Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 99 (2008) 2406–2443
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

Model checking in errors-in-variables regression

Weixing Song

Department of Statistics, Kansas State University, Manhattan, KS 66502, USA

Received 28 January 2007
Available online 10 March 2008

Abstract

This paper discusses a class of minimum distance tests for fitting a parametric regression model to a class of regression functions in the errors-in-variables model. These tests are based on certain minimized distances between a nonparametric regression function estimator and a deconvolution kernel estimator of the conditional expectation of the parametric model being fitted. The paper establishes the asymptotic normality of the proposed test statistics under the null hypothesis and that of the corresponding minimum distance estimators. We also prove the consistency of the proposed tests against a fixed alternative and obtain the asymptotic distributions for general local alternatives. Simulation studies show that the testing procedures are quite satisfactory in the preservation of the finite sample level and in terms of a power comparison.

AMS 1991 subject classifications: primary 62G08; secondary 62G10

Keywords: Errors-in-variables model; Deconvolution kernel estimator; Minimum distance; Lack-of-fit test

1. Introduction

In the errors-in-variables regression model of interest here, one observes Z, Y obeying the model

$$Y = \mu(X) + \varepsilon, \qquad Z = X + u, \tag{1.1}$$

where X is the unobservable d-dimensional random design variables. The random variables (X, u, ε) are assumed to be mutually independent, with u being d-dimensional and ε being one-dimensional having $E(\varepsilon) = 0$, $E(u) = 0$. The marginal densities of X and u will be denoted as f_X, f_u respectively. For the sake of identifiability, the density f_u is assumed to be known. This is

E-mail address: weixing@ksu.edu.

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2008.02.034

a common and standard assumption in the literature of the errors-in-variables regression models. The density f_X and the distribution of ε may not be known.

Errors-in-variables regression models have received continuing attention in the statistical literature. For some literature reviews, see [17,1,16,2,4,13–15,6,7] and the references therein. Most of the existing literature has focused on the estimation problem. The lack-of-fit testing problem has not been discussed thoroughly. Only some sporadic results on this topic can be found in the literature. See [16,5] for some informal lack-of-fit tests in the linear errors-in-variables regression model. The problem of interest in this paper is to develop tests for the hypothesis

$$H_0 : \mu(x) = \theta_0' r(x), \quad \text{for some } \theta_0 \in \mathbb{R}^q, \quad \text{versus } H_1 : H_0 \text{ is not true}, \tag{1.2}$$

in the model (1.1).

Many interesting and profound results, on the other hand, are available for the above testing problem in the absence of errors in the independent variables, that is, for the ordinary regression models. For instance, [10–12,18,21,23,22], among others, give such results.

For the errors-in-variables model, in the case of $r(x) = x$, Zhu, Cui and Ng [25] found a necessary and sufficient condition for the linearity of the conditional expectation $E(\varepsilon|Z)$ with respect to Z. Based on this fact, they constructed a score type lack-of-fit test. This test is of limited use since the normality assumptions on the design variable and measurement errors are rather restrictive. Zhu, Song and Cui [24] and Cheng and Kukush [8] independently extended the method of Zhu, Cui and Ng to deal with a polynomial errors-in-variables model without assuming the normality. The model checking problem for general $r(x)$ was studied by Zhu and Cui [26]. After correcting for the bias of the conditional expectation given Z of least square residuals, they construct a score type test based on the modified residuals, but the theoretical arguments still require the density function of X to be known up to an unknown parameter. This restriction will be removed in the current developments. Cheng and Kukush [8] does not require the density function of X to be known, but their procedure puts very strict restrictions on the moments of the predictor and the measurement error, also their procedure is computationally extensive.

The paper is organized as follows. Section 2 introduces the construction of the test. Section 3 states the needed assumptions and the main results. A multidimensional extension of Lemma A.1 in [20] and the consistency and the asymptotic power of the test against certain fixed alternatives and a class of local alternatives are also stated in Section 3. Section 4 includes some results from finite sample simulation studies. The conclusion and some further discussion on the MD test are present in Section 5. All the technical proofs are postponed to Section 6.

2. Construction of test

The way for constructing tests here is to first recognize that the independence of X and ε and $E(\varepsilon) = 0$ imply that $\nu(z) = E(Y|Z = z) = E(\mu(X)|Z = z)$. Thus one can consider the new regression model $Y = \nu(Z) + \zeta$ in which the error ζ is uncorrelated with Z and has mean 0. The problem of testing for H_0 is now transformed to a test for $\nu(z) = \nu_{\theta_0}(z)$, where $\nu_\theta(z) = \theta' E(r(X)|Z = z)$.

A very important question related to the above transformation is: Are the two hypotheses, $H_{10} : \mu(x) = m_\theta(x)$, for all x, and $H_{20} : \nu(z) = \nu_\theta(z)$, for all z, equivalent? The answer is generally negative, but note that $E(m_1(X)|Z = z) = E(m_2(X)|Z = z)$ is equivalent to $\int m_1(x) f_X(x) f_u(z - x)\mathrm{d}x = \int m_2(x) f_X(x) f_u(z - x)\mathrm{d}x$ for all z, hence if $f_u(z - \cdot)$, as a distribution family with parameter $z \in \mathbb{R}^d$, forms a complete family, then these two

hypotheses are indeed equivalent. This is the case, for example, for double exponential and normal distributions.

For any z for which $f_Z(z) > 0$, we have $\nu(z) = \int \mu(x) f_X(x) f_u(z-x)\mathrm{d}x / f_Z(z)$. If f_X is known then f_Z is known and hence ν_θ is known except for θ. Let $Q(z) = E(r(X)|Z = z)$. Now suppose (Y_i, Z_i), $i = 1, 2, \ldots, n$ are independently and identically distributed copies of (Y, Z) from model (1.1), h is a bandwidth only depending on n and $K_{hi}(z) = K((z - Z_i)/h)/h^d$ for any kernel function K and bandwidth h. If we define

$$\bar{T}_n(\theta) = \int_{\mathcal{I}} \left[\frac{1}{nf_Z(z)} \sum_{i=1}^{n} K_{hi}(z)(Y_i - \theta' Q(Z_i)) \right]^2 \mathrm{d}G(z), \quad \theta \in \mathbb{R}^q,$$

where G is a σ-finite measure on $\mathbb{R}^d$, $\mathcal{I}$ is a compact subset in $\mathbb{R}^d$, then one can see that, $\bar{T}_n$ indeed is a weighted distance between a nonparametric kernel estimator and a parametric estimator of the regression function $\nu(z)$. Then, we may use $\bar{\theta}_n = \arg\min_{\theta \in \mathbb{R}^q} \bar{T}_n(\theta)$ to estimate θ, and construct the test statistic through $\bar{T}_n(\bar{\theta}_n)$. The same method, called minimum distance (MD) procedure, was used in the recent paper of Koul and Ni [19] (K–N) in the classical regression set up. One can see that, if f_X is known, the above test procedure will be a trivial extension of K–N.

Unfortunately, f_X is generally not known and hence f_Z and Q are unknown. This makes the above procedure infeasible. To construct the test statistic, one needs estimators for f_Z and Q. In this connection the deconvolution kernel density estimators are found to be useful here.

For any density L on $\mathbb{R}^d$, let ϕ_L denote its characteristic function and define

$$L_h(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \exp(-\mathbf{i}t \cdot x) \frac{\phi_L(t)}{\phi_u(t/h)} \mathrm{d}t, \qquad \hat{f}_{Xh}(x) = \frac{1}{nh^d} \sum_{i=1}^{n} L_h\left(\frac{x - Z_i}{h}\right),$$
$$x \in \mathbb{R}^d, \tag{2.1}$$

where $\mathbf{i} = (-1)^{1/2}$. The function $\hat{f}_{Xh}$ is called a deconvolution kernel density estimator and it can be used to estimate f_X. See Masry [9]. Note that $Q(z) = R(z)/f_Z(z)$, where

$$R(z) = \int r(x) f_X(x) f_u(z-x)\mathrm{d}x, \qquad f_Z(z) = \int f_X(x) f_u(z-x)\mathrm{d}x. \tag{2.2}$$

Then, one can estimate $Q(z)$ by $\hat{Q}_n(z) = \hat{R}_n(z)/\hat{f}_{Zh}(z)$, where

$$\hat{R}_n(z) = \int r(x) \hat{f}_{Xh}(x) f_u(z-x)\mathrm{d}x, \qquad \hat{f}_{Zh}(z) = \int \hat{f}_{Xh}(x) f_u(z-x)\mathrm{d}x.$$

At this point, it is worth mentioning that, by the definition of L_h and a direct calculation, one can show $\hat{f}_{Zh}$ is nothing but the classical kernel estimator of f_Z with kernel L and bandwidth h. That is, $\hat{f}_{Zh}(z) = \sum_{i=1}^{n} L((z - Z_i)/h)/nh^d$. Our proposed inference procedures will be based on the analogs of $\bar{T}_n$ where $Q(z)$ is replaced by the above estimator $\hat{Q}_n$, and f_Z is replaced by a classical kernel estimator in which a different kernel other than L may be adopted.

It is well known that the convergence rates of the deconvolution kernel density estimators are slower than that of the classical kernel density estimators. See [9,20] and [14]. This creates extra difficulty when considering the asymptotic behaviors of the analogs of the corresponding MD estimators and test statistics. In fact, the consistency of the corresponding MD estimator is still available, but its asymptotic normality and that of the corresponding MD test statistic may

not be obtained. We overcome this difficulty by using different bandwidths and splitting the full sample, say S, with sample size n into two subsamples, S_1 with size n_1, and S_2 with size n_2, then using the subsample S_2 to estimate f_X hence $Q(z)$ and the subsample S_1 to estimate other quantities. The sample size allocation scheme is stated in Section 3. A more detailed discussion on this sample size allocation scheme can be found in Section 5. Without loss of generality, we will number the observations in S_1 from 1 to n_1, and the observations in S_2 from n_1+1 to n. Also all the integration with respect to G in the following will be over the compact subset $\mathcal{I}$.

To be precise, let

$$\tilde{f}_{Zh_2}(z) = \sum_{i=1}^{n_1} K_{h_2 i}(z)/n_1, \qquad \hat{f}_{Xw}(x) = \sum_{j=n_1+1}^{n} L_w((x-Z_j)/w)/n_2 w^d,$$

$$\hat{R}_{n_2}(z) = \int r(x)\hat{f}_{Xw_1}(x) f_u(z-x)\mathrm{d}x,$$

$$\hat{f}_{Zw_2}(z) = \int \hat{f}_{Xw_2}(x) f_u(z-x)\mathrm{d}x, \qquad \hat{Q}_{n_2}(z) = \hat{R}_{n_2}(z)/\hat{f}_{Zw_2}(z),$$

then define, for $\theta \in \mathbb{R}^q$,

$$M_n(\theta) = \int \left[\frac{1}{n_1 \tilde{f}_{Zh_2}(z)} \sum_{i=1}^{n_1} K_{h_1 i}(z)(Y_i - \theta' \hat{Q}_{n_2}(Z_i)) \right]^2 \mathrm{d}G(z), \tag{2.3}$$

with h_1, h_2 depending on n_1, and w_1 and w_2 depending on n_2. One can easily see that $M_n(\theta)$ is a weighted distance between a nonparametric kernel estimator and a deconvolution kernel estimator of the regression function $\nu(z)$ under the null hypothesis. Then we may use

$$\hat{\theta}_n = \arg \inf_{\theta \in \mathbb{R}^q} M_n(\theta) \tag{2.4}$$

to estimate θ, and construct the test statistic through $M_n(\hat{\theta}_n)$. We first prove the consistency of $\hat{\theta}_n$ for θ, then the asymptotic normality of $\sqrt{n_1}(\hat{\theta}_n - \theta_0)$. Finally, let

$$\hat{\zeta}_i = Y_i - \hat{\theta}_n' \hat{Q}_{n_2}(Z_i), \qquad \hat{C}_n = n_1^{-2} \sum_{i=1}^{n_1} \int K_{h_1 i}^2(z) \hat{\zeta}_i^2 \mathrm{d}\hat{\psi}_{h_2}(z),$$

$$\hat{\Gamma}_n = 2h_1^d n_1^{-2} \sum_{i \neq j=1}^{n_1} \left(\int K_{h_1 i}(z) K_{h_1 j}(z) \hat{\zeta}_i \hat{\zeta}_j \mathrm{d}\hat{\psi}_{h_2}(z) \right)^2, \qquad \mathrm{d}\hat{\psi}_{h_2}(z) := \frac{\mathrm{d}G(z)}{\tilde{f}_{Zh_2}^2(z)}.$$

We prove that the asymptotic null distribution of the normalized test statistic

$$\hat{\mathcal{D}}_n = n_1 h_1^{d/2} \hat{\Gamma}_n^{-1/2} (M_n(\hat{\theta}_n) - \hat{C}_n) \tag{2.5}$$

is standard normal. Consequently, the test that rejects H_0 whenever $|\hat{\mathcal{D}}_n| > z_{\alpha/2}$ is of the asymptotic size α, where z_α is the $100(1-\alpha)\%$ percentile of the standard normal distribution.

3. Assumptions and main results

This section first states the various conditions needed in the subsequent sections. About the errors, the underlying design and the integrating σ-finite measure G, we assume the following:

(**e1**) The random variables $\{(Z_i, Y_i) : Z_i \in \mathbb{R}^d, Y_i \in \mathbb{R}, i = 1, 2, \ldots, n\}$ are independently and identically distributed with the conditional expectation $\nu(z) = E(Y|Z = z)$ satisfying $\int \nu^2 \mathrm{d}G < \infty$, where G is a σ-finite measure on $\mathbb{R}^d$.

(**e2**) $0 < \sigma_\varepsilon^2 = E\varepsilon^2 < \infty$, $E\|r(X)\|^2 < \infty$, where $\|\cdot\|$ denotes the usual Euclidean norm. The function $\delta^2(z) = E[(\theta_0^{\mathrm{T}} r(X) - \theta_0^{\mathrm{T}} Q(Z))^2 | Z = z]$ is a.s. (G) continuous.

(**e3**) $E|\varepsilon|^{2+\delta} < \infty$, $E\|r(X)\|^{2+\delta} < \infty$, for some $\delta > 0$.

(**e4**) $E|\varepsilon|^4 < \infty$, $E\|r(X)\|^4 < \infty$.

(**u**) The density function f_u is continuous and $\int |\phi_u(t)| \mathrm{d}t < \infty$.

(**f1**) The density f_X and its all possible first and second derivatives are continuous and bounded.

(**f2**) For some $\delta_0 > 0$, the density f_Z is bounded below on the compact subset $\mathcal{I}_{\delta_0}$ of $\mathbb{R}^d$, where

$$\mathcal{I}_{\delta_0} = \{y \in \mathbb{R}^d : \max_{1 \le j \le d} |y_j - z_j| \le \delta_0, y = (y_1, \ldots, y_d)', z = (z_1 z_d)', z \in \mathcal{I}\},$$

(**g**) G has a continuous Lebesgue density g.

(**q**) $\Sigma_0 = \int Q(z) Q'(z) \mathrm{d}G(z)$ is positive definite.

About the null model we need to assume the following:

(**m1**) There exist a positive continuous function $J(z)$ and a positive number T_0, such that for all t with $\|t\| > T_0$,

$$\|t\|^{-\alpha} \left\| \frac{\int (r(z - x) - r(z)) \exp(-\mathbf{i} t' x) f_u(x) \mathrm{d}x}{\phi_u(t)} \right\| \le J(z),$$

holds for some $\alpha \ge 0$ and all $z \in \mathbb{R}^d$, and $EJ^2(Z) < \infty$.

(**m2**) $E\|r(Z)\|^2 < \infty$, $EI^2(Z) < \infty$, where $I(z) = \int \|r(x)\| f_u(z - x) \mathrm{d}x$.

About the kernel functions, we assume:

(ℓ) The kernel function L is a density, symmetric around the origin, $\sup_{t \in \mathbb{R}^d} \|t\|^\alpha |\phi_L(t)| < \infty$, for all $t \in \mathbb{R}^d$; moreover, $\int \|v\|^2 L(v) \mathrm{d}v < \infty$ and $\int \|t\|^\alpha |\phi_L(t)| \mathrm{d}t < \infty$ with α as in (m1).

About the bandwidths and sample size we need to assume the following:

(**n**) With n denoting the sample size, let n_1, n_2 be two positive integers such that $n = n_1 + n_2$, $n_2 = [n_1^b]$, $b > 1 + (d + 2\alpha)/4$, where α is as in (m1).

(**h1**) $h_1 \sim n_1^{-a}$, where $0 < a < \min(1/2d, 4/d(d + 4))$. (**h2**) $h_2 = c_1 (\log(n_1)/n_1)^{1/(d+4)}$.

(**w1**) $w_1 = n_2^{-1/(d+4+2\alpha)}$. (**w2**) $w_2 = c_2 (\log(n_2)/n_2)^{1/(d+4)}$.

Assumption (m1) is not so strict as it appears. Some commonly used regression functions such as polynomial and exponential functions indeed satisfy this assumption as shown below.

Example 1. Suppose $d = q$, $r(x) = x$, and $u \sim N_d(0, \Sigma_u)$. Then,

$$\begin{aligned} \left\| \frac{\int (r(z - x) - r(z)) \exp(-it' x) f_u(x) \mathrm{d}x}{\phi_u(t)} \right\| &= \left\| \int x \exp(-it' x) f_u(x) \mathrm{d}x \right\| \cdot \exp\left(\frac{t' \Sigma_u t}{2} \right) \\ &= \left\| \frac{\partial \phi_u(t)}{i \partial t} \right\| \cdot \exp(t' \Sigma_u t/2) \le c\|t\|, \end{aligned}$$

where the constant c depends only on Σ_u. Hence (m1) holds with $\alpha = 1$ and $J(z) = c$.

Example 2. Suppose $d = q = 1$, $r(x) = x^2$, and u has a double exponential distribution with mean 0 and variance σ_u^2. In this case, $\phi_u(t) = 1/(1 + \sigma_u^2 t^2/2)$ and

$$\begin{aligned}\left|\frac{\int (r(z-x) - r(z)) \exp(-\mathbf{i}tx) f_u(x)\mathrm{d}x}{\phi_u(t)}\right| &= \left|\int (-2zx + x^2) \exp(-\mathbf{i}tx) f_u(x)\mathrm{d}x\right| \Big/ |\phi_u(t)| \\ &\le 2|z| \left|\frac{\partial \phi_u(t)}{i\partial t}\right| \Big/ |\phi_u(t)| + \left|\frac{\partial^2 \phi_u(t)}{\partial t^2}\right| \Big/ |\phi_u(t)| \\ &\le 2|z| \frac{\sigma_u^2 |t|}{1 + \sigma_u^2 t^{2/2}} + \frac{\sigma_u^2}{1 + \sigma_u^2 t^{2/2}} \\ &\quad + \frac{2\sigma_u^4 t^2}{(1 + \sigma_u^2 t^{2/2})^2}.\end{aligned}$$

Hence, as $|t| \to \infty$, (m1) holds for $\alpha = 0$ and, $J(z) = 2|z| + 2$. One can easily verify that a similar result holds for $r(x) = x^k$, where k is any positive integer, hence for $r(x)$ being polynomials of x.

Example 3. Suppose $d = q = 1$, $r(x) = \mathrm{e}^x$, and $u \sim N(0, \sigma_u^2)$. Then

$$\begin{aligned}&\left|\int (r(z-x) - r(z)) \exp(-\mathbf{i}tx) f_u(x)\mathrm{d}x\right| = \left|\int (\mathrm{e}^{z-x} - \mathrm{e}^z) \exp(-\mathbf{i}tx) f_u(x)\mathrm{d}x\right| \\ &\quad \le \mathrm{e}^z \left[\left|\int \mathrm{e}^x \mathrm{e}^{\mathrm{i}tx} f_u(x)\mathrm{d}x\right| + |\phi_u(t)|\right] \le c\mathrm{e}^z |\phi_u(t)|,\end{aligned}$$

where c is some positive number depending only on σ_u^2. Hence, (m1) holds for $\alpha = 0$ and, $J(z) = c\mathrm{e}^z$.

Next, we give some general preliminaries needed for the proofs below.

The following lemma is a multidimensional extension of a Stefanski and Carroll [20] result which will be frequently used in the following.

Lemma 3.1. *Suppose $d \ge 1$, and* (f1), (u), (m1), (h1) *hold. Then for any $z \in \mathbb{R}^d$,*

$$\begin{aligned}\|E\hat{R}_{n_2}(z) - R(z)\|^2 \le c w_1^4 I^2(z), \quad E\|\hat{R}_{n_2}(z) - E\hat{R}_{n_2}(z)\|^2 \le \frac{c}{n_2 w_1^d}(J^2(z) w_1^{-2\alpha} \\ + \|r(z)\|^2),\end{aligned}$$

where $R(z)$ is as in (2.2), $I(z)$ is as in (m2)*, $J(z)$ is as in* (m1) *and c is a constant not depending on z, n_2 and w_1.*

By the usual bias and variance decomposition of mean square error, the following inequality is a direct consequence of Lemma 3.1,

$$E\|\hat{R}_{n_2}(z) - R(z)\|^2 \le c w_1^4 I^2(z) + \frac{c}{n_2 w_1^d}(J^2(z) w_1^{-2\alpha} + \|r(z)\|^2).$$

If the bandwidth w_1 is chosen by assumption (w1), then

$$E\|\hat{R}_{n_2}(z) - R(z)\|^2 \le c n_2^{-\frac{4}{d+2\alpha+4}} (I^2(z) + J^2(z) + \|r(z)\|^2). \tag{3.1}$$

In the following, we will write

$$\mathcal{T}(z) = I^2(z) + J^2(z) + \|r(z)\|^2. \tag{3.2}$$

The following lemma will be used repeatedly, which, along with its proof, appears as Theorem 2.2 part (2) in [3]. We state the lemma for a sample size n and a bandwidth h, they may be replaced by n_1 or n_2, h_2 or w_2 according to the context.

Lemma 3.2. *Let $\hat{f}_Z$ be the kernel estimator with a kernel K which satisfies a Lipschitz condition and has bandwidth h. If f_Z is twice continuously differentiable, and the bandwidth h is chosen to be $c_n(\log(n)/n)^{1/(d+4)}$, where $c_n \to c > 0$, then*

$$(\log_k n)^{-1}(n/\log(n))^{2/(d+4)} \sup_{z\in\mathcal{I}} |\hat{f}_Z(z) - f_Z(z)| \to 0 \quad a.s.$$

for any positive integer k and compact set $\mathcal{I}$.

Recall the definitions in (2.3). Because the null model is linear in θ, so the minimizer $\hat{\theta}_n$ has an explicit form obtained by setting the derivative of $M_n(\theta)$ with respect to θ equal to 0, which gives the equation

$$\begin{aligned}&\int \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\hat{Q}_{n_2}(Z_i)\cdot\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\hat{Q}'_{n_2}(Z_i)\mathrm{d}\hat{\psi}_{h_2}(z)\cdot\hat{\theta}_n\\ &= \int \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)Y_i\cdot\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\hat{Q}_{n_2}(Z_i)\mathrm{d}\hat{\psi}_{h_2}(z).\end{aligned} \tag{3.3}$$

Adding and subtracting $\theta_0'\hat{Q}_{n_2}(Z_i)$ from Y_i, and doing some routine arrangement, $\hat{\theta}_n$ will satisfy the following equation:

$$\begin{aligned}&\int \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\hat{Q}_{n_2}(Z_i)\cdot\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\hat{Q}'_{n_2}(Z_i)\mathrm{d}\hat{\psi}_{h_2}(z)\cdot(\hat{\theta}_n-\theta_0)\\ &= \int \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)(Y_i-\theta_0'\hat{Q}_{n_2}(Z_i))\cdot\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\hat{Q}_{n_2}(Z_i)\mathrm{d}\hat{\psi}_{h_2}(z).\end{aligned} \tag{3.4}$$

The above explicit relation between $\hat{\theta}_n - \theta_0$ and other quantities allows us, compared to K–N, to investigate the asymptotic distribution of $\hat{\theta}_n$ without proving the consistency in advance. Most importantly, the separation of $\hat{\theta}_n$ from $\hat{R}_{n_2}(z)$ makes a conditional expectation argument in the following proofs relatively easy.

The asymptotic distributions of $\hat{\theta}_n$ and $M_n(\hat{\theta}_n)$ under the null hypothesis are summarized in the following theorems.

Theorem 3.1. *Suppose H_0,* (e1), (e2), (e3), (u), (f1), (f2), (q), (m1), (m2), *(ℓ),* (n), (h1), (h2), (w1), *and* (w2) *hold, then $\sqrt{n_1}(\hat{\theta}_n - \theta_0) \Longrightarrow N_d(0, \Sigma_0^{-1}\Sigma\Sigma_0^{-1})$, where*

$$\Sigma = \int \tau^2(z)Q(z)Q'(z)g^2(z)/f(z)\mathrm{d}z,$$

Σ_0 is as in condition (q), and $\tau^2(z) = \sigma_\varepsilon^2 + \delta^2(z)$, where σ_ε^2, and $\delta^2(z)$ are as in (e2).

Theorem 3.2. *Suppose H_0, and the conditions* (e1), (e2), (e4), (u), (f1), (f2), (q), (m1), (m2), *(ℓ)*, (n), (h1), (h2), (w1) *and* (w2) *hold, then $\hat{\mathcal{D}}_n \Longrightarrow N(0,1)$, where $\hat{\mathcal{D}}_n$ is as in* (2.5).

We end this section by adding some remarks. First, the MD estimator and testing procedure depend on the choice of the integrating measure. In the classical regression case, K–N provides some guidelines on how to choose G. The same guidelines also apply here. For example, in the one-dimensional case, the asymptotic variance of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ can attain its minimum if G is chosen to be $\hat{f}_{Zh_2}(z)$. As far as the MD test statistic $M_n(\hat{\theta}_n)$ is concerned, the choice of G will depend on the alternatives. In the classical regression case, K–N found that the test has high power against the selected alternatives, if the density function is chosen to be the square of the density estimator of the design variables. The same phenomenon happens in our case. Secondly, since replacing $\hat{\Gamma}_n$ in (2.5) by any other consistent estimator of Γ does not affect the validity of Theorem 3.2, where

$$\Gamma = 2\int (\sigma_e^2(z))^2 g(z)\mathrm{d}\psi(z) \cdot \int \left(\int K(u)K(u+v)\mathrm{d}u\right)^2 \mathrm{d}v, \tag{3.5}$$

$\sigma_e^2(z) = \sigma_\varepsilon^2 + \delta^2(z)$, $\delta^2(z)$ is as in condition (e2), so we can choose some other consistent estimator of Γ, such as

$$\bar{\Gamma}_n = C\int \left(\frac{\sum_{i=1}^{n_1} K_{h_1 i}(z)(Y_i - \hat{\theta}_n'\hat{Q}_{n_2}(Z_i))^2}{n_1 \hat{f}_{Zh_2}(z)}\right)^2 g(z)\mathrm{d}\hat{\psi}_{h_2}(z), \tag{3.6}$$

to make the test computationally efficient, where the constant $C = 2\int[\int K(u)K(u+v)\mathrm{d}u]^2\mathrm{d}v$.

Finally we present some theoretical results about asymptotic power of the proposed tests.

Let $m(x)$ be a Borel measurable real-valued function of $x \in \mathbb{R}^d$, and $H(z) = E(m(X)|Z = z)$ such that $H(z) \in L_2(G)$. We will show that the MD estimator defined by (2.4) converges to some finite constant in probability, then based on this result, one can show the consistency of the MD test against certain fixed alternatives. In fact, we have

Theorem 3.3. *Suppose the conditions in Theorem 3.2 and the alternative hypothesis H_a : $\mu(x) = m(x), \forall x$ hold with the additional assumption that* $\inf_\theta \int [H(z) - \theta' Q(z)]^2 \mathrm{d}G(z) > 0$. *Then, for the MD estimator $\hat{\theta}_n$ defined in* (2.4), $|\hat{\mathcal{D}}_n| \to \infty$ *in probability.*

Now we consider the asymptotic power of the proposed MD tests against the following local alternatives.

$$H_{na} : \mu(x) = \theta_0' r(x) + \gamma_n v(x), \quad \gamma_n = 1\Big/\sqrt{n_1 h_1^{d/2}} \tag{3.7}$$

where $v(x)$ is an arbitrary and known continuous real-valued function with $V(z) = E(v(X)|Z = z) \in L_2(G)$. The following theorem gives asymptotic distribution of the MD test against the local alternative (3.7). This enables us to investigate the asymptotic local power of the MD test.

Theorem 3.4. *Suppose the conditions in Theorem 3.2. Then under the local alternative* (3.7), *we have $\hat{\mathcal{D}}_n \to_d N(\Gamma^{-1/2}D, 1)$, where*

$$D = \int V^2(z)\mathrm{d}G(z) + \int V(z)Q'(z)\mathrm{d}G(z) \cdot \int Q(z)Q'(z)\mathrm{d}G(z) \cdot \int V(z)Q(z)\mathrm{d}G(z)$$

and Γ is as in (3.5).

Table 4.1
Mean and MSE of $\hat{\theta}_n$

	(50, 134)	(100, 317)	(200, 753)	(300, 1250)	(500, 2366)
Mean	1.0103	1.0095	1.0102	1.0105	1.0098
MSE	0.0014	0.0007	0.0004	0.0003	0.0002

4. Monte Carlo simulation

This section contains the results of four simulations corresponding to the following cases: Case 1: $d = q = 1$ and m_θ linear, the measurement error ϵ is chosen to be normal and u double exponential; Case 2: $d = q = 1$ and m_θ linear, the measurement error ϵ and u are chosen to be normal; Case 3: $d = 1, q = 2$ and m_θ a polynomial, the measurement error ϵ is chosen to be normal and u double exponential; Case 4: $d = q = 2$, and m_θ linear, the measurement error ϵ is chosen to be normal and u double exponential. It is easy to check that the models being simulated below satisfy all the conditions stated in Section 3. In each case the Monte Carlo average of $\hat{\theta}_n$, MSE($\hat{\theta}_n$), empirical levels and powers of the MD test are reported. The asymptotic level is taken to be 0.05 in all cases. For any random variable W, we will use $\{W_{jk_j}\}_{k_j=1}^{n_j}$, $j = 1, 2$ to denote the jth subsample S_j from W with sample size n_j. So the full sample is $S_1 \bigcup S_2$. Finally, to make the simulation less time consuming, $\bar{\Gamma}_n$ defined in (3.6) will be used in the test statistic instead of $\hat{\Gamma}_n$. So the value of the test statistic is calculated by $\widehat{\mathcal{D}}_n = n_1 h_1^{1/2} \bar{\Gamma}_n^{-1/2}(M_n(\hat{\theta}_n) - \hat{C}_n)$.

Case 1. In this case, $\{X_{jk_j}\}_{k_j=1}^{n_j}$ are obtained as a random sample from the uniform distribution on $[-1, 1]$, $\{\varepsilon_{jk_j}\}_{k_j=1}^{n_j}$ are obtained as a random sample from the normal distribution $\mathcal{N}(0, 0.1^2)$ and $\{u_{jk_j}\}_{k_j=1}^{n_j}$ are obtained as a random sample from the double exponential distribution with mean 0 and variance 0.01. The parametric model is taken to be $m_\theta(X) = \theta X$, and the true parameter $\theta_0 = 1$. Then (Y_i, Z_i) are generated using the model $Y_{jk_j} = X_{jk_j} + \varepsilon_{jk_j}$, $Z_{jk_j} = X_{jk_j} + u_{jk_j}$, $k_j = 1, 2, \ldots, n_j$, $j = 1, 2$. From Example 2, we know that the assumption (m1) is held for $\alpha = 0$. The kernel functions K and K^* and the bandwidths used in all the simulations are

$$K(z) = K^*(z) = \frac{3}{4}(1 - z^2) I(|z| \le 1), \quad h_1 = a n_1^{-1/3}, \quad h_2 = b n_1^{-1/5} (\log n_1)^{1/5}, \quad (4.1)$$

with some choices for a and b. For the chosen kernel function (4.1), the constant C in $\bar{\Gamma}_n$ is equal to 0.7642. The kernel function used in (2.1) is chosen to be the standard normal, so that the deconvolution kernel function with bandwidth w takes the form $L_w(x) = \exp(-x^2/2)[1 - 0.005(x^2 - 1)/w^2]/\sqrt{2\pi}$, and the bandwidth $w_1 = n_2^{-1/5}$, $w_2 = (\log(n_2)/n_2)^{1/5}$ which are chosen by the assumptions (w1) and (w2). Correspondingly, $\hat{Q}_{n_2}(z) = y\hat{R}_{n_2}(z)/\tilde{f}_{Zw_2}(z)$, where

$$\hat{R}_{n_2}(z) = \int x \hat{f}_{Xw_1}(x) f_u(z - x) \mathrm{d}x, \qquad \tilde{f}_{Zw_2} = \int \hat{f}_{Xw_2}(x) f_u(z - x) \mathrm{d}x.$$

Table 4.1 reports the Monte Carlo mean and the MSE($\hat{\theta}_n$) under H_0 for the sample sizes $(n_1, n_2) = (50, 134), (100, 317), (200, 753), (300, 1250), (500, 2366)$, each repeated 1000 times. One can see that, there appears to be a small bias in $\hat{\theta}_n$ for all chosen sample sizes and as expected, the MSE decreases as the sample size increases.

Table 4.2
Levels and powers of the minimum distance test

Model	(a, b)	(50, 134)	(100, 317)	(200, 753)	(300, 1250)	(500, 2366)
	(0.3, 0.5)	0.003	0.008	0.009	0.020	0.041
	(0.3, 0.8)	0.008	0.014	0.017	0.031	0.053
Model 0	(0.5, 0.5)	0.010	0.011	0.020	0.030	0.049
	(1.0, 0.8)	0.024	0.028	0.026	0.039	0.050
	(1.0, 1.0)	0.028	0.037	0.030	0.048	0.054
	(0.3, 0.5)	0.407	0.865	0.987	0.997	1.000
	(0.3, 0.8)	0.491	0.888	0.990	0.998	1.000
Model 1	(0.5, 0.5)	0.704	0.975	0.999	1.000	1.000
	(1.0, 0.8)	0.921	0.999	1.000	1.000	1.000
	(1.0, 1.0)	0.926	0.997	1.000	1.000	1.000
	(0.3, 0.5)	0.898	0.972	0.999	0.999	1.000
	(0.3, 0.8)	0.919	0.976	0.999	0.999	1.000
Model 2	(0.5, 0.5)	0.985	0.999	0.999	1.000	1.000
	(1.0, 0.8)	0.999	1.000	1.000	1.000	1.000
	(1.0, 1.0)	0.999	1.000	1.000	1.000	1.000
	(0.3, 0.5)	0.774	0.959	0.993	0.998	1.000
	(0.3, 0.8)	0.807	0.964	0.993	0.998	1.000
Model 3	(0.5, 0.5)	0.933	0.966	0.999	1.000	1.000
	(1.0, 0.8)	0.992	1.000	1.000	1.000	1.000
	(1.0, 1.0)	0.988	1.000	1.000	1.000	1.000

To assess the level and power behavior of the $\widehat{\mathcal{D}}_n$ test, we chose the following four alternative models for simulation:

$$\text{Model 0: } Y = X + \varepsilon,$$
$$\text{Model 1: } Y = X + 0.3X^2 + \varepsilon,$$
$$\text{Model 2: } Y = X + 1.4\exp(-0.2X^2) + \varepsilon,$$
$$\text{Model 3: } Y = XI(X \geq 0.2) + \varepsilon.$$

To assess the effect of the choice of (a, b) that appears in the bandwidths on the level and power, we ran the simulations for numerous choices of (a, b), ranging from 0.3 to 1. Table 4.2 reports the simulation results pertaining to $\widehat{\mathcal{D}}_n$ for three choices of (a, b). The simulation results for the other choices were similar to those reported here. Data from Model 0 in this table are used to study the empirical sizes, and from Models 1 to 3 are used to study the empirical powers of the test. These entities are obtained by computing $\#\{|\widehat{\mathcal{D}}_n| \geq 1.96\}/1000$.

From Table 4.2, one sees that the empirical level is sensitive to the choice of (a, b) for moderate sample sizes ($n_1 \leq 200$) but gets closer to the asymptotic level of 0.05 with the increase in the sample size, and hence is stable over the chosen values of (a, b) for large sample sizes. On the other hand the empirical power appears to be far less sensitive to the values of (a, b) for the sample sizes of 100 and more. Even though the theory of the present paper is not applicable to model 3, it was included here to see the effect of the discontinuity in the regression function on the power of the minimum distance test. In our simulation, the discontinuity of the regression has little effect on the power of the minimum distance test.

We also conduct a simulation in which the predictor X follows a normal distribution. The results are similar to the results reported above, hence are omitted.

Table 4.3
Mean and MSE of $\hat{\theta}_n$

	(50, 941)	(100, 3164)	(200, 10643)	(300, 21638)	(500, 52902)
Mean	1.0051	1.0078	1.0085	1.0101	1.0169
MSE	0.0013	0.0007	0.0004	0.0003	0.0004

Table 4.4
Levels and powers of the minimum distance test

Model	(50, 941)	(100, 3164)	(200, 10643)	(300, 21638)	(500, 52902)
Model 0	0.018	0.022	0.029	0.035	0.049
Model 1	0.918	0.999	1.000	1.000	1.000
Model 2	0.999	1.000	1.000	1.000	1.000
Model 3	0.993	1.000	1.000	1.000	1.000

Case 2: The measurement error in this case has normal distribution $\mathcal{N}(0, (0.1)^2)$, x is generated from uniform distribution $U[-1, 1]$ and $\varepsilon \sim N(0, 0.1^2)$. By Example 1 in Section 2, we see the assumption (m1) is satisfied with $\alpha = 1$. Hence, by the sample allocation scheme (n), the sample sizes $n_2 = [n_1]^b$, $b > 7/4$. In the simulation, we choose $b = 7/4 + 0.0001$. The bandwidths are chosen to be

$$h_1 = n_1^{1/3}, \qquad h_2 = (\log(n_1)/n_1)^{1/5}, \qquad w_1 = n_2^{-1/7}, \qquad w_2 = (\log(n_2)/n_2)^{1/5}$$

by the assumptions (h1), (h2), (w1) and (w2). The kernel functions K, K^* are the same as in the first case, while the density function L has a Fourier transform given by $\phi_L(t) = \max\{(1-t^2)^3, 0\}$, the corresponding deconvolution kernel function then takes the form

$$L_w(x) = \frac{1}{\pi}\int_0^1 \cos(tx)(1-t^2)^3 \exp(0.005t^2/w^2)\mathrm{d}t.$$

Table 4.3 reports the Monte Carlo mean and the MSE of the MD estimator $\hat{\theta}_n$ under H_0. One can see that, there appears to be a small bias in $\hat{\theta}_n$ for all chosen sample sizes and as expected, the MSE decreases as the sample size increases.

To assess the level and power behavior of the $\widehat{\mathcal{D}}_n$ test, we chose the following four alternative models for simulation:

Model 0: $Y = X + \varepsilon$,
Model 1: $Y = X + 0.3X^2 + \varepsilon$,
Model 2: $Y = X + 1.4\exp(-0.2X^2) + \varepsilon$,
Model 3: $Y = XI(X \geq 0.2) + \varepsilon$.

Table 4.4 reports the simulation results pertaining to $\widehat{\mathcal{D}}_n$. Data from Model 0 in this table are used to study the empirical sizes, and from Models 1 to 3 are used to study the empirical powers of the test.

Case 3: This simulation considers the case of $d = 1, q = 2$. Everything here is the same as in Case 1 except that the null model to test is $m_\theta(X) = \theta_1 X + \theta_2 X^2$. The true parameters are

Table 4.5
Mean and MSE of $\hat{\theta}_n$

	(50, 134)	(100, 317)	(200, 753)	(300, 1250)	(500, 2366)
Mean of $\hat{\theta}_{n1}$	1.0169	1.0144	1.0139	1.0136	1.0128
MSE of $\hat{\theta}_{n1}$	0.0058	0.0031	0.0015	0.0011	0.0007
Mean of $\hat{\theta}_{n2}$	2.0450	2.0452	2.0463	2.0493	2.0473
MSE of $\hat{\theta}_{n2}$	0.0124	0.0076	0.0046	0.0042	0.0033

Table 4.6
Levels and powers of the minimum distance test

Model	(50, 134)	(100, 317)	(200, 753)	(300, 1250)	(500, 2366)
Model 0	0.001	0.009	0.019	0.029	0.046
Model 1	0.297	0.815	0.999	1.000	1.000
Model 2	0.528	0.965	0.999	1.000	1.000
Model 3	0.996	0.999	1.000	1.000	1.000

$\theta_1 = 1, \theta_2 = 2$. It is easy to see that $\hat{R}_{n_2}(z)$ takes the form

$$\hat{R}_{n_2}(z) = \left(\int x \hat{f}_{Xw_1}(x) f_u(z-x)\mathrm{d}x, \int x^2 \hat{f}_{Xw_1}(x) f_u(z-x)\mathrm{d}x \right)'.$$

Table 4.5 reports the Monte Carlo mean and the MSE of the MD estimator $\hat{\theta}_n = (\hat{\theta}_{n1}, \hat{\theta}_{n2})$ under H_0. One can see that, there appears to be a small bias in $\hat{\theta}_n$ for all chosen sample sizes and as expected, the MSE decreases as the sample size increases.

To assess the level and power behavior of the $\widehat{\mathcal{D}}_n$ test, we chose the following four models to simulate data from.

Model 0: $Y = X + 2X^2 + \varepsilon$,
Model 1: $Y = X + 2X^2 + 0.3X^3 + 0.1 + \varepsilon$,
Model 2: $Y = X + 2X^2 + 1.4\exp(-0.2X^2) + \varepsilon$,
Model 3: $Y = X + 2X^2 \sin(X) + \varepsilon$.

Table 4.6 reports the simulation results pertaining to $\widehat{\mathcal{D}}_n$. Data from Model 0 in this table are used to study the empirical sizes, and from Models 1 to 3 are used to study the empirical powers of the test.

Case 4: This simulation considers the case of $d = 2, q = 2$. The null model we want to test is $m_\theta(X) = \theta_1 X_1 + \theta_2 X_2$. X_1 and X_2 are both generated from uniform distribution $U[-1, 1]$, $\varepsilon \sim N(0, 0.1^2)$, and the measurement error is generated from double exponential distribution with mean 0 and variance 0.01. The true parameters are $\theta_1 = 1, \theta_2 = 2$. The kernel functions K and K^* and the bandwidths used in the simulation are

$$K(z_1, z_2) = K^*(z_1, z_2) = \frac{9}{16}(1 - z_1^2)(1 - z_2^2) I(|z_1| \le 1, |z_2| \le 1), \tag{4.2}$$

$h_1 = n_1^{-1/5}, h_2 = n_1^{-1/6}(\log n_1)^{1/6}$. For the chosen kernel function (4.2), the constant C in $\bar{\Gamma}_n$ is equal to 0.292. The kernel function used in (2.1) is chosen to be the bivariate standard normal,

Table 4.7
Mean and MSE of $\hat{\theta}_n$

	(50, 354)	(100, 1001)	(200, 2830)	(300, 5200)	(500, 11188)
Mean of $\hat{\theta}_{n1}$	1.0099	1.0120	1.0115	1.0094	1.0113
MSE of $\hat{\theta}_{n1}$	0.0042	0.0019	0.0011	0.0008	0.0005
Mean of $\hat{\theta}_{n2}$	2.0202	2.0220	2.0213	2.0225	2.0209
MSE of $\hat{\theta}_{n2}$	0.0042	0.0027	0.0014	0.0011	0.0008

Table 4.8
Levels and powers of the minimum distance test

Model	(50, 354)	(100, 1001)	(200, 2830)	(300, 5200)	(500, 11188)
Model 0	0.002	0.012	0.018	0.016	0.038
Model 1	0.908	0.998	1.000	1.000	1.000
Model 2	0.992	0.999	1.000	1.000	1.000
Model 3	0.935	0.996	1.000	1.000	1.000

so the deconvolution kernel function with bandwidth w takes the form

$$L_w(x) = \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right)\left[1 - \frac{0.005(x_1^2 - 1)}{w^2}\right]\left[1 - \frac{0.005(x_2^2 - 1)}{w^2}\right].$$

Since (m1) holds for $\alpha = 0$, so the bandwidths $w_1 = n_2^{-1/6}$, $w_2 = (\log(n_2)/n_2)^{1/6}$ which are chosen by assumption (w1) and (w2). According to the assumption (n) we take $n_2 = n_1^{1.5001}$.

Table 4.7 reports the Monte Carlo mean and the MSE of the MD estimator $\hat{\theta}_n = (\hat{\theta}_{n1}, \hat{\theta}_{n2})$ under H_0. One can see that, there appears to be a small bias in $\hat{\theta}_n$ for all chosen sample sizes and as expected, the MSE decreases as the sample size increases.

To assess the level and power behavior of the $\widehat{\mathcal{D}}_n$ test, we chose the following four models to simulate data from.

Model 0: $Y = X_1 + 2X_2 + \varepsilon$,
Model 1: $Y = X_1 + 2X_2 + 0.3X_1X_2 + 0.9 + \varepsilon$,
Model 2: $Y = X_1 + 2X_2 + 1.4(\exp(-0.2X_1) - \exp(0.7X_2)) + \varepsilon$,
Model 3: $Y = X_1 I(X_2 \geq 0.2) + \varepsilon$.

Table 4.8 reports the simulation results pertaining to $\widehat{\mathcal{D}}_n$. Data from Model 0 in this table are used to study the empirical sizes, and from Models 1 to 3 are used to study the empirical powers of the test.

5. Conclusion and discussion

For the general linear errors-in-variables model, this paper proposes an MD test procedure, based on the minimum distance idea and by exploiting the nature of deconvolution density estimator, to check if the regression function takes a parametric form. As a byproduct, the MD estimator for the regression parameters is also derived. The asymptotic normality of the proposed test statistics under the null hypothesis and that of the corresponding minimum distance estimators are fully discussed. We also prove the consistency of the proposed tests against a

Table 5.1
Double exponential, independent sample with $n_1 = n_2$

Model	(50, 50)	(100, 100)	(200, 200)	(300, 300)	(500, 500)
Model 0	0.008	0.036	0.033	0.038	0.049
Model 1	0.938	1.000	1.000	1.000	1.000
Model 2	1.000	1.000	1.000	1.000	1.000
Model 3	0.990	1.000	1.000	1.000	1.000

Table 5.2
Double exponential, same sample

Model	50	100	200	300	500
Model 0	0.015	0.024	0.036	0.043	0.047
Model 1	0.934	1.000	1.000	1.000	1.000
Model 2	0.999	1.000	1.000	1.000	1.000
Model 3	0.991	1.000	1.000	1.000	1.000

fixed alternative and obtain asymptotic power against a class of local alternatives orthogonal to the parametric model being fitted. The significant contribution we made in this research is the removal of the common assumption in the existing literature that the density function of the design variable is known or known up to some unknown parameters. The price we paid in removing such restrictive assumption is mainly the slow rate of the test procedure, due to the sample size allocation assumption (n).

The simulation studies show that the proposed testing procedures are quite satisfactory in the preservation of the finite sample level and in terms of a power comparison. But in the proof of the above theorems, we need the sample size allocation assumption (n) to ensure that the estimator $\hat{Q}_{n_2}(z)$ has a faster convergence rate. The assumption (n) plays a very important role in the theoretical argument, but it loses attraction to a practitioner. For example, in the simulation case 1 where the measurement error follows a double exponential distribution, the sample size allocation is $n_2 = [n_1^b]$, and $b = 1.2501$. n_2 in the second subsample S_2 increases in a power rate of the sample size n_1 in the first subsample, If $n_1 = 500$, n_2 is at least 2365, the sample size of the full sample is 2865 which is perhaps not easily available in practice. The situation becomes even worse when the measurement error is super-smooth or $d > 1$. For example, in Case 2, the measurement error has a normal distribution, n_2 is at least 52902 if $n_1 = 500$; in Case 4, $d = 2$, n_2 is at least 11188 if $n_1 = 500$.

Then an interesting question arises. What is the small sample behavior of the test procedure if (1) $n_1 = n_2$ and the two subsamples S_1 and S_2 are independent or (2) $n = n_1 = n_2$ and we do not split the sample at all? We have no theory at this point about the asymptotic behavior of $M_n(\hat{\theta}_n)$. For $d = 1$, we only conduct some Monte Carlo simulations here to see the performance of the test procedure (see Tables 5.1–5.4). The simulation results about levels and powers of the MD test appears in the following tables, in which the measurement error follows the same double exponential and normal distributions as in the previous section, the null and alternative models are the same as in Case 1.

To our surprise, the simulation results for the first three cases in which $d = 1$ are very good. There are almost no differences between the simulation results based on our theory and the simulation results by just neglecting the theory. In the Case 4 with $d = 2$, we only conduct the simulation for $S_1 = S_2$, see Table 5.5. The test procedure is conservative for small sample sizes,

Table 5.3
Normal, independent sample with $n_1 = n_2$

Model	(50, 50)	(100, 100)	(200, 200)	(300, 300)	(500, 500)
Model 0	0.013	0.023	0.027	0.035	0.047
Model 1	0.931	0.999	1.000	1.000	1.000
Model 2	1.000	1.000	1.000	1.000	1.000
Model 3	0.984	1.000	1.000	1.000	1.000

Table 5.4
Normal, same sample

Model	50	100	200	300	500
Model 0	0.017	0.019	0.036	0.036	0.051
Model 1	0.954	0.998	1.000	1.000	1.000
Model 2	0.999	1.000	1.000	1.000	1.000
Model 3	0.992	1.000	1.000	1.000	1.000

Table 5.5
Double exponential, same sample, $d = 2$

Model	50	100	200	300	500
Model 0	0.000	0.004	0.010	0.018	0.041
Model 1	0.628	0.996	1.000	1.000	1.000
Model 2	0.994	0.999	1.000	1.000	1.000
Model 3	0.844	0.998	1.000	1.000	1.000

but the empirical level is close to the nominal level 0.05 when sample size reaches 500. This phenomenon suggests to us that by relaxing some conditions, such as (n), even the assumptions on the choices of the bandwidths, Theorems 3.1 and 3.2 may still be valid.

If the null model is polynomial, then the test procedures proposed by Zhu et al. [24] and Cheng and Kukush [8] are more powerful than the MD test constructed in this paper, and hence should be recommended. To illustrate this point, we conduct a small simulation study to compare the performance of Zhu et al.'s score type test, the Cheng and Kukush [8] exponential weighted test and the MD test. Because of the above-mentioned phenomenon, the simulation for the MD test is done by using the same sample, i.e. without sample splitting. The model being simulated is $y = \theta_1 X + \theta_2 X^2 + cX^3 + \varepsilon$, $Z = X + u$, where $X \sim N(0, 1)$, $\varepsilon \sim N(0, 1)$ and $u \sim N(0, 0.2^2)$, the null model corresponds to $c = 0$, the alternative models correspond to $c = 0.3, 0.5, 1$. The simulation result for the sample size 200 is reported in Table 5.6. Simulation results show that the MD test is more conservative and less powerful than the Cheng and Kukush test and the Zhu, Song and Cui test. The Cheng and Kukush test (the quasi-optimal $\lambda = 1.243$ is used in the simulation) is the most powerful among these three tests. This phenomenon is not out of our expectation in that the Zhu, Song and Cui test and the Cheng and Kukush test are basically parametric tests, the MD test, however, is a nonparametric one.

6. Proofs of the main results

Proof of Lemma 3.1. A direct calculation yields that for any $x \in \mathbb{R}^d$, $E\hat{f}_{Xw_1}(x) = \int L(v) f_X(x - vw_1)\mathrm{d}v$. By assumption (f1), there exists a vector $a(x, v)$ such that $f_X(x - vw_1)$

Table 5.6
Comparison of tests

	$c = 0$	$c = 0.3$	$c = 0.5$	$c = 1$
Cheng and Kukush test	0.042	0.862	0.992	1.000
Zhu, Song and Cui test	0.059	0.572	0.743	0.855
MD test	0.036	0.202	0.297	0.554

has a Taylor expansion up to the second order, $f_X(x - vw_1) = f_X(x) - w_1 v' \dot{f}_X(x) + w_1^2 v' \ddot{f}_X(a(x, v))v/2$, where $\dot{f}$ and $\ddot{f}$ are the first- and second-order derivatives of f with respect to its argument. Hence

$$\begin{aligned} E\hat{R}_{n_2}(z) &= \int \int r(x)L(v) f_X(x - vw_1) f_u(z - x)\mathrm{d}v\mathrm{d}x \\ &= \int \int r(x)L(v) f_X(x) f_u(z - x)\mathrm{d}v\mathrm{d}x \\ &\quad - w_1 \int \int r(x)L(v)v' \dot{f}_X(x) f_u(z - x)\mathrm{d}v\mathrm{d}x \\ &\quad + \frac{1}{2} \int \int r(x)L(v)w_1^2 v' \ddot{f}_X(a(v, x))v f_u(z - x)\mathrm{d}v\mathrm{d}x. \end{aligned}$$

Assumption (ℓ) implies that the first term is $\int r(x) f_X(x) f_u(z - x)\mathrm{d}x = R(z)$, the second term vanishes because of $\int v'L(v)\mathrm{d}v = 0$, while the third term is bounded above by $cI(z)$ by assumption (f1), where c is a positive constant depending only on the kernel function L. Therefore, the first claim in the lemma holds.

Note that $\hat{R}_{n_2}(z) - E\hat{R}_{n_2}(z)$ is an average of independently and identically distributed centered random vectors. A routine calculation shows that

$$E\|\hat{R}_{n_2}(z) - E\hat{R}_{n_2}(z)\|^2 \le \frac{1}{n_2 w_1^{2d}} E \left\| \int r(x) L_{w_1}((x - z)/w_1) f_u(z - x)\mathrm{d}x \right\|^2$$

by using the fact that the variance is bounded above by the second moment. Let $D(t, z) = \int r(x) f_u(z - x) \exp(-\mathbf{i}t'x)\mathrm{d}x$. By the definition of the deconvolution kernel L_b, it follows that

$$\begin{aligned} &\frac{1}{w_1^{2d}} E \left\| \int r(x) L_{w_1}((x - z)/w_1) f_u(z - x)\mathrm{d}x \right\|^2 \\ &\quad = \int \int \frac{D'(t, z)D(s, z)\phi_L(tw_1)\phi_L(sw_1)\phi_X(t + s)\phi_u(t + x)}{(2\pi)^{2d}\phi_u(t)\phi_u(s)} \mathrm{d}s\mathrm{d}t. \end{aligned}$$

By changing the variable, $D(t, z) = \exp(-\mathbf{i}t'z) \int r(z - x) f_u(x) \exp(\mathbf{i}t'x)\mathrm{d}x$. Adding and subtracting $r(z)$ from $r(z - x)$ in the integrand, we obtain

$$D(t, z) = \exp(-\mathbf{i}t^{\mathrm{T}}z)\phi_u(z) \left[r(z) + \frac{\int (r(z - x) - r(z)) f_u(x) \exp(\mathbf{i}t'x)\mathrm{d}x}{\phi_u(t)} \right].$$

From assumption (m1), $\|D(t, z)\|$ is bounded above by $|\phi_u(t)| \cdot [\|r(z)\| + J(z)\|t\|^\alpha]$ for all $z \in \mathbb{R}^d$. Hence $E\|\hat{R}_{n_2}(z) - E\hat{R}_{n_2}(z)\|^2$ is bounded above by

$$\frac{c\|r(z)\|^2}{n_2}\int\int|\phi_L(tw_1)\phi_L(sw_1)\phi_u(t+s)|\mathrm{d}t\mathrm{d}s$$
$$+\frac{cJ(z)\|r(z)\|}{n_2}\int\int(\|t\|^\alpha+\|s\|^\alpha)|\phi_L(tw_1)\phi_L(sw_1)\phi_u(t+s)|\mathrm{d}t\mathrm{d}s$$
$$+\frac{cJ^2(z)}{n_2}\int\int\|t\|^\alpha\|s\|^\alpha|\phi_L(tw_1)\phi_L(sw_1)\phi_u(t+s)|\mathrm{d}t\mathrm{d}s. \tag{6.1}$$

Note that for any m, $p=0$ or α, from assumption (ℓ), we have

$$\int\int\|t\|^p\|s\|^m|\phi_L(tw_1)\phi_L(sw_1)\phi_u(t+s)|\mathrm{d}t\mathrm{d}s$$
$$\le w_1^{-p-m-2d}\int\int\|t\|^p\|s\|^m|\phi_L(t)\phi_L(s)\phi_u((t+s)/w_1)|\mathrm{d}t\mathrm{d}s$$
$$\le cw_1^{-p-m-2d}\int\int\|s\|^m|\phi_L(s)||\phi_u((t+s)/w_1)|\mathrm{d}t\mathrm{d}s$$
$$=cw_1^{-p-m-d}\int\|s\|^m|\phi_L(s)|\mathrm{d}s\cdot\int|\phi_u(t)|\mathrm{d}t=cw_1^{-p-m-d}.$$

The second claim in the lemma follows from (6.1) by using the above inequality. □

Proof of Theorem 3.1. To keep the exposition concise, let

$$U_{n_1}(z)=\frac{1}{n_1}\sum_{i=1}^{n_1}K_{h_1i}(z)(Y_i-\theta_0'Q(Z_i)),$$
$$D_n(z)=\frac{1}{n_1}\sum_{i=1}^{n_1}K_{h_1i}(z)(\hat{Q}_{n_2}(Z_i)-Q(Z_i)), \tag{6.2}$$
$$\mu_{n_1}(z)=\frac{1}{n_1}\sum_{i=1}^{n_1}K_{h_1i}(z)Q(Z_i),\qquad \Delta_{n_1}(z)=\frac{1}{\tilde{f}^2_{Zh_2}(z)}-\frac{1}{f^2_Z(z)}.$$

It suffices to show that the matrix before $\hat{\theta}_n-\theta_0$ on the left-hand side of (3.4) converges to Σ_0 in probability, and $\sqrt{n_1}$ times the right-hand side of (3.4) is asymptotically normal with mean vector 0 and covariance matrix Σ.

Consider the second claim first. Adding and subtracting $\theta_0^{\mathrm{T}}Q(Z_i)$ from $Y_i-\theta_0'\hat{Q}_{n_2}(Z_i)$ in the first factor of the integrand, and adding and subtracting $Q(Z_i)$ from $\hat{Q}_{n_2}(Z_i)$ in the second factor of the integrand, replacing $1/\tilde{f}^2_{Zh_2}(z)$ by $1/\tilde{f}^2_{Zh_2}(z)-1/f^2_Z(z)+1/f^2_Z(z)=\Delta_{n_1}(z)+1/f^2_Z(z)$, $\sqrt{n_1}$ times the right-hand side of (3.4) can be written as the sum of the following eight terms.

$$S_{n1}=\sqrt{n_1}\int U_{n_1}(z)D_n(z)\Delta_{n_1}(z)\mathrm{d}G(z),\qquad S_{n2}=\sqrt{n_1}\int U_{n_1}(z)D_n(z)\mathrm{d}\psi(z),$$
$$S_{n3}=\sqrt{n_1}\int U_{n_1}(z)\mu_{n_1}(z)\Delta_{n_1}(z)\mathrm{d}G(z),\qquad S_{n4}=\sqrt{n_1}\int U_{n_1}(z)\mu_{n_1}(z)\mathrm{d}\psi(z),$$
$$S_{n5}=-\sqrt{n_1}\int D_n(z)D_n'(z)\Delta_{n_1}(z)\mathrm{d}G(z)\theta_0,\qquad S_{n6}=-\sqrt{n_1}\int D_n(z)D_n'(z)\mathrm{d}\psi(z)\theta_0,$$
$$S_{n7}=-\sqrt{n_1}\int D_n(z)\mu_{n_1}'(z)\Delta_{n_1}(z)\mathrm{d}G(z)\theta_0,\qquad S_{n8}=-\sqrt{n_1}\int D_n(z)\mu_{n_1}'(z)\mathrm{d}\psi(z)\theta_0.$$

Among these terms, S_{n4} is asymptotically normal with mean vector 0 and covariance matrix Σ. The proof uses the Lindeberg–Feller central limit theorem, and the arguments are exactly

the same as in K–N with $m_{\theta_0}(X_i)$ and $\dot{m}_{\theta_0}(X_i)$ there replaced by $\theta_0^{\mathrm{T}} Q(Z_i)$ and $Q(Z_i)$ here, respectively. The proof is omitted. All the other seven terms are of the order $o_p(1)$. Since the proofs are similar, only $S_{n8} = o_p(1)$ will be shown below for the sake of brevity. We note that by using a similar method as in K–N, we can show $U_{n_1}(z)$ is $O_p(1/\sqrt{n_1 h_1^d})$, which is used in proving $S_{nl} = o_p(1)$ for $l = 1, 2, 3$.

First, notice that the kernel function K has compact support $[-1, 1]^d$, so $K_{h_1 i}$ is not 0 only if the distances between each coordinate pair of Z_i and z are no more than h. On the other hand, the integrating measure has compact support $\mathcal{I}$, so if we define

$$\begin{aligned}\mathcal{I}_{h_1} = \{&y \in \mathbb{R}^d : |y_j - z_j| \le h_1, j = 1, \ldots, d, y = (y_1, \ldots, y_d)', \\ &z = (z_1, \ldots, z_d)', z \in \mathcal{I}\},\end{aligned}$$

then $\mathcal{I}_{h_1}$ is a compact set in $\mathbb{R}^d$, and $K_{h_1 i} = 0$ if $Z_i \notin \mathcal{I}_{h_1}$. Hence, without loss of generality, we can assume all $Z_i \in \mathcal{I}_{h_1}$. Since f_Z is bounded from below on the compact set $\mathcal{I}_{\delta_0}$ by assumption (f2) and $\mathcal{I}_{h_1} \subset \mathcal{I}_{\delta_0}$ for n_1 large enough, from assumption (w2), Lemma 3.2, we obtain

$$\sup_{z \in \mathcal{I}_{h_1}} \left| \frac{f_Z(z)}{\hat{f}_{Zw_2}(z)} - 1 \right| = o\left((\log_k n_2) \left(\frac{\log n_2}{n_2} \right)^{\frac{2}{d+4}} \right) \quad \text{a.s.,} \sup_{z \in \mathcal{I}_{h_1}} \left| \frac{f_Z(z)}{\hat{f}_{Zw_2}(z)} \right| = O_p(1). \tag{6.3}$$

Secondly, we have the following inequality,

$$\|\hat{Q}_{n_2}(Z_i) - Q(Z_i)\| \le \frac{\|\hat{R}_{n_2}(Z_i) - R(Z_i)\|}{f_Z(Z_i)} \cdot \left| \frac{f_Z(Z_i)}{\hat{f}_{Zw_2}(Z_i)} \right| + \left| \frac{f_Z(Z_i)}{\hat{f}_{Zw_2}(Z_i)} - 1 \right| \cdot \|Q(Z_i)\|.$$

Recall the definition of S_{n8}. We have

$$\|S_{n8}\| \le \sqrt{n_1} \|\theta_0\| \int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) \|\hat{Q}_{n_2}(Z_i) - Q(Z_i)\| \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) \|Q(Z_i)\| \mathrm{d}\psi(z).$$

From (6.3) and (6.4), this upper bound satisfies

$$\sqrt{n_1} \cdot O_p(1) \cdot A_{n_1 1} + \sqrt{n_1} \cdot o\left((\log_k n_2) \left(\frac{\log n_2}{n_2} \right)^{\frac{2}{d+4}} \right) \cdot A_{n_1 2}, \tag{6.4}$$

where

$$\begin{aligned} A_{n_1 1} &= \int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) \|\hat{R}_{n_2}(Z_i) - R(Z_i)\| \cdot \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) \|Q(Z_i)\| \mathrm{d}\psi(z) \\ A_{n_1 2} &= \int \left[\frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) \|Q(Z_i)\| \right]^2 \mathrm{d}\psi(z). \end{aligned}$$

By the Cauchy–Schwarz inequality, $A_{n_1 1}^2$ is bounded above by

$$\int \left[\frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) \|\hat{R}_{n_2}(Z_i) - R(Z_i)\| \right]^2 \mathrm{d}\psi(z) \cdot \int \left[\frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) \|Q(Z_i)\| \right]^2 \mathrm{d}\psi(z).$$

Note that

$$E\int\left[\frac{1}{n_1}\sum_{i=1}^{n_1}K_{h_1i}(z)\|\hat{R}_{n_2}(Z_i)-R(Z_i)\|\right]^2\mathrm{d}\psi(z)$$
$$=\int E\left(\frac{1}{n_1^2}\sum_{i,j=1}^{n_1}K_{h_1i}(z)K_{h_1j}(z)E_{S_1}(\|\hat{R}_{n_2}(Z_i)-R(Z_i)\|\|\hat{R}_{n_2}(Z_j)-R(Z_j)\|)\right)$$
$$\mathrm{d}\psi(z).$$

By the Cauchy–Schwarz inequality again, $E_{S_1}(\|\hat{R}_{n_2}(Z_i)-R(Z_i)\|\|\hat{R}_{n_2}(Z_j)-R(Z_j)\|)$ is bounded above by $(E_{S_1}\|\hat{R}_{n_2}(Z_i)-R(Z_i)\|^2)^{1/2}(E_{S_1}\|\hat{R}_{n_2}(Z_j)-R(Z_j)\|^2)^{1/2}$, which in turn, from the independence of the subsamples S_1 and S_2, the choice of bandwidth w_1, and (3.1), is bounded above by $cn_2^{-4/(d+2\alpha+4)}\mathcal{T}^{1/2}(Z_i)\mathcal{T}^{1/2}(Z_j)$, where $\mathcal{T}$ is defined in (3.2). So

$$E\int\left[\frac{1}{n_1}\sum_{i=1}^{n_1}K_{h_1i}(z)\|\hat{R}_{n_2}(Z_i)-R(Z_i)\|\right]^2\mathrm{d}\psi(z)\le cn_2^{-\frac{4}{d+2\alpha+4}}$$
$$\times\int E\left(\frac{1}{n_1}\sum_{i=1}^{n_1}K_{h_1i}(z)\mathcal{T}^{1/2}(Z_i)\right)^2\mathrm{d}\psi(z).$$

Using the similar method as in K–N, together with the assumptions (m1) and (m2), we can show that

$$\int\left(\frac{1}{n_1}\sum_{i=1}^{n_1}K_{h_1i}(z)\mathcal{T}^{1/2}(Z_i)\right)^2\mathrm{d}\psi(z)=O_p(1)$$
$$=\int\left(\frac{1}{n_1}\sum_{i=1}^{n_1}K_{h_1i}(z)\|Q(Z_i)\|\right)^2\mathrm{d}\psi(z).$$

Finally, from (6.4), we obtain $\|S_{n8}\|\le\sqrt{n_1}\cdot O_p(n_2^{-2/(d+2\alpha+4)})+\sqrt{n_1}\cdot o_p((\log_k n_2)(\frac{\log n_2}{n_2})^{\frac{2}{d+4}})$ which is of the order $o_p(1)$ by the assumption (n).

To finish the proof, we only need to show the matrix before $\hat{\theta}_n-\theta_0$ on the left-hand side of (3.4) converges to Σ_0 in probability. Adding and subtracting $Q(Z_i)$ from $\hat{Q}_{n_2}(Z_i)$, this matrix can be written as the sum of the following eight terms.

$$T_{n1}=\int D_n(z)D_n'(z)\Delta_{n_1}(z)\mathrm{d}G(z),\qquad T_{n2}=\int D_n(z)\mu_{n_1}'(z)\Delta_{n_1}(z)\mathrm{d}G(z),$$
$$T_{n3}=\int \mu_{n_1}(z)D_n'(z)\Delta_{n_1}(z)\mathrm{d}G(z),\qquad T_{n4}=\int \mu_{n_1}(z)\mu_{n_1}'(z)\Delta_{n_1}(z)\mathrm{d}G(z),$$
$$T_{n5}=\int D_n(z)D_n'(z)\mathrm{d}\psi(z),\qquad T_{n6}=\int D_n(z)\mu_{n_1}'(z)\mathrm{d}\psi(z),$$
$$T_{n7}=\int \mu_{n_1}(z)D_n'(z)\mathrm{d}\psi(z),\qquad T_{n8}=\int \mu_{n_1}(z)\mu_{n_1}'(z)\mathrm{d}\psi(z).$$

Notice the connection between T_{n1} and S_{n5}, T_{n2}, T_{n3} and S_{n7}, T_{n5} and S_{n6}, T_{n6}, T_{n7} and S_{n8}. By using a similar argument as above, we can verify that $T_{nl}=o_p(1)$ for $l=1,2,3,4,5,6,7$.

From (6.3), and the second fact in (6.5), T_{n4} is also of the order of $o_p(1)$. Finally, employing a similar method as in K–N, we can show T_{n8} converges to Σ_0 in probability, thereby proving the theorem. □

Proof of Theorem 3.2. To state the result precisely, the following notations are needed.

$$\xi_i = Y_i - \theta_0' Q(Z_i), \qquad \zeta_i = Y_i - \theta_0' \hat{Q}_{n_2}(Z_i),$$
$$\tilde{C}_n = n_1^{-2} \sum_{i=1}^{n_1} \int K_{h_1 i}^2(z) \zeta_i^2 \mathrm{d}\psi(z), \qquad \tilde{M}_n(\theta_0) = \int \left[n_1^{-1} \sum_{i=1}^{n_1} K_{h_1 i}(z) \zeta_i \right]^2 \mathrm{d}\psi(z),$$
$$\Gamma = 2 \int (\tau^2(z))^2 g(z) \mathrm{d}\psi(z) \cdot \int \left[\int K(u) K(u+v) \mathrm{d}u \right]^2 \mathrm{d}v,$$

where $\tau^2(z)$ is as in Theorem 3.1. The proof is facilitated by the following five lemmas.

Lemma 6.1. *If* H_0, (e1), (e2), (e4), (u), (f1), (f2), (m1), (m2), *(ℓ)*, (n), (h1), (w1) *and* (w2) *hold, then* $n_1 h_1^{d/2}(\tilde{M}_n(\theta_0) - \tilde{C}_n) \Longrightarrow N_d(0, \Gamma)$.

Proof. Replacing ζ_i by $\xi_i + \theta_0'(Q(Z_i) - \hat{Q}_{n_2}(Z_i))$ in the definition $\tilde{M}_n(\theta_0)$ and expanding the quadratic term, $n_1 h_1^{d/2}(\tilde{M}_n(\theta_0) - \tilde{C}_n)$ can be written as the sum of the following four terms.

$$B_{n1} = \frac{1}{n_1^2} \sum_{i \neq j}^{n_1} \int K_{h_1 i}(z) K_{h_1 j}(z) \xi_i \xi_j \mathrm{d}\psi(z),$$
$$B_{n2} = \frac{1}{n_1^2} \sum_{i \neq j}^{n_1} \int K_{h_1 i}(z) K_{h_1 j}(z) \xi_i \theta_0'(Q(Z_j) - \hat{Q}_{n_2}(Z_j)) \mathrm{d}\psi(z),$$
$$B_{n3} = \frac{1}{n_1^2} \sum_{i \neq j}^{n_1} \int K_{h_1 i}(z) K_{h_1 j}(z) \xi_j \theta_0'(Q(Z_i) - \hat{Q}_{n_2}(Z_i)) \mathrm{d}\psi(z),$$

and $B_{n4} = n_1^{-2} \sum_{i \neq j}^{n_1} \int K_{h_1 i}(z) K_{h_1 j}(z) \theta_0'(Q(Z_i) - \hat{Q}_{n_2}(Z_i)) \theta_0'(Q(Z_j) - \hat{Q}_{n_2}(Z_j)) \mathrm{d}\psi(z)$. Using the similar method as in K–N, one can show that $n_1 h_1^{d/2} B_{n1} \Longrightarrow N_d(0, \Gamma)$. To prove the lemma, it is sufficient to show $n_1 h_1^{d/2} B_{nl} = o_p(1)$ for $l = 2, 3, 4$. We begin with the case of $l = 2$. By (6.3) and the inequality (6.4), and letting $C_{nij} = \sum_{i \neq j}^{n_1} \int K_{h_1 i}(z) K_{h_1 j}(z) \xi_i \mathrm{d}\psi(z)$, B_{n2} is bounded above by the sum $B_{n21} + B_{n22}$, where

$$B_{n21} = O_p(1) \cdot \frac{1}{n_1^2} \sum_{j=1}^{n_1} [\| \hat{R}_{n_2}(Z_j) - R(Z_j) \| \cdot |C_{nij}|],$$
$$B_{n22} = o\left((\log_k n_2) \left(\frac{\log n_2}{n_2} \right)^{\frac{2}{d+4}} \right) \cdot \frac{1}{n_1^2} \sum_{j=1}^{n_1} [\| Q(Z_j) \| \cdot |C_{nij}|].$$

On the one hand, by the conditional expectation argument and inequality (3.1), we have

$$\begin{aligned} E\frac{1}{n_1^2}\sum_{j=1}^{n_1}[\|\hat{R}_{n_2}(Z_j)-R(Z_j)\|\cdot|C_{nij}|] &= E\frac{1}{n_1^2}\sum_{j=1}^{n_1}[E_{S_1}(\|\hat{R}_{n_2}(Z_j)-R(Z_j)\|)\cdot|C_{nij}|] \\ &\le cn_2^{-2/(d+2\alpha+4)}E\left[\frac{1}{n_1^2}\sum_{j=1}^{n_1}\mathcal{T}^{1/2}(Z_j)\cdot|C_{nij}|\right] \\ &= cn_2^{-2/(d+2\alpha+4)}\frac{1}{n_1}E[\mathcal{T}^{1/2}(Z_1)\cdot|C_{ni1}|]. \end{aligned}$$

Now, consider the asymptotic behavior of $E[\mathcal{T}^{1/2}(Z_1)\cdot|C_{ni1}|]$. Instead of considering the expectation, we investigate the second moment. It is easy to see that $E\mathcal{T}(Z_1)C_{ni1}^2$ equals

$$\begin{aligned} &E\mathcal{T}(Z_1)\sum_{i\neq1}\sum_{j\neq1}\int\int K_{h_1i}(z)K_{h_11}(z)K_{h_1j}(y)K_{h_11}(y)\xi_i\xi_j\mathrm{d}\psi(z)\mathrm{d}\psi(y) \\ &\quad=(n_1-1)\int\int E(K_{h_12}(z)K_{h_12}(y)\xi_2^2)\cdot E(K_{h_11}(z)K_{h_11}(y)\mathcal{T}(Z_1))\mathrm{d}\psi(z)\mathrm{d}\psi(y). \end{aligned} \tag{6.5}$$

The second equality is from the independence of ξ_i, $i=1,\ldots,n_1$ and $E\xi_1=0$. But

$$\begin{aligned} E(K_{h_12}(z)K_{h_12}(y)\xi_2^2) &= E(K_{h_12}(z)K_{h_12}(y)(\sigma_\varepsilon^2+\delta^2(Z_2))) \\ &= \frac{1}{h_1^d}\int K(v)K\left(\frac{y-z}{h_1}-v\right)(\sigma_\varepsilon^2+\delta^2(z-h_1v))f_Z(z-h_1v)\mathrm{d}v. \end{aligned}$$

Similarly, we can show that

$$E(K_{h_11}(z)K_{h_11}(y)\mathcal{T}(Z_1))=\frac{1}{h_1^d}\int K(v)K\left(\frac{y-z}{h_1}-v\right)\mathcal{T}(z-h_1v)f_Z(z-h_1v)\mathrm{d}v.$$

Putting back these two expectations in (6.5), and changing variables $y=z+h_1u$, then by the continuity of f_Z, $\delta^2(z)$, $g(z)$, and $\mathcal{T}(z)$, we obtain $E\mathcal{T}(Z_1)C_{ni1}^2=(n_1-1)h_1^{-d}$. Therefore,

$$E\frac{1}{n_1^2}\sum_{j=1}^{n_1}[\|\hat{R}_{n_2}(Z_j)-R(Z_j)\|\cdot|C_{nij}|]=O\left(n_2^{-2/(d+2\alpha+4)}\frac{1}{n_1}\cdot\sqrt{n_1-1}h_1^{-d/2}\right).$$

This, in turn, implies $B_{n21}=O_p(n_1^{-2b/(d+2\alpha+4)-1/2}h_1^{-d/2})$, by assumption (n). Similarly, one can show $n_1^{-2}\sum_{j=1}^{n_1}[\|Q(Z_j)\|\cdot|C_{nij}|]$ is of the order $O_p(n_1^{-1/2}h_1^{-d/2})$. Thus,

$$B_{n22}=o_p((\log_k n_1)(\log n_1/n_1^b)^{2/(d+4)}\cdot n_1^{-1/2}h_1^{-d/2}).$$

Hence

$$n_1h_1^{d/2}|B_{n2}|=O_p\left(n_1^{\frac{1}{2}-\frac{2b}{d+2\alpha+4}}\right)+O_p\left(n_1^{\frac{1}{2}-\frac{2b}{d+4}}\log_k n_1(\log n_1)^{\frac{2}{d+4}}\right)=o_p(1),$$

since $b>(d+2\alpha+4)/4$ by assumption (n).

By exactly the same method as above, we can show that $n_1h_1^{d/2}B_{n3}=o_p(1)$.

It remains to show that $n_1 h_1^{d/2} B_{n4} = o_p(1)$. Note that

$$|B_{n4}| \le \frac{1}{n^2} \sum_{i \ne j}^{n_1} \int K_{h_1 i}(z) K_{h_1 j}(z) \|\theta_0\|^2 \cdot \|\hat{Q}_{n_2}(Z_i) - Q(Z_i)\| \times \|\hat{Q}_{n_2}(Z_j) - Q(Z_j)\| \mathrm{d}\psi(z).$$

From (6.4), the right-hand side of the above inequality is bounded above by the sum

$$O_p(1) \cdot B_{n41} + o_p\left((\log_k n_2)\left(\frac{\log n_2}{n_2}\right)^{\frac{2}{d+4}}\right)(B_{n42} + B_{n43}) + o_p\left((\log_k^2 n_2)\left(\frac{\log n_2}{n_2}\right)^{\frac{4}{d+4}}\right) B_{n44},$$

where

$$B_{n41} = \frac{1}{n_1^2} \sum_{i \ne j}^{n_1} \int K_{h_1 i}(z) K_{h_1 j}(z) \cdot \|\hat{R}_{n_2}(Z_i) - R(Z_i)\| \cdot \|\hat{R}_{n_2}(Z_j) - R(Z_j)\| \mathrm{d}\psi(z),$$

$$B_{n42} = \frac{1}{n_1^2} \sum_{i \ne j}^{n_1} \int K_{h_1 i}(z) K_{h_1 j}(z) \cdot \|\hat{R}_{n_2}(Z_i) - R(Z_i)\| \cdot \|Q(Z_j)\| \mathrm{d}\psi(z),$$

$$B_{n43} = \frac{1}{n_1^2} \sum_{i \ne j}^{n_1} \int K_{h_1 i}(z) K_{h_1 j}(z) \cdot \|\hat{R}_{n_2}(Z_j) - R(Z_j)\| \cdot \|Q(Z_i)\| \mathrm{d}\psi(z),$$

$$B_{n44} = \frac{1}{n_1^2} \sum_{i \ne j}^{n_1} \int K_{h_1 i}(z) K_{h_1 j}(z) \cdot \|Q(Z_i)\| \cdot \|Q(Z_j)\| \mathrm{d}\psi(z).$$

By a conditional expectation argument, the Cauchy–Schwarz inequality, (2.2), and the continuity of f_Z and $T(z)$, we obtain

$$E B_{n41} \le c n_2^{-4/(d+2\alpha+4)} \int E[K_{h_1 1}(z) \mathcal{T}^{1/2}(Z_1)]^2 \mathrm{d}\psi(z) = O(n_2^{-4/(d+2\alpha+4)}).$$

This implies $B_{n41} = O_P(n_2^{-4/(d+2\alpha+4)})$, since $b > (d + 2\alpha + 4)/4$ by assumption (n), so that

$$n_1 h_1^{d/2} \cdot O_p(1) B_{n41} = n_1 h_1^{d/2} \cdot O_p(1) O_P(n_1^{-4b/(d+2\alpha+4)}) = o_p(1).$$

Similarly, we can show

$$B_{n42} = O_P(n_2^{-2/(d+2\alpha+4)}), \qquad B_{n43} = O_P(n_2^{-2/(d+2\alpha+4)}), \qquad B_{n44} = O_P(1).$$

Therefore, for $l = 2, 3$,

$$n_1 h_1^{d/2} o_p\left((\log_k n_2)\left(\frac{\log n_2}{n_2}\right)^{\frac{2}{d+4}}\right) B_{n4l} = o_p(n_1^{1-\frac{2b}{d+4}-\frac{2b}{d+2\alpha+4}} h_1^{d/2} (\log_k n_1)(\log n_1)^{\frac{2}{d+4}})$$

which is of the order $o_p(1)$ by assumption (n). For B_{n44}, we have

$$n_1 h_1^{d/2} \cdot o_p\left((\log_k^2 n_2)\left(\frac{\log n_2}{n_2}\right)^{\frac{4}{d+4}}\right) B_{n44} = o_p(n_1^{1-\frac{4b}{d+4}} h_1^{d/2} (\log_k^2 n_1)(\log n_1)^{\frac{4}{d+4}})$$

which is also of the order $o_p(1)$. Finally, from the above and (6.6), we prove $n_1 h_1^{d/2} B_{n4} = o_p(1)$, thereby proving the lemma.

Lemma 6.2. *In addition to the conditions in Lemma 6.1, suppose* (h2) *also holds, then* $n_1 h_1^{d/2}(M_n(\hat{\theta}_n) - M_n(\theta_0)) = o_p(1)$.

Proof. Recall the definitions of $M_n(\theta)$. Adding and subtracting $n_1^{-1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\theta_0'\hat{Q}_{n_2}(Z_i)$ in the squared integrand of $M_n(\hat{\theta}_n)$, we can write $M_n(\hat{\theta}_n) - M_n(\theta_0)$ as the sum $W_{n1} + 2W_{n2}$, where

$$W_{n1} = \int \left[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)(\theta_0 - \hat{\theta}_n)'\hat{Q}_{n_2}(Z_i)\right]^2 \mathrm{d}\hat{\psi}_{h_2}(z),$$

$$W_{n2} = \int \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\zeta_i \cdot \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)(\theta_0 - \hat{\theta}_n)'\hat{Q}_{n_2}(Z_i)\mathrm{d}\hat{\psi}_{h_2}(z),$$

and $\zeta_i = Y_i - \theta_0'\hat{Q}_{n_2}(Z_i)$. It is easy to see that

$$W_{n1} \le 2\int \left[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)(\theta_0 - \hat{\theta}_n)'(\hat{Q}_{n_2}(Z_i) - Q(Z_i))\right]^2 \mathrm{d}\hat{\psi}_{h_2}(z)$$
$$+ 2\int \left[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)(\theta_0 - \hat{\theta}_n)^{\mathrm{T}} Q(Z_i)\right]^2 \mathrm{d}\hat{\psi}_{h_2}(z).$$

We write the first term on the right-hand side as W_{n11} and the second term as W_{n12}. On the one hand, note that W_{n11} is bounded above by

$$\|\hat{\theta}_n - \theta_0\|^2 \cdot \sup_{z\in\mathcal{I}}\left|\frac{f_Z(z)}{\hat{f}_{Zw_2}(z)}\right| \cdot \int \left[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\|\hat{Q}_{n_2}(Z_i) - Q(Z_i)\|\right]^2 \mathrm{d}\psi(z).$$

By the conditional expectation argument as we used in the previous part, we can show that the integral part is indeed of the order $o_p(1)$. By assumption (w2), the compactness of $\mathcal{I}_{h_1}$, and the asymptotic behavior of $\hat{\theta}_n - \theta_0$ stated in Theorem 3.1, $n_1 h_1^{d/2} W_{n11} = o_p(h_1^{d/2}) = o_p(1)$. On the other hand, W_{n12} is bounded above by

$$\|\hat{\theta}_n - \theta_0\|^2 \cdot \sup_{z\in\mathcal{I}}\left|\frac{f_Z(z)}{\hat{f}_{Zw_2}(z)}\right| \cdot \int \left[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\|Q(Z_i)\|\right]^2 \mathrm{d}\psi(z).$$

Since the integral part is of the order $O_p(1)$, $n_1 h_1^{d/2} W_{n12} = O_p(h_1^{d/2}) = o_p(1)$ is easily obtained. Therefore, $n_1 h_1^{d/2} W_{n1} = o_p(1)$ is proved. Now rewrite W_{n2} as

$$W_{n2} = \int \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\zeta_i \cdot \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\hat{Q}_{n_2}'(Z_i)\mathrm{d}\hat{\psi}_{h_2}(z) \cdot (\theta_0 - \hat{\theta}_n).$$

Note that the integral part of W_{n2} is the same as the expression on the right-hand side of (3.4), thus

$$W_{n2} = (\hat{\theta}_n - \theta_0)' \int \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\hat{Q}_{n_2}(Z_i) \cdot \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\hat{Q}_{n_2}'(Z_i)\mathrm{d}\hat{\psi}_{h_2}(z) \cdot (\theta_0 - \hat{\theta}_n).$$

Therefore, W_{n2} is bounded above by $\|\hat{\theta}_n - \theta_0\|^2 \int [n_1^{-1} \sum_{i=1}^{n_1} K_{h_1 i}(z) \|\hat{Q}_{n_2}(Z_i)\|]^2 \mathrm{d}\hat{\psi}_{h_2}(z)$. Adding and subtracting $Q(Z_i)$ from $\hat{Q}_{n_2}(Z_i)$, it turns out that W_{n2} is further bounded above by the sum $W_{n21} + W_{n22}$, where

$$W_{n21} = 2\|\hat{\theta}_n - \theta_0\|^2 \int \left[n_1^{-1} \sum_{i=1}^{n_1} K_{h_1 i}(z) \|\hat{Q}_{n_2}(Z_i) - Q(z_i)\| \right]^2 \mathrm{d}\hat{\psi}_{h_2}(z),$$

$$W_{n22} = 2\|\hat{\theta}_n - \theta_0\|^2 \int \left[n_1^{-1} \sum_{i=1}^{n_1} K_{h_1 i}(z) \|Q(z_i)\| \right]^2 \mathrm{d}\hat{\psi}_{h_2}(z).$$

Arguing as in W_{n11} and W_{n12}, we can show $n_1 h_1^{d/2} |W_{n21}| = o_p(1)$, $n_1 h_1^{d/2} |W_{n22}| = o_p(1)$. Therefore, $n_1 h_1^{d/2} |W_{n2}| = o_p(1)$. Together with the result $n_1 h_1^{d/2} |W_{n1}| = o_p(1)$, the lemma is proved. □

Lemma 6.3. *If* H_0, (e1), (e2), (u), (f1), (f2), (m1), (m2), *(ℓ)*, (n), (h1), (h2), (w1) *and* (w2) *hold,* $n_1 h_1^{d/2} (M_n(\theta_0) - \tilde{M}_n(\theta_0)) = o_p(1)$.

Proof. Recall the definition of ζ_i and $U_{n_1}(z)$. Note that

$$n_1 h_1^{d/2} |M_n(\theta_0) - \tilde{M}_n(\theta_0)| \le n_1 h_1^{d/2} \sup_{z \in \mathcal{I}} \left| \frac{f_Z^2(z)}{\hat{f}_{Zh_2}^2(z)} - 1 \right| \int \left[\frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) \zeta_i \right]^2 \mathrm{d}\psi(z).$$

Replace ζ_i by $\xi_i + \theta_0'(Q(Z_i) - \hat{Q}_{n_2}(Z_i))$, the integral part of the above inequality can be bounded above by the sum

$$2 \int U_{n_1}^2(z) \mathrm{d}\psi(z) + 2 \int \left[\frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) \theta_0'(Q(Z_i) - \hat{Q}_{n_2}(Z_i)) \right]^2 \mathrm{d}\psi(z).$$

The first term is of the order $O_p((n_1 h_1^d)^{-1/2})$ which is obtained by the similar method as in K–N, while the second term, by the conditional expectation argument, has the same order as

$$\sup_{z \in \mathcal{I}_{h_1}} \left| \frac{f_Z^2(z)}{\hat{f}_{Zw_2}^2(z)} \right| \cdot O(n_2^{-4/(d+2\alpha+4)}) + \sup_{z \in \mathcal{I}_{h_1}} \left| \frac{f_Z^2(z)}{\hat{f}_{Zw_2}^2(z)} - 1 \right|^2 \cdot O_p(1).$$

Therefore, $n_1 h_1^{d/2} |M_n(\theta_0) - \tilde{M}_n(\theta_0)|$ is less than or equal to

$$\begin{aligned} &O_p\left(n_1 h_1^{d/2} \cdot \frac{1}{n h_1^d} \cdot \log_k n_1 (\log n_1/n_1)^{2/(d+4)} \right) \\ &\quad + O_p\left(n_1 h_1^{d/2} \cdot \log_k n_1 (\log n_1/n_1)^{2/(d+4)} \cdot n_1^{-4b/(d+2\alpha+4)} \right) \\ &\quad + O_p\left(n_1 h_1^{d/2} \cdot \log_k n_1 (\log n_1/n_1)^{2/(d+4)} \cdot \log_k^2 n_1 (\log n_1)^{4/(d+4)} n_1^{-4b/(d+4)} \right). \end{aligned}$$

All the three terms are of the order $o_p(1)$ by the assumptions (n), (h1), (h2), (w1) and (w2). Hence the lemma. □

Lemma 6.4. *If* H_0, (e1), (e2), (e4), (u), (f1), (f2), (m1), (m2), *(ℓ)*, (n), (h1), (h2), (w1) *and* (w2) *hold,* $n_1 h_1^{d/2} (\hat{C}_n - \tilde{C}_n) = o_p(1)$.

Proof. Recall the notation $\Delta_{n_1}(z)$ in (6.2). Adding and subtracting $\theta_0'\hat{Q}_{n_2}(Z_i)$ from Y_i in the integrand of hC_n, then expand the quadratic term, then $\hat{C}_n - \tilde{C}_n$ can be rewritten as the sum of $C_{nl}, l = 1, 2, 3, 4, 5$, where

$$C_{n1} = \frac{1}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)(Y_i - \theta_0'\hat{Q}_{n_2}(Z_i))^2\Delta_{n_1}(z)\mathrm{d}\psi(z),$$

$$C_{n2} = \frac{2}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)(Y_i - \theta_0'\hat{Q}_{n_2}(Z_i))(\theta_0 - \hat{\theta}_n)'\hat{Q}_{n_2}(Z_i)\Delta_{n_1}(z)\mathrm{d}\psi(z),$$

$$C_{n3} = \frac{2}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)(Y_i - \theta_0'\hat{Q}_{n_2}(Z_i))(\theta_0 - \hat{\theta}_n)'\hat{Q}_{n_2}(Z_i)\mathrm{d}\psi(z),$$

$$C_{n4} = \frac{1}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)(Y_i - \theta_0'\hat{Q}_{n_2}(Z_i))((\theta_0 - \hat{\theta}_n)'\hat{Q}_{n_2}(Z_i))^2\Delta_{n_1}(z)\mathrm{d}\psi(z),$$

$$C_{n5} = \frac{1}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)(Y_i - \theta_0'\hat{Q}_{n_2}(Z_i))((\theta_0 - \hat{\theta}_n)'\hat{Q}_{n_2}(Z_i))^2\mathrm{d}\psi(z).$$

To prove the lemma, it is enough to prove $n_1 h_1^{d/2} C_{nl} = o_p(1)$ for $l = 1, 2, 3, 4, 5$.

For the case of $l = 1$, first notice that

$$\begin{aligned}|C_{n1}| &\le 2\sup_{z\in\mathcal{I}}|\Delta_{n_1}(z)| \cdot \frac{1}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)\xi_i^2\mathrm{d}\psi(z)\\ &\quad + 2\sup_{z\in\mathcal{I}}|\Delta_{n_1}(z)| \cdot \frac{1}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)(\theta_0'(Q(Z_i) - \hat{Q}_{n_2}(Z_i)))^2\mathrm{d}\psi(z)\\ &= C_{n11} + C_{n12}.\end{aligned}$$

Since $n_1^{-2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)\xi_i^2\mathrm{d}\psi(z) = O_p(1/n_1 h_1^d)$ by a routine expectation argument,

$$n_1 h_1^{d/2}|C_{n11}| = o_p\left(n_1 h_1^{d/2}\cdot(\log_k n_1)(\log n_1)^{2/(d+4)} n_1^{-2/(d+4)}\cdot(n_1 h_1)^{-1}\right) = o_p(1).$$

Second, from the compactness of Θ, we have

$$\begin{aligned}&\frac{1}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)(\theta_0'(Q(Z_i) - \hat{Q}_{n_2}(Z_i)))^2\mathrm{d}\psi(z)\\ &\quad \le O(1)\cdot\frac{1}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)\|Q(Z_i) - \hat{Q}_{n_2}(Z_i)\|^2\mathrm{d}\psi(z).\end{aligned}$$

Again by the conditional expectation argument, the second factor of the above expression has the same order as

$$\begin{aligned}&O_p(n_2^{-4/(d+2\alpha+4)})\cdot\sup_{z\in\mathcal{I}_{h_1}}\left|\frac{f_Z(z)}{\tilde{f}_{Zw_2}(z)}\right|^2\cdot\frac{1}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)\mathcal{T}^2(Z_i)\mathrm{d}\psi(z)\\ &\quad + \sup_{z\in\mathcal{I}_{h_1}}\left|\frac{f_Z(z)}{\tilde{f}_{Zw_2}(z)} - 1\right|^2\cdot\frac{1}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)\|Q(Z_i)\|^2\mathrm{d}\psi(z).\end{aligned}$$

Because

$$\frac{1}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)\mathcal{T}^2(Z_i)\mathrm{d}\psi(z) = O_p(1/n_1 h_1^d) = \frac{1}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)\|Q(Z_i)\|^2\mathrm{d}\psi(z),$$

then from (h2), (w2), (h1), and Lemma 3.2, we get $n_1 h_1^{d/2}|C_{n12}| = o_p(1)$. Hence $n_1 h_1^{d/2}|C_{n1}| = o_p(1)$. Now we will show that $n_1 h_1^{d/2}|C_{n3}| = o_p(1)$. Once we prove this, then $n_1 h_1^{d/2}|C_{n2}| = o_p(1)$ is a natural consequence. In fact,

$$\begin{aligned} C_{n3} = {} & \frac{2}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)(\xi_i + \theta_0^{\mathrm{T}} Q(Z_i) - \theta_0'\hat{Q}_{n_2}(Z_i))\cdot(\theta_0 - \hat{\theta}_n)' \\ & \times(\hat{Q}_{n_2}(Z_i) - Q(Z_i) + Q(Z_i))\mathrm{d}\psi(z). \end{aligned}$$

So $|C_{n3}|$ is bounded above by the sum $2(C_{n31} + C_{n32} + C_{n33} + C_{n34})$, where

$$\begin{aligned} C_{n31} &= \frac{1}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)|\xi_i|\|\theta_0 - \hat{\theta}_n\|\|\hat{Q}_{n_2}(Z_i) - Q(Z_i)\|\mathrm{d}\psi(z), \\ C_{n32} &= \frac{1}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)|\xi_i|\|\theta_0 - \hat{\theta}_n\|\|Q(Z_i)\|\mathrm{d}\psi(z) \\ C_{n33} &= \frac{1}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)\|\theta_0\|\|\theta_0 - \hat{\theta}_n\|\|\hat{Q}_{n_2}(Z_i) - Q(Z_i)\|^2\mathrm{d}\psi(z), \\ C_{n34} &= \frac{1}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)\|\theta_0\|\|\theta_0 - \hat{\theta}_n\|\|\hat{Q}_{n_2}(Z_i) - Q(Z_i)\|\|Q(Z_i)\|\mathrm{d}\psi(z). \end{aligned}$$

It is sufficient to show that $n_1 h_1^{d/2}|C_{n3l}| = o_p(1)$ for $l = 1, 2, 3, 4$. Because the proofs are similar, here we only show $n_1 h_1^{d/2}|C_{n32}| = o_p(1)$, others are omitted for the sake of brevity. In fact, note that $n_1^{-2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)|\xi_i|\|Q(Z_i)\|\mathrm{d}\psi(z) = O_p(1/n_1 h_1^d)$ by an expectation argument, then from $\|\hat{\theta}_n - \theta_0\| = O_p(n_1^{-1/2})$ by Theorem 3.1, we have $n_1 h_1^{d/2}|C_{n32}| = n_1 h_1^{d/2}\|\hat{\theta}_n - \theta_0\|\cdot O_p(1/n_1 h_1^d) = O_p(n_1^{-1/2}h_1^{-d/2})$. Since $n_1^{-1/2}h_1^{-d/2} = n_1^{-1/2+ad/2}$ and $a < 1/2d$ by assumption (h1), the above expression is $o_p(1)$. Similarly, we can show that the same results hold for C_{n4} and C_{n5}. The details are left out. □

Lemma 6.5. *Under the same conditions as in Lemma 6.4,* $\hat{\Gamma}_n - \Gamma = o_p(1)$.

Proof. Recall the notation for ξ_i. Define

$$\tilde{\Gamma}_n = 2h_1^d n_1^{-2}\sum_{i\neq j}^{n_1}\left(\int K_{h_1 i}(z)K_{h_1 j}(z)\xi_i\xi_j\mathrm{d}\hat{\psi}_{h_2}(z)\right)^2.$$

The lemma is proved by showing that

$$\hat{\Gamma}_n - \tilde{\Gamma}_n = o_p(1), \qquad \tilde{\Gamma}_n - \Gamma = o_p(1), \tag{6.6}$$

where Γ is as in (3.5). But the second claim can be shown using the same method as in K–N, so we only prove the first claim. Write $u_n = \hat{\theta}_n - \theta_0$, $r_i = \theta_0^{\mathrm{T}} Q(Z_i) - \hat{\theta}_n'\hat{Q}_{n_2}(Z_i)$. Now $\hat{\Gamma}_n$ can be

expressed as the sum of $\tilde{\Gamma}_n$ and the following terms:

$$\begin{aligned} B_{n1} &= 2h_1^d n_1^{-2} \sum_{i\neq j}^{n_1} \left[\int K_{h_1 i}(z) K_{h_1 j}(z) \xi_i r_j \mathrm{d}\hat{\psi}_{h_2}(z) + \int K_{h_1 i}(z) K_{h_1 j}(z) \xi_j r_i \mathrm{d}\hat{\psi}_{h_2}(z) \right. \\ &\quad \left. + \int K_{h_1 i}(z) K_{h_1 j}(z) r_i r_j \mathrm{d}\hat{\psi}_{h_2}(z) \right]^2, \\ B_{n2} &= 4h_1^d n_1^{-2} \sum_{i\neq j}^{n_1} \left(\int K_{h_1 i}(z) K_{h_1 j}(z) \xi_i \xi_j \mathrm{d}\hat{\psi}_{h_2}(z) \right) \cdot \left(\int K_{h_1 i}(z) K_{h_1 j}(z) \xi_i r_j \mathrm{d}\hat{\psi}_{h_2}(z) \right. \\ &\quad \left. + \int K_{h_1 i}(z) K_{h_1 j}(z) \xi_j r_i \mathrm{d}\hat{\psi}_{h_2}(z) + \int K_{h_1 i}(z) K_{h_1 j}(z) r_i r_j \mathrm{d}\hat{\psi}_{h_2}(z) \right), \end{aligned}$$

so it suffices to show that both terms are of the order $o_p(1)$. Applying the Cauchy–Schwarz inequality to the double sum, one can see that we only need to show the following

$$\begin{aligned} &h_1^d n_1^{-2} \sum_{i\neq j}^{n_1} \left[\int K_{h_1 i}(z) K_{h_1 j}(z) |\xi_i r_j| \mathrm{d}\hat{\psi}_{h_2}(z) \right]^2 = o_p(1) \\ &h_1^d n_1^{-2} \sum_{i\neq j}^{n_1} \left[\int K_{h_1 i}(z) K_{h_1 j}(z) |r_i r_j| \mathrm{d}\hat{\psi}_{h_2}(z) \right]^2 = o_p(1), \\ &h_1^d n_1^{-2} \sum_{i\neq j}^{n_1} \left[\int K_{h_1 i}(z) K_{h_1 j}(z) |\xi_i \xi_j| \mathrm{d}\hat{\psi}_{h_2}(z) \right]^2 = O_p(1). \end{aligned} \tag{6.7}$$

The third claim in (6.7) can be proved by using the same argument as in K–N. Now, consider the first claim above. From Lemma 3.2, we only need to show the claim is true when $\mathrm{d}\hat{\psi}_{h_2}(z)$ is replaced by $\mathrm{d}\psi(z)$. Since r_j has nothing to do with the integration variable, the left-hand side of the first claim after the replacing can be rewritten as

$$h_1^d n_1^{-2} \sum_{i\neq j}^{n_1} |r_j|^2 \left[\int K_{h_1 i}(z) K_{h_1 j}(z) |\xi_i| \mathrm{d}\psi(z) \right]^2. \tag{6.8}$$

Note that $r_j = u_n'(Q(Z_j) - \hat{Q}_{n_2}(Z_j)) - u_n^{\mathrm{T}} Q(Z_j) - \theta_0'(\hat{Q}_{n_2}(Z_j) - Q(Z_j))$, so (6.8) can be bounded above by the sum of the following three terms:

$$\begin{aligned} A_{n1} &= 3h_1^d n_1^{-2} \|u_n\|^2 \sum_{i\neq j}^{n_1} \|\hat{Q}_{n_2}(Z_j) - Q(Z_j)\|^2 \left[\int K_{h_1 i}(z) K_{h_1 j}(z) |\xi_i| \mathrm{d}\psi(z) \right]^2, \\ A_{n2} &= 3h_1^d n_1^{-2} \|u_n\|^2 \sum_{i\neq j}^{n_1} \left[\int K_{h_1 i}(z) K_{h_1 j}(z) |\xi_i| \|Q(Z_j)\| \mathrm{d}\psi(z) \right]^2, \\ A_{n3} &= 3h_1^d n_1^{-2} \|\theta_0\|^2 \sum_{i\neq j}^{n_1} \|\hat{Q}_{n_2}(Z_j) - Q(Z_j)\|^2 \left[\int K_{h_1 i}(z) K_{h_1 j}(z) |\xi_i| \mathrm{d}\psi(z) \right]^2. \end{aligned}$$

$A_{n2} = o_p(1)$ can be shown by the fact that $u_n = \hat{\theta}_n - \theta_0 = o_p(1)$, and that

$$h_1^d n_1^{-2} \sum_{i\neq j}^{n_1} \left[\int K_{h_1 i}(z) K_{h_1 j}(z) |\xi_i| \|Q(Z_j)\| \mathrm{d}\psi(z) \right]^2 = O_p(1)$$

which can be shown by using the same argument as in K–N. Let us consider A_{n3}. Using the inequality (6.4), Lemma 3.2 or (6.3), and the compactness of Θ, it is easy to see A_{n3} is bounded above by the sum $A_{n31} + A_{n32}$, where

$$A_{n31} = O_p(1) \cdot h_1^d n_1^{-2} \sum_{i \neq j}^{n_1} \|\hat{R}_{n_2}(Z_j) - R(Z_j)\|^2 \left[\int K_{h_1 i}(z) K_{h_1 j}(z) |\xi_i| \mathrm{d}\psi(z) \right]^2$$

$$A_{n32} = o_p(1) \cdot h_1^d n_1^{-2} \sum_{i \neq j}^{n_1} \left[\int K_{h_1 i}(z) K_{h_1 j}(z) |\xi_i| \|Q(Z_j)\| \mathrm{d}\psi(z) \right]^2.$$

Apply the conditional expectation argument to the second factor in A_{n31}, using the fact (3.1) and the elementary inequality $a < (1+a)^2$, we can show

$$\begin{aligned} &E\left[h_1^d n_1^{-2} \sum_{i \neq j}^{n_1} \|\hat{R}_{n_2}(Z_j) - R(Z_j)\|^2 \left[\int K_{h_1 i}(z) K_{h_1 j}(z) |\xi_i| \mathrm{d}\psi(z) \right]^2 \right] \\ &\quad = E\left[h_1^d n_1^{-2} \sum_{i \neq j}^{n_1} (E_{S_1} \|\hat{R}_{n_2}(Z_j) - R(Z_j)\|^2) \left[\int K_{h_1 i}(z) K_{h_1 j}(z) |\xi_i| \mathrm{d}\psi(z) \right]^2 \right] \\ &\quad \leq c n_2^{-\frac{4}{d+2\alpha+4}} E\left[h_1^d n_1^{-2} \sum_{i \neq j}^{n_1} \left[\int K_{h_1 i}(z) K_{h_1 j}(z) |\xi_i \parallel 1 + \mathcal{T}(Z_j)| \mathrm{d}\psi(z) \right]^2 \right]. \end{aligned}$$

The expectation of the right-hand side of the above inequality turns out to be $O(1)$ by using the same argument as in K–N. So, $h_1^d n_1^{-2} \sum_{i \neq j}^{n_1} \|\hat{R}_{n_2}(Z_j) - R(Z_j)\|^2 \left[\int K_{h_1 i}(z) K_{h_1 j}(z) |\xi_i| \mathrm{d}\psi(z) \right]^2 = o_p(1)$. This, in turn, implies that the second factor in $A_{n31} = o_p(1)$. The same method as in K–N also leads to $h_1^d n_1^{-2} \sum_{i \neq j}^{n_1} \left[\int K_{h_1 i}(z) K_{h_1 j}(z) |\xi_i| \|Q(Z_j)\| \mathrm{d}\psi(z) \right]^2 = O_p(1)$. Hence $A_{n32} = o_p(1)$. Therefore, $B_{n1} = o_p(1)$, and $B_{n2} = o_p(1)$, thereby proving the first claim in (6.6), hence the lemma. □

Proof of Theorem 3.3. Before we prove the consistency of the MD test, let us consider the convergence of the MD estimator. Under the alternative hypothesis $H_a : \mu(x) = m(x)$, one can verify that the right-hand side of (3.3) can be written as the sum of the following two terms

$$A_{n1} = \int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) m(X_i) \cdot \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) \hat{Q}_{n_2}(Z_i) \mathrm{d}\hat{\psi}_{h_2}(z),$$

$$A_{n2} = \int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) \varepsilon_i \cdot \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) \hat{Q}_{n_2}(Z_i) \mathrm{d}\hat{\psi}_{h_2}(z).$$

Adding and subtracting $Q(Z_i)$ from $\hat{Q}_{n_2}(Z_i)$, on the one hand, one has

$$\begin{aligned} A_{n1} = &\int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) m(X_i) \cdot D_n(z) \mathrm{d}\hat{\psi}_{h_2}(z) + \int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) m(X_i) \\ &\times \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) Q(Z_i) \mathrm{d}\hat{\psi}_{h_2}(z), \end{aligned}$$

where $D_n(z)$ is as in (6.2). The first term of A_{n1} is the order of $o_p(1)$, while the second term converges to $\int H(z)Q(z)\mathrm{d}G(z)$ in probability. On the other hand,

$$A_{n2} = \int \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\varepsilon_i \cdot D_n(z)\mathrm{d}\hat{\psi}_{h_2}(z) + \int \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\varepsilon_i \times \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)Q(Z_i)\mathrm{d}\hat{\psi}_{h_2}(z).$$

Similarly to proving the asymptotic results for $S_{n1} + S_{n2}$ and $S_{n3} + S_{n4}$ in the proof of Theorem 3.1, we can show that both terms of A_{n2} are $o_p(1)$. Recall that $\hat{\theta}_n$ satisfies (3.3), indeed we proved that $\hat{\theta}_n \to \Sigma_0^{-1}\int H(z)Q(z)\mathrm{d}G(z)$.

Adding and subtracting $H(z) = E(m(X)|Z = z)$ from Y_i, $M_n(\hat{\theta}_n)$ can be written as the sum $S_{n1} + S_{n2} + S_{n3}$, where

$$S_{n1} = \int \left[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)(Y_i - H(Z_i))\right]^2 \mathrm{d}\hat{\psi}_{h_2}(z),$$

$$S_{n2} = \int \left[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)(H(Z_i) - \hat{\theta}_n'\hat{Q}_{n2}(Z_i))\right]^2 \mathrm{d}\hat{\psi}_{h_2}(z),$$

$$S_{n3} = 2\int \left[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)(Y_i - H(Z_i))\right] \times \left[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)(H(Z_i) - \hat{\theta}_n'\hat{Q}_{n2}(Z_i))\right]\mathrm{d}\hat{\psi}_{h_2}(z).$$

Define

$$C_n^* = \frac{1}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)(Y_i - H(Z_i))^2\mathrm{d}\hat{\psi}_{h_2}(z),$$

$$\Gamma^* = 2\int (\sigma_*^2(z))^2 g(z)\mathrm{d}\psi(z)\int\left[\int K(u)K(u+v)\mathrm{d}u\right]^2\mathrm{d}v,$$

$$\sigma_*^2(z) = \sigma_\varepsilon^2 + E[(m(X) - H(Z))^2|Z = z] + (H(z) - \theta' Q(z))^2.$$

Similarly to the proof of Theorem 3.2, one can show that $n_1 h_1^{d/2}(S_{n1} - C_n^*) \to N(0, \Gamma^*)$ in distribution. Let $\theta = \Sigma_0^{-1}\int H(z)Q(z)\mathrm{d}G(z)$, adding and subtracting $\theta' Q(Z_i)$ from $H(Z_i) - \hat{\theta}_n'\hat{Q}_{n2}(Z_i)$, then S_{n2} equals the sum of S_{n21}, S_{n22} and S_{n23}, where

$$S_{n21} = \int \left[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)(H(Z_i) - \theta' Q(Z_i))\right]^2 \mathrm{d}\hat{\psi}_{h_2}(z),$$

$$S_{n22} = -2\int \left[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)(H(Z_i) - \theta' Q(Z_i))\right] \times \left[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)(\hat{\theta}_n'\hat{Q}_{n2}(Z_i) - \theta' Q(Z_i))\right]\mathrm{d}\hat{\psi}_{h_2}(z),$$

$$S_{n23} = 2\int \left[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)(\hat{\theta}_n'\hat{Q}_{n2}(Z_i) - \theta' Q(Z_i))\right]^2 \mathrm{d}\hat{\psi}_{h_2}(z).$$

Routine calculation and Lemma 3.2 show that $S_{n21} = \int [H(z) - \theta' Q(z)]^2 \mathrm{d}G(z) + o_p(1)$, while S_{n23} is bounded above by

$$2\int \left[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)(\hat{\theta}_n - \theta)'\hat{Q}_{n2}(Z_i)\right]^2 \mathrm{d}\hat{\psi}_{h_2}(z)$$
$$+ 2\int \left[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)(\hat{Q}_{n2}(Z_i) - Q(Z_i))'\theta\right]^2 \mathrm{d}\hat{\psi}_{h_2}(z),$$

which is $o_p(1)$ by Theorem 3.3 and the asymptotic property of $\hat{Q}_{n2}(Z_i)$ discussed in the proof of Theorem 3.1. Hence by the Cauchy–Schwarz inequality, $S_{n22} = o_p(1)$. $S_{n3} = o_p(1)$ can be obtained by using the Cauchy–Schwarz inequality again.

Note that $C_n = C_n^* + C_{n1} + C_{n2}$, where

$$C_{n1} = \frac{2}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)(Y_i - H(Z_i))(H(Z_i) - \hat{\theta}_n'\hat{Q}_{n2}(Z_i))\mathrm{d}\hat{\psi}_{h_2}(z),$$
$$C_{n2} = \frac{1}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)(H(Z_i) - \hat{\theta}_n'\hat{Q}_{n2}(Z_i))^2\mathrm{d}\hat{\psi}_{h_2}(z).$$

Both C_{n1} and C_{n2} are $o_p(1)$. Hence $C_n - C_n^* = o_p(1)$.

Next, we shall show that $\hat{\Gamma}_n = \Gamma^* + o_p(1)$. To this end, write $e_i = Y_i - H(Z_i)$, $w(Z_i, \hat{\theta}_n) = \hat{\theta}_n'\hat{Q}_{n_2}(Z_i) - H(Z_i)$, then

$$\hat{\Gamma}_n = 2h_1 n_1^{-2}\sum_{i\neq j=1}^{n}\left[\int K_{h_1 i}(z)K_{h_1 j}(z)(e_i - w(Z_i, \hat{\theta}_n))(e_j - w(Z_j, \hat{\theta}_n))\mathrm{d}\hat{\psi}_{h_2}(z)\right]^2.$$

Expanding the square of the integral, one can rewrite $\hat{\Gamma}_n = \sum_{j=1}^{10} A_{nj}$, where

$$A_{n1} = \frac{2h_1^d}{n_1^2}\sum_{i\neq j}\left(\int K_{h_1 i}(z)K_{h_1 j}(z)e_i e_j \mathrm{d}\hat{\psi}_{h_2}(z)\right)^2$$
$$A_{n2} = \frac{2h_1^d}{n_1^2}\sum_{i\neq j}\left(\int K_{h_1 i}(z)K_{h_1 j}(z)e_i w(Z_j, \hat{\theta}_n)\mathrm{d}\hat{\psi}_{h_2}(z)\right)^2$$
$$A_{n3} = \frac{2h_1^d}{n_1^2}\sum_{i\neq j}\left(\int K_{h_1 i}(z)K_{h_1 j}(z)w(Z_i, \hat{\theta}_n)e_j \mathrm{d}\hat{\psi}_{h_2}(z)\right)^2$$
$$A_{n4} = \frac{2h_1^d}{n_1^2}\sum_{i\neq j}\left(\int K_{h_1 i}(z)K_{h_1 j}(z)w(Z_i, \hat{\theta}_n)w(Z_j, \hat{\theta}_n)\mathrm{d}\hat{\psi}_{h_2}(z)\right)^2$$
$$A_{n5} = -\frac{4h_1^d}{n_1^2}\sum_{i\neq j}\left(\int K_{h_1 i}(z)K_{h_1 j}(z)e_i e_j \mathrm{d}\hat{\psi}_{h_2}(z)\right.$$
$$\left.\times \int K_{h_1 i}(z)K_{h_1 j}(z)e_i w(Z_j, \hat{\theta}_n)\mathrm{d}\hat{\psi}_{h_2}(z)\right)$$

$$
\begin{aligned}
A_{n6} &= -\frac{4h_1^d}{n_1^2}\sum_{i\neq j}\left(\int K_{h_1 i}(z)K_{h_1 j}(z)e_i e_j \mathrm{d}\hat{\psi}_{h_2}(z)\right.\\
&\quad\left.\times\int K_{h_1 i}(z)K_{h_1 j}(z)w(Z_i,\hat{\theta}_n)e_j \mathrm{d}\hat{\psi}_{h_2}(z)\right)\\
A_{n7} &= \frac{4h_1^d}{n_1^2}\sum_{i\neq j}\left(\int K_{h_1 i}(z)K_{h_1 j}(z)e_i e_j \mathrm{d}\hat{\psi}_{h_2}(z)\right.\\
&\quad\left.\times\int K_{h_1 i}(z)K_{h_1 j}(z)w(Z_i,\hat{\theta}_n)w(Z_j,\hat{\theta}_n)\mathrm{d}\hat{\psi}_{h_2}(z)\right)\\
A_{n8} &= \frac{4h_1^d}{n_1^2}\sum_{i\neq j}\left(\int K_{h_1 i}(z)K_{h_1 j}(z)e_i w(Z_j,\hat{\theta}_n)\mathrm{d}\hat{\psi}_{h_2}(z)\right.\\
&\quad\left.\times\int K_{h_1 i}(z)K_{h_1 j}(z)w(Z_i,\hat{\theta}_n)e_j \mathrm{d}\hat{\psi}_{h_2}(z)\right)\\
A_{n9} &= -\frac{4h_1^d}{n_1^2}\sum_{i\neq j}\left(\int K_{h_i}(z)K_{h_1 j}(z)e_i w(Z_j,\hat{\theta}_n)\mathrm{d}\hat{\psi}_{h_2}(z)\right.\\
&\quad\left.\times\int K_{h_1 i}(z)K_{h_1 j}(z)w(Z_i,\hat{\theta}_n)w(Z_j,\hat{\theta}_n)\mathrm{d}\hat{\psi}_{h_2}(z)\right)\\
A_{n10} &= -\frac{4h_1^d}{n_1^2}\sum_{i\neq j}\left(\int K_{h_1 i}(z)K_{h_1 j}(z)w(Z_i,\hat{\theta}_n)e_j \mathrm{d}\hat{\psi}_{h_2}(z)\right.\\
&\quad\left.\times\int K_{h_1 i}(z)K_{h_1 j}(z)w(Z_i,\hat{\theta}_n)w(Z_j,\hat{\theta}_n)\mathrm{d}\hat{\psi}_{h_2}(z)\right).
\end{aligned}
$$

By taking the expectation, using Fubini's theorem, we obtain

$$
h_1^d n_1^{-2}\sum_{i\neq j}\left(\int K_{h_1 i}(z)K_{h_1 j}(z)|e_i||e_j|\mathrm{d}\psi(z)\right)^2 = O_p(1), \tag{6.9}
$$

$$
h_1^d n_1^{-2}\sum_{i\neq j}\left(\int K_{h_1 i}(z)K_{h_1 j}(z)|e_i|\mathrm{d}\psi(z)\right)^2 = O_p(1), \quad k=0,1. \tag{6.10}
$$

By Lemma 3.2, (6.9), and arguing as in the proof of Lemma 6.5, one can verify that

$$
A_{n1} \to \Gamma_1^* = 2\int \sigma_e^4(z)g^2(z)/f_Z^2(z)\mathrm{d}z\int\left[\int K(u+v)K(u)\mathrm{d}u\right]^2\mathrm{d}v
$$

in probability, where $\sigma_e^2(z) = E[(Y-H(Z))^2|Z=z] = \sigma_\varepsilon^2 + E[(m(X)-H(Z))^2|Z=z]$. As for A_{n2}, write $\hat{\theta}_n'\hat{Q}_{n2}(Z_j)$ as $(\hat{\theta}_n-\theta+\theta)'(\hat{Q}_{n2}(Z_j)-Q(Z_j)+Q(Z_j))$ and expand the integral by considering $\hat{\theta}_n-\theta_0$, $\hat{Q}_{n_2}(Z_j)-Q(Z_j)$ as a whole, respectively, A_{n2} can be written as the sum of

$$
2h_1^d n_1^{-2}\sum_{i\neq j}\left(\int K_{h_1 i}(z)K_{h_1 j}(z)e_i(\theta' Q(Z_j)-H(Z_j))\mathrm{d}\hat{\psi}_{h_2}(z)\right)^2,
$$

and a remainder. The above term converges to

$$\Gamma_2^* = 2\int \sigma_e^2(z)[H(z)-\theta' Q(z)]^2 g^2(z)/f_Z^2(z)\mathrm{d}z \int \left[\int K(u+v)K(u)\mathrm{d}u\right]^2 \mathrm{d}v,$$

and the remainder equals $o_p(1)$ by the consistency of $\hat{\theta}_n$, and a similar conditional argument on $\hat{Q}_{n_2}(Z_j) - Q(Z_j)$ as in the proof of Lemma 6.1, together with (6.10) with $k = 1$.

The same argument leads to $A_{n3} \to \Gamma_2^*$ in probability.

Adding and subtracting $\theta' Q(Z_i)$ from $w(Z_i, \hat{\theta}_n)$, $\theta' Q(Z_j)$ from $w(Z_j, \hat{\theta}_n)$, arguing as above, one can show that

$$A_{n4} \to \Gamma_3^* = 2\int [H(z)-\theta' Q(z)]^4 g^2(z)/f_Z^2(z)\mathrm{d}z \int \left[\int K(u+v)K(u)\mathrm{d}u\right]^2 \mathrm{d}v$$

in probability. Next, write A_{n5} as the sum of

$$\begin{aligned} A_{n51} &= -\frac{4h_1^d}{n_1^2}\sum_{i\neq j}\Bigg(\int K_{h_1 i}(z)K_{h_1 j}(z)e_i e_j \mathrm{d}\hat{\psi}_{h_2}(z) \\ &\quad \times \int K_{h_1 i}(z)K_{h_1 j}(z)e_i[\theta' Q(Z_j) - H(Z_j)]\mathrm{d}\hat{\psi}_{h_2}(z)\Bigg) \\ A_{n52} &= -\frac{4h_1^d}{n_1^2}\sum_{i\neq j}\Bigg(\int K_{h_1 i}(z)K_{h_1 j}(z)e_i e_j \mathrm{d}\hat{\psi}_{h_2}(z) \\ &\quad \times \int K_{h_1 i}(z)K_{h_1 j}(z)e_i[\hat{\theta}_n' \hat{Q}_{n_2}(Z_j) - \theta' Q(Z_j)]\mathrm{d}\hat{\psi}_{h_2}(z)\Bigg). \end{aligned}$$

By Lemma 3.2, one can verify that $A_{n51} = \tilde{A}_{n51} + o_p(1)$, where

$$\begin{aligned} \tilde{A}_{n51} &= -\frac{4h_1^d}{n_1^2}\sum_{i\neq j}\Bigg(\int K_{h_1 i}(z)K_{h_1 j}(z)e_i e_j \ \mathrm{d}\hat{\psi}_{h_2}(z) \\ &\quad \times \int K_{h_1 i}(z)K_{h_1 j}(z)e_i[\theta' Q(Z_j) - H(Z_j)]\mathrm{d}\hat{\psi}_{h_2}(z)\Bigg). \end{aligned}$$

Clearly, $E\tilde{A}_{n51} = 0$. Arguing as in K–N, one can show that $E(\tilde{A}_{n51}^2) = O(n_1^{-1}h_1^{-d})$. Hence $A_{n51} = o_p(1)$. One can also show that $A_{n52} = o_p(1)$. These results imply $A_{n5} = o_p(1)$. Similarly, we can show that $A_{nj} = o_p(1)$ for $j = 6, 7, 8, 9, 10$.

Note that $\Gamma^* = \Gamma_1^* + 2\Gamma_2^* + \Gamma_3^*$, we obtain $\hat{\Gamma}_n \to \Gamma^*$ in probability.

Finally, we get

$$\hat{\mathcal{D}}_n = n_1 h_1^{d/2}\hat{\Gamma}_n^{-1/2}(S_{n1} - C_n^*) + n_1 h_1^{d/2}\hat{\Gamma}_n^{-1/2}\int [H(z)-\theta' Q(z)]^2 \mathrm{d}G(z) + o_p(n_1 h_1^{d/2})$$

and the theorem follows immediately. □

Proof of Theorem 3.4. Adding and subtracting $\theta_0'\hat{Q}_{n_2}(Z_i)$ from Y_i on the right-hand side of (3.3), then (3.3) becomes (3.4). Adding and subtracting $\theta_0' Q(Z_i)$, $Q(Z_i)$ from Y_i and $\hat{Q}_{n_2}(Z_i)$, respectively, and letting $\xi_i = \varepsilon_i + \theta_0' r(X_i) - \theta_0' Q(Z_i)$, then under the local alternatives (3.7), the right-hand side of (3.4) can be written as the sum of the following six terms

$$B_{n1} = \int \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\xi_i \cdot \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)Q(Z_i)\mathrm{d}\hat{\psi}_{h_2}(z),$$

$$B_{n2} = \int \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\xi_i \cdot D_n(z)\mathrm{d}\hat{\psi}_{h_2}(z),$$

$$B_{n3} = \gamma_n \int \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)v(X_i) \cdot \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)Q(Z_i)\mathrm{d}\hat{\psi}_{h_2}(z),$$

$$B_{n4} = \gamma_n \int \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)v(X_i) \cdot D_n(z)\mathrm{d}\hat{\psi}_{h_2}(z),$$

$$B_{n5} = -\int \theta_0' D_n(z) \cdot D_n(z)\mathrm{d}\hat{\psi}_{h_2}(z),$$

$$B_{n6} = -\int \theta_0' D_n(z) \cdot \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)Q(Z_i)\mathrm{d}\hat{\psi}_{h_2}(z).$$

Note that ξ_i are i.i.d. with mean 0 and finite second moment, arguing as in the proof of the asymptotic normality of S_{n4} in the proof of Theorem 3.1 with $Y_i - \theta_0' Q(Z_i)$ replaced by ξ_i, one can show that $\sqrt{n_1}B_{n1} \sim N_d(0, \Sigma)$ as $n \to \infty$. While $\sqrt{n_1}B_{n2} = o_p(1)$ can be shown in the similar way to showing $S_{n1} + S_{n2} = o_p(1)$, $\sqrt{n_1}B_{n4} = o_p(1)$ and $\sqrt{n_1}B_{n6} = o_p(1)$ can be proven similarly as in proving $S_{n7} + S_{n8}$ in the proof of Theorem 3.1. Then $\sqrt{n_1}B_{n5} = o_p(1)$ as well.

Now let us consider $\sqrt{n_1}B_{n3} = o_p(1)$. Denote

$$\eta_v(z) = E(K_h(z - Z)v(X)), \qquad \eta_Q(z) = E(K_h(z - Z)Q(Z)), \tag{6.11}$$

then

$$\begin{aligned}\sqrt{n_1}B_{n3} = \gamma_n\sqrt{n_1}\int \Bigg[&\left(\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)v(X_i) - \eta_v(z)\right)\\ &+ (\eta_v(z) - f_Z(z)V(Z)) + f_Z(z)V(Z)\Bigg]\\ &\times\Bigg[\left(\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)Q(Z_i) - \eta_Q(z)\right) + \big(\eta_Q(z) - f_Z(z)Q(Z)\big)\\ &+ f_Z(z)VQ(Z)\Bigg]\mathrm{d}\hat{\psi}_{h_2}(z).\end{aligned}$$

Expanding the product will result in nine terms. All terms can be shown to be the order of $o_p(h_1^{d/4})$ except

$$\begin{aligned}\gamma_n\sqrt{n_1}\int f_Z^2(z)V(z)Q(z)\mathrm{d}\hat{\psi}_{h_2}(z) = \gamma_n\sqrt{n_1}\int V(z)Q(z)\mathrm{d}G(z)\\ + \gamma_n\sqrt{n_1}\int V(z)Q(z)[1/\tilde{f}_{Zh_2}^2(z) - 1/f_Z^2(z)]\mathrm{d}G(z).\end{aligned}$$

The second term is $o_p(1)$ by the condition (n), (h1) and Lemma 3.2. Hence

$$\sqrt{n_1}B_{n3} = \sqrt{n_1}\gamma_n\int V(z)Q(z)\mathrm{d}G(z) + o_p(1).$$

Note that the random matrix before $\hat{\theta}_n - \theta_0$ in (3.4) does not depend on the local alternative hypothesis, so it converges to Σ_0 in probability. Thus, we showed that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) - \sqrt{n_1}\gamma_n \int V(z)Q(z)\mathrm{d}G(z) \to_d N_q(0, \Sigma_0^{-1}\Sigma\Sigma_0^{-1}), \tag{6.12}$$

where Σ_0 and Σ are the same as in Theorem 3.1.

Now, let us consider the local power of the MD test. Under the local alternative (3.7), $Y_i = \theta_0' r(X_i) + \gamma_n v(X_i) + \varepsilon_i$. Then

$$\begin{aligned} M_n(\hat{\theta}_n) &= \int \left[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)[Y_i - \hat{\theta}_n' \hat{Q}_{n_2}(Z_i)]\right]^2 \mathrm{d}\hat{\psi}_{h_2}(z) \\ &= \int \left[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)[\theta_0' r(X_i) + \gamma_n v(X_i) + \varepsilon_i - \theta_0' Q(Z_i) \right. \\ &\quad \left. + \theta_0' Q(Z_i) - \hat{\theta}_n' \hat{Q}_{n_2}(Z_i)]\right]^2 \mathrm{d}\hat{\psi}_{h_2}(z). \end{aligned}$$

Expanding the integral, $M_n(\hat{\theta}_n)$ can be written as the sum $\sum_{j=1}^6 T_{nj}$, where

$$\begin{aligned} T_{n1} &= \int \left[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)[\varepsilon_i + \theta_0' r(X_i) - \theta_0' Q(Z_i)]\right]^2 \mathrm{d}\hat{\psi}_{h_2}(z), \\ T_{n2} &= \gamma_n^2 \int \left[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z) v(X_i)\right]^2 \mathrm{d}\hat{\psi}_{h_2}(z), \\ T_{n3} &= \int \left[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)[\theta_0' Q(Z_i) - \hat{\theta}_n' \hat{Q}_{n_2}(Z_i)]\right]^2 \mathrm{d}\hat{\psi}_{h_2}(z), \\ T_{n4} &= 2\gamma_n \int \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)[\varepsilon_i + \theta_0' r(X_i) - \theta_0' Q(Z_i)] \cdot \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z) v(X_i) \mathrm{d}\hat{\psi}_{h_2}(z), \\ T_{n5} &= 2\int \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)[\varepsilon_i + \theta_0' r(X_i) - \theta_0' Q(Z_i)] \\ &\quad \times \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)[\theta_0' Q(Z_i) - \hat{\theta}_n' \hat{Q}_{n_2}(Z_i)] \mathrm{d}\hat{\psi}_{h_2}(z), \\ T_{n6} &= 2\gamma_n \int \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z) v(X_i) \cdot \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)[\theta_0' Q(Z_i) - \hat{\theta}_n' \hat{Q}_{n_2}(Z_i)] \mathrm{d}\hat{\psi}_{h_2}(z). \end{aligned}$$

A simple argument leads to

$$n_1 h_1^{d/2} T_{n2} = \int \left[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z) v(X_i)\right]^2 \mathrm{d}\hat{\psi}_{h_2}(z) \to \int [V(z)]^2 \mathrm{d}G(z) \tag{6.13}$$

in probability.

To deal with T_{n3}, note that

$$n_1 h_1^{d/2} T_{n3} = \int \left[\frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z)[(\hat{\theta}_n - \theta_0 + \theta_0)'[\hat{Q}_{n_2}(Z_i) - Q(Z_i) + Q(Z_i)] - \theta_0' Q(Z_i)] \right]^2 \mathrm{d}\hat{\psi}_{h_2}(z)$$

can be written as the sum of the following six terms

$$T_{n31} = n_1 h_1^{d/2} \int \left[\frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z)(\hat{\theta}_n - \theta_0)'(\hat{Q}_{n_2}(Z_i) - Q(Z_i)) \right]^2 \mathrm{d}\hat{\psi}_{h_2}(z)$$

$$T_{n32} = n_1 h_1^{d/2} \int \left[\frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z)(\hat{\theta}_n - \theta_0)' Q(Z_i) \right]^2 \mathrm{d}\hat{\psi}_{h_2}(z)$$

$$T_{n33} = n_1 h_1^{d/2} \int \left[\frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z)\theta_0'(\hat{Q}_{n_2}(Z_i) - Q(Z_i)) \right]^2 \mathrm{d}\hat{\psi}_{h_2}(z)$$

$$T_{n34} = n_1 h_1^{d/2} \int \left[\frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z)(\hat{\theta}_n - \theta_0)'(\hat{Q}_{n_2}(Z_i) - Q(Z_i)) \right] \times \left[\frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z)(\hat{\theta}_n - \theta_0)' Q(Z_i) \right] \mathrm{d}\hat{\psi}_{h_2}(z)$$

$$T_{n35} = n_1 h_1^{d/2} \int \left[\frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z)(\hat{\theta}_n - \theta_0)'(\hat{Q}_{n_2}(Z_i) - Q(Z_i)) \right] \times \left[\frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z)\theta_0'(\hat{Q}_{n_2}(Z_i) - Q(Z_i)) \right] \mathrm{d}\hat{\psi}_{h_2}(z)$$

$$T_{n36} = n_1 h_1^{d/2} \int \left[\frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z)(\hat{\theta}_n - \theta_0)' Q(Z_i) \right] \times \left[\frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z)\theta_0'(\hat{Q}_{n_2}(Z_i) - Q(Z_i)) \right] \mathrm{d}\hat{\psi}_{h_2}(z).$$

By (6.12), and conditional expectation arguments on $\hat{Q}_{n_2}(Z_i)$, one can show that $T_{n3k} = o_p(1)$ for $k = 1, 3, 4, 5, 6$.

For T_{n32}, we have

$$T_{n32} = n_1 h_1^{d/2} (\hat{\theta}_n - \theta_0)' \int \left[\frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) Q(Z_i) \right] \times \left[\frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) Q'(Z_i) \right] \mathrm{d}\hat{\psi}_{h_2}(z)(\hat{\theta}_n - \theta_0).$$

From (6.12), we see that

$$\sqrt{n_1 h_1^{d/2}}(\hat{\theta}_n - \theta_0) = \gamma_n \sqrt{n_1 h_1^{d/2}} \int V(z)Q(z)\mathrm{d}G(z) + o_p(h_1^{d/4})$$
$$= \int V(z)Q(z)\mathrm{d}G(z) + o_p(1).$$

Therefore, we can show that

$$n_1 h_1^{d/2} T_{n3} = \int V(z)Q'(z)\mathrm{d}G(z) \cdot \int Q(z)Q'(z)\mathrm{d}G(z) \cdot \int V(z)Q(z)\mathrm{d}G(z) + o_p(1). \tag{6.14}$$

To show $T_{n4} = o_p(1)$, recall the notation $\eta_v(z)$ defined in (6.11), and

$$\int \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)[\varepsilon_i + \theta_0' r(X_i) - \theta_0' Q(Z_i)] \cdot EK_{h_1}(z - Z)v(X)\mathrm{d}\psi(z)$$
$$= \frac{1}{n_1}\sum_{i=1}^{n_1} \int K_{h_1 i}(z)EK_{h_1}(z - Z)v(X)\mathrm{d}\psi(z)[\varepsilon_i + \theta_0' r(X_i) - \theta_0' Q(Z_i)] = O_p(1/\sqrt{n_1}),$$

one has

$$T_{n4} = 2\gamma_n \int \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)[\varepsilon_i + \theta_0' r(X_i) - \theta_0' Q(Z_i)] \cdot \eta_v(z)\mathrm{d}\hat{\psi}_{h_2}(z)$$
$$+ 2\gamma_n \int \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)[\varepsilon_i + \theta_0' r(X_i) - \theta_0' Q(Z_i)]$$
$$\times \left[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)v(X_i) - \eta_v(z)\right]\mathrm{d}\hat{\psi}_{h_2}(z)$$
$$= n_1 h_1^{d/2}\gamma_n O_p(1/\sqrt{n_1}) + n_1 h_1^{d/2}\gamma_n O_p(1/(n_1 h_1^d)) = o_p(1)$$

by assumption (h1). Similarly, one can obtain that $n_1 h_1^{d/2} T_{n5} = o_p(1) = n_1 h_1^{d/2} T_{n6}$.

Write $\xi_i = \varepsilon_i + \theta_0' r(X_i) - \theta_0' Q(Z_i)$. Then $\hat{C}_n$ can be written as

$$\frac{1}{n_1^2}\sum_{i=1}^{n_1} \int K_{h_1 i}^2(z)[\xi_i + \gamma_n v(X_i) + (\theta_0 - \hat{\theta}_n)'(Q(Z_i) - \hat{Q}_{n_2}(Z_i))$$
$$+ (\theta_0 - \hat{\theta}_n)'\hat{Q}_{n_2}(Z_i) + (Q(Z_i) - \hat{Q}_{n_2}(Z_i))'\theta_0]^2 \mathrm{d}\hat{\psi}_{h_2}(z)$$

which, by expanding the second square term in the integrand, equals $\hat{C}_{n1} + \hat{C}_{n2}$, where $\hat{C}_{n1} = n_1^{-2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)\xi_i^2 \mathrm{d}\hat{\psi}_{h_2}(z)$ and $\hat{C}_{n2}$ is the remainder. By consistency of $\hat{\theta}_n$ and the conditional argument on $\hat{Q}_{n_2}(Z_i)$, one can verify that $n_1 h_1^{d/2}\hat{C}_{n2} = o_p(1)$.

To see the asymptotic property of $\hat{\Gamma}_n$ under the local alternatives, we use the same technique. Adding and subtracting $\theta_0' Q(Z_i)$, $\theta_0' Q(Z_j)$ from $Y_i - \hat{\theta}_n'\hat{Q}_{n_2}(Z_i)$ and $Y_j - \hat{\theta}_n'\hat{Q}_{n_2}(Z_j)$ respectively, one obtains

$$\hat{\Gamma}_n = \frac{2h_1^d}{n_1^2}\sum_{i\neq j}^{n_1}\left[K_{h_1 i}(z)K_{h_1 j}(z)\xi_i\xi_j \mathrm{d}\hat{\psi}_{h_2}(z)\right]^2 + V_n.$$

One can show that $V_n = o_p(1)$ by the consistency of $\hat{\theta}_n$, the Cauchy–Schwarz inequality on the double sum, and the facts (6.9) and (6.10), while the first term converges to Γ in probability.

Therefore,

$$\hat{\mathcal{D}}_n = n_1 h_1^{d/2} \hat{\Gamma}_n^{-1/2}(T_{n1} - \hat{C}_{n1}) + n_1 h_1^{d/2}(\hat{C}_{n1} - \hat{C}_n) + n_1 h_1^{d/2}(T_{n2} + T_{n3}) + o_p(1),$$

the theorem follows by noting that the second term on the right-hand side of $\hat{\mathcal{D}}_n$ asymptotically has standard normal distribution and (6.13) and (6.14). □

Acknowledgments

The author would like to thank the Associate Editor and two referees for their helpful and critical comments which led to substantial improvement on the presentation of the paper.

References

[1] T.W. Anderson, Estimating linear statistical relationships, Ann. Statist. 12 (1984) 1–45.
[2] P.J. Bickel, Y. Ritov, Efficient estimation in the errors in variables model, Ann. Statist. 15 (2) (1987) 513–540.
[3] D. Bosq, Nonparametric Statistics for Stochastic Processes: Estimation and Prediction, 2nd edition, in: Springer Lecture Notes in Statistics, vol. 110, Springer-Verlag, New York, Inc, 1998.
[4] R.J. Carroll, P. Hall, Optimal rates of convergence for deconvoluting a density, JASA 83 (1988) 1184–1185.
[5] R.J. Carroll, C.H. Spiegelman, Diagnostics for nonlinearity and heteroscedasticity in errors in variables regression, Technometrics 34 (1992) 186–196.
[6] R.J. Carroll, D. Ruppert, L.A. Stefanski, Measurement Error in Nonlinear Models, Chapman & Hall/CRC, Boca Raton, 1995.
[7] C. Cheng, J.W. Van Ness, Statistical Regression with Measurement Error, Arnold, London, 1999.
[8] C.L. Cheng, A.G. Kukush, A goodness-of- fit test for a polynomial errors-in-variables model, Ukrainian Math. J. 56 (2004) 527–543.
[9] E. Masry, Strong consistency and rates for deconvolution of multivariate densities of stationary process, Stochastic Process. Appl. 47 (1993) 53–74.
[10] R.L. Eubank, C.H. Spiegelman, Testing the goodness of fit of a linear model via nonparametric regression techniques, J. Amer. Statist. Assoc. 85 (1990) 387–392.
[11] R.L. Eubank, J.D. Hart, Testing the goodness of fit in regression via order selection criteria, Ann. Statist. 20 (1992) 1412–1425.
[12] R.L. Eubank, J.D. Hart, Commonality of CUMSUM, von Neumann and smoothing based goodness-of-fit tests, Biometrika 80 (1993) 89–98.
[13] J. Fan, On the optimal rates of convergence for nonparametric deconvolution problems, Ann. Statist. 19 (1991) 1257–1272.
[14] J. Fan, Asymptotic normality for deconvolution kernel density estimators, Sankhy $\bar{a}$ Ser. A 53 (1991) 97–110.
[15] J. Fan, K.T. Truong, Nonparametric regression with errors in variables, Ann. Statist. 21 (1993) 1900–1925.
[16] W.A. Fuller, Measurement Error Models, Wiley, New York, 1987.
[17] L.J. Gleser, Estimation in a multivariate "errors in variables" regression model: Large sample results, Ann. Statist. 9 (1) (1981) 24–44.
[18] J.D. Hart, Nonparametric Smoothing and Lack-of-Fit Tests, Springer-Verlag, New York, Inc, 1997.
[19] Hira L. Koul, Pingping Ni, Minimum distance regression model checking, J. Stat. Plann. Inference 119 (1) (2004) 109–141.
[20] L.A. Stefanski, R.J. Carroll, Deconvolution-based score tests in measurement error models, Ann. Statist. 19 (1991) 249–259.
[21] W. Stute, Nonparametric model checks for regression, Ann. Statist. 25 (1997) 613–641.
[22] W. Stute, S. Thies, L.X. Zhu, Model checks for regression: An innovation process approach, Ann. Statist. 26 (1998) 1916–1934.
[23] J.X. Zheng, A consistent test of functional form via nonparametric estimation techniques, J. Econometrics 75 (1996) 263–289.

[24] L.X. Zhu, W.X. Song, H.J. Cui, Testing lack-of-fit for a polynomial errors-in-variables model, Acta Math. Appl. Sin. Engl. Ser. 19 (2003) 353–362.
[25] L.X. Zhu, H.J. Cui, K.W. Ng, Some properties of a lack-of-fit for a linear errors in variables model, Acta Math. Appl. Sinica 20 (4) (2004) 533–540.
[26] L.X. Zhu, H.J. Cui, Testing lack-of-fit for general linear errors in variables models, Statistica Sinica 15 (4) (2005) 1049–1068.