Journal of Statistical Planning and Inference 140 (2010) 1056–1072

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Conditional variance model checking

Hira L. Koul [a,b,1], Weixing Song [a,b,*,2]

[a] Michigan State University, USA
[b] Department of Statistics, Kansas State University, Manhattan, KS66502 USA

ARTICLE INFO

Article history:
Received 22 December 2008
Received in revised form
19 October 2009
Accepted 21 October 2009
Available online 31 October 2009

MSC:
primary 62G08
secondary 62G10

Keywords:
Heteroscedasticity
Lack-of-fit test
Martingale transformation
Brownian motion

ABSTRACT

This paper discusses the problem of fitting a parametric model to the conditional variance function in a class of heteroscedastic regression models. The proposed test is based on the supremum of the Khmaladze type martingale transformation of a certain partial sum process of calibrated squared residuals. Asymptotic null distribution of this transformed process is shown to be the same as that of a time transformed standard Brownian motion. Test is shown to be consistent against a large class of fixed alternatives and to have nontrivial asymptotic power against a class of nonparametric $n^{-1/2}$—local alternatives, where n is the sample size. Simulation studies are conducted to assess the finite sample performance of the proposed test and to make a finite sample comparison with an existing test.

1. Introduction

Regression analysis frequently assumes errors to be homoscedastic, while real data generated from applications often exhibit certain heteroscedastic structures in the sense that the conditional variance of the error, given the design variable, is not a constant. Inference procedures that are optimal under homoscedastic error structure often fail to be so in the presence of heteroscedasticity. Assuming errors are heteroscedastic, the focus of this paper is to fit a family of parametric models to the conditional error variance.

Early work in this area includes some graphical procedures and some formal tests. Most of them are based on the residuals obtained by fitting a model with a completely specified parametric regression and variance functions, see Harrison and McCabe (1979), Breusch and Pagan (1979), White (1980), Koenker and Bassett (1981), Cook and Weisberg (1983), Carroll and Ruppert (1988), and Diblasi and Bowman (1997) and the references therein. Most of these works focus on checking whether heteroscedasticity is present in the regression model or not, i.e., to check whether the variance function is constant or not. When covariate is one dimensional, Dette and Munk (1998) propose a consistent test based on the best L^2 approximation of the variance function by a constant. Inspired by the idea that the problem of testing for heteroscedasticity is equivalent to the problem of testing for pseudo-residuals having a constant mean, Dette (2002) constructs a testing procedure which can detect $1/\sqrt{nh^{1/2}}$ local alternative, where n is the sample size and h is the

* Corresponding author at: Department of Statistics, Kansas State University, Manhattan, KS66502 USA.
E-mail address: weixing@ksu.edu (W. Song).

[1] Research of this author in part supported by the NSF Grant DMS 0704130.
[2] Research of this author is supported by USRG No. 2232 of Kansas State University.

0378-3758/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2009.10.008

bandwidth in the kernel smoothing. Liero (2003) suggests a test statistic using an L_2 distance between the two nonparametric variance estimators constructed under null and alternative models. In the case of multi-dimensional covariates, a Cramér–von Mises type test based on estimated residual empirical process is proposed by Zhu et al. (2001). Asymptotic null distributions of the aforementioned test statistics often depend on the null model being fitted and error and covariate distributions in a complicated fashion. Some re-sampling techniques are then employed to implement these tests, even for the large samples.

Relatively, few procedures for testing the adequacy of a given nonconstant variance function are available in the literature. In the case of one dimensional covariate, Dette et al. (2007) propose a Kolmogorov–Smirnov and a Cramér–von Mises type tests constructed from a stochastic process based on the difference between the empirical processes that are obtained from the standardized nonparametric residuals under the null and the alternative hypothesis. Bootstrap distribution is used to implement the test because of the intractability of its asymptotic null distribution. In some one dimensional nonrandom design cases, Dette and Hetzler (2009) propose a test based on a partial sum process of squares of pseudo-residuals whose null limiting distribution is complicated except when testing for homoscedasticity. Their test has nontrivial asymptotic power against some $n^{-1/2}-$local alternatives. In the case of d-dimensional covariate, $d \geq 1$, Wang and Zhou (2007) propose a kernel type nonparametric test based on the framework of Zheng (1996). They discuss consistency and asymptotic power of their test which can only detect $(nh^{d/2})^{-1/2}$ local alternatives.

In this paper, we propose a new testing procedure to assess the adequacy of fitting the conditional variance function with a possibly nonconstant parametric function in heteroscedastic regression models with one dimensional random design variable. The proposed test is based on Khmaladze type martingale transformation of a marked empirical process of suitably standardized squared residuals. Under the null hypothesis, this process converges weakly to a time transformed Brownian motion in uniform metric. Consequently, any test based on a continuous functional of this process will be asymptotically distribution free. We investigate in some detail the test based on the supremum of this process. This test is shown to be consistent against a large class of fixed nonparametric alternatives, and to have nontrivial asymptotic power against $n^{-1/2}-$local nonparametric alternatives. Finally, the test procedure is easy to program and the computation of the test statistic is very fast.

The paper is organized as follows. Various model assumptions, the testing procedure and the main results appear in Section 2. Section 3 discusses consistency and asymptotic power properties of the proposed test. In these sections we assume that regression function is of a known parametric form. A modification of the proposed test in the case of nonparametric regression function is discussed in Section 4, while Section 5 contains a simulation study to illustrate the finite sample behavior of the proposed test and a finite sample power comparison with the test of Wang and Zhou (2007). Section 6 contains some proofs.

2. Main results

Consider the heteroscedastic regression model

$$Y = \mu(X) + \tau(X)\varepsilon, \tag{2.1}$$

where X, Y are one-dimensional explanatory and response variables, respectively, with X taking values in a compact interval of $\mathbb{R}$, which we take to be $\mathcal{I} := [0,1]$, without loss of generality. The function $\mu(\cdot)$ is a real valued function and $\tau(\cdot)$ is a positive function. The random error ε is assumed to satisfy

(e) $E(\varepsilon|X) = 0$, $E(\varepsilon^2|X) = 1$, $E(\varepsilon^8) < \infty$ and $\mu_4(x) := E((\varepsilon^2-1)^2|X=x) \geq d$ for all $x \in \mathcal{I}$ and some positive constant d.

If $\mu_4(x)$ does not depend on the design variable x, then $E\varepsilon^4 < \infty$ will suffice for the results of this paper to hold.

Under (e), $\mu(x) \equiv E(Y|X=x)$, $\tau^2(x) \equiv \mathrm{Var}(Y|X=x)$. Many existing efficient statistical inference procedures are particularly developed for homoscedastic models where $\tau(x) \equiv c$, a constant. But real data sets often show some heteroscedasticity. In this case, by experience, or by exploring the data set, one can try to fit the variance function with a parametric function. This reasoning is very attractive in practice, since there are many ways to incorporate a specified variance function into inference procedures without loss of much efficiency. Accordingly, Γ be a relatively compact subset of $\mathbb{R}^p$ and $\mathcal{V} := \{v(x;\beta), x \in \mathcal{I}, \beta \in \Gamma\}$ be a parametric family of positive functions, where p is a known positive integer. The problem of interest is to test the hypothesis that $\tau \in \mathcal{V}$, i.e., to test for

$$H_0 : \tau(x) = v(x;\beta_0) \quad \text{for all } x \in \mathcal{I}, \text{ some } \beta_0 \in \Gamma, \tag{2.2}$$

versus the alternative $H_1 : H_0$ is not true. Any test for fitting a parametric model to $\tau(x)$ is likely to be affected, to some extent, by the knowledge about the regression function $\mu(x)$. For the time being we shall assume μ to be parametric. Accordingly, let $\{m(x;\beta), x \in \mathcal{I}, \beta \in \Gamma\}$ be a parametric family of real valued functions, and assume additionally that in (2.1), $\mu(x) = m(x;\beta_0)$, for all x. The case of nonparametric μ is discussed later.

To describe the test statistic, let $\tilde{Y} := (Y - m(X;\beta_0))^2$. Under (2.1) and (e), $E[\tilde{Y}|X] = E(\tau^2(X)\varepsilon^2|X) = \tau^2(X)$. If β_0 is known, we may consider the new regression model

$$\tilde{Y} = \tau^2(X) + \xi, \tag{2.3}$$

where ξ satisfies $E(\xi|X)=0$, so that ξ and X are uncorrelated, and

$$\sigma^2(x) := \mathrm{Var}(\xi|X=x) = \tau^4(x)\mu_4(x). \tag{2.4}$$

Moreover, now testing for H_0 is equivalent to fitting the parametric regression model $\nu^2(x;\beta)$ to the regression function $\tau^2(x)$ in the model (2.3) based on i.i.d. observations $\{(X_i,\tilde{Y}_i), 1\le i\le n\}$ on $(X,\tilde{Y})$. Further, if $\mu_4(x)$ is also known, then tests for H_0 can be based on the partial sum residual process

$$V_n(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\tilde{Y}_i - \nu^2(X_i;\beta_0)}{\sigma(X_i;\beta_0)} I[X_i\le x], \quad x\in\mathcal{I},$$

where $\sigma(X_i;\beta_0)=\nu^2(X_i;\beta_0)\sqrt{\mu_4(X_i)}$. If β_0 and $\mu_4(x)$ are unknown, one uses $\widehat{V}_n$, the modified V_n where β_0 is replaced by an $n^{1/2}$−consistent estimator, and $\mu_4(x)$ is replaced by a consistent nonparametric estimator, such as a Nadaraya–Watson estimator or a local linear estimator. The process V_n is an analog of the partial sum residual empirical process used by von Neumann (1941) to test for the constancy of the regression function, and later by Stute et al. (1998) to construct asymptotically distribution free tests of lack-of-fit of a parametric regression function.

Let F denote the distribution function (d.f.) of X having density f. Under some regularity conditions on the null model, we can show that under H_0, $\widehat{V}_n$ converges weakly, in $D(\mathcal{I})$ and uniform metric, to a continuous centered Gaussian process with a covariance function depending on the null model and F in a complicated way. Hence any test based on this process is practically not easy to implement even asymptotically. It is well known that similar phenomena occur in many other model checking problems. To overcome this difficulty when fitting a regression function, some authors resort to the bootstrap methodology, e.g., Stute and Zhu (2002), Zhu and Ng (2003), Liang (2006), among others, while some authors construct asymptotically distribution free tests through the Khmaladze type transform, cf., Khmaladze (1981, 1988, 1993), of the process $\widehat{V}_n$ as is done in Stute et al. (1998), Khmaladze and Koul (2004), Koul (2006), Koul and Song (2008) and the references therein. The transformed test statistic not only possesses a completely known asymptotic null distribution, but is also easy to implement.

A referee pointed out to us the paper by Dette and Hetzler (2008) which also uses analogous transformation of a partial sum process of squares of certain pseudo-residuals to obtain asymptotically distribution free tests for fitting a parametric function to $\sigma(t)$ in the model $Y_{ni}=\mu(t_{ni})+\sigma(t_{ni})\varepsilon_{ni}$, where t_{ni} is the $i/(n+1)$th quantile of a distribution on $[0,1]$. They also assume finite 8th error moment.

To proceed further, we shall assume the functions m and ν are differentiable with respect to β. Let $\nu_0(x):=\nu(x;\beta_0)$, $\dot{\nu}_0(x):=\partial\nu(x;\beta)/\partial\beta|_{\beta=\beta_0}$, $\sigma_0(x):=\sigma(x;\beta_0)$, $e:=\xi/\sigma_0(X)$ and

$$l(x) := \frac{2\dot{\nu}_0(x)}{\nu_0(x)\sqrt{\mu_4(x)}}, \quad M_s := El(X)l'(X)I[X\ge s]. \tag{2.5}$$

Note that M_s is a $p\times p$ matrix depending on the unknown parameter β_0 and the design d.f. F. For some estimator $\widehat{\beta}_n$ of β_0, a kernel function K and a bandwidth h, let $m_n(x):=m(x,\widehat{\beta}_n)$, $\nu_n(x):=\nu(x;\widehat{\beta}_n)$,

$$\widehat{Y}_i := (Y_i - m_n(X_i))^2, \quad \widehat{\varepsilon}_i = \frac{Y_i-m_n(X_i)}{\nu_n(X_i)}, \quad \widehat{\xi}_i = \widehat{Y}_i - \nu_n^2(X_i),$$

$$\widehat{\mu}_4(x) = \frac{\sum_{j=1}^n K((x-X_j)/h)(\widehat{\varepsilon}_j^2-1)^2}{\sum_{j=1}^n K((x-X_j)/h)}, \quad \widehat{\sigma}(x) = \nu_n^2(x)\sqrt{\widehat{\mu}_4(x)}, \quad \widehat{e}_i = \widehat{\xi}_i/\widehat{\sigma}(X_i),$$

$$\widehat{l}(x) = \frac{2\dot{\nu}_n(x)}{\nu_n(x)\sqrt{\widehat{\mu}_4(x)}}, \quad \widehat{M}_s = n^{-1}\sum_{i=1}^n \widehat{l}(X_i)\widehat{l}'(X_i)I(X_i\ge s). \tag{2.6}$$

Assume $n\ge p$, and

(M) M_s is positive definite matrix for all $s\in[0,1)$.

Then, we can rewrite

$$\widehat{V}_n(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \widehat{e}_i I[X_i\le x], \quad x\in\mathcal{I}.$$

Its Khmaladze type martingale transformation is given by

$$\widehat{W}_n(x) = \widehat{V}_n(x) - \int_{s\le x}\widehat{l}'(s)\widehat{M}_s^{-1}\int_{z\ge s}\widehat{l}(z)\,d\widehat{V}_n(z)\,d\widehat{F}(s), \tag{2.7}$$

where $\widehat{F}$ is the empirical d.f. of X_i's. The proposed tests of H_0 are to be based on $\widehat{W}_n$. We need the following additional assumptions to establish its null weak limit.

(k) The kernel function K is supported on $[-1,1]$, symmetric about 0, and bounded.

(h) The bandwidth $h \sim n^{-a}$ such that $\frac{1}{4} < a < \frac{1}{2}$.

(f) The density function f is Lipschitz continuous on $\mathcal{I}$ and $0 < c_1 \le \inf_{x \in \mathcal{I}} f(x) \le \sup_{x \in \mathcal{I}} f(x) \le c_2 < \infty$ for some constants c_1 and c_2.

(m1) $\mu_4(x)$ is Lipschitz continuous on $\mathcal{I}$.

(m2) For all $x \in \mathcal{I}$, $m(x;\beta)$ is differentiable in β in a neighborhood of β_0 with the derivative $\dot{m}(x;\beta_0)$ such that $E|m(X;\beta_0|^2$, $E\|\dot{m}(X;\beta_0)\|^4$ are finite and for every $k < \infty$, $\sup_{1 \le i \le n, \sqrt{n}\|\beta-\beta_0\| < k} \sqrt{n}|m(X_i;\beta) - m(X_i;\beta_0) - (\beta-\beta_0)'\dot{m}(X_i;\beta_0)| = o_p(1)$.

(m3) For every $x \in \mathcal{I}$, there exist a $p \times p$ square matrix $\ddot{m}(x;\beta_0)$, having finite expectation, and a nonnegative function $h(x;\beta_0)$, with $Eh(X,\beta_0) < \infty$ and satisfying the following: $\forall \delta > 0$, $\exists \zeta > 0$ such that $\|\beta - \beta_0\| \le \zeta$ implies

$$\|\dot{m}(x;\beta) - \dot{m}(x;\beta_0) - \ddot{m}(x;\beta_0)(\beta-\beta_0)\| \le \delta h(x;\beta_0)\|\beta-\beta_0\|, \quad \forall x.$$

(v1) For every $x \in \mathcal{I}$, $v(x;\beta_0)$ is bounded below from 0; $v(x;\beta)$ is differentiable in β in a neighborhood of β_0 with the derivative $\dot{v}(x;\beta)$, such that $Ev^2(X;\beta_0)$, $E\|\dot{v}(X;\beta_0)\|^4$ are finite and for every $k < \infty$,

$$\sup \sqrt{n}|v(X_i;\beta) - v(X_i;\beta_0) - (\beta-\beta_0)'\dot{v}(X_i;\beta_0)| = o_p(1),$$

where the sup is taken over $\{1 \le i \le n, \sqrt{n}\|\beta - \beta_0\| \le k\}$.

(v2) For each $x \in \mathcal{I}$, there exist a $p \times p$ square matrix $\ddot{v}(x;\beta_0)$, having finite expectation, and a nonnegative function $g(x;\beta_0)$, with $Eg(X,\beta_0) < \infty$ and satisfying the following: $\forall \delta > 0$, $\exists \eta > 0$ such that $\|\beta - \beta_0\| \le \eta$ implies

$$\|\dot{v}(x;\beta) - \dot{v}(x;\beta_0) - \ddot{v}(x;\beta_0)(\beta-\beta_0)\| \le \delta g(x;\beta_0)\|\beta-\beta_0\|.$$

(ℓ) $\int \|l'(x)M_x^{-1}\| \, dF(x) < \infty$.

Conditions (k), (h), (m1) are required to guarantee that the nonparametric estimator $\hat{\mu}_4(x)$ has the desired large sample properties. Condition (f) is often used in nonparametric regression. Conditions (m1)–(m3), (v1), (v2) are concerned with the smoothness of the regression and variance functions; (ℓ) is a technical condition to ensure tightness of a sequence of certain stochastic processes. The following theorem gives the needed weak convergence result, where B is standard Brownian motion.

Theorem 2.1. *Suppose the conditions* (e), (k), (h), (f), (M), (m1)–(m3), (v1), (v2) *and* (ℓ) *hold. In addition, suppose the estimator $\hat{\beta}_n$ satisfies, under H_0,*

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = O_p(1). \tag{2.8}$$

Then, for every $x_0 < 1$, $\widehat{W}_n(x) \Longrightarrow B \circ F(x)$, in $D([0,x_0])$ and uniform metric.

A computational formula for calculating $\widehat{W}_n$ is as follows. Let $X_{(i)}$ denote i th order statistics among $X_i, 1 \le i \le n$, and $\hat{e}_{(i)}$, $\hat{l}_{(i)}$'s denote the permutated $\hat{e}_i$, $\hat{l}_i$'s with respect to the ordered X_i's. Then, with $X_{(n+1)} = 1$, $\widehat{W}_n(x) = 0$, for $x < X_{(1)}$, and

$$\widehat{W}_n(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{k} \hat{e}_{(i)} \left[1 - \frac{1}{n} \sum_{j=1}^{i} \hat{l}_{(j)}' \widehat{M}_{(j)}^{-1} \hat{l}_{(i)} \right], \quad X_{(k)} \le x < X_{(k+1)}, \ k = 1, \ldots, n,$$

where

$$\hat{l}_{(j)} = \frac{2\dot{v}_n(X_{(j)})}{v_n(X_{(j)})\sqrt{\hat{\mu}_4(X_{(j)})}}, \quad \widehat{M}_{(j)} = n^{-1} \sum_{i=j}^{n} \hat{l}_{(i)} \hat{l}_{(i)}', \ j = 1, \ldots, n.$$

As in Stute et al. (1998), it is recommended to take x_0 to be the 99% quantile of the empirical d.f. $\hat{F}$. Then from Theorem 2.1,

$$\sup_{0 \le x \le x_0} \left| \frac{\widehat{W}_n(x)}{\{\hat{F}(x_0)\}^{1/2}} \right| = \sup_{0 \le x \le x_0} \left| \frac{\widehat{W}_n(x)}{0.995} \right| \longrightarrow_D \sup_{0 \le x \le x_0} \left| B\left(\frac{F(x)}{F(x_0)} \right) \right| = \sup_{0 \le u \le 1} |B(u)|.$$

The last equality is understood as equal in distribution. Consequently, the test that rejects H_0 whenever $\sup_{0 \le x \le x_0} |\widehat{W}_n(x)|/0.995 > b_\alpha$ is of the asymptotic size α, where b_α is such that $P(\sup_{0 \le u \le 1} |B(u)| > b_\alpha) = \alpha$.

Remark 2.1. The implementation of this test requires $\sqrt{n}$–consistent estimator $\hat{\beta}_n$ under the null hypothesis. Such estimators may be found by least squares or pseudo-likelihood methods, among others, under certain smoothness conditions on the regression function and the variance function. Discussion on the least square estimators can be found in Jennrich (1969), Wu (1981), Amemiya (1985), and the references therein. Carroll and Ruppert (1988) provide a comprehensive accounts for estimating the parameters in heteroscedastic regression models, and the pseudo-likelihood method has especially been proven to be simple and effective. See also Wang and Zhou (2007) for a brief discussion on this method. In this paper, we will presume the existence of such estimators without providing the exact procedure and formulae for them.

3. Consistency and local power

In this section, we shall show that, under some regularity conditions, the above test is consistent for certain fixed alternatives and has nontrivial asymptotic power against a large class of $1/\sqrt{n}-$local alternatives.

3.1. Consistency

Let $\nu_1 \notin \mathcal{V}$ be a known positive function defined on $\mathcal{I}$. Consider the alternative

$$H_a : \tau(x) = \nu_1(x), \quad \forall x \in \mathcal{I}. \tag{3.1}$$

We continue to assume that the regression function is of the parametric form $m(x;\beta)$ and that the estimator $\widehat{\beta}_n$ used in the test statistic now satisfies $\sqrt{n}(\widehat{\beta}_n - \beta_a) = O_p(1)$, for some $\beta_a \in \Gamma$, under the alternative H_a. We still use β_0 to denote the true value of β under H_a.

Let $Y_i^a = m(X_i;\beta_a) + \nu(X_i;\beta_a)\varepsilon_i$. For the sake of brevity, let $\nabla m(x) = m(x;\beta_0) - m(x;\beta_a)$, $\nabla\nu(x) = \nu_1(x) - \nu(x;\beta_a)$, A_s be same as M_s in the null case except that β_0 is replaced by β_a, and

$$\phi(x) := 2E\left[\frac{\nabla\nu(X)\dot{\nu}(X;\beta_a)I[X \ge x]}{\nu^2(X;\beta_a)\mu_4(X)}\right], \quad H(x) = [(\nabla m(x))^2 + (\nabla\nu(x))^2]\dot{\nu}(x;\beta_a),$$

$$\rho(x) := 2E\left[\frac{H(x)I[X \ge x]}{\nu^3(X;\beta_a)\mu_4(X)}\right],$$

$$K(x) := 2E\left(\frac{\nabla\nu(X)I[X \le x]}{\nu(X;\beta_a)\mu_4(X)}\right) - 2\int_{s \le x} l(s)'A_s^{-1}\phi(s)\,dF(s),$$

$$D_1(x) := E\left[\frac{(\nabla m(X))^2 + (\nabla\nu(X))^2}{\nu^2(X;\beta_a)\sqrt{\mu_4(X)}}I[X \le x]\right], \quad D_2(x) := \int_{s \le x} l'(s)A_s^{-1}\rho(s)\,dF(s),$$

$$\mathcal{D}(x) := D_1(x) - D_2(x). \tag{3.2}$$

Now assume that

(a1) The conditions (m1)–(m3), (v1), (v2), and (ℓ) hold with β_0 replaced with β_a.
(a2) $Em^2(X,\beta_0)$ and $E\nu_1^2(X)$ are finite.
(a3) $d = \sup_{0 \le x \le x_0} |K(x) + \mathcal{D}(x)| > 0$.

Then we have the following consistency result of the supremum test.

Theorem 3.1. *Suppose the conditions* (e), (k), (h), (f), (m1) *and* (a1)–(a3) *hold. Then, for every* $x_0 < 1$, *the test that rejects* H_0 *whenever* $\sup_{0 \le x \le x_0} |\widehat{W}_n(x)/\{\widehat{F}(x_0)\}^{1/2}| > b_\alpha$ *is consistent for* H_a.

Remark. Although the term $\nabla m(x)$ appears in $K(x) + \mathcal{D}(x)$, but it is indeed a direct consequence of misspecification of the variance function. Since we assume the regression function has the parametric form of $m(x;\beta)$ in advance, so any estimator of β based on the null model may not be a consistent estimator of the true parameter under H_a. Therefore, condition (a3) actually requires that ν_0 and the alternative ν_1 should be different on $x \le x_0$.

3.2. Local power

Let $\delta(x)$ be a measurable function such that $E\delta^2(X) < \infty$ and $\delta \notin \mathcal{V}$. Now we shall study asymptotic power of the proposed test against the local alternatives

$$H_{\rm loc} : \tau(x) = \nu_0(x) + n^{-1/2}\delta(x), \quad x \in \mathcal{I}. \tag{3.3}$$

Let $m_0(x) := m(x;\beta_0)$. Under $H_{\rm loc}$, the regression model (2.1) becomes

$$Y_i = m_0(X_i) + [\nu_0(X_i) + n^{-1/2}\delta(X_i)]\varepsilon_i.$$

We shall still assume that the estimator $\widehat{\beta}_n$ used in the test statistic satisfy the $\sqrt{n}-$consistency under $H_{\rm loc}$: $\sqrt{n}(\widehat{\beta}_n - \beta_0) = O_p(1)$. Define

$$\psi(y) = E\frac{\delta(X)l(X)I[X \ge y]}{\nu_0(X)\sqrt{\mu_4(X)}},$$

$$D(x) = 2E\frac{\delta(X)I[X\le x]}{\nu_0(X)\sqrt{\mu_4(X)}} - 2\int_{y\le x} l'(y)M_y^{-1}\psi(y)\,dF(y). \tag{3.4}$$

Then we have the following theorem.

Theorem 3.2. *Suppose the assumptions of Theorem* 2.1 *hold. Then under the local alternative* (3.3), *for any* $x_0 < 1$,

$$\lim_{n\to\infty} P\left(\sup_{0\le x\le x_0}\left|\frac{\widehat{W}_n(x)}{\{\widehat{F}(x_0)\}^{1/2}}\right| > b_\alpha\right) = P\left(\sup_{0\le x\le x_0}\frac{|B\circ F(x)+D(x)|}{\{F(x_0)\}^{1/2}}\ge b_\alpha\right).$$

4. Case of nonparametric $\mu(x)$

In the above discussion we had assumed that the regression function has a known parametric form. To broaden the applicability of the proposed inference procedure it is desirable to relax this assumption when checking the adequacy of the variance function. We shall now assume μ to be purely a nonparametric function. As pointed out by Wang and Zhou (2007), the main advantage of considering the regression model with unknown regression function is to avoid the likely adverse effects of a misspecified parametric regression function on checking the adequacy of the variance function.

The above test procedure is modified as follows in this situation. To begin with, we need a nonparametric estimator for $\mu(x)$. For convenience, we shall use the Nadaraya–Watson estimator. The main results of this paper continue to hold with any other estimator of μ that satisfies the uniform consistency property (6.5) given below, in particular, for the local linear estimator.

Let $h=h_n$ be a bandwidth sequence, and K be a density kernel. The Nadaraya–Watson kernel estimator of $\mu(x)$ is $\widehat{\mu}(x) = \sum_{i=1}^n K((x-X_i)/h)Y_i/\sum_{i=1}^n K((x-X_i)/h)$.

Now we will assume

(μ) The regression function $\mu(x)$ is Lipschitz continuous on $\mathcal{I}$, i.e., $|\mu(x+u)-\mu(x)|\le c|u|$, for some constant $c<\infty$ and all $u\in\mathbb{R}, x\in\mathcal{I}$.

Replace $m(x;\widehat{\beta})$ by $\widehat{\mu}(x)$ in the statistic $\widehat{W}_n$. The weak convergence of this modified statistic $\widehat{W}_n$ is given in the following:

Theorem 4.1. *Suppose conditions* (e), (k), (h), (f), (m1), (v1), (v2), (l), (μ) *and* (2.8) *holds. Then, under* H_0, *for every* $x_0<1$, $\widehat{W}_n(x)\Longrightarrow B\circ F(x)$, *in* $D([0,x_0])$ *and uniform metric.*

5. Numerical simulations

Simulation studies are conducted in this section to evaluate the finite sample performance of the proposed test procedure. We choose x_0 to be the 99th percentile of the empirical distribution function $\widehat{F}$. For the significance levels $\alpha = 0.05, 0.025, 0.01$, the critical values b_α obtained from the distribution of $\sup_{0\le u\le 1}|B(u)|$ are 2.24241, 2.49771, 2.80705, respectively. See, e.g., Khmaldze and Koul (2004). In the following simulation, for various sample size, we repeated the above procedure 1000 times and the empirical size and power are computed by using $\#\{\sup_{0\le x\le x_0}|\widehat{W}_n(x)/\{\widehat{F}(x_0)\}^{1/2}|\ge b_\alpha\}/1000$.

Simulation 1: We generate samples from the following four models:

Model 0: $Y_i = \beta_1+\beta_2X_i+\sqrt{\beta_3+\beta_4X_i}\varepsilon_i$,

Model 1: $Y_i = \beta_1+\beta_2X_i+\sqrt{\beta_3+\beta_4X_i+3X_i^2}\varepsilon_i$,

Model 2: $Y_i = \beta_1+\beta_2X_i+\sqrt{\beta_3+\beta_4X_i+1.5X_i^2}\varepsilon_i$,

Model 3: $Y_i = \beta_1+\beta_2X_i+\sqrt{\beta_3+\beta_4X_i+0.5X_i^2}\varepsilon_i$,

where $X\sim U[-1,1]$. The null model to be fitted is $\nu(x;\beta)=\sqrt{\beta_3+\beta_4x}$. The true parameters are chosen to be $\beta_1=1,\ \beta_2=2$, $\beta_3=0.5$ and $\beta_4=0.25$. Note that for these values of β_3,β_4, $\nu(x;\beta)=\beta_3+\beta_4x>0$, for all $-1\le x\le 1$. Data from model 0 are used to study the empirical level, while from models 1 to 3 are used to study the empirical power of the test. Note that model 1 has a greater departure from the null model 0 than model 2, and model 3 has the smallest departure from the null model. In the simulation, the random errors are chosen from standard normal and standard double exponential distributions. Since the simulation results are similar for both cases, only results from normal random errors are reported here.

As a comparison, we also simulated the test of Wang and Zhou (2007) with the asymptotic variance replaced by a consistent estimator. We compared the two tests at the nominal level of 0.05. Since asymptotic null distribution of this statistic is N(0, 1), we used 1.96 as the critical value for this test. The β's were estimated by the least squares method. According to the simulation study, both tests are conservative (empirical level around $0.01-0.02$). This is not surprising in

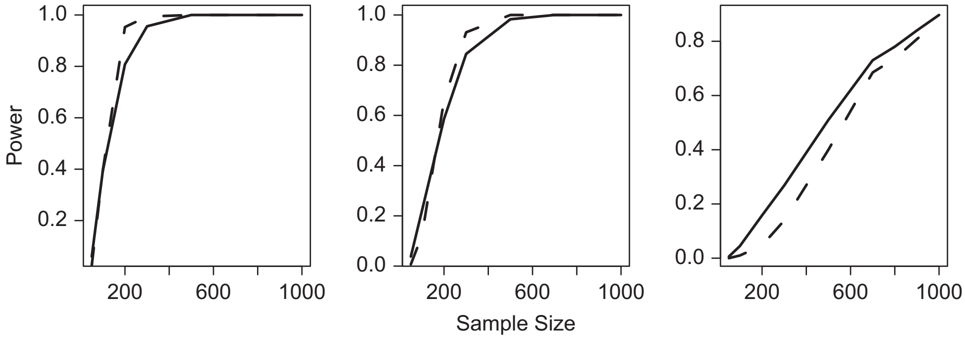

Fig. 1. Power curves for the three alternative models.

Table 1
Monte Carlo level and power of $\widehat{W}_n$ test.

α−level	Model\\n	50	100	200	300	500
0.05	Model 0	0.019	0.034	0.023	0.032	0.033
	Model 1	0.011	0.028	0.168	0.558	0.972
	Model 2	0.342	0.701	0.922	0.987	1.000
0.025	Model 0	0.012	0.023	0.007	0.012	0.014
	Model 1	0.005	0.008	0.058	0.281	0.838
	Model 2	0.254	0.625	0.867	0.972	0.999
0.01	Model 0	0.003	0.006	0.002	0.006	0.008
	Model 1	0.002	0.002	0.009	0.083	0.530
	Model 2	0.168	0.526	0.823	0.938	0.995

that the nonparametric smoothing test statistic often provides less accurate approximation to the asymptotic theory for small to moderate sample sizes.

Figure 1 gives empirical power curves for the three alternative models. In each graph, the solid line corresponds to the power of our test and the dashed line to the power of the modified Wang and Zhou's test. One can see that when the alternative models are farther away from the null model, our test is comparable to Wang and Zhou's test, but our test behaves better when the sample size is small. When the departure is small as in model 3, our test is simply uniformly better than Wang and Zhou's test over all the selected sample sizes. This finding supports the theoretical fact that our test can detect smaller departure ($1/\sqrt{n}$ local alternatives) than Wang and Zhou's test ($1/\sqrt{nh^{1/2}}$ local alternatives).

Simulation 2: A simulation on a nonlinear variance function is conducted here. The data sets are generated from the following regression models:

Model 0: $Y_i = \beta_1 + \beta_2 X_i + \exp(\beta_3 X_i)\varepsilon_i$,
Model 1: $Y_i = \beta_1 + \beta_2 X_i + \exp(\beta_3 X_i - 6(X_i - 0.5)^2)\varepsilon_i$,
Model 2: $Y_i = \beta_1 + \beta_2 X_i + \exp(\beta_3 X_i - 1.5(\sin(2\pi X_i))^2)\varepsilon_i$.

Thus, here the null hypothesis is $H_0 : v(x;\beta) = \exp(\beta_3 x)$. Data from model 0 are used to study the empirical level, while from models 1 to 2 are used to study the empirical power of the test. In the simulation, $\varepsilon \sim N(0,1)$, $X \sim \text{Uniform}(0,1)$, and $\beta_1 = 1, \beta_2 = 2$, $\beta_3 = -0.25$. Table 1 illustrates the simulation results.

The simulation study shows that the empirical levels are all less than the nominal levels in all the chosen cases, hence the proposed test is conservative for all chosen sample sizes. It has a small power against the alternative model 1 when sample sizes are small, but the power improves as the sample size increases. The overall performance of the test against the alternative model 2 is satisfactory, as it has good power (at least 80%) for reasonable sample sizes (200 or above). This is not surprising because the variance function of the alternative 1 is closer than that of the alternative 2 to the variance function of the null model.

6. Proofs of main result

To prove Theorem 2.1, we need three lemmas. Throughout in this section, $u_p(1)$ is a sequence of stochastic processes that tends to zero, uniformly over its time domain, in probability.

Lemma 6.1. *Suppose ξ and U are random variables with $E(\xi|U)=0$, $0<E\xi^2<\infty$. Let $\sigma^2(u)=E(\xi^2|U=u)$, $L(u)=E\sigma^2(U)I[U\le u]$, $u\in\mathcal{I}$. Let (ξ_i,U_i), $1\le i\le n$ be i.i.d. copies of (ξ,U). Define*

$$U_n(u)=\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i I[U_i\le u],\quad u\in\mathcal{I}.$$

Assume L to be continuous. Then $U_n\Longrightarrow B\circ L$ in $D(\mathcal{I})$ and uniform metric.

The proof of this lemma uses Theorems 12.6, 15.5, in Billingsley (1968). Details are similar to those appearing in Stute et al. (1998).

To state the next lemma, let U be a continuous r.v. with d.f. G. Let $\ell(u)$ be a vector of q functions with $E\|\ell(U)\|^2<\infty$. Assume the matrix $C_u:=E\ell(U)\ell'(U)I(U\ge u)$ is positive definite for all $u\in\mathbb{R}$. For a real valued function $\gamma\in L_2(\mathbb{R},G)$ define the transforms

$$\mathcal{T}_\gamma(u):=\int_{y\le u}\gamma(y)\ell'(y)C_y^{-1}\,dG(y)\ell(u),\quad \mathcal{K}_\gamma(u):=\gamma(u)-\mathcal{T}_\gamma(u).$$

The following lemma is from Proposition 4.1 of Khmaladze and Koul (2004) and Lemma 9.1 of Koul (2006), which in turn has origin in Khmaladze (1993).

Lemma 6.2. *Under the above set up,*

$$E\mathcal{K}_\gamma(U)\ell'(U)=0,\quad \forall\gamma\in L_2(\mathbb{R},G),\tag{6.1}$$

$$E\mathcal{K}\gamma_1(U)\mathcal{K}_{\gamma_2}(U)=E\gamma_1(U)\gamma_2(U),\quad \forall\gamma_1,\gamma_2\in L_2(\mathbb{R},G).\tag{6.2}$$

Remark 6.1. Let ξ be a r.v. such that $E(\xi|U)=0$, $E\xi^2<\infty$, $\tau^2(u):=E(\xi^2|U=u)>0$, for all u. Then the covariance of the process $W_\gamma(\xi,U):=[\xi/\tau(U)]\mathcal{K}_\gamma(U)$, as a process in $\gamma\in L_2(\mathbb{R},G)$, is like that of $B_\gamma(G)$, where B_γ is a Brownian motion in γ. Hence, if (ξ_i,U_i), $1\le i\le n$, are i.i.d. copies of (ξ,U), then by the classical CLT, the finite dimensional distributions of $n^{-1/2}\sum_{i=1}^n W_\gamma(\xi_i,U_i)$, as γ varies, will converge weakly to those of $B_\gamma(G)$.

To proceed further, we need to introduce

$$S_n(x):=\frac{1}{\sqrt{n}}\sum_{i=1}^n\psi(\varepsilon_i)L(X_i)(\sqrt{\hat\mu_4(X_i)}-\sqrt{\mu_4(X_i)}I(X_i\le x),\quad x\in\mathbb{R},$$

where ψ and L is some measurable functions. The following lemma is useful when dealing with nonparametric $\mu_4(x)$.

Lemma 6.3. *Suppose* (e), (h), (f) *and* (k) *hold. Then, for any measurable function L with $EL^2(X)<\infty$ and for $\psi(\varepsilon)=\varepsilon$ or ε^2-1,*

$$\sup_{x\in\mathcal{I}}|S_n(x)|=o_p(1).\tag{6.3}$$

Moreover,

$$\max_{1\le i\le n}|\sqrt{\hat\mu_4(X_i)}-\sqrt{\mu_4(X_i)}|=o_p(1).\tag{6.4}$$

The proof of this lemma is given at the end of this section. Here we note that because $\mu_4(x)$ is bounded below from 0, (6.4) implies

$$\sup_{1\le i\le n}\left|\frac{\sqrt{\mu_4(X_i)}}{\sqrt{\hat\mu_4(X_i)}}-1\right|=o_p(1).\tag{6.5}$$

This is used in the sequel.

Proof of Theorem 2.1. Recall (2.5) and let

$$W_n(x)=V_n(x)-\int_{s\le x}l'(s)M_s^{-1}\int_{z\ge s}l(z)\,dV_n(z)\,dF(s).$$

The proof consists of the following two steps.

(a) $W_n\Longrightarrow B\circ F$ in $D(\mathcal{I})$ and in uniform metric.
(b) Under the null hypothesis, for every $x_0<1$, $\sup_{0\le x\le x_0}|\widehat{W}_n(x)-W_n(x)|=o_p(1)$.

Proof of part (*a*). Use assumptions (m1), (v1) and (v2) to conclude $E\|l(X)\|^2<\infty$ and with the aid of the Lindeberg–Feller CLT, that all finite dimensional distributions of W_n converge weakly to those of $B\circ F$. Thus the claim (a) would follow if we prove the tightness of W_n. By Lemma 6.1 applied to with $\xi=e$, $U=X$, the process V_n is tight. It remains to prove the tightness of the second term in W_n. Denote it by W_{2n}. Note that

$$W_{2n}(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i \int_{s\le x} l'(s)M_s^{-1} l(X_i)I[X_i \ge s]\,dF(s).$$

Let $\psi(x) = \int_{s\le x} \|l'(s)M_s^{-1}\|dF(s)$, $x \in \mathcal{I}$. By condition (ℓ), $0 < a := \psi(1) < \infty$. Because F is continuous, the function $\Psi := \psi/a$ is a strictly increasing continuous d.f. on $\mathcal{I}$. Moreover, for $x_1 < x \le x_2$,

$$\begin{aligned}
E[W_{2n}(x_2) - W_{2n}(x_1)]^2 &= E\left[\int_{x_1 < s \le x_2} l'(s)M_s^{-1} l(X)I[X \ge s]\,dF(s)\right]^2 \\
&= \int\int_{x_1\le s_1\le x_2}\int_{x_1\le s_2\le x_2} l'(s_1)M_{s_1}^{-1} l(y)I[y \ge s_1 \vee s_2]l'(s_2)M_{s_2}^{-1} l(y)\,dF(s_1)\,dF(s_2)\,dF(y) \\
&= \int_{x_1\le s_1\le x_2}\int_{x_1\le s_2\le x_2} l'(s_1)M_{s_1}^{-1} M_{s_1\vee s_2} M_{s_2}^{-1} l(s_2)\,dF(s_1)\,dF(s_2) \le \|M\|_\infty a^2[\Psi(x_2) - \Psi(x_1)]^2.
\end{aligned}$$

This bound, together with Theorems 12.3 of Billingsley (1968), imply that W_{2n} is tight in uniform metric on $\mathcal{C}[0,1]$. This complete the proof of part (a).

Proof of part (*b*). For the sake of brevity, let $\widehat{l}_i = \widehat{l}(X_i), l_i = l(X_i)$, and

$$\widehat{U}_n(y) := \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{e}_i\widehat{l}_i I[X_i \ge y], \quad U_n(y) := \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i l_i I[X_i \ge y],$$

$$\Delta_n(X_i) := \frac{\upsilon_0^2(X_i)\sqrt{\mu_4(X_i)}}{\upsilon_n^2(X_i)\sqrt{\widehat{\mu}_4(X_i)}} - 1.$$

Then,

$$\widehat{W}_n(x) = \widehat{V}_n(x) - \int_{s\le x}\widehat{l}'(s)\widehat{M}_s^{-1}\widehat{U}_n(s)\,d\widehat{F}(s),$$

$$W_n(x) = V_n(x) - \int_{s\le x} l'(s)M_s^{-1}U_n(s)dF(s). \tag{6.6}$$

Now, $\widehat{V}_n(x)$ can be written as the sum of $V_n(x)$ and the following seven terms,

$$W_{1n1}(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{[m_n(X_i) - m_0(X_i)]^2}{\upsilon_0^2(X_i)\sqrt{\mu_4(X_i)}} I[X_i \le x],$$

$$W_{1n2}(x) = -\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\frac{[Y_i - m_0(X_i)][m_n(X_i) - m_0(X_i)]}{\upsilon_0^2(X_i)\sqrt{\mu_4(X_i)}} I[X_i \le x],$$

$$W_{1n3}(x) = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\upsilon_n^2(X_i) - \upsilon_0^2(X_i)}{\upsilon_0^2(X_i)\sqrt{\mu_4(X_i)}} I[X_i \le x],$$

$$W_{1n4}(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i\Delta_n(X_i)I[X_i \le x],$$

$$W_{1n5}(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{[m_n(X_i) - m_0(X_i)]^2}{\upsilon_0^2(X_i)\sqrt{\mu_4(X_i)}}\Delta_n(X_i) I[X_i \le x],$$

$$W_{1n6}(x) = -\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\frac{[Y_i - m_0(X_i)][m_n(X_i) - m_0(X_i)]}{\upsilon_0^2(X_i)\sqrt{\mu_4(X_i)}}\Delta_n(X_i) I[X_i \le x],$$

$$W_{1n7}(x) = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\upsilon_n^2(X_i) - \upsilon_0^2(X_i)}{\upsilon_0^2(X_i)\sqrt{\mu_4(X_i)}}\Delta_n(X_i) I[X_i \le x].$$

We shall first sketch the proof of the claim $\sup_{x\in\overline{\mathbb{R}}}|W_{1nj}(x)| = o_p(1)$ for $j = 1,2,4,5,6,7$.

Consider the case of $j=1$. Adding and subtracting $(\widehat{\beta}_n-\beta_0)'\dot{m}_\beta(X_i,\beta_0)$ from $m_n(X_i)-m_0(X_i)$, $W_{1n1}(x)$ is uniformly bounded above by the sum terms

$$\frac{2}{\sqrt{n}}\sum_{i=1}^n\frac{[m_n(X_i)-m_0(X_i)-(\widehat{\beta}_n-\beta_0)'\dot{m}_\beta(X_i,\beta_0)]^2}{\upsilon_0^2(X_i)\sqrt{\mu_4(X_i)}}+\frac{2\|(\widehat{\beta}_n-\beta_0)\|^2}{\sqrt{n}}\sum_{i=1}^n\frac{\|\dot{m}_\beta(X_i,\beta_0)\|^2}{\upsilon_0^2(X_i)\sqrt{\mu_4(X_i)}}.$$

By assumptions (e), (m1), (m2), (v2) and (2.8), both of these two summands are $o_p(1)$.

Next, consider the case of $j=2$. Adding and subtracting $(\widehat{\beta}_n-\beta_0)'\dot{m}_\beta(X_i,\beta_0)$ from $m_n(X_i)-m_0(X_i)$, $W_{1n2}(x)$ equals the sum

$$-\frac{2}{\sqrt{n}}\sum_{i=1}^n\frac{\varepsilon_i[m_n(X_i)-m_0(X_i)-(\widehat{\beta}_n-\beta_0)'\dot{m}_\beta(X_i,\beta_0)]}{\upsilon_0(X_i)\sqrt{\mu_4(X_i)}}I[X_i\le x],$$

$$-\frac{2(\widehat{\beta}_n-\beta_0)'}{\sqrt{n}}\sum_{i=1}^n\frac{\varepsilon_i\dot{m}_\beta(X_i,\beta_0)}{\upsilon_0(X_i)\sqrt{\mu_4(X_i)}}I[X_i\le x]. \tag{6.7}$$

Conditions (e), (m1), (2.8) and the LLNs imply the first term of the sum (6.7) to be $u_p(1)$. A Glivenko–Cantelli type argument shows that

$$\sup_{x\in\mathcal{I}}\left|\frac{1}{n}\sum_{i=1}^n\frac{\varepsilon_i\dot{m}_\beta(X_i,\beta_0)I[X_i\le x]}{\upsilon_0(X_i)\sqrt{\mu_4(X_i)}}\right|=o_p(1).$$

Hence, in view of (2.8), the second term in (6.7) is also $u_p(1)$, thereby completing the proof of $\sup_{x\in\mathcal{I}}|W_{1n2}(x)|=o_p(1)$.

Now, we shall show that $\sup_{x\in\mathcal{I}}|W_{1n4}(x)|=o_p(1)$. Note that

$$\begin{aligned}W_{1n4}(x)=&-\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i\left[\frac{(\upsilon_n^2(X_i)-\upsilon_0^2(X_i))(\sqrt{\widehat{\mu}_4(X_i)}-\sqrt{\mu_4(X_i)})}{\upsilon_0^2(X_i)\sqrt{\mu_4(X_i)}}\right]I[X_i\le x]-\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i\left[\frac{\upsilon_n^2(X_i)-\upsilon_0^2(X_i)}{\upsilon_0^2(X_i)}\right]I[X_i\le x]\\&-\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i\left[\frac{\sqrt{\widehat{\mu}_4(X_i)}-\sqrt{\mu_4(X_i)}}{\sqrt{\mu_4(X_i)}}\right]I[X_i\le x]\\&-\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i\left[\frac{(\upsilon_n^2(X_i)-\upsilon_0^2(X_i))(\sqrt{\widehat{\mu}_4(X_i)}-\sqrt{\mu_4(X_i)})}{\upsilon_0^2(X_i)\sqrt{\mu_4(X_i)}}\right]\Delta_n(X_i)I[X_i\le x]\\&-\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i\left[\frac{\upsilon_n^2(X_i)-\upsilon_0^2(X_i)}{\upsilon_0^2(X_i)}\right]\Delta_n(X_i)I[X_i\le x]-\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i\left[\frac{\sqrt{\widehat{\mu}_4(X_i)}-\sqrt{\mu_4(X_i)}}{\sqrt{\mu_4(X_i)}}\right]\Delta_n(X_i)I[X_i\le x],\end{aligned} \tag{6.8}$$

so it suffices to show that all terms on the right are $u_p(1)$. Because the details are similar, here we verify this only for the first term. Adding and subtracting $\upsilon_0(X_i)$ from $\upsilon_n(X_i)$, the first term in (6.8), without the minus sign, can be written as

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i\left[\frac{(\upsilon_n(X_i)-\upsilon_0(X_i))^2(\sqrt{\widehat{\mu}_4(X_i)}-\sqrt{\mu_4(X_i)})}{\upsilon_0^2(X_i)\sqrt{\mu_4(X_i)}}\right]I[X_i\le x]\\&\quad+\frac{2}{\sqrt{n}}\sum_{i=1}^n e_i\left[\frac{(\upsilon_n(X_i)-\upsilon_0(X_i))(\sqrt{\widehat{\mu}_4(X_i)}-\sqrt{\mu_4(X_i)})}{\upsilon_0(X_i)\sqrt{\mu_4(X_i)}}\right]I[X_i\le x].\end{aligned} \tag{6.9}$$

Adding and subtracting $(\widehat{\beta}_n-\beta_0)'\dot{\upsilon}_0(X_i)$ from $\upsilon_n(X_i)-\upsilon_0(X_i)$, the second term in (6.9) can be written as

$$\begin{aligned}&\frac{2}{\sqrt{n}}\sum_{i=1}^n e_i\left[\frac{(\upsilon_n(X_i)-\upsilon_0(X_i)-(\widehat{\beta}_n-\beta_0)'\dot{\upsilon}_0(X_i))(\sqrt{\widehat{\mu}_4(X_i)}-\sqrt{\mu_4(X_i)})}{\upsilon_0(X_i)\sqrt{\mu_4(X_i)}}\right]I[X_i\le x]\\&\quad+\frac{2(\widehat{\beta}_n-\beta_0)'}{\sqrt{n}}\sum_{i=1}^n e_i\left[\frac{\dot{\upsilon}_0(X_i)(\sqrt{\widehat{\mu}_4(X_i)}-\sqrt{\mu_4(X_i)})}{\upsilon_0(X_i)\sqrt{\mu_4(X_i)}}\right]I[X_i\le x].\end{aligned}$$

By (e), (v1), and (6.4), the first term above is bounded above by

$$\max_i|\upsilon_n(X_i)-\upsilon_0(X_i)-(\widehat{\beta}_n-\beta_0)'\dot{\upsilon}_0(X_i)|\cdot\max_i\left|\sqrt{\widehat{\mu}_4(X_i)}-\sqrt{\mu_4(X_i)}\right|\cdot\frac{2c}{\sqrt{n}}\sum_{i=1}^n|\varepsilon_i^2-1|,$$

which is $o_p(1)$, where c is some positive constant. Consistency of $\widehat{\beta}_n$ and (6.3) in Lemma 6.3 applied to $L(x)\equiv\dot{\upsilon}_0(x)/[\upsilon_0(x)\sqrt{\mu_4(x)}]$ imply the second term is also $u_p(1)$. Hence the second term in (6.9) is $u_p(1)$. Similarly one can show the same holds for the first term in (6.9). This completes the proof of $\sup_{x\in\mathcal{I}}|W_{1n4}(x)|=o_p(1)$.

Note that assumption (v1) implies $\max_i |\nu_n^2(X_i)-\nu_0^2(X_i)| = o_p(1)$. This fact together with (6.4), and the fact that $\mu_4(x)$ and $\nu_0^2(x)$ are bounded below from 0, we obtain

$$\max_i |\Delta_n(X_i)| = o_p(1). \tag{6.10}$$

This fact and an argument similar to the above will yield $\sup_{x\in\mathcal{I}} |W_{1nj}(x)| = o_p(1),\ j=5,6,7$.

Finally, consider $W_{1n3}(x)$. Let $\alpha_n = \sqrt{n}(\widehat{\beta}_n - \beta_0)$. Using (v1) and (2.8), verify that

$$W_{1n3}(x) = -\frac{2\alpha_n'}{n}\sum_{i=1}^n \frac{\dot{\nu}_0(X_i)}{\nu_0(X_i)\sqrt{\mu_4(X_i)}} I[X_i \le x] + u_p(1),$$

which, by the finiteness of the second moment of $\|\dot{\nu}_0(X)\|$ and a Glivenko–Cantelli type argument, implies

$$W_{1n3}(x) = -E(l'(X)I[X\le x])\alpha_n + u_p(1).$$

Therefore,

$$\sup_{x\in\mathcal{I}} |\widehat{V}_n(x) - V_n(x) + E(l'(X)I[X\le x])\alpha_n| = o_p(1). \tag{6.11}$$

Next, consider the difference $\widehat{U}_n(y) - U_n(y)$. By definition of $\widehat{\sigma}(X)$ in (2.6),

$$\widehat{U}_n(y) = \frac{2}{\sqrt{n}}\sum_{i=1}^n \frac{[Y_i - m_n(X_i)]^2 - \nu_n^2(X_i)}{\nu_n^3(X_i)\widehat{\mu}_4(X_i)} \dot{\nu}_n(X_i) I[X_i \ge y].$$

By Lemma 6.3 and (2.8), one obtains

$$\widehat{U}_n(y) - U_n(y) = -\frac{2}{\sqrt{n}}\sum_{i=1}^n \frac{\nu_n^2(X_i) - \nu_0^2(X_i)}{\nu_0^3(X_i)\mu_4(X_i)} \dot{\nu}_0(X_i) I(X_i \ge y) + u_p(1).$$

Add and subtract $\nu_0(X_i)$ from $\nu_n(X_i)$, and use assumption (v1), to show that

$$\frac{2}{\sqrt{n}}\sum_{i=1}^n \frac{\nu_n^2(X_i) - \nu_0^2(X_i)}{\nu_0^3(X_i)\mu_4(X_i)} \dot{\nu}_0(X_i) I(X_i \ge y) = 4E\frac{\dot{\nu}_0(X)\dot{\nu}_0'(X) I(X\ge y)}{\nu_0^2(X_i)\mu_4(X)} + u_p(1).$$

Hence, by the definition of M_y, we have

$$\sup_{y\in\mathcal{I}} \|\widehat{U}_n(y) - U_n(y) + M_y\alpha_n\| = o_p(1). \tag{6.12}$$

Again using conditions (m1)–(m3), (v1), (v2), one can show that $\sup_{y\in\mathcal{I}} \|\widehat{M}_y - M_y\| = o_p(1)$, where $\widehat{M}_y$ is defined in (2.6). Consequently,

$$\sup_{0\le y\le x_0} \|\widehat{M}_y^{-1} - M_y^{-1}\| = o_p(1) \tag{6.13}$$

by the positive definiteness of M_y for all $y\in[0,1)$.

For convenience, let $\widehat{W}_{2n}(x)$, $W_{2n}(x)$ denote the second terms on the right hand sides of $\widehat{W}_n(x)$, $W_n(x)$ in (6.6). Then

$$\begin{aligned}\widehat{W}_{2n}(x) &= \int_{s\le x} \widehat{l}(s)'\widehat{M}_s^{-1}\widehat{U}_n(s)\,d\widehat{F}(s) = \int_{s\le x} \frac{2\dot{\nu}_n(s)}{\nu_n(x)\sqrt{\widehat{\mu}_4}}[\widehat{M}_s^{-1} - M_s^{-1} + M_s^{-1}]\cdot[\widehat{U}_n(s) - U_n(s) + U_n(s)]\,d\widehat{F}(s)\\ &= \int_{s\le x} l'(s)M_s^{-1}U_n(s)\,dF(s) - El'(X)I[X\le x]\alpha_n + u_p(1) = W_{2n}(x) - E(l'(X)I[X\le x])\alpha_n + u_p(1),\end{aligned}$$

by (v2), (6.12) and (6.13). Therefore, by (6.11) and the relation

$$\widehat{W}_n(x) - W_n(x) = \widehat{V}_n(x) - V_n(x) - [\widehat{W}_{2n}(x) - W_{2n}(x)],$$

we obtain $\sup_{0\le x\le x_0} |\widehat{W}_n(x) - W_n(x)| = o_p(1)$, thereby completing the proof of Theorem 2.1. □

Remark 6.2. Note that in the above proof, the only place we needed to restrict the supremum to $[0,x_0]$ is to establish (6.13). If $\ell(s)$, and hence M_s, are free from the null parameter β and depend on the other unknown entities in a simpler fashion, then one does not need this result and a suitable analog of $\widehat{W}_n$ will converge weakly to $B\circ F$ in $D(\mathcal{I})$. Consider, for example, the case where $h := \dot{\nu}_0/\nu_0$ is known completely, $\mu_4(x)\equiv c$, a positive constant, and the design d.f. F is known. Then $l(x) = 2h(x)/\sqrt{c}$, and $\widehat{l}(x) = 2h(x)/\sqrt{\widehat{c}}$, where $\widehat{c}$ is a consistent estimator of c. Assume $H_s := \int_{y\ge s} h(y)h(y)'\,dF(s)$ is positive definite for all $0\le s<1$. Note that now H_s is known and $M_s^{-1} = 4^{-1}cH_s^{-1}$, exists for each $s<1$. Then, $\widehat{M}_s^{-1} = 4^{-1}\widehat{c}H_s^{-1}$, and

$$\widehat{W}_{2n}(x) = \int_{s\le x} \widehat{l}'(s)\widehat{M}_s^{-1}\int_{z\ge s}\widehat{l}(z)\,d\widehat{V}_n(z)\,dF(s) = \int_{s\le x} h(s)'H_s^{-1}\int_{y\ge s} h(y)\widehat{V}_n(dy)\,dF(s).$$

Thus here the transformation to construct tests for H_0 is

$$\tilde{W}_n(x) := \hat{V}_n(x) - \int_{s \le x} h(s)' H_s^{-1} \int_{y \ge s} h(y) \hat{V}_n(dy)\, dF(s).$$

The above proof of Theorem 2.1 readily shows that here $\sup_{x \in \mathcal{I}} |\tilde{W}_n(x) - W_n(x)| = o_p(1)$, and hence $\tilde{W}_n \Longrightarrow B \circ F$ in $D(\mathcal{I})$ and uniform metric.

Proof of Theorem 3.1. Let $Y_i^a = m(X_i;\beta_a) + v(X_i;\beta_a)\varepsilon_i$, then we rewrite

$$\hat{e}_i = \frac{(Y_i^a - m_n(X_i))^2 - v_n^2(X_i)}{v_n^2(X_i)\sqrt{\hat{\mu}_4(X_i)}} + \frac{2(Y_i - Y_i^a)(Y_i^a - m_n(X_i))}{v_n^2(X_i)\sqrt{\hat{\mu}_4(X_i)}} + \frac{(Y_i - Y_i^a)^2}{v_n^2(X_i)\sqrt{\hat{\mu}_4(X_i)}}.$$

Denote the three terms on the right hand side by $\hat{e}_{i1}^a$, $\hat{e}_{i2}^a$ and $\hat{e}_{i3}^a$, respectively. Then $\widehat{W}_n(x)$ is the sum of the following three terms:

$$W_{1n}^a(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{e}_{i1}^a \left(I[X_i \le x] - \int_{s \le x} \hat{l}'(s) \widehat{M}_s^{-1} I[X_i \ge s]\, d\hat{F}(s) \hat{l}(X_i) \right),$$

$$W_{2n}^a(x) = \frac{2}{\sqrt{n}} \sum_{i=1}^n \hat{e}_{i2}^a \left(I[X_i \le x] - \int_{s \le x} \hat{l}'(s) \widehat{M}_s^{-1} I[X_i \ge s]\, d\hat{F}(s) \hat{l}(X_i) \right),$$

$$W_{3n}^a(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{e}_{i3}^a \left(I[X_i \le x] - \int_{s \le x} \hat{l}'(s) \widehat{M}_s^{-1} I[X_i \ge s]\, d\hat{F}(s) \hat{l}(X_i) \right).$$

Exactly the same argument as before yields $W_{1n}^a(x) \Longrightarrow B \circ F(x)$. Adding and subtracting $m(X_i;\beta_a)$ from $Y_i^a - m_n(X_i)$, $W_{2n}^a(x)$ can be written as the sum $A_{1n}(x) + A_{2n}(x)$, where

$$\begin{aligned} A_{1n}(x) &= \frac{2}{\sqrt{n}} \sum_{i=1}^n \frac{(Y_i - Y_i^a)(Y_i^a - m(X_i;\beta_a))}{v_n^2(X_i)\sqrt{\hat{\mu}_4(X_i)}} I[X_i \le x] - \frac{2}{\sqrt{n}} \sum_{i=1}^n \frac{(Y_i - Y_i^a)(Y_i^a - m(X_i;\beta_a))}{v_n^2(X_i)\sqrt{\hat{\mu}_4(X_i)}} \cdot \int_{s \le x} \hat{l}'(s) \widehat{M}_s^{-1} I[X_i \ge s]\, d\hat{F}(s) \hat{l}(X_i) \\ &= A_{1n1}(x) - A_{1n2}(x), \end{aligned}$$

$$\begin{aligned} A_{2n}(x) &= \frac{2}{\sqrt{n}} \sum_{i=1}^n \frac{(Y_i - Y_i^a)(m(X_i;\beta_a) - m_n(X_i))}{v_n^2(X_i)\sqrt{\hat{\mu}_4(X_i)}} I[X_i \le x] - \frac{2}{\sqrt{n}} \sum_{i=1}^n \frac{(Y_i - Y_i^a)(m(X_i;\beta_a) - m_n(X_i))}{v_n^2(X_i)\sqrt{\hat{\mu}_4(X_i)}} \\ &\quad \cdot \int_{s \le x} \hat{l}'(s) \widehat{M}_s^{-1} I[X_i \ge s]\, d\hat{F}(s) \hat{l}(X_i) \\ &= A_{2n1}(x) + A_{2n2}(x). \end{aligned}$$

Note that (6.10) still holds when β_0 is replaced by β_a. Then by (6.10) and the assumption (v2), one verifies

$$n^{-1/2} A_{1n1}(x) = K_1(x) + u_p(1), \tag{6.14}$$

where $K_1(x)$ is defined in Section 3.1. Let

$$V_n^a(s) = \frac{1}{n} \sum_{i=1}^n \frac{(Y_i - Y_i^a)(Y_i^a - m(X_i;\beta_a))}{v_n^2(X_i)\sqrt{\hat{\mu}_4(X_i)}} \hat{l}(X_i) I[X_i \ge s].$$

Then by (6.10), assumption (m2), (m3), (v1), (v2), the definition of Y_i^a, and (6.4) in Lemma 6.3, one verifies

$$V_n^a(s) = \phi(s) + u_p(1), \tag{6.15}$$

where $\phi(s)$ is defined in Section 3.1.

Using exactly the same argument as in the null case, one can verify that under the alternative hypothesis (3.1), $\sup_{s \le x_0} \|\widehat{M}_s^{-1} - A_s^{-1}\| = o_p(1)$. Note that $n^{-1/2} A_{1n2}(x) = 2\int_{s \le x_0} \hat{l}'(s) \widehat{M}_s^{-1} V_n(s)\, d\hat{F}(s)$, so by adding and subtracting $\dot{v}(s;\beta_a)$ from $\dot{v}(s;\hat{\beta}_n)$ in $\hat{l}(s)$, M_y^{-1} from $\widehat{M}_y^{-1}$, using assumption (m2), (m3), (v1), (v2), (6.10) and (6.15), we can show that

$$n^{-1/2} A_{1n2}(x) = K_2(x) + o_p(1) \tag{6.16}$$

uniformly for $x \le x_0$, where $K_2(x)$ is defined in Section 3.1. From the consistency of $\hat{\beta}_n$ to β_a, one can finally show $\sup_{0 \le x \le x_0} |A_{2n}(x)| = o_p(1)$. This, together with (6.16) implies that

$$\sup_{0 \le x \le x_0} \| n^{-1/2} W_{2n}^a(x) - [K_1(x) - K_2(x)] \| = o_p(1). \tag{6.17}$$

Now, consider $W_{3n}^a(x)$. Define

$$H_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{(Y_i - Y_i^a)^2}{v_n^2(X_i)\sqrt{\hat{\mu}_4(X_i)}} \hat{l}(X_i) I[X_i \ge x].$$

An argument similar to one used in deducing (6.15) yields

$$H_n(x) = 2E\frac{H(x)I[X \ge x]}{\nu^3(X;\beta_a)\mu_4(X)} + o_p(1) = \rho(x) + u_p(1),$$

where $H(x)$ is defined in Section 3.1.

Note that $n^{-1/2}W_{3n}^a(x)$ can be written as

$$\frac{1}{n}\sum_{i=1}^n \frac{(Y_i - Y_i^a)^2}{\nu_n^2(X_i)\sqrt{\hat\mu_4(X_i)}}\left(I[X_i \le x] - \int_{s\le x}\hat l'(s)\widehat M_s^{-1}H_n(s)\,d\hat F(s)\right). \tag{6.18}$$

The first term in the above expression can be written as

$$\frac{1}{n}\sum_{i=1}^n \frac{(Y_i - Y_i^a)^2}{\nu^2(X_i;\beta_a)\sqrt{\mu_4(X_i)}}I[X_i \le x] \quad + \frac{1}{n}\sum_{i=1}^n \frac{(Y_i - Y_i^a)^2}{\nu^2(X_i;\beta_a)\sqrt{\mu_4(X_i)}}\left[\frac{\nu^2(X_i;\beta_a)\sqrt{\mu_4(X_i)}}{\nu_n^2(X_i)\sqrt{\hat\mu_4(X_i)}} - 1\right]I[X_i \le x].$$

Condition (a1) implies that

$$\sup_{0\le x\le x_0}\left|\frac{1}{n}\sum_{i=1}^n \frac{(Y_i - Y_i^a)^2}{\nu^2(X_i;\beta_a)\sqrt{\mu_4(X_i)}}I[X_i \le x] - D_1(x)\right| = o_p(1), \tag{6.19}$$

where $D_1(x)$ is as in (3.2). Also, (a1) and (6.10) imply that

$$\sup_{x\in\mathcal{I}}\left|\frac{1}{n}\sum_{i=1}^n \frac{(Y_i - Y_i^a)^2}{\nu^2(X_i;\beta_a)\sqrt{\mu_4(X_i)}}\left[\frac{\nu^2(X_i;\beta_a)\sqrt{\mu_4(X_i)}}{\nu_n^2(X_i)\sqrt{\hat\mu_4(X_i)}} - 1\right]I[X_i \le x]\right| = o_p(1).$$

Therefore, one can show that the second term on the right end of (6.18) satisfies

$$\sup_{x\in\mathcal{I}}\left|\int_{s\le x}\hat l'(s)\widehat M_s^{-1}H_n(s)\,d\hat F(s) - D_2(x)\right| = o_p(1), \tag{6.20}$$

where $D_2(x)$ is as in (3.2). Combining (6.19) and (6.20), we get

$$\sup_{0\le x\le x_0}|n^{-1/2}W_{3n}^a(x) - [D_1(x) - D_2(x)]| = o_p(1).$$

Therefore,

$$\sup_{0\le x\le x_0}|W_n(x)| = \sup_{0\le x\le x_0}|W_{1n}^a(x) + W_{2n}^a(x) + W_{3n}^a(x)| \ge \sup_{0\le x\le x_0}|W_{2n}^a(x) + W_{3n}^a(x)| - \sup_{0\le x\le x_0}|W_{1n}^a(x)| \ge \sqrt{n}\sup_{0\le x\le x_0}|n^{-1/2}W_{2n}^a(x) + n^{-1/2}W_{3n}^a(x)| - \sup_{0\le x\le x_0}|W_{1n}^a(x)| \to \infty \quad \text{in probability,}$$

which implies the consistency, hence the theorem. □

Proof of Theorem 3.2. Denote $Y_i^L = m(X_i;\beta_0) + \nu(X_i;\beta_0)\varepsilon_i$. Then

$$\hat e_i = \frac{(Y_i^L - m_n(X_i))^2 - \nu_n^2(X_i)}{\nu_n^2(X_i)\sqrt{\hat\mu_4(X_i)}} + \frac{2(Y_i - Y_i^L)(Y_i^L - m_n(X_i))}{\nu_n^2(X_i)\sqrt{\hat\mu_4(X_i)}} + \frac{(Y_i - Y_i^L)^2}{\nu_n^2(X_i)\sqrt{\hat\mu_4(X_i)}}.$$

Denoting the three terms on the right end by $\hat e_{i1}^L$, $\hat e_{i2}^L$ and $\hat e_{i3}^L$, respectively. Then the test statistic is the sum of the following three terms:

$$W_{1n}^L(x) = \frac{1}{\sqrt n}\sum_{i=1}^n \hat e_{i1}^L\left(I[X_i \le x] - \int_{s\le x}\hat l'(s)\widehat M_s^{-1}I[X_i \ge s]\,d\hat F(s)\hat l(X_i)\right),$$

$$W_{2n}^L(x) = \frac{2}{\sqrt n}\sum_{i=1}^n \hat e_{i2}^L\left(I[X_i \le x] - \int_{s\le x}\hat l'(s)\widehat M_s^{-1}I[X_i \ge s]\,d\hat F(s)\hat l(X_i)\right),$$

$$W_{3n}^L(x) = \frac{1}{\sqrt n}\sum_{i=1}^n \hat e_{i3}^L\left(I[X_i \le x] - \int_{s\le x}\hat l'(s)\widehat M_s^{-1}I[X_i \ge s]\,d\hat F(s)\hat l(X_i)\right).$$

Exactly same argument as before leads to $W_{1n}^L(x) \Longrightarrow B\circ F(x)$.

Next, consider $W_{2n}^L(x)$. Note that $Y_i - Y_i^L = \delta(X_i)\varepsilon_i/\sqrt n$, so

$$W_{2n}^L(x) = \frac{2}{n}\sum_{i=1}^n \frac{\delta(X_i)\varepsilon_i(Y_i^L - m_n(X_i))}{\nu_n^2(X_i)\sqrt{\hat\mu_4(X_i)}}(I[X_i \le x] - \int_{s\le x}\hat l'(s)\widehat M_s^{-1}I[X_i \ge s]d\hat F(s)\hat l(X_i)\Big).$$

Adding and subtracting $m_0(X_i)$ from $Y_i-m_n(X_i)$, $W_{2n}^L(x)$ can be written as the sum of the following two terms:

$$B_{1n}(x)=\frac{2}{n}\sum_{i=1}^n\frac{\delta(X_i)\varepsilon_i(Y_i^L-m_0(X_i))}{\nu_n^2(X_i)\sqrt{\widehat{\mu}_4(X_i)}}\left(I[X_i\le x]-\int_{s\le x}\widehat{l}'(s)\widehat{M}_s^{-1}I[X_i\ge s]\,d\widehat{F}(s)\widehat{l}(X_i)\right),$$

$$B_{2n}(x)=\frac{2}{n}\sum_{i=1}^n\frac{\delta(X_i)\varepsilon_i(m_0(X_i)-m_n(X_i))}{\nu_n^2(X_i)\sqrt{\widehat{\mu}_4(X_i)}}\left(I[X_i\le x]-\int_{s\le x}\widehat{l}'(s)\widehat{M}_s^{-1}I[X_i\ge s]\,d\widehat{F}(s)\widehat{l}(X_i)\right).$$

Use assumptions (m1) and (v2) to verify $\sup_{0\le x\le x_0}|B_{2n}(x)|=o_p(1)$. To study $B_{1n}(x)$, let

$$T_n(y):=\frac{2}{n}\sum_{i=1}^n\frac{\delta(X_i)\nu_0(X_i)\varepsilon_i^2}{\nu_n^2(X_i)\sqrt{\widehat{\mu}_4(X_i)}}\widehat{l}(X_i)I[X_i\ge y].$$

Similar argument as before leads to

$$T_n(y)=2E\frac{\delta(X)l(X)}{\nu_0(X)\sqrt{\mu_4(X)}}I[X\ge y]+o_p(1)=2\psi(y)+u_p(1),\tag{6.21}$$

where $\psi(y)$ is given in (3.4). Note that

$$B_{1n}(x)=\frac{2}{n}\sum_{i=1}^n\frac{\delta(X_i)\nu_0(X_i)\varepsilon_i^2}{\nu_n^2(X_i)\sqrt{\widehat{\mu}_4(X_i)}}I[X_i\le x]-\int_{s\le x}\widehat{l}'(s)\widehat{M}_s^{-1}V_n(s)\,d\widehat{F}(s).$$

From (6.10), (6.5) and (6.21), one can show that $\sup_{0\le x\le x_0}|B_{1n}(x)-D(x)|=o_p(1)$, where $D(x)$ is given in (3.4). This implies $\sup_{0\le x\le x_0}|W_{2n}^L(x)-D(x)|=o_p(1)$.

As for $W_{3n}^L(x)$, note that

$$W_{3n}^L(x)=\frac{1}{n\sqrt{n}}\sum_{i=1}^n\frac{\delta^2(X_i)\varepsilon_i^2}{\nu_n^2(X_i)\sqrt{\widehat{\mu}_4(X_i)}}\left(I[X_i\le x]-\int_{s\le x}\widehat{l}'(s)\widehat{M}_s^{-1}I[X_i\ge s]\,d\widehat{F}(s)\widehat{l}(X_i)\right).$$

Then usual argument shows that $\sup_{0\le x\le x_0}|W_{3n}^L(x)|=o_p(1)$. Finally, by summarizing the above results, we obtain $\sup_{0\le x\le x_0}|W_n(x)|=\sup_{0\le x\le x_0}|W_{1n}^L(x)+D(x)|+o_p(1)$. Hence the theorem. □

Proof of Theorem 4.1. In the following proof, we will assume $\mu_4(x)$ is a constant, which is estimated by the simple average of $(\varepsilon_i^2-1)^2$. Easy to see, it is an $\sqrt{n}$–consistent estimator for $\mu_4(x)$. Through this simplification, the effect of the nonparametric estimator of the regression function on the argument can be clearly understood, and the reader will not be trapped into clumsy notations and involved expressions. Of course, if $\mu_4(x)$ does depend on x, then Lemma 6.3 should be used to deal with the incurred complications.

The main steps of the proof are the same as in the proof of part (b) in Theorem 2.1. But some necessary modification need to be made because of the presence of a nonparametric estimator of the regression function. For the sake of brevity, we only present the proof of $\sup_{0\le x\le x_0}|W_{1n2}(x)|=o_p(1)$, the arguments for the other claims can be adjusted similarly.

Let

$$\widehat{g}(X_i)=\frac{1}{nh}\sum_{j=1}^n K\left(\frac{X_j-X_i}{h}\right)Y_j,\quad \widehat{f}(X_i)=\frac{1}{nh}\sum_{j=1}^n K\left(\frac{X_j-X_i}{h}\right)$$

and $g(x)=m(x)f(x)$. Then, for each i,

$$\widehat{m}_n(X_i)-m(X_i)=\frac{\widehat{g}(X_i)-g(X_i)}{f(X_i)}-\frac{[\widehat{g}(X_i)-g(X_i)][\widehat{f}(X_i)-f(X_i)]}{\widehat{f}(X_i)f(X_i)}+\frac{m(X_i)[\widehat{f}(X_i)-f(X_i)]^2}{\widehat{f}(X_i)f(X_i)}-\frac{m(X_i)[\widehat{f}(X_i)-f(X_i)]}{f(X_i)}.\tag{6.22}$$

An argument similar to the one used for computing mean square error of the estimator, along with conditions (e), (k), (h), (f), and (μ), yield

$$\frac{1}{n}\sum_{i=1}^n[\widehat{g}(X_i)-g(X_i)]^2=O_p[(\log n)^4/(nh)+h^2]=\frac{1}{n}\sum_{i=1}^n[\widehat{f}(X_i)-f(X_i)]^2.\tag{6.23}$$

See, e.g., Zhu et al. (2001) for details, where there is a typo in that $1/n$ is missing in the left and right hand side.

From (6.22) and the consistency of $\widehat{\mu}_4$ to μ_4, uniformly in $x \in \mathcal{I}$,

$$W_{1n2}(x) = -\frac{1}{\sqrt{n\widehat{\mu}_4}}\sum_{i=1}^n \frac{\varepsilon_i[I(X_i \le x) - F(x)][\widehat{g}_n(X_i) - g(X_i)]}{\upsilon_0^2(X_i)f(X_i)} + \frac{1}{\sqrt{n\widehat{\mu}_4}}\sum_{i=1}^n \frac{\varepsilon_i m(X_i)[I(X_i \le x) - F(x)][\widehat{f}(X_i) - f(X_i)]}{\upsilon_0^2(X_i)f(X_i)}$$
$$- F(x)\frac{1}{\sqrt{n\widehat{\mu}_4}}\sum_{i=1}^n \frac{\varepsilon_i[\widehat{g}_n(X_i) - g(X_i)]}{\upsilon_0^2(X_i)f(X_i)} + F(x)\frac{1}{\sqrt{n\widehat{\mu}_4}}\sum_{i=1}^n \frac{\varepsilon_i m(X_i)[\widehat{f}(X_i) - f(X_i)]}{\upsilon_0^2(X_i)f(X_i)} + O_p[(\log n)^4/(h\sqrt{n}) + h^2\sqrt{n}].$$

Because of the consistency of $\widehat{\mu}_4$ for μ_4, showing the first term above is $u_p(1)$ is equivalent to showing that

$$M(x) =: \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\varepsilon_i[I(X_i \le x) - F(x)][\widehat{g}_n(X_i) - g(X_i)]}{\upsilon_0^2(X_i)f(X_i)} = u_p(1). \tag{6.24}$$

Details of the proof of this claim, however, are similar to those for proving $J_1(x) = u_p(1)$ appearing in Zhu et al. (2001, p. 1248). We thus do not reproduce them here. Their proof in turn is based on the works of Nolan and Pollard (1987). Similarly, one show the second term in the above decomposition of $W_{1n2}(x)$ is $u_p(1)$.

To show the third and fourth term of $W_{1n2}(x)$ are $u_p(1)$, one only needs to show

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\varepsilon_i[\widehat{g}_n(X_i) - g(X_i)]}{\upsilon_0^2(X_i)f(X_i)} = o_p(1) \quad \text{and} \quad \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\varepsilon_i m_0(X_i)[\widehat{f}(X_i) - f(X_i)]}{\upsilon_0^2(X_i)f(X_i)} = o_p(1) \tag{6.25}$$

because of $0 \le F(x) \le 1$.

Let

$$S_{n1} = \frac{1}{nh\sqrt{n}}\sum_{i=1}^n\sum_{j=1}^n K\left(\frac{X_j - X_i}{h}\right)\frac{\varepsilon_i\varepsilon_j\upsilon_0(X_j)}{\upsilon_0^2(X_i)f(X_i)},$$

$$S_{n2} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\varepsilon_i}{\upsilon_0^2(X_i)f(X_i)}\left[\frac{1}{nh}\sum_{j=1}^n K\left(\frac{X_j - X_i}{h}\right)m(X_j) - g(X_i)\right].$$

Then

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\varepsilon_i[\widehat{g}_n(X_i) - g(X_i)]}{\upsilon_0^2(X_i)f(X_i)} = S_{n1} + S_{n2}.$$

We shall show that both S_{n1} and S_{n2} are $o_p(1)$. Consider

$$ES_{n1}^2 = E\left[\frac{1}{nh\sqrt{n}}\sum_{i=1}^n K(0)\frac{\varepsilon_i^2}{\upsilon_0(X_i)f(X_i)} + \frac{1}{nh\sqrt{n}}\sum_{i \ne j}^n K\left(\frac{X_i - X_j}{h}\right)\frac{\varepsilon_i\varepsilon_j\upsilon_0(X_j)}{\upsilon_0^2(X_i)f(X_i)}\right]^2,$$

which is bounded above by the sum

$$2E\left[\frac{1}{nh\sqrt{n}}\sum_{i=1}^n \frac{K(0)\varepsilon_i^2}{\upsilon_0(X_i)f(X_i)}\right]^2 + 8E\left[\frac{1}{nh\sqrt{n}}\sum_{i<j} K\left(\frac{X_i - X_j}{h}\right)\frac{\varepsilon_i\varepsilon_j\upsilon_0(X_j)}{\upsilon_0^2(X_i)f(X_i)}\right]^2. \tag{6.26}$$

By conditions (e), (k), (f), and υ_0 being bounded away from zero, and by the Cauchy–Schwarz inequality, the first term in (6.26) is bounded above by $CE\varepsilon^4/nh^2 \to 0$. The same assumptions and direct calculations show that the second term is also of the order $(nh^2)^{-1} \to 0$. Therefore, $ES_{n1}^2 = o(1)$, which implies $S_{n1} = o_p(1)$.

Under the same conditions as mentioned above, by a conditioning argument, and using the fact $(X_i, \varepsilon_i), 1 \le i \le n$ are independent random vectors,

$$ES_{n2}^2 = \frac{1}{n}\sum_{i=1}^n E\frac{\varepsilon_i^2}{\upsilon_0^4(X_i)f^2(X_i)}\left[\frac{1}{nh}\sum_{j=1}^n K\left(\frac{X_j - X_i}{h}\right)m(X_j) - g(X_i)\right]^2 \le CE\left[\frac{1}{nh}\sum_{j=1}^n K\left(\frac{X_j - X_1}{h}\right)m(X_j) - g(X_1)\right]^2 = o(1),$$

by a routing bias and variance decomposition argument, where C is some positive constant. This implies that $S_{n2} = o_p(1)$. This completes the proof of the first claim in (6.25). The second claim in (6.25) can be proved using the similar argument as in proving $S_{n2} = o_p(1)$, thereby completing the proof of $\sup_{x\in\mathcal{I}}|W_{1n2}(x)| = o_p(1)$. □

Proof of Lemma 6.3. Use the elementary equality $\sqrt{a} - \sqrt{b} = (a-b)/(\sqrt{a} + \sqrt{b})$, to obtain

$$S_n(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\psi(\varepsilon_i)L(X_i)(\widehat{\mu}_4(X_i) - \mu_4(X_i))I(X_i \le x)}{\sqrt{\widehat{\mu}_4(X_i)}} + \sqrt{\mu_4(X_i)}.$$

Let $K_{hij} = K((X_i - X_j)/h)/h$, and $\widehat{f}(x) = \sum_{j=1}^n K((x - X_j)/h)/(nh)$. Then

$$\widehat{\mu}_4(X_i) = \frac{\sum_{j=1}^n K_{hij}[(Y_j - m_n(X_j))^2/\nu_n^2(X_j) - 1]^2}{n\widehat{f}(X_i)}. \tag{6.27}$$

Let $\Delta m_n(X_j) = m_0(X_j) - m_n(X_j)$, $\Delta\nu_n^2(X_j) = \nu_0^2(X_j) - \nu_n^2(X_j)$ and $D_n(X_j) = \nu_0^2(X_j)/\nu_n^2(X_j) - 1$. Then

$$\begin{aligned}\left[\frac{(Y_j - m_n(X_j))^2}{\nu_n^2(X_j)} - 1\right]^2 &= \left[\frac{(\nu_0(X_j)\varepsilon_j - \Delta m_n(X_j))^2 - \nu_0^2(X_j) - \Delta\nu_n^2(X_j)}{\nu_0^2(X_j)}(D_n(X_j)+1)\right]^2 \\ &= \left[\frac{\nu_0^2(X_j)(\varepsilon_j^2 - 1) + (\Delta m_n(X_j))^2 - 2\varepsilon_j\nu_0(X_j)\Delta m_n(X_j) - \Delta\nu_n^2(X_j)}{\nu_0^2(X_j)}(D_n(X_j)+1)\right]^2 \\ &= \left[(\varepsilon_j^2 - 1) + \frac{(\Delta m_n(X_j))^2}{\nu_0^2(X_j)} - \frac{2\varepsilon_j\Delta m_n(X_j)}{\nu_0(X_j)} - \frac{\Delta\nu_n^2(X_j)}{\nu_0^2(X_j)}\right]^2 (D_n(X_j)+1)^2.\end{aligned}$$

Expanding the square terms, the above expression, and hence $\widehat{\mu}_4(X_i)$ can be written as a sum of 30 terms. For example, one of these terms is $\tilde{\mu}_4(X_i) := \sum_{j=1}^n K_{hij}[\varepsilon_j^2 - 1]^2/n\widehat{f}(X_i)$. Accordingly, the following is one term in $S_n(x)$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\psi(\varepsilon_i)L(X_i)(\tilde{\mu}_4(X_i) - \mu_4(X_i))I(X_i \le x)}{2\sqrt{\mu_4(X_i)}},$$

which can be rewritten as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\psi(\varepsilon_i)L(X_i)(\tilde{\mu}_4(X_i) - \mu_4(X_i))[I(X_i \le x) - F(x)]}{2\sqrt{\mu_4(X_i)}} + \frac{F(x)}{\sqrt{n}}\sum_{i=1}^n \frac{\psi(\varepsilon_i)L(X_i)(\tilde{\mu}_4(X_i) - \mu_4(X_i))}{2\sqrt{\mu_4(X_i)}}.$$

Similar to the proof of (6.24) and (6.25) below, one can show the above two terms are $u_p(1)$. Using conditions (m2), (m3), (v1), (v2) and the $\sqrt{n}$–consistency of $\widehat{\beta}_n$, one can also show the other 29 terms are all $u_p(1)$. Therefore, (6.3) holds.

To prove (6.4), use assumption (e) to conclude

$$|\sqrt{\widehat{\mu}_4(X_i)} - \sqrt{\mu_4(X_i)}| = \frac{|\widehat{\mu}_4(X_i) - \mu_4(X_i)|}{2\sqrt{\mu_4(X_i)}} \le c|\widehat{\mu}_4(X_i) - \mu_4(X_i)|, \quad 1 \le i \le n,$$

where $c = 1/2d$. Note that

$$E\sum_{i=1}^n(\widehat{f}(X_i) - f(X_i))^4 = E\sum_{i=1}^n\left[\frac{1}{nh}\sum_{j=1}^n K\left(\frac{X_i - X_j}{h}\right) - f(X_i)\right]^4 \le 8nE\left[\frac{1}{nh}\sum_{j=2}^n K\left(\frac{X_1 - X_j}{h}\right) - f(X_1)\right]^4 + \frac{8n}{(nh)^4}K^4(0).$$

Let $H(X_1, X_j) = h^{-1}K((X_1 - X_j)/h)$, then

$$E\left[\frac{1}{nh}\sum_{j=2}^n K\left(\frac{X_1 - X_j}{h}\right) - f(X_1)\right]^4 \le 8E\left\{\frac{1}{n}\sum_{j=2}^n(H(X_1, X_j) - E[H(X_1, X_j)|X_1])\right\}^4 + 8E\left(\frac{n-1}{n}E[H(X_1, X_j)|X_1] - f(X_1)\right)^4.$$

Routine argument show that the first term on the right is $O(1/(nh)^3) + O(1/(nh)^2)$, and the second term $O(h^4)$. Therefore, from assumption (h),

$$E\sum_{i=1}^n(\widehat{f}(X_i) - f(X_i))^4 = O(1/n^3h^4) + O(1/n^2h^3) + O(1/nh^2) + O(nh^4) = o_p(1).$$

Hence

$$\max_{1 \le i \le n}|\widehat{f}(X_i) - f(X_i)|^4 \le \sum_{i=1}^n(\widehat{f}(X_i) - f(X_i))^4 = o_p(1), \tag{6.28}$$

which, in turn, implies $\max_{1 \le i \le n}|\widehat{f}(X_i) - f(X_i)| = o_p(1)$. Similar arguments lead to

$$\max_{1 \le i \le n}\left|\frac{1}{nh}\sum_{j=1}^n K\left(\frac{X_i - X_j}{h}\right)(\varepsilon_i^2 - 1)^2 - f(X_i)E((\varepsilon^2 - 1)^2|X = X_i)\right| = o_p(1),$$

$$\max_{1 \le i \le n}|[\widehat{f}(X_i) - f(X_i)]E((\varepsilon^2 - 1)^2|X = X_i)| = o_p(1). \tag{6.29}$$

In turn, (6.27)–(6.29), and assumption (f) imply $\max_{1 \le i \le n}|\widehat{\mu}_4(X_i) - \mu_4(X_i)| = o_p(1)$, which readily implies (6.4). □

Acknowledgments

Authors would like to thank the two referees for their constructive comments.

References

Amemiya, T., 1985. Advanced Econometrics. Harvard University Press, Cambridge, MA.
Breusch, T.S., Pagan, A.R., 1979. Assessing the adequacy of variance function in heteroscedastic regression models. Biometrics 63, 1218–1225.
Billingsley, P., 1968. Convergence of Probability Measure. Wiley, New York, London, Sydney.
Carroll, R.J., Ruppert, D., 1988. Transformation and Weighting in Regression. Chapman & Hall, New York.
Cook, R.D., Weisberg, S., 1983. Diagnostics for heteroscedasticity in regression. Biometrika 70, 1–10.
Dette, H., 2002. A consistent test for heteroscedasticity in nonparametric regression based on the kernel method. J. Statist. Plann. Inference 103, 311–329.
Dette, H., Munk, A., 1998. Testing heteroscedasticity in nonparametric regression. J. Roy. Statist. Soc. B 60, 693–708.
Dette, H., Neumeyer, N., Keilegom, I.V., 2007. A new test for the parametric form of the variance function in non-parametric regression. J. Roy. Statist. Soc. B 69, 903–917.
Dette, H., Hetzler, B., 2009. A simple test for the parametric form of the variance function in nonparametric regression. Ann. Inst. Statist. Math. 61 (4), 861–886.
Dette, H., Hetzler, B., 2008. A martingale-transform goodness-of-fit test for the form of the conditional variance 〈http://arXiv.org/abs/0809.4914?context=stat〉.
Diblasi, A., Bowman, A., 1997. Testing for constant variance in a linear model. Statist. Probab. Lett. 33, 95–103.
Harrison, M.J., McCabe, B.P.M., 1979. A test for heteroscedasticity based on least squares residuals. J. Amer. Statist. Assoc. 74, 494–500.
Jennrich, R.J., 1969. Asymptotic properties of non-linear least squares estimators. Ann. Math. Statist. 40, 633–643.
Khmaladze, E.V., 1981. Martingale approach in the theory of goodness-of-fit tests. Theory Probab. Appl. 26, 240–257.
Khmaladze, E.V., 1988. An innovation approach to goodness-of-fit tests in $\mathbb{R}^m$. Ann. Statist. 16, 1503–1516.
Khmaladze, E.V., 1993. Goodness of fit problem and scanning innovation martingales. Ann. Statist. 21 (2), 798–829.
Khmaladze, E.V., Koul, H.L., 2004. Martingale transforms goodness-of-fit tests in regression models. Ann. Statist. 32, 995–1034.
Koenker, R., Bassett, G., 1981. Robust tests for heteroscedasticity based on regression quantiles. Econometrica 50, 43–61.
Koul, H.L., 2006. Model diagnostics via martingale transforms: a brief review, Frontiers in Statistics. Imperial College Press, London, pp. 183–206.
Koul, H.L., Song, W., 2008. Model checking with Berkson measurement errors. J. Statist. Plann. Inference 138, 1615–1628.
Liang, H., 2006. Checking linearity of non-parametric component in partially linear models with an application in systemic inflammatory response syndrome study. Statist. Methods Medical Res. 15, 273–284.
Liero, H., 2003. Testing homoscedasticity in nonparametric regression. Nonparametric Statist. 15 (1), 31–51.
Nolan, D., Pollard, D., 1987. U-process: rates of convergecne. Ann. Statist. 15, 780–799.
Stute, W., Thies, S., Zhu, L.X., 1998. Model checks for regression: an innovation process approach. Ann. Statist. 26, 1916–1934.
Stute, W., Zhu, L.X., 2002. Model checks for generalized linear models. Scand. J. Statist. 29, 535–545.
von Neumann, J., 1941. Distribution of the ratio of the mean square successive difference to the variance. Ann. Math. Statist. 12, 367–395.
Wang, L., Zhou, X.H., 2007. 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 for heteroscedasticity. Econometrica 48, 817–838.
Wu, C.-F., 1981. Asymptotic theory of nonlinear least squares estimation. Ann. Statist. 9, 501–513.
Zheng, J.X., 1996. A consistent test of functional form via nonparametric estimation techniques. J. Econometrics 75, 263–289.
Zhu, L.X., Fujikoshi, Y., Naito, K., 2001. Heteroscedasticity checks for regression models. Sci. China Ser. A 44 (10), 1236–1252.
Zhu, L.X., Ng, K.W., 2003. Checking the adequacy of a partial linear model. Statist. Sinica 13, 763–781.