathcal{C}$ is a compact set in $\mathbb{R}^d$, G is a weighting measure with $\mathcal{C}$ being a compact subset of its support, K is a kernel function, $K_h(\cdot) = K(\cdot/h)$, and h,w are the bandwidths. In real applications, β and θ are usually unknown. In Samarakoon and Song (2011), β is estimated in advance, θ is then estimated by $\hat{\theta}_n = \arg\min_{\theta\in\Theta} T_n(\hat{\beta}_n,\theta)$ and the test statistic is constructed from $T_n(\hat{\beta}_n,\hat{\theta}_n)$. The integral in $T_n(\hat{\beta}_n,\hat{\theta}_n)$ usually does not have a tractable form. Therefore, one has to approximate the integration using some numerical methods to implement the test. These numerical methods either take a long execution time because of the complex iterations or provide unstable results because of the subjectivity of choosing some tuning parameters in the algorithms. Zheng (1996) provided a nonparametric smoothing test for checking the adequacy of mean function forms. This test has a close connection with Koul and Ni's (2004) minimum distance method, but it does not need to calculate any integrations. Wang and Zhou (2006) applied Zheng's (1996) method to test the hypothesis (2.2). Denote $\xi_i = (Y_i - m(X_i;\beta_0))^2 - v(X_i;\beta_0,\theta_0)$. Note that under H_0,

$$E(\xi_i|X_i) = 0 \quad \text{and} \quad E[\xi_i E(\xi_i|X_i) f(X_i)] = 0 \text{ for } i = 1,2,\ldots,n,$$

while under H_a, since $E(\xi_i|X_i) = v(X_i) - v(X_i;\beta_0,\theta_0)$, it is clear that

$$E[\xi_i E(\xi_i|X_i) f(X_i)] = E[((E(\xi_i|X_i))^2) f(X_i)] = E[(v(X_i) - v(X_i;\beta_0,\theta_0))^2 f(X_i)] > 0. \tag{2.4}$$

Applying Zheng's (1996) idea, Wang and Zhou's (2006) test is based on the quantity

$$n^{-1}\sum_{i=1}^{n}\xi_i E(\xi_i|X_i)f(X_i), \tag{2.5}$$

which is a sample analogue of $E[\xi_i E(\xi_i|X_i)f(X_i)]$. Replacing $E(\xi_i|X_i)$, $f(X_i)$ with the leave-one-out Nadaraya–Watson kernel estimates as follows:

$$\hat{E}(\xi_i|X_i)=\frac{1}{(n-1)\hat{f}(X_i)}\sum_{j\neq i}\frac{1}{h^d}K_h(i,j)e_j,\quad \hat{f}(X_i)=\frac{1}{(n-1)}\sum_{j\neq i}\frac{1}{h^d}K_h(i,j), \tag{2.6}$$

where $e_i=(Y_i-m(X_i;\hat{\beta}_n))^2-v(X_i;\hat{\beta}_n,\hat{\theta}_n)$, $i=1,2,\ldots,n$, $\hat{\beta}_n$ and $\hat{\theta}_n$ are any $\sqrt{n}$-consistent estimators of β_0 and θ_0, the true values of β and θ under the null hypothesis, respectively, and $K_h(i,j)=K((X_i-X_j)/h)$, Wang and Zhou's (2006) test is then constructed from the following quantity:

$$Z_n=\frac{1}{n(n-1)}\sum_{i\neq j}\frac{1}{h^d}K_h(i,j)e_ie_j.$$

Similar to the question raised in Song and Du (2011) when checking the adequacy of mean function, we wonder why not use the empirical version of the second term in (2.4) to build the test statistic? An attractive feature of the empirical version of $E(E^2(\xi|X)f(X))$ is that the variance of this empirical version will be less than that of (2.5) used in Wang and Zhou (2006), which is derived from the following fact:

$$E(E^4(\xi|X)f^2(X))\leq EE(\xi^2|X)E^2(\xi|X)f^2(X)=E\xi^2E^2(\xi|X)f^2(X) \tag{2.7}$$

by applying Cauchy–Schwartz inequality. So if a new test is constructed based on the standardized sample analogue of the second term in (2.4), comparing to Zheng's test which uses the standardized sample analogue of the first term in (2.4) as the test statistic, we will find that these two test statistics might have similar numerators based on the first equality in (2.4), while the new test statistic has a smaller denominator than Wang and Zhou's (2006) test statistic. This implies that the new test might be more powerful than Wang and Zhou's (2006) test. Although that the variance of the population version (2.7) is smaller than that of (2.5) does not necessarily imply their empirical counterparts posses the same relationship, in particular, after replacing all unknown quantities with the estimators, but it is intuitively appealing to investigate the actual performance of the new test. Comparing with Wang and Zhou's (2006) test, the new test statistic is relatively complicated, in particular, the appearance of the kernel estimator of $f(x)$ in the denominator needs some extra conditions to avoid the possible asymptotic negligibility at the boundary points and the possible numeric instability when $f(x)$ is small. In real applications, if we are not sure whether or not these conditions hold for $f(x)$, then special attention should be paid when employing the proposed method. But except that, the new test shares the same advantages as Wang and Zhou's (2006) test.

Using the leave-one-out estimators in (2.6), the sample analogue of $E[(E(\xi|X))^2f(X)]$ is given by

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{(n-1)h^d}\sum_{j\neq i}K_h(i,j)e_j\right]^2\hat{f}^{-1}(X_i).$$

By expanding the square term, it can be written as

$$\frac{1}{n(n-1)^2h^{2d}}\sum_{i=1}^{n}\left[\sum_{j\neq i}\sum_{k\neq i}K_h(i,j)K_h(i,k)e_je_k\right]\hat{f}^{-1}(X_i). \tag{2.8}$$

Similar to the leave-one-out technique in (2.6), we drop all the terms with $k=j$ from the third sum in (2.8), accordingly, change one $1/(n-1)$ into $1/(n-2)$. Then the test statistic we are proposing has the form

$$Z_n=\frac{1}{n(n-1)(n-2)h^{2d}}\sum_{i=1}^{n}\left[\sum_{j\neq i}\sum_{k\neq i,j}K_h(i,j)K_h(i,k)e_je_k\right]\hat{f}^{-1}(X_i). \tag{2.9}$$

Denote $\dot{m}(x;\beta)$ as the derivative of $m(x;\beta)$ with respect to β, and $\dot{v}_\beta(x;\beta,\theta)$, $\dot{v}_\theta(x;\beta,\theta)$ be the derivatives of $v(x;\beta,\theta)$ with respect to β and θ, respectively. The following is a list of technical assumptions needed for proving the main results in the paper.

(C1) The design variable X has a compact support I and its density function $f(x)$ is continuous and satisfies $\min_{x\in I}f(x)\geq c$, where c is a positive constant.
(C2) $m(x;\beta)$, $v(x;\beta,\theta)$ and their derivatives $\dot{m}(x;\beta)$, $\dot{v}_\beta(x;\beta,\theta)$, $\dot{v}_\theta(x;\beta,\theta)$ are continuous in x for all β and θ.
(C3) $E(\varepsilon^4|X=x)$ is continuous in x.
(C4) For any $\sqrt{n}$-consistent estimator of β_0,

$$\sup_{1\leq i\leq n}|m(X_i;\hat{\beta}_n)-m(X_i;\beta_0)-(\hat{\beta}_n-\beta_0)'\dot{m}(X_i;\beta_0)|=O_p(1/n).$$

(C5) For any $\sqrt{n}$-consistent estimators $\hat{\beta}_n$, $\hat{\theta}_n$ of β_0 and θ_0, respectively,

$$\sup_{1\le i\le n}\left|\nu(X_i;\hat{\beta}_n,\hat{\theta}_n)-\nu(X_i;\beta_0,\theta_0)-(\hat{\beta}_n-\beta_0')\dot{\nu}_\beta(X_i;\beta_0,\theta_0)-(\hat{\theta}_n-\theta_0)'\dot{\nu}_\theta(X_i;\beta_0,\theta_0)\right|=O_p(1/n).$$

(C6) The nonnegative kernel function K is bounded, continuous, symmetric about 0, and $\int K(u)\,du=1$.
(C7) The bandwidth h is chosen so that $h\to 0$ and $nh^{2d}\to\infty$ as $n\to\infty$.

Condition (C1) is a typical restriction that avoids a nonparametric estimator of $f(x)$ from vanishing near the boundary of the design space; Conditions (C4) and (C5) might appear stronger, but if the second derivatives of $m(x,\beta)$, $\nu(x;\beta,\theta)$ with respect to β and θ are bounded in a neighborhood of β_0, θ_0, then (C4) and (C5) hold. Conditions (C6) and (C7) are the typical assumptions adopted in nonparametric smoothing literature.

3. Main results

Denote

$$\tau^2(x;\beta,\theta)=E(\xi^2|X=x)=\nu^2(x;\beta,\theta)[E(\varepsilon^4|X=x)-1]. \tag{3.1}$$

The asymptotic distribution of Z_n under the null hypothesis is given in the following theorem.

Theorem 3.1. *Assume that the conditions* (C1)–(C7) *hold, then under* H_0 *in* (2.2), $nh^{d/2}Z_n\Longrightarrow N(0,\sigma^2)$, *where*

$$\sigma^2=2\int\left[\int K(u+v)K(v)\,dv\right]^2du\cdot\int[\tau^2(x;\beta_0,\theta_0)]^2f^2(x)\,dx. \tag{3.2}$$

Let $H(u)=\int K(u+v)K(v)\,dv$, which is the convolution of K. If K is a nonnegative, bounded, continuous, and symmetric density function, so is H. Then σ^2 can be consistently estimated by $\hat{\sigma}^2$, where

$$\hat{\sigma}^2=\frac{2}{n(n-1)}\sum_{i\ne j}\frac{1}{h^d}H^2\left(\frac{x_i-x_j}{h}\right)e_i^2e_j^2.$$

Thus, the test that rejects H_0 whenever,

$$T_n=\frac{nh^{d/2}|Z_n|}{\hat{\sigma}}>z_{\alpha/2} \tag{3.3}$$

is of the asymptotic size α, where z_α is the $(1-\alpha)$ 100th percentile of the standard normal distribution.

The result above is similar to that in Wang and Zhou (2006) except for the first integration in σ^2. The integration in Wang and Zhou's (2006) result is $\int K^2(v)\,dv$. Note that H is the convolution of K, by Cauchy–Schwartz inequality, one can easily show that $\int H^2(v)\,dv\le\int K^2(v)\,dv$. That is, our test has a smaller asymptotic variance than that of Wang and Zhou's (2006) test.

Although the motivation of the current research is to construct a more precise test by modifying Wang and Zhou's test, it turns out that there are some interesting connections with Samarakoon and Song's (2011) minimum distance test based on (2.3). In fact, if we choose $w=h$, $dG(x)=\hat{f}_h(x)\,dF_n(x)$ in (2.3), where $F_n(x)$ is the empirical cumulative distribution function of X_i's, then after a slight and obvious modification, $T_n(\hat{\beta}_n,\hat{\theta}_n)$ is simply Z_n defined in (2.9).

The proof of Theorem 3.1, which is postponed to Section 5, shows that

$$nh^{d/2}Z_n=\frac{1}{(n-1)h^{d/2}}\sum_{j\ne k}H\left(\frac{X_j-X_k}{h}\right)\varepsilon_j\varepsilon_k+o_p(1):=V_n+o_p(1). \tag{3.4}$$

This also gives an interesting connection between our test and Wang and Zhou (2006)'s test: Our test is asymptotically equivalent to Wang and Zhou's (2006) test with the kernel function K replaced with the convolution H of K, nevertheless, our test is more powerful.

If one wants to construct a test based on V_n in (3.4) with the random errors ε_i's replaced by the residual e_i's, denoted it as $\hat{V}_n$, that is, we will rejects H_0 whenever

$$R_n=nh^{d/2}|\hat{V}_n|/\hat{\sigma}>z_{\alpha/2},$$

then the conditions needed for the asymptotic theory can be greatly simplified. For example, (C1) can be removed, and (C7) can be changed to $nh^d\to\infty$.

Typically, nonparametric tests are design to be omnibus, in the sense that they are consistent against a very wide class of fixed alternatives. A test is said to be consistent against a given alternative if the power of the test under that alternative tends to 1 as sample size tends to ∞. Let $\nu_1(x)$ be a known positive real valued function such that

$\nu_1 \notin \{\nu(x;\beta,\theta) : (\beta,\theta) \in \Gamma \times \Theta\}$. Consider the alternative hypothesis

$$H_a : E((Y-m(X;\beta))^2 | X=x) = \nu_1(x). \tag{3.5}$$

Under the null hypothesis, we have assumed that estimator $\hat{\theta}_n$ is $\sqrt{n}$-consistent for the true parameter θ_0. Would this estimator still have the similar property under the alternative hypothesis H_a? The question is of interest in its own right. In the classic regression setup, Jennrich (1969) and White (1981, 1982) showed that, under some mild regularity conditions, the nonlinear least squares estimator converges in probability and is asymptotically normal even in the presence of model misspecification. Suppose the true value of β under H_a is still β_0, the estimator $\hat{\beta}_n$ is usually not a consistent estimator for β_0. But under some regularity conditions, it might be a consistent estimator of some other value, say β_a and $\hat{\beta}_n$ is asymptotically normal. In the following, we simply assume that $\sqrt{n}(\hat{\theta}_n-\theta_a)=O_p(1)$ and $\sqrt{n}(\hat{\beta}_n-\beta_a)=O_p(1)$ under the alternative H_a for some $\beta_a \in \Gamma, \theta_a \in \Theta$. We will not justify this assumption rigorously here.

The following theorem states the asymptotic property of the test statistic under H_a.

Theorem 3.2. *Suppose the conditions* (C1)–(C7) *hold with* β_0 *and* θ_0 *replaced by* β_a *and* θ_a. *Then under the alternative hypothesis* H_a *in* (3.5),

$$Z_n \to E[(m(X;\beta_0)-m(X;\beta_a))^2+\nu_1(X)-\nu(X;\beta_a,\theta_a)]^2 f(X)$$

in probability and,

$$\hat{\sigma}^2 \to \int \left[\int K(u+v)K(v)\,dv\right]^2 du \cdot \int [\tau^2(x;\beta_a,\theta_a)+(m(x;\beta_0)-m(x;\beta_a))^2+(\nu_1(x)-\nu(x;\beta_a,\theta_a))]^2 f^2(x)\,dx \tag{3.6}$$

in probability, where $\tau^2(x;\beta,\theta)$ *is defined by* (3.1).

The consistency of the test is thus implied by the positiveness of

$$E[(m(X;\beta_0)-m(X;\beta_a))^2+\nu_1(X)-\nu(X;\beta_a,\theta_a)]^2 f(X)$$

and the finiteness of the right hand side of (3.6).

Sometimes, it would also be desirable to investigate how sensitive the test is to local alternatives. For this purpose, let $\delta(x)$ be a positive real valued continuous function in x, and consider the following local alternative:

$$H_{Loc} : \nu(x) = \nu(x;\beta_0,\theta_0)+c_n\delta(x), \quad \forall x \in I. \tag{3.7}$$

Under H_{Loc}, the regression model has the form

$$Y = m(X;\beta_0)+\sqrt{\nu(X;\beta_0,\theta_0)+c_n\delta(X)}\varepsilon.$$

The following theorem states that the proposed test has nontrivial power against a sequence of local alternatives which approaches to the null hypothesis at the rate of $1/\sqrt{nh^{d/2}}$.

Theorem 3.3. *Suppose* (C1)–(C7) *hold*, $c_n = 1/\sqrt{nh^{d/2}}$. *Then under* H_{Loc} *defined in* (3.7), $nh^{d/2}Z_n/\hat{\sigma} \Rightarrow N(\mu,1)$, *where* $\mu = E[\delta^2(x)f(x)]/\sigma$, *and* σ *is defined in* (3.2).

4. Simulation

This section investigates the finite sample performance of the proposed test through a Monte Carlo simulation study. We generate samples from the following models:

$$\text{Model 0}: \quad Y = \beta_1+\beta_2 X+\sqrt{\theta_1+\theta_2 X}\varepsilon,$$

$$\text{Model 1}: \quad Y = \beta_1+\beta_2 X+\sqrt{\theta_1+\theta_2 X+0.5X^2}\varepsilon,$$

$$\text{Model 2}: \quad Y = \beta_1+\beta_2 X+\sqrt{\theta_1+\theta_2 X+0.8X^2}\varepsilon,$$

$$\text{Model 3}: \quad Y = \beta_1+\beta_2 X+\sqrt{\theta_1+\theta_2 X+X^2}\varepsilon.$$

The data from model 0 are used to study the empirical level, while the data from models 1 to 3 are used to study the empirical power of the test. In the simulation, $X \sim U(-3,3)$, $\beta_1=1, \beta_2=2, \theta_1=2$ and $\theta_2=0.1$. Two types of error distributions are considered, $\varepsilon \sim N(0,1)$ and $\varepsilon \sim U(-\sqrt{3},\sqrt{3})$. The kernel function K is chosen to be the standard normal and the bandwidth is set to be $h=an^{-1/3}$ where a is a positive constant and the sample sizes are chosen to be $n=100, 200, 300, 400, 500, 800$. In the simulation, we chose $a=0.5, 0.8$, and 1 to see the influence of the bandwidth on the power of the test. For all scenarios, the nominal significance level is chosen to be 0.05, and the test is repeated 500 times. The empirical size and power are computed by using $\#\{T_n \geq 1.96\}/500$ with T_n being defined in (3.3). Tables 1–3 show the empirical levels and powers of the proposed test.

Table 1
Empirical smoothing lack-of-fit test, $a=1$.

	100	200	300	400	500	800
$\varepsilon \sim N(0,1)$						
Model 0	0.014	0.022	0.020	0.042	0.024	0.024
Model 1	0.056	0.342	0.666	0.844	0.942	0.998
Model 2	0.134	0.650	0.926	0.982	1.000	1.000
Model 3	0.224	0.748	0.960	1.000	1.000	1.000
$\varepsilon \sim U(-\sqrt{3},\sqrt{3})$						
Model 0	0.008	0.002	0.012	0.014	0.022	0.016
Model 1	0.358	0.872	0.990	0.998	1.000	1.000
Model 2	0.602	0.998	1.000	1.000	1.000	1.000
Model 3	0.730	0.966	1.000	1.000	1.000	1.000

Table 2
Empirical smoothing lack-of-fit test, $a=0.8$.

	100	200	300	400	500	800
$\varepsilon \sim N(0,1)$						
Model 0	0.028	0.036	0.030	0.046	0.034	0.052
Model 1	0.072	0.306	0.602	0.832	0.936	0.998
Model 2	0.160	0.600	0.892	0.978	1.000	1.000
Model 3	0.222	0.734	0.958	0.994	1.000	1.000
$\varepsilon \sim U(-\sqrt{3},\sqrt{3})$						
Model 0	0.018	0.014	0.016	0.016	0.020	0.024
Model 1	0.370	0.876	0.992	0.998	1.000	1.000
Model 2	0.640	0.968	1.000	1.000	1.000	1.000
Model 3	0.748	0.996	1.000	1.000	1.000	1.000

Table 3
Empirical smoothing lack-of-fit test, $a=0.5$.

	100	200	300	400	500	800
$\varepsilon \sim N(0,1)$						
Model 0	0.028	0.038	0.030	0.048	0.038	0.056
Model 1	0.042	0.168	0.428	0.668	0.864	0.986
Model 2	0.096	0.434	0.754	0.944	0.990	1.000
Model 3	0.126	0.552	0.872	0.974	0.998	1.000
$\varepsilon \sim U(-\sqrt{3},\sqrt{3})$						
Model 0	0.014	0.014	0.020	0.026	0.028	0.028
Model 1	0.244	0.760	0.972	0.996	1.000	1.000
Model 2	0.448	0.944	0.998	1.000	1.000	1.000
Model 3	0.538	0.966	1.000	1.000	1.000	1.000

The simulation study shows that most empirical levels are less than the nominal level 0.05 and hence the proposed test is conservative for the chosen sample sizes and for both error types. The empirical powers against all alternative models get larger when the sample sizes get larger. From the above simulation, one can see that if ε has a uniform distribution, the empirical levels and powers are not very sensitive to the choices of a values, but for normal error, the value of a has some effects on the empirical levels and powers. A further extensive study on the choice of bandwidth is necessary and this is left for future research.

We also conduct a simulation study using a bootstrap method as it generally provides more accurate approximation to the distribution of the test statistic than asymptotic normal theory does when the sample size is small to moderate. The bootstrap method we use in this study is same as that of in Samarakoon and Song (2011). We use 400 bootstrap samples per run to obtain the critical value c_α^*. The empirical size and power are computed by using $\#\{T_n \geq c_\alpha^*\}/500$. Table 4 shows the empirical level and power of the test for $a=1$. Similar pattern as in Table 1 can be found in Table 4, but the empirical levels are all close to the nominal level 0.05.

As a comparison, we also carry out a simulation study for the tests proposed by Samarakoon and Song (2011), Wang and Zhou (2006) using bootstrap. The simulation results for $a=1$ are shown in Tables 5 and 6. Comparing Table 4 with Tables 5 and 6, we can see that the proposed test is more powerful than the tests proposed by Samarakoon and Song (2011), Wang and Zhou (2006), when sample size is bigger.

Table 4
Empirical smoothing lack-of-fit test, bootstrap, $a=1$.

	100	200	300	400	500	800
$\varepsilon\sim N(0,1)$						
Model 0	0.062	0.046	0.054	0.055	0.043	0.048
Model 1	0.120	0.390	0.716	0.880	0.968	0.994
Model 2	0.174	0.694	0.932	0.999	1.000	1.000
Model 3	0.260	0.782	0.976	0.998	1.000	1.000
$\varepsilon\sim U(-\sqrt{3},\sqrt{3})$						
Model 0	0.061	0.059	0.062	0.052	0.050	0.050
Model 1	0.474	0.950	0.996	0.998	1.000	1.000
Model 2	0.744	0.996	1.000	1.000	1.000	1.000
Model 3	0.790	1.000	1.000	1.000	1.000	1.000

Table 5
Minimum distance test, bootstrap $a=1$.

	100	200	300	400	500	800
$\epsilon\sim N(0,1)$						
Model 0	0.040	0.028	0.016	0.024	0.018	0.034
Model 1	0.148	0.350	0.582	0.730	0.834	0.988
Model 2	0.224	0.520	0.782	0.926	0.964	1.000
Model 3	0.270	0.606	0.892	0.968	0.988	1.000
$\epsilon\sim U(-\sqrt{3},\sqrt{3})$						
Model 0	0.064	0.042	0.058	0.044	0.024	0.042
Model 1	0.384	0.856	0.980	0.998	1.000	1.000
Model 2	0.598	0.966	0.996	1.000	1.000	1.000
Model 3	0.688	0.980	1.000	1.000	1.000	1.000

Table 6
Wang and Zhou's test, bootstrap, $a=1$.

	100	200	300	400	500	800
$\epsilon\sim N(0,1)$						
Model 0	0.038	0.036	0.028	0.046	0.020	0.030
Model 1	0.210	0.424	0.578	0.715	0.834	0.976
Model 2	0.243	0.537	0.710	0.791	0.903	0.989
Model 3	0.256	0.630	0.865	0.903	0.980	1.000
$\epsilon\sim U(-\sqrt{3},\sqrt{3})$						
Model 0	0.048	0.062	0.046	0.050	0.040	0.056
Model 1	0.384	0.624	0.868	0.915	0.948	1.000
Model 2	0.420	0.752	0.890	0.953	0.991	1.000
Model 3	0.711	0.890	0.983	0.992	1.000	1.000

5. Proof of the main results

Proof of Theorem 3.1. Adding and subtracting $m(X_i;\beta_0)$ and $\nu(X_i;\beta_0,\theta_0)$ from e_i, e_i can be written as

$$\begin{aligned} e_i &= (Y_i-m(X_i;\beta_0)+m(X_i;\beta_0)-m(X_i;\hat\beta))^2-\nu(X_i;\beta_0,\theta_0)+\nu(X_i;\beta_0,\theta_0)-\nu(X_i;\hat\beta,\hat\theta) \\ &= \xi_i+(\Delta m_i)^2-2\Delta m_i(Y_i-m(X_i;\beta_0))-\Delta\nu_i, \end{aligned}$$

where $\Delta m_i = m(X_i;\hat\beta)-m(X_i;\beta_0)$ and $\Delta\nu_i=\nu(X_i;\hat\beta,\hat\theta)-\nu(X_i;\beta_0,\theta_0)$. Then Z_n can be further written as the sum of

$$Z_{nl}=\frac{1}{n(n-1)(n-2)h^{2d}}\sum_{i=1}^{n}\left[\sum_{j\neq i}\sum_{k\neq i,j}K_h(i,j)K_h(i,k)\right]\hat f^{-1}(X_i)P_{jk,l}$$

for $l=1,2,\ldots,10$ with,

$$P_{jk,1}=\xi_j\xi_k,\quad P_{jk,2}=2\xi_j(\Delta m_k)^2,\quad P_{jk,3}=-2\xi_j\Delta\nu_k,\quad P_{jk,4}=\Delta\nu_j\Delta\nu_k,$$

$$P_{jk,5}=(\Delta m_j)^2(\Delta m_k)^2,\quad P_{jk,6}=-2(\Delta m_j)^2\Delta\nu_k,\quad P_{jk,7}=4\Delta m_j\Delta\nu_k(Y_j-m(X_j;\beta_0)),$$

$$P_{jk,8} = -4\xi_j \Delta m_k (Y_k - m(X_k;\beta_0)), \quad P_{jk,9} = -4\Delta m_k (\Delta m_j)^2 (Y_k - m(X_k;\beta_0)),$$

$$P_{jk,10} = 4\Delta m_j \Delta m_k (Y_j - m(X_j,\beta_0))(Y_k - m(X_k;\beta_0)).$$

In the following, we use $\tilde{Z}_{nl}$ to denote Z_{nl} when $\hat{f}(X_i)$ is replaced by $f(X_i)$ for all l.

Note that

$$nh^{d/2}\tilde{Z}_{n1} = \frac{1}{(n-1)(n-2)h^{3d/2}} \sum_{i=1}^{n} \left[\sum_{j\neq i} \sum_{k\neq i,j} K_h(i,j) K_h(i,k) \xi_j \xi_k \right] f^{-1}(X_i)$$
$$= \frac{1}{(n-1)h^{3d/2}} \sum_{j=1}^{n} \sum_{k\neq j} \left[\frac{1}{(n-2)} \sum_{i\neq j,k} K_h(i,j) K_h(i,k) f^{-1}(X_i) \right] \xi_j \xi_k.$$

Denote $H_h(k,j) = \int K(u+(X_k-X_j)/h)K(u)\,du$. In fact, $H_h(k,j) = H((X_k-X_j)/h)$. We have $E[K_h(i,j)K_h(i,k)f^{-1}(X_i)|X_j,X_k] = h^d H_h(k,j)$. Now we can write, $nh^{d/2}\tilde{Z}_{n1} = A_{n1} + A_{n2}$ where

$$A_{n1} = \frac{1}{(n-1)h^{3d/2}} \sum_{j=1}^{n} \sum_{k\neq j} \left[\frac{1}{(n-2)} \sum_{i\neq j,k} K_h(i,j) K_h(i,k) f^{-1}(X_i) - h^d H_h(k,j) \right] \xi_j \xi_k,$$

$$A_{n2} = \frac{1}{(n-1)h^{d/2}} \sum_{j=1}^{n} \sum_{k\neq j} H_h(k,j) \xi_j \xi_k.$$

Using the expectation-variance argument, we can show that A_{n1} is the order of $o_p(1)$. It is clear that $EA_{n1} = 0$. To consider the second moment of A_{n1}, let

$$G_h(j,k) = \frac{1}{(n-2)} \sum_{i\neq j,k} K_h(i,j) K_h(i,k) f^{-1}(X_i) - h^d H_h(k,j),$$

which is symmetric in (j, k). Then A_{n1} can be rewritten as

$$A_{n1} = \frac{1}{(n-1)h^{3d/2}} \sum_{j=1}^{n} \sum_{k\neq j} G_h(j,k) \xi_j \xi_k.$$

The independence of ξ_j and ξ_k when $j\neq k$, and $E(\xi|X) = 0$ imply

$$EA_{n1}^2 = \frac{1}{(n-1)^2 h^{3d}} E\left(\sum_{j=1}^{n} \sum_{k\neq j} G_h(j,k) \xi_j \xi_k \right)^2 = \frac{1}{(n-1)^2 h^{3d}} E \sum_{j=1}^{n} \sum_{k\neq j} G_h^2(j,k) \xi_j^2 \xi_k^2$$
$$= \frac{1}{(n-1)^2 h^{3d}} \sum_{j=1}^{n} \sum_{k\neq j} E G_h^2(j,k) \xi_j^2 \xi_k^2 = \frac{n}{(n-1)h^{3d}} E G_h^2(1,2) \tau_0^2(X_1) \tau_0^2(X_2),$$

where $\tau_0^2(x) = \tau^2(x;\beta_0,\theta_0)$ is defined by (3.1). Conditioning on (X_1,X_2), $G_h(1,2)$ is a sum of i.i.d. centered random variables. Therefore the expectation of $G_h^2(1,2)\tau_0^2(X_1)\tau_0^2(X_2)$ equals

$$E\left[\frac{1}{(n-2)} \sum_{i\neq 1,2} K_h(i,1) K_h(i,2) f^{-1}(X_i) - h^d H_h(1,2) \right]^2 \tau_0^2(X_1) \tau_0^2(X_2)$$
$$= \frac{1}{(n-2)} E[K_h(3,1)K_h(3,2)f^{-1}(X_3) - h^d H_h(1,2)]^2 \tau_0^2(X_1)\tau_0^2(X_2)$$
$$\leq \frac{1}{(n-2)} E K_h^2(3,1) K_h^2(3,2) f^{-2}(X_3) \tau_0^2(X_1) \tau_0^2(X_2)$$
$$= \frac{1}{n-2} \iiint K^2\left(\frac{x-y}{h}\right) K^2\left(\frac{x-z}{h}\right) \tau_0^2(y) \tau_0^2(z) f^{-1}(x) f(y) f(z)\,dx\,dy\,dz$$
$$= \frac{h^{2d}}{n-2} \iiint K^2(u) K^2(v) \tau_0^2(x-uh) \tau_0^2(x-vh) f^{-1}(x) f(x-uh) f(x-vh)\,dx\,du\,dv.$$

From the continuity and boundedness of f, τ_0^2, K by (C1), (C3), (C6), we have

$$EG_h^2(1,2)\tau_0^2(X_1)\tau_0^2(X_2) = O(h^{2d}/(n-2)). \tag{5.1}$$

So $EA_{n1}^2 = o(1)$ from (C7). This implies $nh^{d/2}\tilde{Z}_{n1} = A_{n2} + o_p(1)$.

To show that $nh^{d/2}\tilde{Z}_{n2} = o_p(1)$, denote

$$d_{ni} = m(X_i;\hat{\beta}_n) - m(X_i;\beta_0) - (\hat{\beta}_n - \beta_0)'\dot{m}(X_i;\beta_0) = \Delta m_i - (\hat{\beta}_n - \beta_0)'\dot{m}(X_i;\beta_0).$$

By (C4), $\sup_{1\le i\le n}|d_{ni}|=O_p(1/n)$. Using the notation d_{ni} and $a_n=1/(n(n-1)(n-2)h^{2d})$, $\tilde{Z}_{n2}$ is the sum of $\tilde{Z}_{n21}+\tilde{Z}_{n22}+\tilde{Z}_{n23}$ where

$$\tilde{Z}_{n21}=2a_n\sum_{i=1}^{n}\left[\sum_{j\ne k}\sum_{k\ne i,j}K_h(i,j)K_h(i,k)\xi_j d_{nk}^2\right]f^{-1}(X_i),$$

$$\tilde{Z}_{n22}=4a_n\sum_{i=1}^{n}\left[\sum_{j\ne k}\sum_{k\ne i,j}K_h(i,j)K_h(i,k)\xi_j d_{nk}(\hat{\beta}_n-\beta_0)'\dot{m}(X_k;\beta_0)\right]f^{-1}(X_i),$$

$$\tilde{Z}_{n23}=2a_n\sum_{i=1}^{n}\left[\sum_{j\ne k}\sum_{k\ne i,j}K_h(i,j)K_h(i,k)\xi_j(\hat{\beta}_n-\beta_0)'\dot{m}(X_k;\beta_0)\dot{m}'(X_k;\beta_0)(\hat{\beta}_n-\beta_0)\right]f^{-1}(X_i).$$

Notice that $|\tilde{Z}_{n21}|$ is bounded above by

$$2\sup_{1\le k\le n}|d_{nk}|^2\cdot a_n\sum_{i=1}^{n}\left[\sum_{j\ne k}\sum_{k\ne i,j}K_h(i,j)K_h(i,k)|\xi_j|\right]f^{-1}(X_i).$$

The expectation of the second term is further bounded by

$$\frac{1}{h^{2d}}E[K_h(1,2)K_h(1,3)E(|\xi_2||X_2)f^{-1}(X_1)]\le\frac{1}{h^{2d}}\iint K\left(\frac{x-y}{h}\right)K\left(\frac{x-z}{h}\right)\tau_0(y)f(y)f(z)\,dx\,dy\,dz$$
$$=\iint K(u)K(v)\tau_0(x-uh)f(x-vh)f(x-uh)\,dx\,du\,dv=O(1).$$

Therefore, $nh^{d/2}\tilde{Z}_{n21}=nh^{d/2}\cdot O_p(1/n^2)\cdot O(1)=o_p(1)$.

Similarly, one can show that $nh^{d/2}\tilde{Z}_{n22}=o_p(1)$ and $nh^{d/2}\tilde{Z}_{n23}=o_p(1)$ by the $\sqrt{n}$-consistency of $\hat{\beta}_n$, and the boundedness of $\|\dot{m}(x;\beta_0)\|$. Therefore, we obtain $nh^{d/2}\tilde{Z}_{n2}=o_p(1)$.

Next, let us show that $nh^{d/2}\tilde{Z}_{n3}=o_p(1)$. Adding and subtracting $(\hat{\beta}_n-\beta_0)'\dot{v}_\beta(X_k;\beta_0,\theta_0)$ and $(\hat{\theta}_n-\theta_0)'\dot{v}_\theta(X_k;\beta_0,\theta_0)$ from $\Delta v_k=v(X_k;\hat{\beta},\hat{\theta})-v(X_k;\beta_0,\theta_0)$ and denote

$$u_{nk}=\Delta v_k-(\hat{\beta}_n-\beta_0)'\dot{v}_\beta(X_k;\beta_0,\theta_0)-(\hat{\theta}_n-\theta_0)'\dot{v}_\theta(X_k;\beta_0,\theta_0),$$

we can write $nh^{d/2}\tilde{Z}_{n3}$ as the sum of the following three terms:

$$B_{n1}=-\frac{2(\hat{\beta}_n-\beta_0)'}{(n-1)(n-2)h^{3d/2}}\sum_{i=1}^{n}\left[\sum_{j\ne k}\sum_{k\ne i,j}K_h(i,j)K_h(i,k)\right]f^{-1}(X_i)\xi_j\dot{v}_\beta(X_k;\beta_0,\theta_0),$$

$$B_{n2}=-\frac{2(\hat{\theta}_n-\theta_0)'}{(n-1)(n-2)h^{3d/2}}\sum_{i=1}^{n}\left[\sum_{j\ne k}\sum_{k\ne i,j}K_h(i,j)K_h(i,k)\right]f^{-1}(X_i)\xi_j\dot{v}_\theta(X_k;\beta_0,\theta_0),$$

$$B_{n3}=-\frac{2}{(n-1)(n-2)h^{3d/2}}\sum_{i=1}^{n}\left[\sum_{j\ne k}\sum_{k\ne i,j}K_h(i,j)K_h(i,k)\right]f^{-1}(X_i)\xi_j u_{nk}.$$

By adding and subtracting $h^dH_h(j,k)$, B_{n1} can be written as the sum of $B_{n11}+B_{n12}$, where

$$B_{n11}=-\frac{2(\hat{\beta}_n-\beta_0)'}{(n-1)h^{3d/2}}\sum_{j\ne k}\left[\frac{1}{(n-2)}\sum_{i\ne j,k}K_h(i,j)K_h(i,k)f^{-1}(X_i)-h^dH_h(j,k)\right]\xi_j\dot{v}_\beta(X_k;\beta_0,\theta_0),$$

$$B_{n12}=-\frac{2(\hat{\beta}_n-\beta_0)'}{(n-1)h^{d/2}}\sum_{j\ne k}H_h(j,k)\xi_j\dot{v}_\beta(X_k;\beta_0,\theta_0).$$

Let $\dot{v}_{\beta l}(X_k;\beta_0,\theta_0)$ denote the l-th element of the $p\times 1$ vector $\dot{v}_\beta(X_k;\beta_0,\theta_0)$. Note that

$$E\left[\frac{1}{(n-1)h^{3d/2}}\sum_{j\ne k}\left[\frac{1}{n-2}\sum_{i\ne j,k}K_h(i,j)K_h(i,k)f^{-1}(X_i)-h^dH_h(j,k)\right]\xi_j\dot{v}_{\beta l}(X_k;\beta_0,\theta_0)\right]^2$$
$$=\frac{n}{(n-1)^2h^{3d}}E\left[\sum_{k=2}^{n}\left(\frac{1}{n-2}\sum_{i\ne 1,k}K_h(i,1)K_h(i,k)f^{-1}(X_i)-h^dH_h(1,k)\right)\dot{v}_{\beta l}(X_k;\beta_0,\theta_0)\right]^2\tau_0^2(X_1)$$

$$= \frac{n(n-1)}{(n-1)^2h^{3d}}E\left(\frac{1}{n-2}\sum_{i\neq 1,2}K_h(i,1)K_h(i,2)f^{-1}(X_i)-h^dH_h(1,2)\right)^2\dot{v}^2_{\beta l}(X_2;\beta_0,\theta_0)\tau_0^2(X_1)$$
$$+\frac{n(n-1)(n-2)}{(n-1)^2h^{3d}}E\left(\frac{1}{n-2}\sum_{i\neq 1,2}K_h(i,1)K_h(i,2)f^{-1}(X_i)-h^dH_h(1,2)\right)\dot{v}_{\beta l}(X_2;\beta_0,\theta_0)$$
$$\left(\frac{1}{n-2}\sum_{i\neq 1,3}K_h(i,1)K_h(i,3)f^{-1}(X_i)-h^dH_h(1,3)\right)\dot{v}_{\beta l}(X_3;\beta_0,\theta_0)\tau_0^2(X_1).$$

Similar to the proof of (5.1), by the continuity of $\dot{v}_\beta(x;\beta_0,\theta_0)$, $\tau_0^2(x)$, one can prove that the expectation in the first term on the right is $O(h^{2d}/(n-2))$. Therefore the first term on the right hand side is $O_p(1/(nh^d))$ which is $o_p(1)$. By a lengthy but trivial argument, one can show that the second term on the right is $O_p(1)$. By the $\sqrt{n}$-consistency of $\hat{\beta}_n$, $B_{n11}=O_p(1/\sqrt{n})O_p(1)=o_p(1)$. According to Lemma 3.3b in Zheng (1996),

$$\frac{1}{n(n-1)h^d}\sum_{j=1}^n\sum_{j\neq k}H_h(k,j)\xi_j\dot{v}_\beta(X_k;\beta_0)=O_p(1/\sqrt{n}).$$

Therefore $B_{n12}=nh^{d/2}O_p(1/\sqrt{n})O_p(1/\sqrt{n})=o_p(1)$. Hence B_{n1} is $o_p(1)$.

By the $\sqrt{n}$-consistency of $\hat{\theta}_n$, and the continuity of $\dot{v}_\theta(x;\beta_0,\theta_0)$, one can similarly show that $B_{n2}=o_p(1)$.

To show that $B_{n3}=o_p(1)$, note that $|B_{n3}|$ is bounded above by

$$\sup_{1\le k\le n}|u_{nk}|\cdot\frac{2}{(n-1)h^{3d/2}}\sum_{j=1}^n\sum_{k\neq j}\left[\frac{1}{n-2}\sum_{i\neq j,k}K_h(i,j)K_h(i,k)|\xi_j|f^{-1}(X_i)\right].$$

It can be shown that the expectation of the second term is

$$\frac{2n(n-1)}{(n-1)h^{3d/2}}E\left[\frac{K_h(1,2)K_h(1,3)}{f(X_1)}E(|\xi||X_2)\right]=O(nh^{d/2}).$$

By (C5), $\sup_{1\le k\le n}|u_{nk}|=O_p(1/n)$, which implies $B_{n3}=O_p(1/n)O_p(nh^{d/2})=o_p(1)$. Therefore, we have $nh^{d/2}\tilde{Z}_{n3}=o_p(1)$.

Using similar arguments, one can show that $nh^{d/2}\tilde{Z}_{nl}=o_p(1)$, for all $l=4,5,\ldots,10$.

Note that the above results are obtained by replacing $\hat{f}(x)$ with $f(x)$. Next we shall show that similar results hold for $\hat{f}(x)$. Denote

$$C_n=\frac{nh^{d/2}}{n(n-1)(n-2)h^{2d}}\sum_{i\neq j,k}K_h(i,j)K_h(i,k)f^{-1}(X_i)\left[\frac{f(X_i)}{\hat{f}(X_i)}-1\right]\xi_j\xi_k.$$

Let

$$M_n(X_j,X_k)=\frac{1}{(n-2)h^{2d}}\sum_{i\neq j,k}K_h(i,j)K_h(i,k)f^{-1}(X_i)\left[\frac{f(X_i)}{\hat{f}(X_i)}-1\right].$$

Then

$$C_n=\frac{h^{d/2}}{(n-1)}\sum_{j\neq k}M_n(X_j,X_k)\xi_j\xi_k.$$

The symmetry of $M_n(X_j,X_k)$ and the assumptions of error terms imply

$$EC_n^2=\frac{4h^d}{(n-1)^2}\sum_{j<k}EM_n^2(X_j,X_k)\tau_0^2(X_j)\tau_0^2(X_k)=O(h^d)EM_n^2(X_1,X_2)\tau_0^2(X_1)\tau_0^2(X_2).$$

Note that

$$EM_n^2(X_1,X_2)\tau_0^2(X_1)\tau_0^2(X_2)=E\left[\frac{1}{(n-2)h^{2d}}\sum_{i=3}^nK_h(i,1)K_h(i,2)f^{-1}(X_i)\left[\frac{f(X_i)}{\hat{f}(X_i)}-1\right]\right]^2\tau_0^2(X_1)\tau_0^2(X_2)$$
$$\le\frac{1}{(n-2)h^{4d}}\sum_{i=3}^nEK_h^2(i,1)K_h^2(i,2)f^{-2}(X_i)\left[\frac{f(X_i)}{\hat{f}(X_i)}-1\right]^2\tau_0^2(X_1)\tau_0^2(X_2)$$
$$=\frac{1}{h^{4d}}EK_h^2(3,1)K_h^2(3,2)f^{-2}(X_3)\left[\frac{f(X_3)}{\hat{f}(X_3)}-1\right]^2\tau_0^2(X_1)\tau_0^2(X_2). \quad (5.2)$$

From the condition (C1), we can see that the last expectation in (5.2) has the same order as

$$\frac{1}{h^{4d}} EK_h^2(3,1)K_h^2(3,2)f^{-2}(X_3)[f(X_3)-\hat{f}(X_3)]^2\tau_0^2(X_1)\tau_0^2(X_2). \tag{5.3}$$

Notice that

$$\hat{f}(X_3)-f(X_3) = \frac{1}{nh^d}\sum_{j=4}^{n} K_h(j,3) - \frac{1}{h^d}E\left[K_h(4,3)|X_3\right] + \frac{1}{h^d}E\left[K_h(4,3)|X_3\right] - f(X_3)$$
$$+\frac{1}{nh^d}K_h(1,3)+\frac{1}{nh^d}K_h(2,3)+\frac{1}{nh^d}K_h(1,1).$$

Therefore, (5.3) can be bounded above by the sum of the following three terms:

$$3E\frac{K_h^2(3,1)K_h^2(3,2)}{h^{4d}f^2(X_3)}\left[\frac{1}{nh^d}\sum_{j=4}^{n}K_h(j,3)-\frac{1}{h^d}E\left[K_h(4,3)|X_3\right]\right]^2\tau_0^2(X_1)\tau_0^2(X_2),$$

$$3E\frac{K_h^2(3,1)K_h^2(3,2)}{h^{4d}f^2(X_3)}\left[\frac{1}{h^d}E[K_h(4,3)|X_3]-f(X_3)\right]^2\tau_0^2(X_1)\tau_0^2(X_2),$$

$$3E\frac{K_h^2(3,1)K_h^2(3,2)}{h^{4d}f^2(X_3)}\left[\frac{1}{nh^d}K_h(1,3)+\frac{1}{nh^d}K_h(2,3)+\frac{1}{nh^d}K_h(1,1)\right]^2\tau_0^2(X_1)\tau_0^2(X_2).$$

By changing variables when calculating the above expectations, one can show that the first term has the order of $O(1/nh^{3d})$, the second term has the order of $O(1)$, and the third one has the order of $O(1/n^2h^{4d})$. Therefore,

$$C_n^2 = O_p(1/nh^{2d})+O_p(h^d)+O_p(1/n^2h^{3d}).$$

By condition (C7), we have $C_n = o_p(1)$.

Now we consider the following term:

$$D_n = \frac{2nh^{d/2}}{n(n-1)(n-2)h^{2d}}\sum_{i\neq j,k}K_h(i,j)K_h(i,k)f^{-1}(X_i)\left[\frac{f(X_i)}{\hat{f}(X_i)}-1\right]\xi_j(\Delta m_k)^2.$$

Using the same definition of $M_n(X_j,X_k)$, we can write

$$D_n = \frac{2h^{d/2}}{(n-1)}\sum_{j\neq k}M_n(X_j,X_k)\xi_j(\Delta m_k)^2.$$

Note that $\Delta m_k = d_{nk}+(\hat{\beta}_n-\beta_0)'\dot{m}(X_k;\beta_0)$, D_n can be written as the sum of $D_{n1}+D_{n2}+D_{n3}$ where

$$D_{n1} = \frac{2h^{d/2}}{(n-1)}\sum_{j\neq k}M_n(X_j,X_k)\xi_j d_{nk}^2,$$

$$D_{n2} = \frac{4h^{d/2}}{(n-1)}\sum_{j\neq k}M_n(X_j,X_k)\xi_j d_{nk}(\hat{\beta}_n-\beta_0)'\dot{m}(X_k;\beta_0),$$

$$D_{n3} = \frac{2h^{d/2}}{(n-1)}\sum_{j\neq k}M_n(X_j,X_k)\xi_j(\hat{\beta}_n-\beta_0)'\dot{m}(X_k;\beta_0)\dot{m}'(X_k;\beta_0)(\hat{\beta}_n-\beta_0).$$

By (C4) and the $\sqrt{n}$-consistency of $\hat{\beta}_n$, we have

$$|D_{n1}| \le O_p\left(\frac{h^{d/2}}{n^2}\right)\cdot\frac{1}{n-1}\sum_{j\neq k}|M_n(X_j,X_k)|\cdot|\xi_j|,$$

$$|D_{n2}| \le O_p\left(\frac{h^{d/2}}{n\sqrt{n}}\right)\cdot\frac{1}{n-1}\sum_{j\neq k}|M_n(X_j,X_k)|\cdot|\xi_j|\cdot\|\dot{m}(X_k;\beta_0)\|,$$

$$|D_{n3}| \le O_p\left(\frac{h^{d/2}}{n}\right)\cdot\frac{1}{n-1}\sum_{j\neq k}|M_n(X_j,X_k)|\cdot|\xi_j|\cdot\|\dot{m}(X_k;\beta_0)\|^2.$$

But for any $l=0,1,2$,

$$E\left[\frac{1}{n-1}\sum_{j\neq k}|M_n(X_j,X_k)|\cdot|\xi_j|\cdot\|\dot m(X_k;\beta_0)\|^l\right]\le nE|M_n(X_1,X_2)|\cdot\tau_0(X_1)\cdot\|\dot m(X_2;\beta_0)\|^l.$$

Similar arguments as in discussing (5.2), we have $E|M_n(X_1,X_2)|\tau_0(X_1)\|\dot m(X_2;\beta_0)\|^l=O_p(1)$. Therefore, $D_{n1}=o_p(1)$, $D_{n2}=o_p(1)$ and $D_{n3}=o_p(1)$ from (C7), which imply $D_n=o_p(1)$. Using similar methods, one can deal with all other modified terms.

In summary, we finally get

$$nh^{d/2}Z_n=\frac{1}{(n-1)h^{d/2}}\sum_{j=1}^{n}\sum_{k\neq j}H_h(k,j)\xi_j\xi_k+o_p(1).$$

From Lemma 3.3a in Zheng (1996), $nh^{d/2}Z_n\Rightarrow N(0,\sigma^2)$, where σ^2 is defined in (3.2). □

Proof of Theorem 3.2. For convenience, write $m_0(x)=m(x;\beta_0), m_a(x)=m(x;\beta_a), \nu_a(x)=\nu(x;\beta_a,\theta_a)$, $Y_i^a=m(X_i;\beta_a)+\sqrt{\nu_a(X_i)}\varepsilon_i$, $Y_i=m(X_i;\beta_0)+\sqrt{\nu_1(X_i)}\varepsilon_i$, $e_i^a=(Y_i^a-m(X_i;\hat\beta_n))^2-\nu(X_i;\hat\beta_n,\hat\theta_n)$ and $a_n=1/(n(n-1)(n-2)h^{2d})$. By adding and subtracting Y_i^a from Y_i, Z_n can be written as the sum of the following six terms:

$$U_{n1}=a_n\sum_{i\neq j\neq k}K_h(i,j)K_h(i,k)(Y_j-Y_j^a)^2(Y_k-Y_k^a)^2\hat f^{-1}(X_i),$$

$$U_{n2}=4a_n\sum_{i\neq j\neq k}K_h(i,j)K_h(i,k)(Y_j-Y_j^a)^2(Y_k-Y_k^a)(Y_k^a-m(X_k;\hat\beta_n))\hat f^{-1}(X_i),$$

$$U_{n3}=2a_n\sum_{i\neq j\neq k}K_h(i,j)K_h(i,k)(Y_j-Y_j^a)^2e_k^a\hat f^{-1}(X_i),$$

$$U_{n4}=4a_n\sum_{i\neq j\neq k}K_h(i,j)K_h(i,k)(Y_j-Y_j^a)(Y_j^a-m(X_j;\hat\beta_n))(Y_k-Y_k^a)(Y_k^a-m(X_k;\hat\beta_n))\hat f^{-1}(X_i),$$

$$U_{n5}=4a_n\sum_{i\neq j\neq k}K_h(i,j)K_h(i,k)(Y_j-Y_j^a)(Y_j^a-m(X_j;\hat\beta_n))e_k^a\hat f^{-1}(X_i),$$

$$U_{n6}=a_n\sum_{i\neq j\neq k}K_h(i,j)K_h(i,k)e_j^ae_k^a\hat f^{-1}(X_i).$$

Since $Y_i-Y_i^a=m(X_i;\beta_0)-m(X_i;\beta_a)+(\sqrt{\nu_1(X_i)}-\sqrt{\nu_a(X_i)})\varepsilon_i$, and denote now $\Delta m_i=m(X_i;\beta_0)-m(X_i;\beta_a)$, $\Delta\nu_i=(\sqrt{\nu_1(X_i)}-\sqrt{\nu_a(X_i)})$, U_{n1} can be written as the sum of the following six terms:

$$U_{n11}=a_n\sum_{i\neq j\neq k}K_h(i,j)K_h(i,k)(\Delta m_j)^2(\Delta m_k)^2\hat f^{-1}(X_i),$$

$$U_{n12}=4a_n\sum_{i\neq j\neq k}K_h(i,j)K_h(i,k)(\Delta m_j)^2\Delta m_k\Delta\nu_k\varepsilon_k\hat f^{-1}(X_i),$$

$$U_{n13}=2a_n\sum_{i\neq j\neq k}K_h(i,j)K_h(i,k)(\Delta m_j)^2(\Delta\nu_k)^2\varepsilon_k^2\hat f^{-1}(X_i),$$

$$U_{n14}=4a_n\sum_{i\neq j\neq k}K_h(i,j)K_h(i,k)\Delta m_j\Delta m_k\Delta\nu_j\Delta\nu_k\varepsilon_j\varepsilon_k\hat f^{-1}(X_i),$$

$$U_{n15}=4a_n\sum_{i\neq j\neq k}K_h(i,j)K_h(i,k)\Delta m_j\Delta\nu_j(\Delta\nu_k)^2\varepsilon_j\varepsilon_k^2\hat f^{-1}(X_i),$$

$$U_{n16}=a_n\sum_{i\neq j\neq k}K_h(i,j)K_h(i,k)(\Delta\nu_j)^2(\Delta\nu_k)^2\varepsilon_j^2\varepsilon_k^2\hat f^{-1}(X_i).$$

With the assumption on ε, one can show that $U_{n11}\to E[(m(X;\beta_0)-m(X;\beta_a))^4f(X)]$, $U_{n13}\to 2E[(m(X;\beta_0)-m(X;\beta_a))^2(\sqrt{\nu_1(X)}-\sqrt{\nu_a(X)})^2f(X)]$, $U_{n16}\to E[(\sqrt{\nu_1(X)}-\sqrt{\nu_a(X)})^4f(X)]$, and all other terms are of order $o_p(1)$. Hence,

$$U_{n1}\to E[(m(X;\beta_0)-m(X;\beta_a))^2+(\sqrt{\nu_1(X)}-\sqrt{\nu_a(X)})^2]^2f(X).$$

Using the $\sqrt{n}$-consistent of $\hat\beta_n$ to β_a under H_a and (C4), we can write U_{n2} as the sum of the following two terms and a remainder of order $o_p(1)$:

$$U_{n21} = a_n \sum_{i \neq j \neq k} K_h(i,j) K_h(i,k) (\Delta m_j)^2 \Delta v_k \sqrt{v_a(X_i)} \varepsilon_k^2 \hat{f}^{-1}(X_i),$$

$$U_{n22} = a_n \sum_{i \neq j \neq k} K_h(i,j) K_h(i,k) (\Delta v_j)^2 \Delta v_k \sqrt{v_a(X_i)} \varepsilon_j^2 \varepsilon_k^2 \hat{f}^{-1}(X_i),$$

one can show that U_{n2} converges to

$$4E[(m(X;\beta_0)-m(X;\beta_a))^2(\sqrt{v_1(X)}-\sqrt{v_a(X)})+(\sqrt{v_1(X)}-\sqrt{v_a(X)})^3]\sqrt{v_a(X)}f(X)$$

in probability.

Using the condition (C4) and the $\sqrt{n}$-consistency of $\hat{\beta}_n$ to β_a, we can also write U_{n4} as the sum of the following term and a remainder of order $o_p(1)$:

$$U_{n41} = a_n \sum_{i \neq j \neq k} K_h(i,j) K_h(i,k) \Delta v_j \Delta v_k \sqrt{v_a(X_j)} \sqrt{v_a(X_k)} \varepsilon_j^2 \varepsilon_k^2 \hat{f}^{-1}(X_i).$$

We can show that

$$U_{n4} \to 4E[(\sqrt{v_1(X)}-\sqrt{v_a(X)})^2 v_a(X)]f(X).$$

Using the same methods used in null case, one can show that U_{n3}, U_{n5}, and U_{n6} are all of the order $o_p(1)$. After doing some algebraic manipulations, we can show that

$$Z_n \to E[(m(X;\beta_0)-m(X;\beta_a))^2+(v_1(X)-v_a(X))]^2 f(X)$$

in probability.

Finally, similar to Lemma 3.4 in Zheng (1996) and Theorem 3.1 in Song and Du (2011), we have

$$\hat{\sigma}^2 \to \int \left[\int K(u+v)K(v)dv\right]^2 du. \int [\tau^2(x;\beta_a,\theta_a)+(m(x;\beta_0)-m(x;\beta_a))^2+(v_1(x)-v(x;\beta_a,\theta_a))]^2 f^2(x)\, dx$$

in probability. □

Proof of Theorem 3.3. The proof is similar to those in the proofs of Theorems 3.1 and 3.2, the details are omitted here for the sake of brevity. □

References

Breusch, T.S., Pagan, A.R., 1979. A simple test for heteroscedasticity and random coefficient variation. Econometrica 47, 1287–1294.
Dette, H., Munk, A., 1998. Testing heteroscedasticity in nonparametric regression. Journal of the Royal Statistical Society Series B 60, 693–708.
Dette, H., 2002. A consistent test for heteroscedasticity in nonparametric regression based on the kernel method. Journal of Statistical Planning and Inference 103, 311–330.
Dette, H., Hetzler, B., 2009a. A simple test for the parametric form of the variance function in nonparametric regression. Annals of the Institute of Statistical Mathematics 61, 861–886.
Dette, H., Hetzler, B., 2009b. Khmaladze transformation of integrated variance processes with applications to goodness-of-fit testing. Mathematical Methods of Statistics 18, 97–116.
Dette, H., Marchlewski, M., 2010. A robust test for homoscedasticity in nonparametric regression. Journal of Nonparametric Statistics 22, 723–736.
Dette, H., Neumeyer, N., Keilegom, I., 2007. A new test for the parametric form of the variance function in non-parametric regression. Journal of the Royal Statistical Society Series B 69, 903–917.
Diblasi, A., Bowman, A., 1997. Testing for constant variance in a linear models. Statistics and Probability Letters 33, 95–103.
Harrison, M.J., Mccabe, B.P.M., 1979. A test of heteroscedasticity based on least squares residuals. Journal of the American Statistical Association 74, 494–500.
Jennrich, R.I., 1969. Asymptotic properties of non-linear least squares estimators. Annals of Mathematical Statistics 40, 633–643.
Koenker, R., Bassett, G., 1981. Robust tests for heteroscedasticity based on regression quantiles. Econometrica 50, 43–61.
Koul, H.L., Ni, P., 2004. Minimum distance regression model checking. Journal of Statistical Planning and Inference 119, 109–141.
Liero, H., 2003. Testing homoscedasticity in nonparametric regression. Journal of Nonparametric Statistics 15, 31–51.
Samarakoon, N., Song, W., 2011. Minimum distance conditional variance function checking in heteroscedastic regression models. Journal of Multivariate Analysis 102, 579–600.
Song, W., Du, J., 2011. A note on testing the functional forms for regression functions via nonparametric smoothing. Canadian Journal of Statistics 39, 108–125.
Wang, L., Zhou, X., 2006. Assessing the adequacy of variance function in heteroscedastic regression models. Biometrics 63, 1218–1225.
White, H., 1980. A heteroscedasticity-consistent covariance matrix estimator and a direct test of heteroscedasticity. Econometrica 48, 817–838.
White, H., 1981. Consequences and detection of misspecified nonlinear regression models. Journal of the American Statistical Association 76, 419–433.
White, H., 1982. Maximum likelihood estimation of misspecified models. Econometrica 50, 1–25.
You, J., Chen, G., 2005. Testing heteroscedasticity in partially linear regression models. Statistics and Probability Letters 73, 61–70.
Zheng, J.X., 1996. A consistent test of functional form via nonparametric estimation techniques. Journal of Econometrics 75, 263–289.
Zhu, L., Fujikoshi, Y., Naito, K., 2001. Heteroscedasticity checks for regression models. Science in China Series A: Mathematics 44, 1236–1252.