Journal of Multivariate Analysis 102 (2011) 579–600

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Minimum distance conditional variance function checking in heteroscedastic regression models

Nishantha Samarakoon, Weixing Song *

Kansas State University, Manhattan, United States

A R T I C L E   I N F O

Article history:
Received 11 February 2010
Available online 25 November 2010

AMS subject classifications:
primary 62G08
secondary 62G10

Keywords:
Kernel estimator
Lack-of-fit test
Heteroscedasticity
Variance function
L_2 distance

A B S T R A C T

This paper discusses a class of minimum distance tests for fitting a parametric variance function in heteroscedastic regression models. These tests are based on certain minimized L_2 distances between a nonparametric variance function estimator and the parametric variance function estimator. The paper establishes the asymptotic normality of the proposed test statistics and that of the corresponding minimum distance estimator under the fitted model. These estimators turn out to be $\sqrt{n}$-consistent. Consistency of this sequence of tests at some fixed alternatives and asymptotic power under some local nonparametric alternatives are also discussed. Some simulation studies are conducted to assess the finite sample performance of the proposed test.

Published by Elsevier Inc.

1. Introduction

Regression analysis frequently assumes homoscedasticity, while real data generated from an application often exhibit certain heteroscedastic structure. The importance of detecting heteroscedasticity is now widely recognized among researchers and practitioners, in that the efficient statistical inference for the regression analysis should take the heteroscedasticity into account, when the homoscedasticity assumption fails. This may result in some proper transformation of the data, or modified adaptive inference procedures, such as weighted least squares, modified likelihood procedures, or the choice of a variable bandwidth if nonparametric smoothing is used.

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 [12,2,21,14,5,4,11], and the references therein. Most of the aforementioned works, also including much recent research, focus on checking whether the variance function is constant or not, or the testing of homoscedasticity. For when the covariate is one dimensional, Dette and Munk [9] proposed a test based on the best L_2 approximation of the variance function by a constant, and the resulting test is shown to be consistent. Inspired by the idea that the problem of testing heteroscedasticity is equivalent to the problem of testing pseudo-residuals for a constant mean, Dette [6] constructed a testing procedure which can detect $1/\sqrt{nh^{1/2}}$ local alternatives, where n is the sample size and h is the bandwidth in the kernel smoothing. Liero [18] suggested a test statistic using a L_2 distance between nonparametric variance estimations in both null and alternative models. In the multidimensional covariate case, a Cramer–von Mises type of test based on cumulative estimated residuals was proposed by Zhu et al. [25]. The asymptotic distributions of the aforementioned test statistics are usually complicated and are not asymptotically distribution free (ADF). Some resampling procedures are employed to find the critical or p-values. Compared to the research on the testing of homoscedasticity, even fewer rigorous procedures for testing the adequacy of a given variance function are proposed in the literature. For when the covariate is one dimensional,

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

0047-259X/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.jmva.2010.11.003

Dette et al. [10] proposed a Kolmogorov–Smirnov and a Cramer–von Mises type of test from the difference between the empirical residual processes under the null and the alternative hypothesis. Recently, using the Khamaladze transformation, Dette and Hetzler [8] considered a standardized version of the empirical process of pseudo-residuals, which was proposed in [7]; some asymptotically distribution-free tests are obtained. A similar test procedure for a random design case is proposed by Koul and Song [17]. For the multivariate covariate case, Wang and Zhou [20] proposed a kernel type nonparametric test based on the framework of Zheng [24]. The consistency and local power were discussed. The test can detect $1/\sqrt{nh^{d/2}}$ local alternatives, where n is the sample size, d is the dimension of the covariate, and h is the bandwidth in the kernel smoothing.

In this paper, we shall propose a new test procedure for assessing the adequacy of fitting the variance function with a parametric function in the heteroscedastic regression model. Specifically, consider the following regression model:

$$Y = m(X;\beta) + \sqrt{v(X)}\varepsilon, \tag{1.1}$$

where Y is a one-dimensional response variable, X is a d-dimensional explanatory variable, β is a p-dimensional vector of unknown parameters, $v(\cdot)$ is the variance function, and ε is the random error with

$$E(\varepsilon|X) = 0, \qquad E(\varepsilon^2|X) = 1. \tag{1.2}$$

We are interested in testing the hypothesis

$$H_0 : v(x) = v(x;\beta,\theta), \quad \text{for some } \theta \in \Theta,\ \beta \in \mathbb{R}^p, \tag{1.3}$$

where Θ is a compact subset in $\mathbb{R}^q$. From (1.1) and (1.2), we have $E[(Y - m(X;\beta))^2|X] = v(X)$. If β is known, then testing the hypothesis (1.3) is equivalent to testing the regression function in the model

$$(Y - m(X;\beta))^2 = v(X) + \xi \tag{1.4}$$

in which $(Y - m(X;\beta))^2$ is viewed as the response, and $\xi = (Y - m(X;\beta))^2 - E((Y - m(X;\beta))^2|X)$ which is uncorrelated with X. Like Koul and Ni [15] (KN), we can build the test statistic based on the quantities

$$T_n^*(\beta,\theta) = \int_{\mathcal{C}} \left[\frac{\sum_{i=1}^n K_h(x - X_i)(Y_i - m(X_i;\beta))^2}{\sum_{i=1}^n K_w(x - X_i)} - v(x;\beta,\theta) \right]^2 \mathrm{d}G(x) \tag{1.5}$$

and

$$T_n(\beta,\theta) = \int_{\mathcal{C}} \left[\frac{\sum_{i=1}^n K_h(x - X_i)[(Y_i - m(X_i;\beta))^2 - v(X_i;\beta,\theta)]}{\sum_{i=1}^n K_w(x - X_i)} \right]^2 \mathrm{d}G(x), \tag{1.6}$$

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, $K_h(\cdot) = h^{-d}K(\cdot/h)$, and h is the bandwidth. In real applications, β is usually unknown. A natural way to proceed is to replace β with an estimator. Indeed, many estimating procedures can provide estimators, say $\hat{\beta}_n$, of β, such that $\sqrt{n}(\hat{\beta}_n - \beta_0) \Rightarrow N(0, \Sigma_{\beta_0,\theta_0})$ in distribution, where $\Sigma_{\beta_0,\theta_0}$ is a $p \times p$ positive definite matrix that may depend on the true parameters β_0 and θ_0. Using one such estimator, θ can be estimated by either one of the following:

$$\theta_n^* = \arg\min_{\theta\in\Theta} T_n^*(\hat{\beta}_n,\theta), \qquad \hat{\theta}_n = \arg\min_{\theta\in\Theta} T_n(\hat{\beta}_n,\theta). \tag{1.7}$$

The integrand in T_n^* is not centered, and the nonparametric estimator has a non-negligible asymptotic bias, so the statistic $T_n^*(\hat{\beta}_n,\theta_n^*)$ does not have a desirable asymptotic distribution under the null hypothesis. Moreover, we can show that θ_n^* is consistent, but the sequence $\sqrt{n}(\theta_n^* - \theta_0)$ may not even be tight. A similar phenomenon occurs when testing the regression function in the traditional regression models. See KN. To overcome this difficulty, we will use $T_n(\hat{\beta}_n,\hat{\theta}_n)$ to construct the test statistic.

The rest of the paper is organized as follows. Section 2 states various assumptions. The main results, including the consistency and asymptotic normality of the minimum distance (MD) estimator, and asymptotic normality of the MD test statistic, together with the consistency, and the local power of the MD test, are discussed in Section 3. Section 4 contains simulation studies, and the proofs of the main results are included in Section 5.

2. Assumptions

From now on, $\dot{m}(x,\beta)$ will denote the derivative of m with respect to β, $\dot{v}_\beta(x,\beta,\theta)$ the derivative of v with respect to β, and $\dot{v}_\theta(x,\beta,\theta)$ the derivative of v with respect to θ. The assumptions needed are listed below.

As regards the errors, the underlying design variables and the weighting measure G, we assume:

(e1) The random variables $\{(X_i, Y_i) : X_i \in \mathbb{R}^d, Y_i \in \mathbb{R}, i = 1, 2, \ldots, n\}$ are i.i.d. with the regression functions $E(Y|X = x) = m(x, \beta)$ and $E((Y - m(X, \beta))^2|X = x) = v(x)$ satisfying $\int v^2(x)\mathrm{d}G(x) < \infty$, where G is a σ-finite measure on $\mathbb{R}^d$.

(e2) $E[((Y - m(X, \beta))^2 - v(X))^2] < \infty$, and the function $\tau(x) = E[((Y - m(X, \beta))^2 - v(X))^2|X = x]$ is a.s. (G) continuous on $\mathcal{C}$.

(e3) $E[(Y - m(X, \beta))^2 - v(X)]^{2+\delta} < \infty$ for some $\delta > 0$.

(e4) $E[(Y - m(X, \beta))^2 - v(X)]^4 < \infty$.

(f1) X has a uniformly continuous density f that is bounded below on $\mathcal{C}$.

(f2) The density function f is twice continuously differentiable.

(g) G has a continuous density function g.

As regards the kernel function and the bandwidth, we assume:

(k) The kernel function K is a positive symmetric square integrable densities on $[-1, 1]^d$; it also satisfies the Lipschitz condition.

(h1) $h, w \to 0, nh^{2d}, nw^{2d} \to \infty$ as $n \to \infty$.

(h2) $w \sim n^{-a}$, where $a < \min(1/2d, 4/d(d + 4))$.

As regards the regression function m, we shall assume:

(m1) For any fixed x, $m(x; \beta)$ is differentiable with respect to β and its derivative is square integrable, that is $E\|\dot{m}(X; \beta)\|^2 < \infty$, $\int_{\mathcal{C}} \|\dot{m}(x; \beta)\|^4\mathrm{d}G(x) < \infty$.

(m2) For any $\sqrt{n}$ consistent estimator of β_0,

$$\sqrt{n} \sup_{1\le i\le n} |m(X_i; \hat{\beta}_n) - m(X_i; \beta_0) - (\hat{\beta}_n - \beta_0)'\dot{m}(X_i; \beta_0)| = o_p(1).$$

As regards the variance function v, we make the following assumptions:

(v1) For all β and θ, $v(x; \beta, \theta)$, $\dot{v}_\beta(x; \beta, \theta)$, and $\dot{v}_\theta(x; \beta, \theta)$ are a.s. continuous in x with respect to the integrating measure G.

(v2) The parametric family of variance functions $v(x; \beta_0, \theta)$ is identifiable with respect to θ, that is, if $v(x; \beta_0, \theta_1) = v(x; \beta_0, \theta_2)$ for almost all x (G), then $\theta_1 = \theta_2$.

(v3) $v(x; \beta, \theta)$ is Lipschitz continuous with respect to β and θ. That is, for some positive continuous function ℓ on $\mathcal{C}$, and for some $\alpha > 0$,

$$|v(x; \beta_1, \theta_1) - v(x; \beta_2, \theta_2)| \le \ell(x)[\|\beta_1 - \beta_2\|^\alpha + \|\theta_1 - \theta_2\|^\alpha]$$

holds for all $\beta_1, \beta_2, \theta_1$ and θ_2.

(v4)

$$\limsup_{n\to\infty} P\left(\sup \frac{|v(X_i; \beta, \theta) - v(X_i; \beta, \theta_0) - \dot{v}'_\theta(X_i; \beta, \theta_0)(\theta - \theta_0)|}{\|\theta - \theta_0\|} \ge \varepsilon\right) = 0$$

where the supremum is taken over the set $\{1 \le i \le n; \beta \in \Gamma; \sqrt{nh^d}\|\theta - \theta_0\| \le k\}$ for any $k > 0$, and

$$\limsup_{n\to\infty} P\left(\sup \frac{|v(X_i; \beta, \theta) - v(X_i; \beta_0, \theta) - \dot{v}'_\beta(X_i; \beta_0, \theta)(\beta - \beta_0)|}{\|\beta - \beta_0\|} \ge \varepsilon\right) = 0$$

where the supremum is taken over the set $\{1 \le i \le n; \theta \in \Theta; \sqrt{n}\|\beta - \beta_0\| \le k\}$ for any $k > 0$.

(v5)

$$\limsup_{n\to\infty} P\left(\sup h^{-d/2}\|\dot{v}_\theta(X_i; \beta, \theta) - \dot{v}_\theta(X_i; \beta, \theta_0)\| \ge \varepsilon\right) = 0$$

where the supremum is taken over the set $\{1 \le i \le n; \beta \in \Gamma; \sqrt{nh^d}\|\theta - \theta_0\| \le k\}$ for any $k > 0$, and

$$\limsup_{n\to\infty} P\left(\sup h^{-d/2}\|\dot{v}_\beta(X_i; \beta, \theta) - \dot{v}_\beta(X_i; \beta_0, \theta)\| \ge \varepsilon\right) = 0$$

where the supremum is taken over the set $\{1 \le i \le n; \theta \in \Theta; \sqrt{nh^d}\|\beta - \beta_0\| \le k\}$ for any $k > 0$.

Conditions (e1), (e2), (f1), (g), (k), (h1), (m1), (m2), and (v1)–(v3) suffice for the consistency of the MD estimator $\hat{\theta}_n$, while these plus conditions (e3), (f2), (v4), (v5) and (h2) are needed for the asymptotic normality of $\hat{\theta}_n$. To show the asymptotic normality of the test statistic, we need assumptions (e1), (e2), (e4), (f1), (f2), (k), (h1), (h2), (m1), (m2) and (v1)–(v5). It is easy to see that (h2) implies $nw^{2d} \to \infty$ in (h1). (m1) and (m2) are two smooth conditions for the regression function. (v2) is a condition for identifiability of θ, and (v1), (v3)–(v5) are smooth conditions for the variance function. Note that the conditions (v1)–(v5) are trivially satisfied by the model $v(x; \beta, \theta) = (\beta', \theta')r(x)$ provided that the components of r are continuous on $\mathcal{C}$.

It is well known that under conditions (f1), (k), (h1), and (h2), we have

$$\sup_{x\in\mathcal{C}}|\hat{f}_h(x)-f(x)|=o_p(1),\qquad \sup_{x\in\mathcal{C}}|\hat{f}_w(x)-f(x)|=o_p(1),\qquad \sup_{x\in\mathcal{C}}\left|\frac{f(x)}{\hat{f}_w(x)}-1\right|=o_p(1). \tag{2.1}$$

See [19]. These conclusions are often used in the proofs. In the sequel, "$\Longrightarrow_d$" means convergence in distribution.

3. Main results

3.1. Consistency and asymptotic normality of $\hat{\theta}_n$

The following theorem states the consistency of the estimators θ_n^* and $\hat{\theta}_n$.

Theorem 3.1. *Suppose* (e1), (e2), (f1), (k), (m1), (m2), (v1)–(v3), (h1) *and* (h2) *hold. Then under* H_0, $\theta_n^*\to\theta_0$, $\hat{\theta}_n\to\theta_0$ *in probability.*

To show the asymptotic normality of the minimum distance (MD) estimator $\hat{\theta}_n$, we shall assume that $\hat{\beta}_n$ has the following approximate linear expansion:

$$\sqrt{n}(\hat{\beta}_n-\beta_0)=\frac{1}{\sqrt{n}}\sum_{i=1}^n L(Y_i,X_i;\beta_0,\theta_0)+o_p(1) \tag{3.1}$$

with $EL(Y,X;\beta_0,\theta_0)=0$, $\Sigma_L=EL(Y,X;\beta_0,\theta_0)L'(Y,X;\beta_0,\theta_0)>0$, and

$$E\|L(Y,X;\beta_0,\theta_0)\|^{2+\delta}<\infty. \tag{3.2}$$

In fact, estimators from quite a few standard estimation procedures have such expansions; among them are the least squares, weighted least squares and quasi-likelihood procedures. Therefore, this assumption is not a stringent one. We also need the following condition on L:

(l) $\rho(x)=E[(\varepsilon^2-1)L(Y,X;\beta_0,\theta_0)|X=x]$ is a.s. continuous in x with respect to integrating measure G.

In many cases, L is a linear function of ε, so $\rho(x)$ may be equal to 0 if additionally $E(\varepsilon^3|X=x)=0$ is assumed. This is the case in linear regression.

Now define

$$\Pi=\int_{\mathcal{C}} v_\theta(x;\beta_0,\theta_0)v_\beta'(x;\beta_0,\theta_0)\mathrm{d}G(x),\qquad \Sigma_0=\int_{\mathcal{C}}\dot{v}_\theta(x;\beta_0,\theta_0)\dot{v}_\theta'(x;\beta_0,\theta_0)\mathrm{d}G(x), \tag{3.3}$$

$$\Omega=\int\frac{\tau(x)v^2(x;\beta_0,\theta_0)\dot{v}_\theta(x;\beta_0,\theta_0)\dot{v}_\theta'(x;\beta_0,\theta_0)g^2(x)}{f^2(x)}\mathrm{d}x,$$

$$M=\int\rho(x)v(x;\beta_0,\theta_0)\dot{v}_\theta'(x;\beta_0,\theta_0)g(x)\mathrm{d}x.$$

The following theorem states the asymptotic normality of $\hat{\theta}_n$.

Theorem 3.2. *Suppose* (e1)–(e3), (f1), (f2), (g), (v1)–(v5), (m1)–(m2), (h2) *and* (l) *hold. Then under* H_0, $\sqrt{n}(\hat{\theta}_n-\theta_0)\Longrightarrow_d N(0,\Sigma_0^{-1}\Sigma\Sigma_0^{-1})$ *where* $\Sigma=\Omega+\Pi\Sigma_L\Pi+\Pi M+M'\Pi$.

If $\rho(x)=0$ in (l), then $M=0$, and the asymptotic variance of $\hat{\theta}_n$ is simply $\Omega+\Pi\Sigma_L\Pi$.

3.2. Asymptotic normality of the MD test statistic

To present the asymptotic normality of the MD statistic $T_n(\hat{\beta}_n,\hat{\theta}_n)$, we define

$$C_n(\beta,\theta)=\frac{1}{n^2}\sum_{i=1}^n\int_{\mathcal{C}}K_h^2(x-X_i)[(Y_i-m(X_i;\beta))^2-v(X_i;\beta,\theta)]^2\mathrm{d}\hat{\psi}_w(x)$$

$$\Gamma_n(\beta,\theta)=\frac{2h^d}{n^2}\sum_{i\neq j}\left(\int_{\mathcal{C}}K_h(x-X_i)K_h(x-X_j)\xi_i(\beta,\theta)\xi_j(\beta,\theta)\mathrm{d}\hat{\psi}_w(x)\right)^2$$

$$\Gamma=2\int_{\mathcal{C}}\frac{\tau^2(x)g^2(x)}{f^2(x)}\mathrm{d}x\cdot\int\left(\int K(u)K(u+v)\mathrm{d}u\right)^2\mathrm{d}v, \tag{3.4}$$

where

$$\xi(\beta,\theta)=(Y-m(X;\beta))^2-v(X;\beta,\theta), \tag{3.5}$$

$\mathrm{d}\hat{\psi}_w(x)=\mathrm{d}G(x)/\hat{f}_w^2(x)$, and $\hat{f}_w(x)=n^{-1}\sum_{i=1}^n K_h(x-X_i)$. We have the following theorem.

Theorem 3.3. *Suppose* (e1), (e2), (e4), (f1), (f2), (g), (v1)–(v5), (m1)–(m2), (h2) *and* (l) *hold. Then under* H_0*, we have* $nh^{d/2}\Gamma_n^{-1/2}(\hat\beta_n,\hat\theta_n)(T_n(\hat\beta_n,\hat\theta_n)-C_n(\hat\beta_n,\hat\theta_n))\Longrightarrow_d N(0,1)$.

Thus, the test that rejects H_0 whenever $|nh^{d/2}\Gamma_n^{-1/2}(\hat\beta_n,\hat\theta_n)(T_n(\hat\beta_n,\hat\theta_n)-C_n(\hat\beta_n,\hat\theta_n))|\ge z_{\alpha/2}$ is of the asymptotic size α, where z_α is the $100(1-\alpha)\%$ percentile of the standard normal distribution.

Remark 1. The test statistic has a relatively complicated form, which makes the implementation of the test procedure not easy. In particular, the integrations in T_n, Γ_n, C_n usually have no tractable expressions, so Riemann-sum approximations are necessary for carrying out the test. But the test statistic can be simplified by choosing a proper weighting measure G, and using an approximately equivalent expression for $\hat\Gamma_n$. For example, choose $dG(x) = g(x)dx = \hat f_w^2(x)dx$; then $T_n(\hat\beta_n,\hat\theta_n)$, $C_n(\hat\beta_n,\hat\theta_n)$ can be simplified as

$$T_n(\hat\beta_n,\hat\theta_n)=\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)\xi_i(\hat\beta_n,\hat\theta_n)\right]^2 dx,\qquad C_n(\hat\beta_n,\hat\theta_n)=\frac{1}{n^2}\sum_{i=1}^n\int_{\mathcal{C}}K_h^2(x-X_i)\xi_i^2(\hat\beta_n,\hat\theta_n)dx,$$

where $\xi_i(\beta,\theta)$ is defined by (3.5). From (3.4), the definition of $\tau^2(x)$, and $g(x)=\hat f_w(x)^2$, a simpler consistent estimator of Γ is given by

$$\hat\Gamma_n = 2\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)\xi_i^2(\hat\beta_n,\hat\theta_n)\right]^2 dx\int\left(\int K(u)K(u+v)du\right)^2 dv.$$

When the sample size is small to moderate, the bootstrap often provides a more accurate approximation to the distribution of the test statistic than the asymptotic normal theory does. The following is a simple bootstrap algorithm for implementing the MD test procedure. It consists of six steps:

(1) For a given random sample of observations, obtain a $\sqrt{n}$-consistent estimator $\hat\beta_n$ of β under the null hypothesis. Such an estimator can be found by using a least squares procedure, pseudo-likelihood procedure, etc.
(2) Obtain the MD estimator $\hat\theta_n$ of θ by minimizing $T_n(\hat\beta_n,\theta)$ under the null hypothesis.
(3) Define $\widehat{\varepsilon}_i=[Y_i-m(X_i;\hat\beta_n)]/\sqrt{v(X_i;\hat\beta_n,\hat\theta_n)}$, $i=1,2,\ldots,n$. Center and standardize these residuals such that they have mean 0 and variance 1.
(4) Obtain a bootstrap sample from the standardized residuals in Step (3), denote them as $\widehat{\varepsilon}_i^*$, $i=1,2,\ldots,n$, and define $Y_i^*=m(X_i;\hat\beta_n)+\sqrt{v(X_i;\hat\beta_n,\hat\theta_n)}\widehat{\varepsilon}_i^*$, $i=1,2,\ldots,n$.
(5) For the bootstrap sample (X_i,Y_i^*), $i=1,2,\ldots,n$, calculate the estimator $\hat\beta_n^*$ as in Step (1) and the MD estimator $\hat\theta_n^*$ as in Step (2) under the null hypothesis. Let $\xi_i^*(\hat\beta_n^*,\hat\theta_n^*)=(Y_i^*-m(X_i;\hat\beta_n^*))^2-v(X_i;\hat\beta_n^*,\hat\theta_n^*)$. Then the bootstrap version of the test statistic is $T_n^*=\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)\xi_i^*(\hat\beta_n^*,\hat\theta_n^*)\right]^2 dx$.
(6) Repeat Steps (4) and (5) a sufficiently large number of times. For a specified significance level of the test, the critical value is then determined as the appropriate quantile of the bootstrap distribution of the test statistic.

Remark 2. Another important issue for implementing the MD test procedure is the choice of the compact set $\mathcal{C}$. A general recommendation in this regard is to select $\mathcal{C}$ such that $\mathcal{C}$ covers most of the X-range. For example, if $X\sim N(0,1)$, then we can choose $\mathcal{C}=[-3,3]$.

Remark 3. The proof of Theorem 3.3, which will be given in Section 5, indicates that the asymptotic normality of the MD test statistic continues to hold if the MD estimator $\hat\theta_n$ is replaced by any other $\sqrt{n}$-consistent estimator.

Remark 4. The implementation of the proposed test requires one to calculate the integrations in the test statistic. In the one-dimensional case, one can use the Riemann-sum approximation. But the Riemann-sum approximation might not be reliable in high dimensional integration; the curse of dimensionality will make this even worse if the sample size is small. In fact, by selecting the weighting measure G such that $dG(x)=\hat f_w^2(x)dx$, and after some algebra, we have

$$T_n=\frac{1}{n^2}\sum_{i,j=1}^n\xi_i(\hat\beta_n,\hat\theta_n)\xi_j(\hat\beta_n,\hat\theta_n)\int_{\mathcal{C}}K_h(x-X_i)K_h(x-X_j)dx.$$

For some kernel functions, such as the Epanechnikov kernel, the integrations in the above expression do have explicit forms, even in high dimensional design space. In this case, one really does not have to use Riemann sums to approximate the integrations. The same is true for $C_n(\hat\beta_n,\hat\theta_n)$ and $\hat\Gamma_n$. Another way to proceed is to use the weighting measure $dG(x)=$

$\hat{f}_w^2(x)\mathrm{d}F_n(x)$, where $F_n(x)$ is the empirical CDF of the design variables $X_i,\ i = 1, 2, \ldots, n$. By doing this, we will have

$$T_n = \frac{1}{n}\sum_{j=1}^{n}\left[\frac{1}{n}\sum_{i=1}^{n} K_h(X_j - X_i)\xi_i(\hat{\beta}_n, \hat{\theta}_n)\right]^2 I(X_j \in \mathcal{C})$$

$$C_n(\hat{\beta}_n, \hat{\theta}_n) = \frac{1}{n^3}\sum_{i,j=1}^{n} \xi_i^2(\hat{\beta}_n, \hat{\theta}_n)K_h^2(X_j - X_i)I(X_j \in \mathcal{C})$$

and

$$\hat{\Gamma}_n = \frac{2}{n}\sum_{j=1}^{n}\left[\frac{1}{n}\sum_{i=1}^{n} K_h(X_j - X_i)\xi_i^2(\hat{\beta}_n, \hat{\theta}_n)\right]^2 I(X_j \in \mathcal{C})\int\left(\int K(u)K(u+v)\mathrm{d}u\right)^2 \mathrm{d}v.$$

Then, again, one does not to have to use Riemann sums to approximate the integrations and to worry about the curse of dimensionality.

3.3. Consistency and local power of the MD test

In this section, we show that, under some regularity conditions, the MD test is consistent for certain fixed alternatives, and has non-trivial asymptotic power against a large class of $1/\sqrt{nh^{d/2}}$-local nonparametric alternatives.

3.3.1. Consistency

Let $v_1(x)$ be a known positive and real valued function such that $v_1 \notin \{v(x; \beta, \theta) : \beta \in \Gamma, \theta \in \Theta\}$. Consider the alternative hypothesis $H_a : v(x) = v_1(x)$, for all $x \in \mathbb{R}^d$. Suppose the true value of β under H_a is β_a, the estimator $\hat{\beta}_n$ is usually not a consistent estimator for β_0. But under some regularity conditions, it is a consistent estimator of some other value, say β_a; moreover, $\hat{\beta}_n$ is still asymptotically normal. See [13,22,23] for a further discussion. The MD estimators $\hat{\theta}_n$ have the same property. Koul and Song [16] discussed a similar question in the regression model with Berkson measurement error, but their argument also applies to the current set-up. So, without loss of generality, we assume now that the estimators $\hat{\beta}_n$ and the MD estimator $\hat{\theta}_n$ used in the test statistic satisfy

$$\sqrt{n}(\hat{\beta}_n - \beta_a) = O_p(1), \qquad \sqrt{n}(\hat{\theta}_n - \theta_a) = O_p(1) \tag{3.6}$$

for some $\beta_a \in \Gamma$, $\theta_a \in \Theta$. Define $m_0(x) = m(x; \beta_0)$, $m_a(x) = m(x; \beta_a)$, $v_a(x) = v(x; \beta_a, \theta_a)$, and

$$\begin{aligned}\Delta = {} & \int_{\mathcal{C}}([m_0(x) - m_a(x)]^2 + [v_1(x) - v_a(x)]^2)^2\mathrm{d}G(x) + 4\int_{\mathcal{C}}[m_0(x) - m_a(x)]^2(\sqrt{v_1(x)} - \sqrt{v_a(x)})\sqrt{v_a(x)}\mathrm{d}G(x) \\ & + 4\int_{\mathcal{C}}(\sqrt{v_1(x)} - \sqrt{v_a(x)})^2\sqrt{v_a(x)}\mathrm{d}G(x) + 4\int_{\mathcal{C}}(\sqrt{v_1(x)} - \sqrt{v_a(x)})^3\sqrt{v_a(x)}\mathrm{d}G(x).\end{aligned}$$

Then the following indicates the consistency of the MD test procedure.

Theorem 3.4. *Suppose the conditions in Theorem 3.3 hold with β_0, θ_0 replaced by β_a, θ_a. Then, under H_a, if (3.6) holds and $\Delta > 0$, for $0 < \alpha < 1$, the test that rejects H_0 whenever $|nh^{d/2}\Gamma_n^{-1/2}(\hat{\beta}_n, \hat{\theta}_n)(T_n(\hat{\beta}_n, \hat{\theta}_n) - C_n(\hat{\beta}_n, \hat{\theta}_n))| \geq z_{\alpha/2}$ is consistent for H_a.*

3.3.2. Local power

Sometimes it is desirable to investigate the performance of a test for local alternatives, since the consistency tells us nothing about the power when the sample size is relatively small. Let $\delta(x)$ be a real valued function such that $\int_{\mathcal{C}}\delta^2(x)\mathrm{d}G(x) < \infty$. Here we shall study the asymptotic power of the proposed MD test against the local alternatives

$$H_{\mathrm{Loc}} : v(x) = v(x; \beta_0, \theta_0) + c_n\delta(x), \quad \forall x \in \mathbb{R}^d. \tag{3.7}$$

Under H_{Loc}, the regression model has the form of $Y = m(X; \beta_0) + \sqrt{v(X; \beta_0, \theta_0) + c_n\delta(X)}\varepsilon$. We will assume that the estimators $\hat{\beta}_n, \hat{\theta}_n$ used in the test statistic have the same asymptotic properties as in the null case. Then we have the following theorem.

Theorem 3.5. *Suppose the conditions in Theorem 3.3 hold and $c_n = 1/\sqrt{nh^{d/2}}$. Under the local alternatives H_{Loc},*

$$nh^{d/2}\Gamma_n^{-1/2}(\hat{\beta}_n, \hat{\theta}_n)(T_n(\hat{\beta}_n, \hat{\theta}_n) - C_n(\hat{\beta}_n, \hat{\theta}_n)) \Longrightarrow_d N\left(\Gamma^{-1/2}\int_{\mathcal{C}}\delta^2(x)\mathrm{d}G(x), 1\right).$$

Table 1
Mean and MSE of $\hat{\theta}_1$.

	100	200	300	400	500	800	1000
Mean	1.9747	1.9759	1.9778	1.9929	1.9981	1.9932	1.9914
MSE	0.0970	0.0443	0.0316	0.0259	0.0181	0.0123	0.0091

Table 2
Mean and MSE of $\hat{\theta}_2$.

	100	200	300	400	500	800	1000
Mean	0.1002	0.0941	0.1055	0.1037	0.0968	0.0968	0.0986
MSE	0.1151	0.0567	0.0397	0.0291	0.0263	0.0136	0.0116

Table 3
Empirical size and power for $a = 1$.

Model \ n	100	200	300	400	500	800	1000
Model 0	0.044	0.044	0.046	0.036	0.048	0.042	0.044
Model 1	0.124	0.194	0.256	0.316	0.414	0.568	0.650
Model 2	0.236	0.334	0.466	0.608	0.658	0.872	0.930
Model 3	0.298	0.416	0.616	0.712	0.806	0.950	0.958

4. Simulation study

To investigate the finite sample performance of the MD test procedure, we generate the sample from the following models:

$$\text{Model 0: } Y = \beta_1 + \beta_2 X + \sqrt{\theta_1 + \theta_2 X}\varepsilon,$$
$$\text{Model 1: } Y = \beta_1 + \beta_2 X + \sqrt{\theta_1 + \theta_2 X + 0.5X^2}\varepsilon,$$
$$\text{Model 2: } Y = \beta_1 + \beta_2 X + \sqrt{\theta_1 + \theta_2 X + 0.8X^2}\varepsilon,$$
$$\text{Model 3: } Y = \beta_1 + \beta_2 X + \sqrt{\theta_1 + \theta_2 X + X^2}\varepsilon.$$

The data from model 0 are used to study the empirical level, while data from models 1–3 are used to study the empirical power of the test. In the simulation, $\varepsilon \sim N(0, 1), X \sim N(0, 1), \beta_1 = 1, \beta_2 = 2, \theta_1 = 2$ and $\theta_2 = 0.1$. The kernel function K is chosen to be $K(u) = 3(1 - u^2)I(|u| \le 1)/4$; thus, the integral $\int \left[\int K(u)K(u + v)\mathrm{d}u\right]^2 \mathrm{d}v = 0.4338$. The bandwidth h is chosen to be $an^{-1/3}$, where a is some positive constant, and the sample sizes are taken to be $n = 100, 200, 300, 400, 500, 800$ and 1000. The compact set $\mathcal{C}$ is chosen to be $[-3, 3]$ and the integration is approximated by a Riemann sum with $[-3, 3]$ being equally divided into 300 subintervals. The test is calculated with 500 simulation runs with the nominal level of 0.05. The simulated level thus has a Monte Carlo error of $(0.05 \times 0.95/500)^{1/2} \approx 1\%$. We use 400 bootstrap samples per run to obtain the critical value c_α^*. The empirical size and power are computed by using $\#\{T_n(\hat{\beta}_n, \hat{\theta}_n) \ge c_\alpha^*\}/500$.

For $a = 1$, Tables 1 and 2 report the MD estimates of θ_1 and θ_2. For various sample sizes, the mean of the MD estimate for θ_1 is around the true value 2, and the MSE, as we expected, decreases when sample size increases. The same is true for the MD estimate of θ_2.

Table 3 shows the empirical size and power of the MD test when $a = 1$. The simulation study shows that the empirical levels are close to the nominal level for all the cases chosen. The empirical powers against all alternative models get bigger when sample size gets larger. For fixed sample size, the alternative model 1 has the smallest power; then the power becomes bigger when the alternative model becomes further apart (as the coefficient of x^2 changes from 0.5 to 1) from the null model.

To see the effect of the bandwidth on the performance of the MD test, we also conduct simulation studies for $a = 0.8$. Table 4 shows the simulation result. Comparing to the case of $a = 1$, the simulation results do not vary too much. But the difference between these two simulations does indicate that the bandwidth may have some influence on the test when sample sizes are smaller or moderate.

We also conduct a simulation study when the design variable has two dimensions. The data are generated from the models $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \sqrt{\theta_0 + \theta_1 X_1 + \theta_2 X_2 + b(X_1^2 + X_2^2)}\varepsilon$. The samples from models with $b = 0$ are used to study the empirical level, while data from models with $b = 0.5, 0.8$ and 1 are used to study the empirical power of the test. In the simulation, $\varepsilon \sim N(0, 1), X_1 \sim N(0, 1), X_2 \sim N(0, 1), \beta_0 = 1, \beta_1 = 2, \beta_2 = 1, \theta_0 = 2$ and $\theta_1 = \theta_2 = 0.1$. The kernel function K is chosen to be the product Epanechnikov kernel, i.e. $K(u, v) = 9(1 - u^2)(1 - v^2)I(|u| \le 1)I(|v| \le 1)/16$. The bandwidth h is chosen to be $n^{-1/5}$, and the sample sizes are taken to be $n = 200, 300, 400$, and 500. The weighting measure

Table 4
Empirical size and power for $a = 0.8$.

Model \ n	100	200	300	400	500	800	1000
Model 0	0.044	0.038	0.042	0.028	0.044	0.032	0.042
Model 1	0.130	0.170	0.252	0.312	0.358	0.526	0.610
Model 2	0.210	0.330	0.452	0.556	0.600	0.846	0.912
Model 3	0.266	0.386	0.552	0.636	0.764	0.934	0.936

Table 5
Empirical size and power for the two-dimensional case.

Model \ n	100	200	300	400	500
$b = 0$	0.020	0.022	0.018	0.038	0.028
$b = 0.5$	0.102	0.124	0.102	0.114	0.154
$b = 0.8$	0.176	0.198	0.200	0.236	0.258
$b = 1$	0.260	0.196	0.290	0.472	0.384

is chosen to be $\mathrm{d}G(x) = \hat{f}_w^2(x)\mathrm{d}F_n(x)$ to make the computation easier, where $F_n(x)$ is the empirical CDF of (X_1, X_2). Like for the one-dimensional case, the test is calculated with 500 simulation runs and the nominal level of 0.05. We use 400 bootstrap samples per run to obtain the critical value c_α^*. The empirical size and power are computed by using $\#\{T_n(\hat{\beta}_n, \hat{\theta}_n) \ge c_\alpha^*\}/500$. Table 5 shows that the proposed tests are pretty conservative for all chosen models and sample sizes.

5. Proofs of the main results

We shall use $\tilde{C}_n(\beta, \theta)$ to denote $C_n(\beta, \theta)$ when $\mathrm{d}\hat{\psi}_w(x)$ is replaced by $\mathrm{d}\psi(x) = \mathrm{d}G(x)/f^2(x)$, with the same understanding for $\tilde{\Gamma}_n(\beta, \theta)$. For the sake of convenience, let us define

$$\begin{aligned}
\mu_n(x;\beta) &= \frac{1}{n}\sum_{i=1}^n K_h(x - X_i)(Y_i - m(X_i;\beta))^2, \\
\eta_n(x;\beta,\theta) &= \frac{1}{n}\sum_{i=1}^n K_h(x - X_i)v(X_i;\beta,\theta), \\
\dot{\eta}_n(x;\beta,\theta) &= \frac{1}{n}\sum_{i=1}^n K_h(x - X_i)\dot{v}_\theta(X_i;\beta,\theta).
\end{aligned} \tag{5.1}$$

Before proving Theorem 3.1, we will state some lemmas first.

Lemma 5.1. *Suppose* (e1), (e2), (f1), (m1), (m2), (v1)–(v3), (h1), *and* (h2) *hold. Then under* H_0,

(a) $\tilde{\theta}_n = \arg\min_{\theta\in\Theta} T_n^*(\beta_0, \theta)$ *is a consistent estimator of* θ_0;
(b) $\sup_{\theta\in\Theta} |T_n^*(\hat{\beta}_n, \theta) - T_n^*(\beta_0, \theta)| = o_p(1)$,

where T_n^* *is defined in* (1.5).

Proof. The proof of (a) is similar to that of Corollary 3.1 in KN, and is hence omitted here for the sake of brevity. Now let us show the validity of (b). Let

$$A_{n1} = \int_{\mathcal{C}} \left[\frac{\mu_n(x;\hat{\beta}_n) - \mu_n(x;\beta_0)}{\hat{f}_w(x)}\right]^2 \mathrm{d}G(x), \qquad A_{n2}(\theta) = \int_{\mathcal{C}} [v(x;\hat{\beta}_n,\theta) - v(x;\beta_0,\theta)]^2 \mathrm{d}G(x). \tag{5.2}$$

Then $T_n^*(\hat{\beta}_n, \theta)$ can be written as the sum of $T_n^*(\beta_0, \theta)$, A_{n1}, $A_{n2}(\theta)$ and some three other terms which are bounded above, using the Cauchy–Schwarz inequality, by $2\sqrt{A_{n1}A_{n2}(\theta)}$, $2\sqrt{A_{n1}T_n^*(\beta_0,\theta)}$, and $2\sqrt{A_{n2}(\theta)T_n^*(\beta_0,\theta)}$, respectively. Therefore, it is enough to show that $A_{n1} = o_p(1)$, $\sup_{\theta\in\Theta}|A_{n2}(\theta)| = o_p(1)$ and $\sup_{\theta\in\Theta}|T_n^*(\beta_0,\theta)| = O_p(1)$.

Adding and subtracting $m(X_i;\beta_0)$ to and from $Y_i - m(X_i;\hat{\beta}_n)$, one can show that A_{n1} is bounded above by the following two terms:

$$A_{n11} = 2\int_{\mathcal{C}} \left[\frac{\sum_{i=1}^n K_h(x - X_i)(m(X_i;\hat{\beta}_n) - m(X_i;\beta_0))^2}{\sum_{i=1}^n K_w(x - X_i)}\right]^2 \mathrm{d}G(x),$$

$$A_{n12} = 8\int_{\mathcal{C}} \left[\frac{\sum_{i=1}^{n} K_h(x - X_i)(Y_i - m(X_i;\beta_0))(m(X_i;\hat\beta_n) - m(X_i;\beta_0))}{\sum_{i=1}^{n} K_w(x - X_i)} \right]^2 \mathrm{d}G(x).$$

Let

$$e_{ni} = m(X_i;\hat\beta_n) - m(X_i;\beta_0) - (\hat\beta_n - \beta_0)'\dot m(X_i;\beta_0). \tag{5.3}$$

Then

$$\begin{aligned} A_{n11} &= \int_{\mathcal{C}} \left[\frac{\sum_{i=1}^{n} K_h(x - X_i)(e_{ni} + (\hat\beta_n - \beta_0)'\dot m(X_i;\beta_0))^2}{\sum_{i=1}^{n} K_w(x - X_i)} \right]^2 \mathrm{d}G(x) \\ &\le 8 \sup_{1\le i\le n} |e_{ni}|^4 \int_{\mathcal{C}} [\hat f_h(x)/\hat f_w(x)]^2 \mathrm{d}G(x) + 8\|\hat\beta_n - \beta_0\|^4 \int_{\mathcal{C}} \left[n^{-1}\sum_{i=1}^{n} K_h(x - X_i)\|\dot m(X_i;\beta_0)\|^2/\hat f_w(x) \right]^2 \mathrm{d}G(x) \\ &= o_p(n^{-2})O_p(1) + O_p(n^{-2})O_p(1) = o_p(1) \end{aligned}$$

from conditions (f1), (m1), (m2), (k), (h1), (h2), the $\sqrt{n}$-consistency of $\hat\beta_n$, and the fact of (2.1). Similarly, one can show that $A_{n12} = o_p(1)$ from conditions (m1)–(m2), (k), (h1), the $\sqrt{n}$-consistency of $\hat\beta_n$, and the fact (2.1). $\sup_{\theta\in\Theta} A_{n2}(\theta) = o_p(1)$ can be obtained by using condition (v3) and the $\sqrt{n}$-consistency of $\hat\beta_n$. Finally, $\sup_{\theta\in\Theta} T_n^*(\beta_0, \theta) = O_p(1)$ can shown using (v1) and the following:

$$\int_{\mathcal{C}} \left[\frac{\sum_{i=1}^{n} K_h(x - X_i)(Y_i - m(X_i;\beta_0))^2}{\sum_{i=1}^{n} K_w(x - X_i)} \right]^2 \mathrm{d}G(x) = O_p(1). \tag{5.4}$$

Then the desired result (b) follows. □

To state the second lemma, let $L_2(G)$ denote a class of square integrable real valued functions on $\mathbb{R}^d$ with respect to G. Define

$$\rho(v_1, v_2) = \int_{\mathcal{C}} [v_1(x) - v_2(x)]^2 \mathrm{d}G(x), \quad v_1,\ v_2 \in L_2(G)$$

and the map

$$\mathcal{M}(u) = \arg\min_{\theta\in\Theta} \rho(u, v(x;\beta_0,\theta)), \quad u \in L_2(G).$$

Lemma 5.2. *Let v satisfy conditions* (v1)–(v3). *Then the following hold.*

(a) *$\mathcal{M}(u)$ always exists, $\forall u \in L_2(G)$.*
(b) *If $\mathcal{M}(u)$ is unique, then $\mathcal{M}$ is continuous at u in the sense that for any sequence of $u_n \in L_2(G)$ converging to u in $L_2(G)$, $\mathcal{M}(u_n) \to \mathcal{M}(u)$, i.e.,*

$$\rho(u_n, u) \to 0 \quad \text{implies } \mathcal{M}(u_n) \to \mathcal{M}(u), \text{ as } n \to \infty.$$

(c) *$\mathcal{M}(v(x;\theta,\beta_0)) = \theta$, uniquely $\forall\theta \in \Theta$.*

The proof of this lemma is similar to that of Theorem 1 of Beran [1], and hence is omitted.

Proof of Theorem 3.1. We shall use part (b) of Lemma 5.2 with $u_n(x) = v(x;\beta_0,\theta_n^*)$ and $u(x) = v(x;\beta_0,\theta_0)$. Note that $\theta_n^* = \mathcal{M}(u_n)$, $\theta_0 = \mathcal{M}(u)$, uniquely by (v2). It thus suffices to show that

$$\rho(u_n, u) = \int_{\mathcal{C}} [v(x;\beta_0,\theta_n^*) - v(x;\beta_0,\theta_0)]^2 \mathrm{d}G(x) = o_p(1). \tag{5.5}$$

In fact, adding and subtracting $\mu_n(x;\beta_0)/\hat f_w(x)$ in the brackets of the above integral, $\rho(u_n, u)$ can be bounded above by the sum

$$2\int_{\mathcal{C}} [\mu_n(x;\beta_0)/\hat f_w(x) - v(x;\beta_0,\theta_n^*)]^2 \mathrm{d}G(x) + 2\int_{\mathcal{C}} [\mu_n(x;\beta_0)/\hat f_w(x) - v(x;\beta_0,\theta_0)]^2 \mathrm{d}G(x).$$

The second term is of the order of $o_p(1)$, using an argument similar to that used for proving $C_{n2}(\theta_0) = o_p(1)$ in KN, while the first term will be bounded above by the sum of

$$B_{n1} = 6\int_{\mathcal{C}} [v(x; \beta_0, \theta_n^*) - v(x; \hat{\beta}_n, \theta_n^*)]^2 dG(x),$$

$$B_{n2} = 6\int_{\mathcal{C}} [\mu_n(x; \hat{\beta}_n)/\hat{f}_w(x) - v(x; \hat{\beta}_n, \theta_n^*)]^2 dG(x),$$

$$B_{n3} = 6\int_{\mathcal{C}} [\mu_n(x; \hat{\beta}_n)/\hat{f}_w(x) - \mu_n(x; \beta_0)/\hat{f}_w(x)]^2 dG(x).$$

Lipschitz condition (v3) and the $\sqrt{n}$-consistency of $\hat{\beta}_n$ imply that $B_{n1} = o_p(1)$. To show that $B_{n2} = o_p(1)$, note that, from part (b) of Lemma 5.1, $\sup_{\theta\in\Theta} |T_n^*(\hat{\beta}_n, \theta) - T_n^*(\beta_0, \theta)| = o_p(1)$, and therefore, $T_n^*(\hat{\beta}_n, \theta_n^*) - T_n^*(\beta_0, \theta_n^*) = o_p(1)$, $T_n^*(\hat{\beta}_n, \tilde{\theta}_n) - T_n^*(\beta_0, \tilde{\theta}_n) = o_p(1)$, where $\tilde{\theta}_n$ is defined in part (a) of Lemma 5.1. Hence

$$T_n^*(\hat{\beta}_n, \theta_n^*) - T_n^*(\hat{\beta}_n, \tilde{\theta}_n) = T_n^*(\beta_0, \theta_n^*) - T_n^*(\beta_0, \tilde{\theta}_n) + o_p(1). \tag{5.6}$$

By the definition of θ_n^* and $\tilde{\theta}_n$, the left hand side of (5.6) is nonpositive, and the difference $T_n^*(\beta_0, \theta_n^*) - T_n^*(\beta_0, \tilde{\theta}_n)$ on the right hand side is nonnegative. Hence $T_n^*(\beta_0, \theta_n^*) - T_n^*(\beta_0, \tilde{\theta}_n) = o_p(1)$. Notice that $T_n^*(\beta_0, \tilde{\theta}_n) \le T_n^*(\beta_0, \theta_0) = o_p(1)$; then we have $T_n^*(\beta_0, \tilde{\theta}_n) = o_p(1)$, but this implies $T_n^*(\hat{\beta}_n, \theta_n^*) = o_p(1)$ or $B_{n2} = o_p(1)$. Finally, notice that $B_{n3} = A_{n1}$, where A_{n1} is defined in (5.2); then from the proof of Lemma 5.1, we have $A_{n1} = o_p(1)$, and this is also so for B_{n3}. Therefore, (5.5) is proved, and hence θ_n^* is a consistent estimator of θ_0. □

Now, let us show the consistency of $\hat{\theta}_n$. Again we will use part (b) of Lemma 5.2 but with $u_n(x) = v(x; \beta_0, \hat{\theta}_n)$ and $u(x) = v(x; \beta_0, \theta_0)$. Note that $\hat{\theta}_n = \mathcal{M}(u_n)$, $\theta_0 = \mathcal{M}(u)$, uniquely by (v2). It thus suffices to show that

$$\rho(u_n, u) = \int_{\mathcal{C}} [v(x; \beta_0, \hat{\theta}_n) - v(x; \beta_0, \theta_0)]^2 dG(x) = o_p(1). \tag{5.7}$$

Adding and subtracting $v(x; \hat{\beta}_n, \hat{\theta}_n)$, $\mu_n(x; \hat{\beta}_n)/\hat{f}_w(x)$, $\mu_n(x; \beta_0)/\hat{f}_w(x)$ in the brackets of the above integral, $\rho(u_n, u)$ can be bounded above by the sum of the following four terms:

$$C_{n1} = 4\int_{\mathcal{C}} [v(x; \beta_0, \hat{\theta}_n) - v(x; \hat{\beta}_n, \hat{\theta}_n)]^2 dG(x),$$

$$C_{n2} = 4\int_{\mathcal{C}} [\mu_n(x; \hat{\beta}_n)/\hat{f}_w(x) - v(x; \hat{\beta}_n, \hat{\theta}_n)]^2 dG(x),$$

$$C_{n3} = 4\int_{\mathcal{C}} [\mu_n(x; \hat{\beta}_n)/\hat{f}_w(x) - \mu_n(x; \beta_0)/\hat{f}_w(x)]^2 dG(x),$$

$$C_{n4} = 4\int_{\mathcal{C}} [\mu_n(x; \beta_0)/\hat{f}_w(x) - v(x; \beta_0, \theta_0)]^2 dG(x).$$

Lipschitz condition (v3) and the $\sqrt{n}$-consistency of $\hat{\beta}_n$ imply that $C_{n1} = o_p(1)$. Noting that the integral in C_{n3} is simply A_{n1} defined in (5.2), we have $C_{n3} = o_p(1)$. Also, $C_{n4} = o_p(1)$ is obvious. In the following, we shall show that C_{n2} is of the order of $o_p(1)$. But this is implied by the following claim:

$$\sup_{\theta\in\Theta} |T_n(\hat{\beta}_n, \theta) - T_n^*(\hat{\beta}_n, \theta)| = o_p(1). \tag{5.8}$$

To show (5.8), adding and subtracting $\eta_n(x; \hat{\beta}_n, \theta)/\hat{f}_w(x)$ in the brackets of the integrand in $T_n^*(\hat{\beta}_n, \theta)$, one can show that $|T_n(\hat{\beta}_n, \theta) - T_n^*(\hat{\beta}_n, \theta)| \le D_n(\theta) + 2D_n^{1/2}(\theta)T_n^{1/2}(\hat{\beta}_n, \theta)$, where $D_n(\theta) = \int_{\mathcal{C}} [\eta_n(x; \hat{\beta}_n, \theta)/\hat{f}_w(x) - v(x; \hat{\beta}_n, \theta)]^2 dG(x)$. Therefore, it suffices to show that

$$\sup_{\theta\in\Theta} D_n(\theta) = o_p(1), \qquad \sup_{\theta\in\Theta} T_n(\hat{\beta}_n, \theta) = O_p(1).$$

For this purpose, adding and subtracting $\eta(x; \beta_0, \theta)/\hat{f}_w(x)$, $v(x; \beta_0, \theta)$ in the brackets of the integrand in $D_n(\theta)$, one can show that $D_n(\theta)$ is bounded above by the sum $3D_{n1}(\theta) + 3D_{n2}(\theta) + 3D_{n3}(\theta)$, where

$$D_{n1}(\theta) = \int_{\mathcal{C}} \{[\eta_n(x; \hat{\beta}_n, \theta) - \eta_n(x; \beta_0, \theta)]/\hat{f}_w(x)\}^2 dG(x),$$

$$D_{n2}(\theta) = \int_{\mathcal{C}} \{\eta_n(x; \beta_0, \theta)/\hat{f}_w(x) - v(x; \beta_0, \theta)\}^2 dG(x), \tag{5.9}$$

$$D_{n3}(\theta) = \int_{\mathcal{C}} [v(x; \hat{\beta}_n, \theta) - v(x; \beta_0, \theta)]^2 dG(x).$$

From condition (v3), one can show that $D_{n1}(\theta) = \|\hat{\beta}_n - \beta_0\|^{2\alpha} \cdot O_p(1)$, and $D_{n3}(\theta) \le \|\hat{\beta}_n - \beta_0\|^{2\alpha} \int_{\mathcal{C}} \|l(x)\|^2 \mathrm{d}G(x)$; therefore, by the $\sqrt{n}$-consistency of $\hat{\beta}_n$, both $D_{n1}(\theta)$ and $D_{n3}(\theta)$ are of the order of $o_p(1)$ uniformly for $\theta \in \Theta$. The proof of $\sup_{\theta\in\Theta} D_{n2}(\theta)$ is similar to the proof of $\sup_{\theta\in\Theta} C_{n2}(\theta) = o_p(1)$ in KN. This concludes the proof of $\sup_{\theta\in\Theta} D_n(\theta) = o_p(1)$. To show $\sup_{\theta\in\Theta} T_n(\hat{\beta}_n, \theta) = O_p(1)$, note that $T_n(\hat{\beta}_n, \theta)$ is indeed bounded above by $3A_{n1} + 3T_n(\beta_0, \theta) + 3D_{n1}(\theta)$, where A_{n1} is defined in (5.2). We have already shown that $A_{n1} = o_p(1)$, and $\sup_{\theta\in\Theta} D_{n1}(\theta) = o_p(1)$, so we only have to show that $\sup_{\theta\in\Theta} T_n(\beta_0, \theta) = O_p(1)$, but this can be done by using an argument similar to that of KN. Hence the theorem follows. □

The proof of Theorem 3.2 needs the following lemma, which along with its proof appears as Theorem 2.2 part (2) in [3].

Lemma 5.3. *Let $\hat{f}_w(x)$ be the kernel estimate associated with a kernel K which satisfies a Lipschitz condition. If* (f2) *holds and $w = a_n(\log n/n)^{1/(d+4)}$, where $a_n \to a_0 > 0$, then for any positive integer k,*

$$\frac{n^{2/(d+4)}}{(\log n)^{2/(d+4)} \log_k n} \sup_{\mathcal{C}} |\hat{f}_w(x) - f(x)| \to 0$$

almost surely.

Proof of Theorem 3.2. Our first step is to show that

$$nh^d \|\hat{\theta}_n - \theta_0\|^2 = O_p(1). \tag{5.10}$$

For this purpose, let

$$H_n(\theta) = \int_{\mathcal{C}} \left(\frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) [v(X_i; \hat{\beta}_n, \theta) - v(X_i; \hat{\beta}_n, \theta_0)] \right)^2 \mathrm{d}\hat{\psi}_w(x).$$

We claim that $nh^d H_n(\hat{\theta}_n) = O_p(1)$. To see this, note that

$$\begin{aligned} H_n(\hat{\theta}_n) &\le 2 \int_{\mathcal{C}} \left(\frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) [(Y_i - m(X_i; \hat{\beta}_n))^2 - v(X_i; \hat{\beta}_n, \hat{\theta}_n)] \right)^2 \mathrm{d}\hat{\psi}_w(x) \\ &\quad + 2 \int_{\mathcal{C}} \left(\frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) [(Y_i - m(X_i; \hat{\beta}_n))^2 - v(X_i; \hat{\beta}_n, \theta_0)] \right)^2 \mathrm{d}\hat{\psi}_w(x) \\ &= 2T_n(\hat{\beta}_n, \hat{\theta}_n) + 2T_n(\hat{\beta}_n, \theta_0) \le 4T_n(\hat{\beta}_n, \theta_0). \end{aligned}$$

Therefore, it is sufficient to show that

$$nh^d T_n(\hat{\beta}_n, \theta_0) = O_p(1). \tag{5.11}$$

Adding and subtracting $(Y_i - m(X_i; \beta_0))^2$, $v(x; \beta_0, \theta_0)$ from $(Y_i - m(X_i; \hat{\beta}_n))^2 - v(x; \hat{\beta}_n, \theta_0)$ in $T_n(\hat{\beta}_n, \theta_0)$, one can show that $T_n(\hat{\beta}_n, \theta_0)$ will be bounded above by $3A_{n1} + 3T_n(\beta_0, \theta_0) + 3D_{n1}(\theta_0)$, where A_{n1} is defined in (5.2) and $D_{n1}(\theta)$ is given in (5.9). Since $A_{n1} = O_p(1/n)$ from the proof of Lemma 5.1, we have $nh^d A_{n1} = O_p(h^d) = o_p(1)$. Note that $D_{n1}(\theta_0)$ is bounded above by $2D_{n11}(\theta_0) + 2D_{n12}(\theta_0)$, where

$$D_{n11}(\theta_0) = \int_{\mathcal{C}} \left[\frac{\frac{1}{n}\sum_{i=1}^n K_h(x - X_i)[v(X_i; \hat{\beta}_n, \theta_0) - v(X_i; \beta_0, \theta_0) - (\hat{\beta}_n - \beta_0)' \dot{v}_\beta(X_i; \beta_0, \theta_0)]}{\hat{f}_w(x)} \right]^2 \mathrm{d}G(x)$$

and

$$D_{n12}(\theta_0) = \int_{\mathcal{C}} \left[\frac{\frac{1}{n}\sum_{i=1}^n K_h(x - X_i)[(\hat{\beta}_n - \beta_0)' \dot{v}_\beta(X_i; \beta_0, \theta_0)]}{\hat{f}_w(x)} \right]^2 \mathrm{d}G(x).$$

It is easy to see that $D_{n11}(\theta_0)$ is bounded above by

$$\sup_{1\le i\le n} |v(X_i; \hat{\beta}_n, \theta_0) - v(X_i; \beta_0, \theta_0) - (\hat{\beta}_n - \beta_0)' \dot{v}_\beta(X_i; \beta_0, \theta_0)| \cdot \int_{\mathcal{C}} \left[\frac{\hat{f}_h(x)}{\hat{f}_w(x)} \right]^2 \mathrm{d}G(x)$$

which has the order $o_p(n^{-1})$ by (v4). By the Cauchy–Schwarz inequality, $D_{n12}(\theta)$ is bounded above by

$$\|\hat{\beta}_n - \beta_0\|^2 \int_{\mathcal{C}} \left[\frac{\frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i) \left\| \dot{v}_\beta(X_i; \beta_0, \theta_0) \right\|}{\hat{f}_w(x)} \right]^2 \mathrm{d}G(x)$$

which is $O_p(1/n)$ by (v1) and the $\sqrt{n}$-consistency of $\hat{\beta}_n$. Therefore, $nh^d D_{n1}(\theta_0) = nh^d O_p(1/n) = o_p(1)$. So, we only have to show that $nh^d T_n(\beta_0, \theta_0) = O_p(1)$.

Let $\Delta_n(x) = f^2(x)/\hat{f}_w^2(x) - 1$. Then $nh^d T_n(\beta_0, \theta_0)$ is bounded above by the sum of

$$Q_{n1} = nh^d \int_{\mathcal{C}} \left(\frac{\sum_{i=1}^{n} K_h(x - X_i) v(X_i; \beta_0, \theta_0)(\varepsilon_i^2 - 1)}{f(x)} \right)^2 \mathrm{d}G(x),$$

$$Q_{n2} = nh^d \int_{\mathcal{C}} \left(\frac{\sum_{i=1}^{n} K_h(x - X_i) v(X_i; \beta_0, \theta_0)(\varepsilon_i^2 - 1)}{f(x)} \right)^2 \Delta_n(x)\mathrm{d}G(x).$$

Note that $\varepsilon_i^2 - 1$ are i.i.d. with mean 0, so

$$E \int_{\mathcal{C}} \left(\frac{1}{nh^d} \sum_{i=1}^{n} K_h(x - X_i) v(X_i; \beta_0, \theta_0)(\varepsilon_i^2 - 1) \right)^2 \mathrm{d}G(x) = \frac{1}{nh^{2d}} \int_{\mathcal{C}} EK_h^2(x - X) v^2(X; \beta_0, \theta_0) \tau(X) \mathrm{d}\psi(x) \tag{5.12}$$

where $\mathrm{d}\psi(x) = \mathrm{d}G(x)/f^2(x)$. Then from conditions (e2), (f1) and (v1), one can show that the right end of (5.12) is the order of $O_p(1/nh^d)$. Hence $Q_{n1} = O_p(1)$. Realizing that $|Q_{n2}| \le \sup_{x\in\mathcal{C}} |\Delta_n(x)| \cdot Q_{n1}$, then from (2.1), we have $Q_{n2} = o_p(1)$. These imply that $nh^d T_n(\beta_0, \theta_0) = O_p(1)$; hence $nh^d H_n(\hat{\theta}_n) = O_p(1)$. By a proof similar to that of (4.6) in KN, one can show that

$$\liminf_{n\to\infty} P\left(H_n(\hat{\theta}_n)/\|\hat{\theta}_n - \theta_0\|^2 \ge \frac{1}{2} \inf_{\|b\|=1} b' \Sigma_0 b \right) = 1, \tag{5.13}$$

where Σ_0 is defined in (3.3). Then claim (5.10) will then follow from (5.13), $nh^d H_n(\hat{\theta}_n) = O_p(1)$, $\Sigma_0 > 0$, and the fact that $nh^d H_n(\hat{\theta}_n) = nh^d \|\hat{\theta}_n - \theta_0\|^2 \cdot [H_n(\hat{\theta}_n)/\|\hat{\theta}_n - \theta_0\|^2]$.

In the following, we shall prove the asymptotic normality of $\hat{\theta}_n$.

Since θ_0 is an interior point of Θ, by the consistency of $\hat{\theta}_n$, for sufficiently large n, $\hat{\theta}_n$ will be in the interior of Θ, so $\dot{T}_{n,\theta}(\hat{\beta}_n, \hat{\theta}_n) = 0$, where $\dot{T}_{n,\theta}(\hat{\beta}_n, \hat{\theta}_n)$ is the derivative of $T_n(\hat{\beta}_n, \theta)$ with respect to θ, evaluated at $\theta = \hat{\theta}_n$. This is equivalent to

$$\int_{\mathcal{C}} [\mu_n(x; \hat{\beta}_n) - \eta_n(x; \hat{\beta}_n, \hat{\theta}_n)] \dot{\eta}_n(x; \hat{\beta}_n, \hat{\theta}_n) \mathrm{d}\hat{\psi}_w(x) = 0.$$

Adding and subtracting $\eta_n(x; \hat{\beta}_n, \theta_0)$ to and from $\mu_n(x; \hat{\beta}_n) - \eta_n(x; \hat{\beta}_n, \hat{\theta}_n)$, the above can be written as

$$\int_{\mathcal{C}} [\mu_n(x; \hat{\beta}_n) - \eta_n(x; \hat{\beta}_n, \theta_0)] \dot{\eta}(x; \hat{\beta}_n, \hat{\theta}_n) \mathrm{d}\hat{\psi}_w(x) = \int_{\mathcal{C}} [\eta_n(x; \hat{\beta}_n, \hat{\theta}_n) - \eta_n(x; \hat{\beta}_n, \theta_0)] \dot{\eta}(x; \hat{\beta}_n, \hat{\theta}_n) \mathrm{d}\hat{\psi}_w(x). \tag{5.14}$$

Denote the left hand side as L_n, and the right hand side as R_n.

Note that L_n can be written as the sum $L_{n1} + L_{n2} + L_{n3}$, where

$$L_{n1} = \int_{\mathcal{C}} [\mu_n(x; \hat{\beta}_n) - \mu_n(x; \beta_0)] \dot{\eta}_n(x; \hat{\beta}_n, \hat{\theta}_n) \mathrm{d}\hat{\psi}_w(x),$$

$$L_{n2} = \int_{\mathcal{C}} [\mu_n(x; \beta_0) - \eta_n(x; \beta_0, \theta_0)] \dot{\eta}_n(x; \hat{\beta}_n, \hat{\theta}_n) \mathrm{d}\hat{\psi}_w(x),$$

$$L_{n3} = \int_{\mathcal{C}} [\eta_n(x; \beta_0, \theta_0) - \eta_n(x; \hat{\beta}_n, \theta_0)] \dot{\eta}_n(x; \hat{\beta}_n, \hat{\theta}_n) \mathrm{d}\hat{\psi}_w(x).$$

For L_{n1}, we have

$$\begin{aligned} L_{n1} &= 2\int_{\mathcal{C}} \frac{1}{n}\sum_{i=1}^{n} K_h(x-X_i)(Y_i - m(X_i;\beta_0))(m(X_i;\beta_0) - m(X_i;\hat\beta_n))\dot\eta_n(x;\hat\beta_n,\hat\theta_n)\mathrm{d}\hat\psi_w(x) \\ &\quad + \int_{\mathcal{C}} \frac{1}{n}\sum_{i=1}^{n} K_h(x-X_i)(m(X_i;\beta_0) - m(X_i;\hat\beta_n))^2\dot\eta_n(x;\hat\beta_n,\hat\theta_n)\mathrm{d}\hat\psi_w(x) \\ &= L_{n11} + L_{n12}. \end{aligned}$$

Recalling the notation e_{ni} in (5.3), we have

$$\begin{aligned} L_{n11} &= -2\int_{\mathcal{C}} \left(\frac{1}{n}\sum_{i=1}^{n} K_h(x-X_i)(Y_i - m(X_i;\beta_0))e_{ni}\right)\cdot\dot\eta_n(x;\hat\beta_n,\hat\theta_n)\mathrm{d}\hat\psi_w(x) \\ &\qquad -2\int_{\mathcal{C}} \dot\eta_n(x;\hat\beta_n,\hat\theta_n)\left(\frac{1}{n}\sum_{i=1}^{n} K_h(x-X_i)(Y_i - m(X_i;\beta_0))\dot m'_\beta(X_i;\beta_0)\right)\mathrm{d}\hat\psi_w(x)(\hat\beta_n - \beta_0). \end{aligned}$$

Noticing that

$$\dot\eta_n(x;\hat\beta_n,\hat\theta_n) = \dot\eta_n(x;\hat\beta_n,\hat\theta_n) - \dot\eta_n(x;\hat\beta_n,\theta_0) + \dot\eta_n(x;\hat\beta_n,\theta_0) - \dot\eta_n(x;\beta_0,\theta_0) + \dot\eta_n(x;\beta_0,\theta_0),$$

then by condition (v5) and the fact (2.1), one can show that

$$\int_{\mathcal{C}} \|\dot\eta_n(x;\hat\beta_n,\hat\theta_n)\|^2\mathrm{d}\hat\psi_w(x) = \int_{\mathcal{C}} \|\dot v_\theta(x;\beta_0,\theta_0)\|^2\mathrm{d}G(x) + o_p(1) \tag{5.15}$$

which is $O_p(1)$. Therefore, by the Cauchy–Schwarz inequality,

$$\begin{aligned} & n\left\|\int_{\mathcal{C}} \left(\frac{1}{n}\sum_{i=1}^{n} K_h(x-X_i)(Y_i - m(X_i;\beta_0))e_{ni}\right)\cdot\dot\eta_n(x;\hat\beta_n,\hat\theta_n)\mathrm{d}\hat\psi_w(x)\right\|^2 \\ &\le n\int_{\mathcal{C}} \left(\frac{1}{n}\sum_{i=1}^{n} K_h(x-X_i)(Y_i - m(X_i;\beta_0))e_{ni}\right)^2\mathrm{d}\hat\psi_w(x)\int_{\mathcal{C}} \|\dot\eta_n(x;\hat\beta_n,\hat\theta_n)\|^2\mathrm{d}\hat\psi_w(x) \\ &= n\sup_{1\le i\le n}|e_{ni}|^2\cdot O_p(1) = o_p(1) \end{aligned}$$

from condition (m2). Similarly, one can show that

$$\sqrt{n}\int_{\mathcal{C}} \dot\eta_n(x;\hat\beta_n,\hat\theta_n)\left(\frac{1}{n}\sum_{i=1}^{n} K_h(x-X_i)(Y_i - m(X_i;\beta_0))\dot m'_\beta(X_i;\beta_0)\right)\mathrm{d}\hat\psi_w(x)(\hat\beta_n - \beta_0) = o_p(1)$$

on noting that $\sqrt{n}(\hat\beta_n - \beta_0) = O_p(1)$ and the following fact:

$$\int_{\mathcal{C}} \left\|\frac{1}{n}\sum_{i=1}^{n} K_h(x-X_i)(Y_i - m(X_i;\beta_0))\dot m'_\beta(X_i;\beta_0)\right\|^2\mathrm{d}\hat\psi_w(x) = O_p(1/nh^d)$$

which can be shown by using the fact (2.1) and an expectation and variance argument. Therefore, $\sqrt{n}L_{n11} = o_p(1)$. Using the Cauchy–Schwarz inequality and the conditions (m1) and (m2) on L_{n12}, one can obtain that $\sqrt{n}L_{n12} = o_p(1)$. Thus, we have proved that

$$\sqrt{n}L_{n1} = o_p(1). \tag{5.16}$$

Now, let us consider L_{n2}. For convenience, define $U_n(x) = \mu_n(x;\beta_0) - \eta_n(x;\beta_0,\theta_0)$. Adding and subtracting $\dot v_\theta(X_i;\beta_0,\theta_0)$ to and from $\dot v_\theta(X_i;\hat\beta_n,\hat\theta_n)$ in $\dot\eta_n(x;\hat\beta_n,\hat\theta_n)$, L_{n2} can be written as

$$\begin{aligned} L_{n2} &= \int_{\mathcal{C}} U_n(x)\cdot\frac{1}{n}\sum_{i=1}^{n} K_h(x-X_i)(\dot v_\theta(X_i;\hat\beta_n,\hat\theta_n) - \dot v_\theta(X_i;\beta_0,\theta_0))\mathrm{d}\hat\psi_w(x) \\ &\quad + \int_{\mathcal{C}} U_n(x)\cdot\frac{1}{n}\sum_{i=1}^{n} K_h(x-X_i)\dot v_\theta(X_i;\beta_0,\theta_0)\mathrm{d}\hat\psi_w(x) \\ &= L_{n21} + L_{n22}. \end{aligned}$$

In the following, we shall show that $\sqrt{n}L_{n21} = o_p(1)$. In fact

$$L_{n21} = \int_{\mathcal{C}} U_n(x)\cdot\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)(\dot{\upsilon}_\theta(X_i;\hat{\beta}_n,\hat{\theta}_n) - \dot{\upsilon}_\theta(X_i;\beta_0,\theta_0))\left(\frac{f^2(x)}{\hat{f}_w(x)} - 1\right)\mathrm{d}\psi(x)$$
$$+ \int_{\mathcal{C}} U_n(x)\cdot\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)(\dot{\upsilon}_\theta(X_i;\hat{\beta}_n,\hat{\theta}_n) - \dot{\upsilon}_\theta(X_i;\beta_0,\theta_0))\mathrm{d}\psi(x)$$
$$= L'_{n21} + L''_{n21}.$$

Using the Cauchy–Schwarz inequality, the second term is bounded above by the square roof of

$$\int_{\mathcal{C}} U_n^2(x)\mathrm{d}\psi(x)\cdot\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)(\dot{\upsilon}_\theta(X_i;\hat{\beta}_n,\hat{\theta}_n) - \dot{\upsilon}_\theta(X_i;\beta_0,\theta_0))\right]^2\mathrm{d}\psi(x)$$

which is again bounded above by

$$\int_{\mathcal{C}} U_n^2(x)\mathrm{d}\psi(x)\cdot\sup\|\dot{\upsilon}_\theta(X_i;\hat{\beta}_n,\hat{\theta}_n) - \dot{\upsilon}_\theta(X_i;\beta_0,\theta_0)\|^2\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)\right]^2\mathrm{d}\psi(x).$$

Since $\|\dot{\upsilon}_\theta(X_i;\hat{\beta}_n,\hat{\theta}_n) - \dot{\upsilon}_\theta(X_i;\beta_0,\theta_0)\|$ is bounded above by the sum of $\|\dot{\upsilon}_\theta(X_i;\hat{\beta}_n,\hat{\theta}_n) - \dot{\upsilon}_\theta(X_i;\beta_0,\hat{\theta}_n)\|$ and $\|\dot{\upsilon}_\theta(X_i;\beta_0,\hat{\theta}_n) - \dot{\upsilon}_\theta(X_i;\beta_0,\theta_0)\|$. By (v5), both terms are $o_p(h^{d/2})$. Therefore, from the fact that $\int_{\mathcal{C}} U_n^2(x)\mathrm{d}\psi(x) = O_p(1/nh^d)$, $\sqrt{n}L''_{n21} = \sqrt{n}\cdot O_p(1/\sqrt{nh^d})\cdot o_p(h^{d/2}) = o_p(1)$. $\sqrt{n}L'_{n21} = o_p(1)$ will then follow from a similar argument together with the fact in (2.1). Hence we have proved

$$\sqrt{n}L_{n21} = o_p(1). \tag{5.17}$$

As for L_{n22}, we have

$$L_{n22} = \int_{\mathcal{C}} U_n(x)\cdot\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)\dot{\upsilon}_\theta(X_i;\beta_0,\theta_0)\left(\frac{f^2(x)}{\hat{f}_w^2(x)} - 1\right)\mathrm{d}\psi(x)$$
$$+ \int_{\mathcal{C}} U_n(x)\cdot\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)\dot{\upsilon}_\theta(X_i;\beta_0,\theta_0)\mathrm{d}\psi(x) = L'_{n22} + L''_{n22}.$$

By the Cauchy–Schwarz inequality,

$$\left\|L'_{n22}\right\|^2 \le \int_{\mathcal{C}} U_n^2(x)\mathrm{d}\psi(x)\cdot\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)\dot{\upsilon}_\theta(X_i;\beta_0,\theta_0)\right]^2\mathrm{d}\psi(x)\cdot\sup_{x\in\mathcal{C}}\left|\frac{f^2(x)}{\hat{f}_w^2(x)} - 1\right|^2,$$

so,

$$n\left\|L'_{n22}\right\|^2 = n\cdot O_p(1/nh^d)\cdot o((\log_k n)^2(\log n/n)^{4/(d+4)}) = o_p((\log_k n)^2(\log n)^{4/(d+4)}n^{ad-4/(d+4)})$$

which is $o_p(1)$. Therefore, $\sqrt{n}L'_{n22} = o_p(1)$. This, together with the result (5.17), implies

$$\sqrt{n}L_{n2} = \sqrt{n}\int_{\mathcal{C}} U_n(x)\cdot\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)\dot{\upsilon}_\theta(X_i;\beta_0,\theta_0)\mathrm{d}\psi(x) + o_p(1)$$
$$= \sqrt{n}\int_{\mathcal{C}} U_n(x)\dot{\eta}_h(x)\mathrm{d}\psi(x) + o_p(1), \tag{5.18}$$

where

$$\dot{\eta}_h(x) = EK_h(x-X)\dot{\upsilon}_\theta(X;\beta_0,\theta_0). \tag{5.19}$$

Finally, let us consider L_{n3}. Adding and subtracting $\upsilon(x;\beta_0,\theta_0)$ to and from $\upsilon(x;\hat{\beta}_n,\theta_0)$, and defining $r_{ni} = \upsilon(X_i;\hat{\beta}_n,\theta_0) - \upsilon(X_i;\beta_0,\theta_0) - (\hat{\beta}_n-\beta_0)'\dot{\upsilon}_\beta(X_i;\beta_0,\theta_0)$, we have

$$\sqrt{n}L_{n3} = -\sqrt{n}\int_{\mathcal{C}}\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)r_{ni}\cdot\dot{\eta}_n(x;\hat{\beta}_n,\hat{\theta}_n)\mathrm{d}\hat{\psi}_w(x)$$
$$- \sqrt{n}\int_{\mathcal{C}}\dot{\eta}_n(x;\hat{\beta}_n,\hat{\theta}_n)\dot{\eta}'_{n\beta}(x;\beta_0,\theta_0)\mathrm{d}\hat{\psi}_w(x)(\hat{\beta}_n-\beta_0). \tag{5.20}$$

Condition (v4) and some routine arguments can show that the first term on the right hand side of (5.20) is of the order of $o_p(1)$, and the second term equals

$$\int_{\mathcal{C}} \dot{\eta}_n(x;\beta_0,\theta_0)\dot{\eta}'_{n\beta}(x;\beta_0,\theta_0)\mathrm{d}\psi(x)\sqrt{n}(\hat{\beta}_n-\beta_0)+o_p(1).$$

Note that

$$\int_{\mathcal{C}} \dot{\eta}_n(x;\beta_0,\theta_0)\dot{\eta}'_{n\beta}(x;\beta_0,\theta_0)\mathrm{d}\psi(x)=\Pi+o_p(1),$$

where Π is defined by (3.3). Then we have

$$\sqrt{n}L_{n3}=\Pi\sqrt{n}(\hat{\beta}_n-\beta_0)+o_p(1). \tag{5.21}$$

Combining (5.16), (5.18) and (5.21), we have

$$\sqrt{n}L_n=\sqrt{n}\int_{\mathcal{C}} U_n(x)\dot{\eta}_h(x)\mathrm{d}\psi(x)-\Pi\sqrt{n}(\hat{\beta}_n-\beta_0)+o_p(1). \tag{5.22}$$

Let $s_{ni}=(\varepsilon_i^2-1)v(X_i;\beta_0,\theta_0)\int_{\mathcal{C}}K_h(x-X_i)\dot{\eta}_h(x)\mathrm{d}\psi(x)$, $t_{ni}=\Pi L(Y_i,X_i;\beta_0,\theta_0)$, where L is defined in (3.1). Then $\sqrt{n}L_n=n^{-1/2}\sum_{i=1}^n(s_{ni}-t_{ni})$. For convenience, we shall give the proof here only for the case of $p=q=1$. For the multidimensional case, the result can be proved by using the Wald scheme and applying the same argument. Note that $\{s_{ni}-t_{ni}, i\le 1\le n\}$ are i.i.d. centered random variables for each n. By the Lindeberg–Feller CLT, it suffices to show that as $n\to\infty$,

$$E(s_{n1}-t_{n1})^2\to\Sigma, \tag{5.23}$$

$$E(s_{n1}-t_{n1})^2I[|s_{n1}-t_{n1}|>\lambda\sqrt{n}]\to 0,\quad \text{for } \forall\lambda>0, \tag{5.24}$$

where Σ is defined in Theorem 3.2. The proofs of (5.23) and (5.24) are straightforward, and hence omitted here for the sake of brevity. Hence

$$\sqrt{n}L_n\Rightarrow N(0,\Sigma)\quad\text{in distribution.} \tag{5.25}$$

Now let us consider the term R_n. In the following, we shall show that $R_n=H_n(\hat{\theta}_n-\theta_0)$ with $H_n=\Sigma_0+o_p(1)$, where Σ_0 is defined in (3.3). To see this, let

$$d_{ni}=v(X_i;\hat{\beta}_n,\hat{\theta}_n)-v(X_i;\hat{\beta}_n,\theta_0)-\dot{v}'_\theta(X_i;\hat{\beta}_n,\theta_0)(\hat{\theta}_n-\theta_0). \tag{5.26}$$

Then R_n can be written as the sum $R_{n1}+R_{n2}$, where

$$R_{n1}=\int_{\mathcal{C}}\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)d_{ni}\cdot\dot{\eta}_n(x;\hat{\beta}_n,\hat{\theta}_n)\mathrm{d}\hat{\psi}_w(x),$$

$$R_{n2}=\int_{\mathcal{C}}\dot{\eta}_n(x;\hat{\beta}_n,\hat{\theta}_n)\dot{\eta}'_n(x;\hat{\beta}_n,\theta_0)\mathrm{d}\hat{\psi}_w(x)(\hat{\theta}_n-\theta_0).$$

From (v5), one can show that $\sqrt{n}R_{n1}=o_p(1)\sqrt{n}(\hat{\theta}_n-\theta_0)$. Note that the usual calculation shows that

$$\int_{\mathcal{C}}\dot{\eta}_n(x;\hat{\beta}_n,\hat{\theta}_n)\dot{\eta}'_n(x;\hat{\beta}_n,\theta_0)\mathrm{d}\hat{\psi}_w(x)=\Sigma_0+o_p(1).$$

Hence $\sqrt{n}R_{n2}=(\Sigma_0+o_p(1))\sqrt{n}(\hat{\theta}_n-\theta_0)$. Therefore, $\sqrt{n}R_n=(\Sigma_0+o_p(1))\sqrt{n}(\hat{\theta}_n-\theta_0)$. This, together with (5.25), proves the theorem. □

Proof of Theorem 3.3. The proof of Theorem 3.3 consists of several lemmas. □

Lemma 5.4. *Suppose all the conditions in Theorem 3.3 hold; then*

(i) $nh^{d/2}[T_n(\hat{\beta}_n,\hat{\theta}_n)-T_n(\hat{\beta}_n,\theta_0)]=o_p(1)$;

(ii) $nh^{d/2}[T_n(\hat{\beta}_n,\theta_0)-T_n(\beta_0,\theta_0)]=o_p(1)$;

(iii) $nh^{d/2}[T_n(\beta_0,\theta_0)-\tilde{T}_n(\beta_0,\theta_0)]=o_p(1)$.

Proof. Adding and subtracting $v(X_i;\hat\beta_n,\theta_0)$ to and from $\xi_i(\hat\beta_n,\hat\theta_n)=(Y_i-m(X_i;\hat\beta_n))^2-v(X_i;\hat\beta_n,\hat\theta_n)$, expanding the square term, $T_n(\hat\beta_n,\hat\theta_n)$ can be written as the sum $T_n(\hat\beta_n,\theta_0)+Q_{n1}-2Q_{n2}$, or $T_n(\hat\beta_n,\hat\theta_n)-T_n(\hat\beta_n,\theta_0)=Q_{n1}-2Q_{n2}$, where

$$Q_{n1}=\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)(v(X_i;\hat\beta_n,\hat\theta_n)-v(X_i;\hat\beta_n,\theta_0))\right]^2 \mathrm{d}\hat\psi_w(x),$$

$$Q_{n2}=\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)\hat\xi_i\right]\cdot\left[\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)(v(X_i;\hat\beta_n,\hat\theta_n)-v(X_i;\hat\beta_n,\theta_0))\right]\mathrm{d}\hat\psi_w(x).$$

So it suffices to show that

$$nh^{d/2}Q_{n1}=o_p(1),\qquad nh^{d/2}Q_{n2}=o_p(1). \tag{5.27}$$

Recalling the notation d_{ni} defined in (5.26), we have $Q_{n1}\le 2Q_{n11}+2Q_{n12}$, where

$$Q_{n11}=\int_{\mathcal{C}}\left[n^{-1}\sum_{i=1}^n K_h(x-X_i)d_{ni}\right]^2\mathrm{d}\hat\psi_w(x),$$

$$Q_{n12}=\int_{\mathcal{C}}\left[n^{-1}\sum_{i=1}^n K_h(x-X_i)\dot v_\theta'(X_i;\hat\beta_n,\theta_0)(\hat\theta_n-\theta_0)\right]^2\mathrm{d}\hat\psi_w(x).$$

From (v4), we obtain that

$$Q_{n11}\le\|\hat\theta_n-\theta_0\|^2\sup_{1\le i\le n}\frac{|d_{ni}|^2}{\|\hat\theta_n-\theta_0\|^2}\int_{\mathcal{C}}\hat f_h^2(x)\mathrm{d}\hat\psi_w(x)=o_p(1/n),$$

and from (v5), we obtain that

$$\begin{aligned}Q_{n12}&\le 2\|\hat\theta_n-\theta_0\|^2\int_{\mathcal{C}}\left[n^{-1}\sum_{i=1}^n K_h(x-X_i)\|\dot v_\theta(X_i;\hat\beta_n,\theta_0)-\dot v_\theta(X_i;\beta_0,\theta_0)\|\right]^2\mathrm{d}\hat\psi_w(x)\\&\quad+2\|\hat\theta_n-\theta_0\|^2\int_{\mathcal{C}}\left[n^{-1}\sum_{i=1}^n K_h(x-X_i)\|\dot v_\theta(X_i;\beta_0,\theta_0)\|\right]^2\mathrm{d}\hat\psi_w(x)=O_p(1/n).\end{aligned}$$

These imply the first statement in (5.27). Now let us consider Q_{n2}. By adding and subtracting $\dot v_\theta'(X_i;\hat\beta_n,\theta_0)(\hat\theta_n-\theta_0)$ to and from the $v(X_i;\hat\beta_n,\hat\theta_n)-v(X_i;\hat\beta_n,\theta_0)$, we can write Q_{n2} as the sum $Q_{n21}+Q_{n22}$, where

$$Q_{n21}=\int_{\mathcal{C}}[\mu_n(x;\hat\beta_n)-\eta_n(x;\hat\beta_n,\hat\theta_n)]\cdot\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)d_{ni}\mathrm{d}\hat\psi_w(x),$$

$$Q_{n22}=(\hat\theta_n-\theta_0)'\int_{\mathcal{C}}[\mu_n(x;\hat\beta_n)-\eta_n(x;\hat\beta_n,\hat\theta_n)]\cdot\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)\dot v_\theta(X_i;\hat\beta_n,\theta_0)\mathrm{d}\hat\psi_w(x).$$

By the Cauchy–Schwarz inequality, assumption (v4), and (5.11),

$$\|Q_{n21}\|^2\le T_n(\hat\beta_n,\theta_0)\|\hat\theta_n-\theta_0\|^2\sup_{1\le i\le n}\left(\frac{d_{ni}}{\|\hat\theta_n-\theta_0\|}\right)^2\int_{\mathcal{C}}\hat f_w^2(x)\mathrm{d}\hat\psi_w(x)=o_p(1/n^2h^d).$$

Therefore, $nh^{d/2}Q_{n21}=nh^{d/2}o_p(1/\sqrt{n^2h^d})=o_p(1)$. Note that Q_{n22} can be written as the sum $Q_{n22}'-Q_{n22}''$, where

$$Q_{n22}'=(\hat\theta_n-\theta_0)'\int_{\mathcal{C}}[\mu_n(x;\hat\beta_n)-\eta_n(x;\hat\beta_n,\hat\theta_n)]\cdot\dot\eta_n(X_i;\hat\beta_n,\hat\theta_n)\mathrm{d}\hat\psi_w(x)$$

$$Q_{n22}''=(\hat\theta_n-\theta_0)'\int_{\mathcal{C}}[\mu_n(x;\hat\beta_n)-\eta_n(x;\hat\beta_n,\hat\theta_n)]\cdot[\dot\eta_n(X_i;\hat\beta_n,\hat\theta_n)-\dot\eta_n(X_i;\hat\beta_n,\theta_0)]\mathrm{d}\hat\psi_w(x).$$

By the Cauchy–Schwarz inequality, it is easy to see that

$$|Q_{n22}''|^2\le\|\hat\theta_n-\theta_0\|^2\sup\|\dot\eta_n(X_i;\hat\beta_n,\hat\theta_n)-\dot\eta_n(X_i;\hat\beta_n,\theta_0)\|^2\cdot T_n(\hat\beta_n,\theta_0)\cdot\int_{\mathcal{C}}\hat f_w^2(x)\mathrm{d}\hat\psi_w(x).$$

From condition (v5), the $\sqrt{n}$-consistency of $\hat{\theta}_n$, one obtains $|Q''_{n22}|^2 = O_p(1/n^2)$. Therefore, $nh^{d/2}Q''_{n22} = nh^{d/2}o_p(1/n) = o_p(h^{d/2}) = o_p(1)$. As for Q'_{n22}, note that the integral is same as the left hand side of (5.14); hence

$$Q'_{n22} = (\hat{\theta}_n - \theta_0)' \int_{\mathcal{C}} [\eta_n(x; \hat{\beta}_n, \hat{\theta}_n) - \eta_n(x; \hat{\beta}_n, \theta_0)]\dot{\eta}_n(x; \hat{\beta}_n, \hat{\theta}_n)\mathrm{d}\hat{\psi}_w(x).$$

Adding and subtracting $\dot{\eta}_n(x; \hat{\beta}_n, \theta_0)$, $\dot{\eta}_n(x; \beta_0, \theta_0)$ to and from $\dot{\eta}_n(x; \hat{\beta}_n, \hat{\theta}_n)$, Q''_{n22} can be written as the sum $Q'_{n221} + Q'_{n222} + Q'_{n223}$, where

$$Q'_{n221} = (\hat{\theta}_n - \theta_0)' \int_{\mathcal{C}} [\eta_n(x; \hat{\beta}_n, \hat{\theta}_n) - \eta_n(x; \hat{\beta}_n, \theta_0)][\dot{\eta}_n(x; \hat{\beta}_n, \hat{\theta}_n) - \dot{\eta}_n(x; \hat{\beta}_n, \theta_0)]\mathrm{d}\hat{\psi}_w(x),$$
$$Q'_{n222} = (\hat{\theta}_n - \theta_0)' \int_{\mathcal{C}} [\eta_n(x; \hat{\beta}_n, \hat{\theta}_n) - \eta_n(x; \hat{\beta}_n, \theta_0)][\dot{\eta}_n(x; \hat{\beta}_n, \theta_0) - \dot{\eta}_n(x; \beta_0, \theta_0)]\mathrm{d}\hat{\psi}_w(x),$$
$$Q'_{n223} = (\hat{\theta}_n - \theta_0)' \int_{\mathcal{C}} [\eta_n(x; \hat{\beta}_n, \hat{\theta}_n) - \eta_n(x; \hat{\beta}_n, \theta_0)]\dot{\eta}_n(x; \beta_0, \theta_0)\mathrm{d}\hat{\psi}_w(x).$$

Then from conditions (v4), (v5) and (2.1), we can show that all $nh^{d/2}D_{n1}$, $nh^{d/2}D_{n2}$ and $nh^{d/2}D_{n3}$ are $o_p(1)$. That is, $nh^{d/2}Q'_{n22} = o_p(1)$. Therefore, $nh^{d/2}Q_{n22} = o_p(1)$, and $nh^{d/2}Q_{n2} = o_p(1)$ which is the second claim in (5.27). Thus we have proved (i). Like in the proof of Lemmas 5.2 and 5.3 in KN, one can show (ii) and (iii). Then the desired result follows. □

Lemma 5.5. *Suppose all the conditions in Theorem 3.3 hold; then*

(i) $nh^{d/2}[C_n(\hat{\beta}_n, \hat{\theta}_n) - C_n(\hat{\beta}_n, \theta_0)] = o_p(1)$,

(ii) $nh^{d/2}[C_n(\hat{\beta}_n, \theta_0) - C_n(\beta_0, \theta_0)] = o_p(1)$,

(iii) $nh^{d/2}[C_n(\beta_0, \theta_0) - \tilde{C}_n(\beta_0, \theta_0)] = o_p(1)$.

Proof. Adding and subtracting $v(X_i; \hat{\beta}_n, \theta_0)$ to and from $\hat{\xi}_i = (Y_i - m(X_i; \hat{\beta}_n))^2 - v(X_i; \hat{\beta}_n, \hat{\theta}_n)$, $C_n(\hat{\beta}_n, \hat{\theta}_n)$ can be written as the sum $C_n(\hat{\beta}_n, \theta_0) + 2B_{n1} + B_{n2}$, where

$$B_{n1} = \frac{1}{n^2}\sum_{i=1}^n \int_{\mathcal{C}} K_h^2(x - X_i)[(Y_i - m(X_i; \hat{\beta}_n))^2 - v(X_i; \hat{\beta}_n, \theta_0)][v(X_i; \hat{\beta}_n, \theta_0) - v(X_i; \hat{\beta}_n, \hat{\theta}_n)]\mathrm{d}\hat{\psi}_w(x),$$
$$B_{n2} = \frac{1}{n^2}\sum_{i=1}^n \int_{\mathcal{C}} K_h^2(x - X_i)[v(X_i; \hat{\beta}_n, \theta_0) - v(X_i; \hat{\beta}_n, \hat{\theta}_n)]^2\mathrm{d}\hat{\psi}_w(x).$$

We can see that B_{n2} is bounded above by the sum $B_{n21} + B_{n22}$, where

$$B_{n21} = \frac{2}{n^2}\sum_{i=1}^n \int_{\mathcal{C}} K_h^2(x - X_i)d_{ni}^2\mathrm{d}\hat{\psi}_w(x),$$
$$B_{n22} = \frac{2}{n^2}\sum_{i=1}^n \int_{\mathcal{C}} K_h^2(x - X_i)[\dot{v}'_\theta(X_i; \hat{\beta}_n, \theta_0)(\hat{\theta}_n - \theta_0)]^2\mathrm{d}\hat{\psi}_w(x).$$

By (v4), and the $\sqrt{n}$-consistency of $\hat{\theta}_n$,

$$B_{n21} \le \frac{2}{n^2}\sup_{1\le i\le n}\frac{|d_{ni}|^2}{\|\hat{\theta}_n - \theta_0\|^2}\cdot\|\hat{\theta}_n - \theta_0\|^2\sum_{i=1}^n \int_{\mathcal{C}} K_h^2(x - X_i)\mathrm{d}\hat{\psi}_w(x).$$

Since

$$\frac{1}{n^2}\sum_{i=1}^n \int_{\mathcal{C}} K_h^2(x - X_i)\mathrm{d}\hat{\psi}_w(x) = O_p(1/nh^d),$$

we have $nh^{d/2}|B_{n21}| = nh^{d/2}o_p(1)O_p(1/n)O_p(1/nh^d) = o_p(1)$. For B_{n22}, we have

$$B_{n22} \le \frac{4}{n^2}\|\hat{\theta}_n - \theta_0\|^2\sum_{i=1}^n \int_{\mathcal{C}} K_h^2(x - X_i)\|\dot{v}(X_i; \hat{\beta}_n, \theta_0) - \dot{v}(X_i; \beta_0, \theta_0)\|^2\mathrm{d}\hat{\psi}_w(x)$$
$$+ \frac{4}{n^2}\|\hat{\theta}_n - \theta_0\|^2\sum_{i=1}^n \int_{\mathcal{C}} K_h^2(x - X_i)\|\dot{v}(X_i; \beta_0, \theta_0)\|^2\mathrm{d}\hat{\psi}_w(x).$$

From (v5), and the $\sqrt{n}$-consistency of $\hat{\theta}_n, \hat{\beta}_n$, one can see that the first term is $o_p(1/n^2)$, and the second term is $O_p(1/n)O_p(1/nh^d)$. Therefore, $nh^{d/2}B_{n22} = o_p(1)$. This implies $nh^{d/2}B_{n2} = o_p(1)$. As for B_{n1}, by adding and subtracting $(Y_i - m(X_i; \beta_0))^2 - v(X_i; \beta_0, \theta_0)$ to and from $(Y_i - m(X_i; \hat{\beta}_n))^2 - v(X_i; \hat{\beta}_n, \theta_0)$, it can be written as the sum $B_{n11} + B_{n12} + B_{n13}$, where

$$B_{n11} = \frac{1}{n^2}\sum_{i=1}^{n}\int_{\mathcal{C}} K_h^2(x - X_i)[(Y_i - m(X_i; \hat{\beta}_n))^2 - (Y_i - m(X_i; \beta_0))^2]V_n(x)\mathrm{d}\hat{\psi}_w(x),$$

$$B_{n12} = \frac{1}{n^2}\sum_{i=1}^{n}\int_{\mathcal{C}} K_h^2(x - X_i)[(Y_i - m(X_i; \beta_0))^2 - v(X_i; \beta_0, \theta_0)]V_n(x)\mathrm{d}\hat{\psi}_w(x),$$

$$B_{n13} = \frac{1}{n^2}\sum_{i=1}^{n}\int_{\mathcal{C}} K_h^2(x - X_i)[v(X_i; \beta_0, \theta_0) - v(X_i; \hat{\beta}_n, \theta_0)]V_n(x)\mathrm{d}\hat{\psi}_w(x),$$

and $V_n(x) = v(X_i; \hat{\beta}_n, \theta_0) - v(X_i; \hat{\beta}_n, \hat{\theta}_n)$. B_{n11} can be written as the sum $B'_{n11} + B''_{n11}$, where

$$B'_{n11} = \frac{1}{n^2}\sum_{i=1}^{n}\int_{\mathcal{C}} K_h^2(x - X_i)[m(X_i; \hat{\beta}_n) - m(X_i; \beta_0)]^2 V_n(x)\mathrm{d}\hat{\psi}_w(x),$$

$$B''_{n11} = -\frac{2}{n^2}\sum_{i=1}^{n}\int_{\mathcal{C}} K_h^2(x - X_i)[Y_i - m(X_i; \beta_0)][m(X_i; \hat{\beta}_n) - m(X_i; \beta_0)]V_n(x)\mathrm{d}\hat{\psi}_w(x).$$

Noticing that

$$m(X_i; \hat{\beta}_n) - m(X_i; \beta_0) = e_{ni} + \dot{m}'(X_i; \beta_0)(\hat{\beta}_n - \beta_0), \tag{5.28}$$

and

$$V_n(x) = -d_{ni} - \dot{v}'_\theta(X_i; \hat{\beta}_n, \theta_0)(\hat{\theta}_n - \theta_0), \tag{5.29}$$

then from (m2), (v4), (v5), and the $\sqrt{n}$-consistency of $\hat{\theta}_n, \hat{\beta}_n$, one can show that $nh^{d/2}B'_{n11} = o_p(1)$. Notice that $n^{-2}\sum_{i=1}^{n} \int_{\mathcal{C}} K_h^2(x - X_i)|Y_i - m(X_i; \beta_0)|\mathrm{d}G(x) = O_p(1/nh^d)$; then a similar argument leads to $B''_{n11} = O_p(1/n)O_p(1/nh^d)$, so $nh^{d/2}B''_{n11} = o_p(1)$. This implies $nh^{d/2}B_{n11} = o_p(1)$. Using (5.29) again, we have $nh^{d/2}|B_{n12}| = o_p(1)$. By condition (v4) and (5.29), $nh^{d/2}|B_{n13}| = nh^{d/2}O_p(1/n)O_p(1/nh^d) = O_p(1/nh^{d/2}) = o_p(1)$. Therefore $nh^{d/2}B_{n1} = o_p(1)$, or (i) holds.

Finally, the claims (ii) and (iii) can be shown like in the proof of Lemma 5.4 in KN. □

Lemma 5.6. *Suppose all the conditions in Theorem 3.3 hold; then*

$$\text{(i) } \Gamma_n(\hat{\beta}_n, \hat{\theta}_n) - \Gamma_n(\beta_0, \theta_0) = o_p(1), \qquad \text{(ii) } \Gamma_n(\beta_0, \theta_0) - \tilde{\Gamma}_n(\beta_0, \theta_0) = o_p(1).$$

Proof. By the definition of $\hat{\xi}_i$ and ξ_i, and defining $t_i = (Y_i - m(X_i; \hat{\beta}_n))^2 - (Y_i - m(X_i; \beta_0))^2$, $s_i = v(X_i; \hat{\beta}_n, \hat{\theta}_n) - v(X_i; \beta_0, \theta_0)$, we then have $\hat{\xi}_i = \xi_i + t_i - s_i$. Hence

$$\hat{\xi}_i\hat{\xi}_j = \xi_i\xi_j + \xi_i t_j - \xi_i s_j + t_i\xi_j + t_i t_j - t_i s_j - s_i\xi_j - s_i t_j + s_i s_j.$$

For convenience, define $\delta_{ij} = \hat{\xi}_i\hat{\xi}_j - \xi_i\xi_j$ and $K_{hi} = K_h(x - X_i)$. Then

$$\begin{aligned}\Gamma_n(\hat{\beta}_n, \hat{\theta}_n) = {} & \frac{2h^d}{n^2}\sum_{i\neq j}\left(\int_{\mathcal{C}} K_{hi}K_{hj}\xi_i\xi_j\mathrm{d}\hat{\psi}_w(x)\right)^2 + \frac{2h^d}{n^2}\sum_{i\neq j}\left(\int_{\mathcal{C}} K_{hi}K_{hj}\delta_{ij}\mathrm{d}\hat{\psi}_w(x)\right)^2 \\ & + \frac{4h^d}{n^2}\sum_{i\neq j}\left(\int_{\mathcal{C}} K_{hi}K_{hj}\xi_i\xi_j\mathrm{d}\hat{\psi}_w(x)\right)\left(\int_{\mathcal{C}} K_{hi}K_{hj}\delta_{ij}\mathrm{d}\hat{\psi}_w(x)\right).\end{aligned}$$

Noting that the first term is just $\Gamma_n(\beta_0, \theta_0)$, we have

$$\begin{aligned}|\Gamma_n(\hat{\beta}_n, \hat{\theta}_n) - \Gamma_n(\beta_0, \theta_0)| \leq {} & \frac{2h^d}{n^2}\sum_{i\neq j}\left(\int_{\mathcal{C}} K_{hi}K_{hj}|\delta_{ij}|\mathrm{d}\hat{\psi}_w(x)\right)^2 \\ & + \frac{4h^d}{n^2}\sum_{i\neq j}\left(\int_{\mathcal{C}} K_{hi}K_{hj}|\xi_i\xi_j|\mathrm{d}\hat{\psi}_w(x)\right)\left(\int_{\mathcal{C}} K_{hi}K_{hj}|\delta_{ij}|\mathrm{d}\hat{\psi}_w(x)\right).\end{aligned} \tag{5.30}$$

To proceed, we need the following facts which can be proved using an argument similar to that in KN. For the sake of brevity, details are omitted.

$$\frac{h^d}{n^2}\sum_{i\neq j}\left(\int_{\mathcal{C}} K_{hi}K_{hj}|\xi_i\xi_j|\mathrm{d}\hat{\psi}_w(x)\right)^2 = O_p(1), \tag{5.31}$$

$$\frac{h^d}{n^2}\sum_{i\neq j}\left(\int_{\mathcal{C}} K_{hi}K_{hj}|\xi_i|\mathrm{d}\hat{\psi}_w(x)\right)^2 = O_p(1), \tag{5.32}$$

$$\frac{h^d}{n^2}\sum_{i\neq j}\left(\int_{\mathcal{C}} K_{hi}K_{hj}|\xi_i|k(X_i)\mathrm{d}\hat{\psi}_w(x)\right)^2 = O_p(1), \tag{5.33}$$

$$\frac{h^d}{n^2}\sum_{i\neq j}\left(\int_{\mathcal{C}} K_{hi}K_{hj}\mathrm{d}\hat{\psi}_w(x)\right)^2 = O_p(1), \tag{5.34}$$

where $k(x)$ is such that $\int_{\mathcal{C}} k^2(x)\mathrm{d}G(x) < \infty$. Note that the first term on the right hand side of (5.30) is bounded above by eight terms, such as

$$\frac{8h^d}{n^2}\sum_{i\neq j}\left(\int_{\mathcal{C}} K_{hi}K_{hj}|\xi_i t_j|\mathrm{d}\hat{\psi}_w(x)\right)^2, \qquad \frac{8h^d}{n^2}\sum_{i\neq j}\left(\int_{\mathcal{C}} K_{hi}K_{hj}|\xi_i s_j|\mathrm{d}\hat{\psi}_w(x)\right)^2,$$

etc. All eight terms can be shown as $o_p(1)$. Since the proofs are similar, we only show that the first term above is $o_p(1)$. Since

$$t_i = (m(X_i;\hat{\beta}_n) - m(X_i;\beta_0))^2 - 2(Y_i - m(X_i;\beta_0))(m(X_i;\hat{\beta}_n) - m(X_i;\beta_0)),$$

we have that $2h^d n^{-2}\sum_{i\neq j}\left(\int_{\mathcal{C}} K_{hi}K_{hj}|\xi_i t_j|\mathrm{d}\hat{\psi}_w(x)\right)^2$ will be bounded above by the following two terms:

$$\frac{8h^d}{n^2}\sum_{i\neq j}\left(\int_{\mathcal{C}} K_{hi}K_{hj}|\xi_i(Y_i - m(X_i;\beta_0))|\,|m(X_i;\hat{\beta}_n) - m(X_i;\beta_0)|\mathrm{d}\hat{\psi}_w(x)\right)^2 \tag{5.35}$$

and

$$\frac{4h^d}{n^2}\sum_{i\neq j}\left(\int_{\mathcal{C}} K_{hi}K_{hj}|\xi_i|\,|m(X_i;\hat{\beta}_n) - m(X_i;\beta_0)|^2\mathrm{d}\hat{\psi}_w(x)\right)^2. \tag{5.36}$$

By (m2) and (5.33), we can show that (5.35) has the order $O_p(1/n)$, and (5.32) has the order $O_p(1/n^2)$. Hence $2h^d n^{-2}\sum_{i\neq j}\left(\int_{\mathcal{C}} K_{hi}K_{hj}|\xi_i t_j|\mathrm{d}\hat{\psi}_w(x)\right)^2 = o_p(1)$.

By applying the Cauchy–Schwarz inequality to the double sum, we can also show that the second term on the right hand side of (5.30) is $o_p(1)$. Hence we have proven the first claim of this lemma. Like in the proof of Lemma 5.5 in KN, one can show that (ii) holds. □

Lemma 5.7. *Suppose* (e1), (e2), (e4), (f1), (g), (k), (h1) *and* (v1) *hold; then*

$$nh^{d/2}(\tilde{T}_n(\beta_0,\theta_0) - \tilde{C}_n(\beta_0,\theta_0)) \Longrightarrow_d N(0,\Gamma).$$

Proof. Details of the proof of this theorem are similar to those of the proof of Lemma 5.1 in KN with obvious modifications. □

Proof of Theorem 3.4. Let $Y_i^a = m(X_i;\beta_a) + \sqrt{v_a(X_i)}\varepsilon_i$. Define $K_{hi}(x) = K_h(x - X_i)$, and $K_{wi}(x) = K_w(x - X_i)$. Adding and subtracting Y_i^a to and from Y_i in $T_n(\hat{\beta}_n,\theta_n)$, it can be written as the sum of $T_{n1} + 4T_{n2} + T_{n3} + 4T_{n4}$ and a remainder, where

$$T_{n1} = \int_{\mathcal{C}}\left[\frac{\sum_{i=1}^n K_{hi}(x)(Y_i - Y_i^a)^2}{\sum_{i=1}^n K_{wi}(x)}\right]^2 \mathrm{d}G(x),$$

$$T_{n2} = \int_{\mathcal{C}}\left[\frac{\sum_{i=1}^n K_{hi}(x)(Y_i - Y_i^a)(Y_i^a - m(X_i;\hat{\beta}_n))}{\sum_{i=1}^n K_{wi}(x)}\right]^2 \mathrm{d}G(x),$$

$$T_{n3} = \int_{\mathcal{C}} \left[\frac{\sum_{i=1}^{n} K_{hi}(x)[(Y_i^a - m(X_i; \hat{\beta}_n))^2 - v(X_i; \hat{\beta}_n, \hat{\theta}_n)]}{\sum_{i=1}^{n} K_{wi}(x)} \right]^2 \mathrm{d}G(x),$$

$$T_{n4} = \int_{\mathcal{C}} \left[\frac{\sum_{i=1}^{n} K_{hi}(x)(Y_i - Y_i^a)^2}{\sum_{i=1}^{n} K_{wi}(x)} \right] \left[\frac{\sum_{i=1}^{n} K_{hi}(x)(Y_i - Y_i^a)(Y_i^a - m(X_i; \hat{\beta}_n))}{\sum_{i=1}^{n} K_{wi}(x)} \right] \mathrm{d}G(x).$$

Using the Cauchy–Schwarz inequality, one can show that the remainder is of the order of $o_p(1)$. Note that under H_a, $Y_i - Y_i^a = m(X_i; \beta_0) - m(X_i; \beta_a) + \left[\sqrt{v_1(X_i)} - \sqrt{v_a(X_i)}\right] \varepsilon_i$. Then T_{n1} can be written as the sum $T_{n11} + T_{n12} + T_{n13}$ and a remainder, where

$$T_{n11} = \int_{\mathcal{C}} \left[\frac{\sum_{i=1}^{n} K_{hi}(x)(m_0(X_i) - m_a(X_i))^2}{\sum_{i=1}^{n} K_{wi}(x)} \right]^2 \mathrm{d}G(x),$$

$$T_{n12} = \int_{\mathcal{C}} \left[\frac{\sum_{i=1}^{n} K_{hi}(x)(\sqrt{v_1(X_i)} - \sqrt{v_0(X_i)})^2 \varepsilon_i^2}{\sum_{i=1}^{n} K_{wi}(x)} \right]^2 \mathrm{d}G(x),$$

$$T_{n13} = 2 \int_{\mathcal{C}} \left[\frac{\sum_{i=1}^{n} K_{hi}(x)(m_0(X_i) - m_a(X_i))^2}{\sum_{i=1}^{n} K_{wi}(x)} \right] \cdot \left[\frac{\sum_{i=1}^{n} K_{hi}(x)(\sqrt{v_1(X_i)} - \sqrt{v_0(X_i)})^2 \varepsilon_i^2}{\sum_{i=1}^{n} K_{wi}(x)} \right] \mathrm{d}G(x).$$

While $T_{n11} \to \int_{\mathcal{C}} [m_0(x) - m_a(x)]^4 \mathrm{d}G(x)$, $T_{n12} \to \int_{\mathcal{C}} [\sqrt{v_1(x)} - \sqrt{v_0(x)}]^4 \mathrm{d}G(x)$, and $T_{n13} \to 2 \int_{\mathcal{C}} [m_0(x) - m_a(x)]^2 [\sqrt{v_1(x)} - \sqrt{v_0(x)}]^2 \mathrm{d}G(x)$, the remainder term converges to 0 in probability. So

$$T_{n1} \to \int_{\mathcal{C}} \left([m_0(x) - m_a(x)]^2 + [\sqrt{v_1(x)} - \sqrt{v_0(x)}]^2 \right)^2 \mathrm{d}G(x) \tag{5.37}$$

in probability.

Now, let us consider T_{n2}. Define $m_n(x) = m(x; \hat{\beta}_n)$. By the definition of Y_i^a, T_{n2} can be written as the sum of T_{n21} and a remainder, where

$$T_{n21} = \int_{\mathcal{C}} \left[\frac{\sum_{i=1}^{n} K_{hi}(x)[\sqrt{v_1(X_i)} - \sqrt{v_a(X_i)}]\sqrt{v_a(X_i)}\varepsilon_i^2}{\sum_{i=1}^{n} K_{wi}(x)} \right]^2 \mathrm{d}G(x).$$

Condition (m2), the $\sqrt{n}$-consistency of $\hat{\beta}_n$, and the Cauchy–Schwarz inequality imply the remainder term is $o_p(1)$, and a routing argument leads to $T_{n21} = \int_{\mathcal{C}} [\sqrt{v_1(x)} - \sqrt{v_a(x)}]^2 v_a(x) \mathrm{d}G(x) + o_p(1)$. Hence $T_{n2} \to \int_{\mathcal{C}} [\sqrt{v_1(x)} - \sqrt{v_a(x)}]^2 v_a(x) \mathrm{d}G(x)$ in probability. As for T_{n3}, by arguments similar to those used in proving Theorem 3.3, one can show that

$$nh^{d/2}(T_{n3} - C_n^a) \Longrightarrow N(0, \Gamma_a), \tag{5.38}$$

where

$$C_n^a = \frac{1}{n^2} \sum_{i=1}^{n} \int_{\mathcal{C}} K_h^2(x - X_i)[(Y_i^a - m(X_i; \hat{\beta}_n))^2 - v(X_i; \hat{\beta}_n, \hat{\theta}_n)]^2 \mathrm{d}\hat{\psi}_w(x),$$

and Γ_a is the same as Γ in the null case except for β_0 and θ_0 being replaced by β_a and θ_a, respectively.

Using the definition of Y_i^a, T_{n4} can be written as a sum of twelve terms. One can show that all other terms are negligible in probability, except for the following two terms:

$$B_{n1} = \int_{\mathcal{C}} \left[\frac{\sum_{i=1}^{n} K_{hi}(x)(m_0(X_i) - m_a(X_i))^2}{\sum_{i=1}^{n} K_{wi}(x)} \right] \cdot \left[\frac{\sum_{i=1}^{n} K_{hi}(x)(\sqrt{v_1(X_i)} - \sqrt{v_a(X_i)})\sqrt{v_a(X_i)}\varepsilon_i^2}{\sum_{i=1}^{n} K_{wi}(x)} \right] \mathrm{d}G(x)$$

$$B_{n2} = \int_{\mathcal{C}} \left[\frac{\sum_{i=1}^{n} K_{hi}(x)(\sqrt{v_1(X_i)} - \sqrt{v_a(X_i)})^2\varepsilon_i^2}{\sum_{i=1}^{n} K_{wi}(x)} \right] \cdot \left[\frac{\sum_{i=1}^{n} K_{hi}(x)(\sqrt{v_1(X_i)} - \sqrt{v_a(X_i)})\sqrt{v_a(X_i)}\varepsilon_i^2}{\sum_{i=1}^{n} K_{wi}(x)} \right] \mathrm{d}G(x).$$

In fact, one can show that

$$B_{n1} = \int_{\mathcal{C}} [m_0(x) - m_a(x)]^2(\sqrt{v_1(x)} - \sqrt{v_a(x)})\sqrt{v_a(x)}\mathrm{d}G(x) + o_p(1),$$

$$B_{n2} = \int_{\mathcal{C}} [\sqrt{v_1(x)} - \sqrt{v_a(x)}]^3\sqrt{v_a(x)}\mathrm{d}G(x) + o_p(1).$$

That is

$$T_{n4} = \int_{\mathcal{C}} [m_0(x) - m_a(x)]^2(\sqrt{v_1(x)} - \sqrt{v_a(x)})\sqrt{v_a(x)}\mathrm{d}G(x) + \int_{\mathcal{C}} [\sqrt{v_1(x)} - \sqrt{v_a(x)}]^3\sqrt{v_a(x)}\mathrm{d}G(x) + o_p(1). \tag{5.39}$$

By some simple algebra, one can show that

$$T_{n1} + 4T_{n2} + 4T_{n4} = \Delta + o_p(1). \tag{5.40}$$

Under the alternative hypothesis H_1, $C_n(\hat{\beta}_n, \hat{\theta}_n)$ can be written as C_n^a plus a remainder which can be shown to be a negligible term, while Γ_n, after adding and subtracting Y_i^a to and from Y_i, Y_j^a from Y_j, can be written as a sum of terms bounded in probability. The details are similar to those of Koul and Ni [15], and hence omitted here for the sake of brevity.

Combining the results from (5.40), and the asymptotics of $\hat{\Gamma}(\hat{\beta}_n, \hat{\theta}_n)$ and $C_n(\hat{\beta}_n, \hat{\theta}_n)$, one can see that $nh^{d/2}\Gamma_n^{-1/2}(\hat{\beta}_n, \hat{\theta}_n)$ $(T_n(\hat{\beta}_n, \hat{\theta}_n) - C_n(\hat{\beta}_n, \hat{\theta}_n)) = nh^{d/2}\Gamma_n^{-1/2}(T_{n1} + 4T_{n2} + 4T_{n4}) + o_p(nh^{d/2})$ which tends to ∞ as long as $\Delta > 0$. This implies the consistency of the MD test. Hence the theorem. □

Proof of Theorem 3.5. Details of the proof of this theorem are similar to those for Theorem 3.3 with obvious modification. □

Acknowledgments

The authors are grateful to the Editors and referees for providing constructive criticisms that helped to improve the presentation of the paper.

References

[1] R.J. Beran, Minimum Hellinger distance estimates for parametric models, Ann. Statist. 5 (1977) 445–463.
[2] T.S. Breusch, A.R. Pagan, A simple test for heteroscedasticity and random coefficient variation, Econometrica 47 (1979) 1287–1294.
[3] D. Bosq, Nonparametric Statistics for Stochastic Processes—Estimation and Prediction, 2nd ed., Springer, 1998.
[4] R.J. Carroll, D. Ruppert, Transformation and Weighting in Regression, Chapman and Hall, New York, 1988.
[5] R.D. Cook, S. Weisberg, Diagnostics for heteroscedasticity in regression, Biometrika 70 (1983) 1–10.
[6] H. Dette, A consistent test for heteroscedasticity in nonparametric regression based on the kernel method, J. Statist. Plann. Inference 103 (2002) 311–329.
[7] H. Dette, B. Hetzler, A simple test for parametric form of the variance function in nonparametric regression, Ann. Inst. Statist. Math. 16 (2009) 861–886.
[8] H. Dette, B. Hetzler, Khamaladze transformation of integrated variance processes with applications to goodness-of-fit testing, Math. Methods Statist. 18 (2) (2009) 97–116.
[9] H. Dette, A. Munk, Testing heteroscedasticity in nonparametric regression, J. R. Stat. Soc. Ser. B 60 (1998) 693–708.
[10] H. Dette, N. Neumeyer, I.V. Keilegom, A new test for the parametric form of the variance function in non-parametric regression, J. R. Stat. Soc. Ser. B 69 (2007) 903–917.
[11] A. Diblasi, A. Bowman, Testing for constant variance in a linear model, Statist. Probab. Lett. 33 (1997) 95–103.
[12] M.J. Harrison, B.P.M. McCabe, A test for heteroscedasticity based on least squares residuals, J. Amer. Statist. Assoc. 74 (1979) 494–500.
[13] R.I. Jennrich, Asymptotic properties of non-linear least squares estimators, Ann. Math. Statist. 40 (1969) 633–643.
[14] R. Koenker, G. Bassett, Robust tests for heteroscedasticity based on regression quantiles, Econometrica 50 (1981) 43–61.
[15] H.L. Koul, P. Ni, Minimum distance regression model checking, J. Statist. Plann. Inference 119 (1) (2004) 109–141.
[16] H.L. Koul, W. Song, Model checking with Berkson measurement errors, J. Statist. Plann. Inference 138 (2009) 1615–1628.
[17] H.L. Koul, W. Song, Conditional variance model checking, J. Statist. Plann. Inference 140 (4) (2010) 1056–1072.
[18] H. Liero, Testing homoscedasticity in nonparametric regression, J. Nonparametr. Stat. 15 (1) (2003) 31–51.

[19] Y.P. Mack, B.W. Silverman, Weak and strong uniform consistency of kernel regression estimates, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 61 (1982) 405–415.
[20] L. Wang, X.H. Zhou, Assessing the adequacy of variance function in heteroscedastic regression models, Biometrics 63 (2007) 1218–1225.
[21] H. White, A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity, Econometrica 48 (1980) 817–838.
[22] H. White, Consequences and detection of misspecified nonlinear regression models, JASA 76 (1981) 419–433.
[23] H. White, Maximum likelihood estimation of misspecified models, Econometrica 50 (1982) 1–25.
[24] J.X. Zheng, A consistent test of functional form via nonparametric estimation techniques, J. Econometrics 75 (1996) 263–289.
[25] L.X. Zhu, Y. Fujikoshi, K. Naito, Heteroscedasticity checks for regression models, Sci. China Ser. A 44 (10) (2001) 1236–1252.