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

\[
Y = \mu(X) + \varepsilon, \quad Z = X + u,
\]  

where \( X \) is the unobservable \( d \)-dimensional random design variables. The random variables \( (X, u, \varepsilon) \) are assumed to be mutually independent, with \( u \) being \( d \)-dimensional and \( \varepsilon \) 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.

E-mail address: weixing@ksu.edu.

0047-259X/5 - see front matter © 2008 Elsevier Inc. All rights reserved.
a common and standard assumption in the literature of the errors-in-variables regression models. The density $f_X$ and the distribution of $\varepsilon$ 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

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

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 $\varepsilon$ and $E(\varepsilon) = 0$ imply that $v(z) = E(Y|Z = z) = E(\mu(X)|Z = z)$. Thus one can consider the new regression model $Y = v(Z) + \zeta$ in which the error $\zeta$ is uncorrelated with $Z$ and has mean 0. The problem of testing for $H_0$ is now transformed to a test for $v(z) = v_{\theta_0}(z)$, where $v_{\theta_0}(z) = \theta^t E(r(X)|Z = z)$.

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 $H_{20} : v(z) = v_\theta(z)$, 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 $\int m_1(x)f_X(x)f_\varepsilon(z-x)dx = \int m_2(x)f_X(x)f_\varepsilon(z-x)dx$ for all $z$, hence if $f_\varepsilon(z-x)$, 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 $v(z) = \int \mu(x)f_X(x)f_u(z-x)dx/f_Z(z)$. If $f_X$ is known then $f_Z$ is known and hence $v_0$ is known except for $\theta$. 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 $K_{h_i}(z) = K((z - Z_i)/h)/h^d$ for any kernel function $K$ and bandwidth $h$. If we define

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

where $G$ is a $\sigma$-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 $v(z)$. Then, we may use $\hat{\theta}_n = \arg\min_{\theta\in\mathbb{R}^d} \bar{T}_n(\theta)$ to estimate $\theta$, and construct the test statistic through $\bar{T}_n(\hat{\theta}_n)$. 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 setup. 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 $\phi_L$ denote its characteristic function and define

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

$$x \in \mathbb{R}^d,$$  \hspace{1cm} (2.1)

where $i = (-1)^{1/2}$. 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

$$R(z) = \int r(x)f_X(x)f_u(z-x)dx, \quad f_Z(z) = \int f_X(x)f_u(z-x)dx.$$  \hspace{1cm} (2.2)

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

$$\hat{R}_n(z) = \int r(x)\hat{f}_{Xh}(x)f_u(z-x)dx, \quad \hat{f}_{Zh}(z) = \int \hat{f}_{Xh}(x)f_u(z-x)dx.$$  \hspace{1cm}

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 \( I \).

To be precise, let

\[
\tilde{f}_{Z h_2}(z) = \sum_{i=1}^{n_1} K_{h_2 i}(z)/n_1, \quad \tilde{f}_{X w}(x) = \sum_{j=n_1+1}^{n} L_w((x - Z_j)/w)/n_2 w^d,
\]

\[
\hat{R}_{n_2}(z) = \int r(x) \tilde{f}_{X w_1}(x) f_u(z - x) dx,
\]

\[
\hat{f}_{Z w_2}(z) = \int \tilde{f}_{X w_2}(x) f_u(z - x) dx, \quad \hat{Q}_{n_2}(z) = \hat{R}_{n_2}(z)/\hat{f}_{Z w_2}(z),
\]

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

\[
M_n(\theta) = \int \left[ \frac{1}{n \hat{f}_{Z h_2}(z)} \sum_{i=1}^{n_1} K_{h_1 i}(z) (Y_i - \theta' \hat{Q}_{n_2}(Z_i)) \right]^2 dG(z), \tag{2.3}
\]

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

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

\[
\hat{\Gamma}_n = 2h_1^d n_1^{-2} \sum_{i \neq j=1}^{n_1} \left( \int K_{h_1 i}(z) K_{h_1 j}(z) \hat{\zeta}_i \hat{\zeta}_j d\hat{\psi}_{h_2}(z) \right)^2, \quad d\hat{\psi}_{h_2}(z) := \frac{dG(z)}{f_{Z h_2}(z)}.
\]

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

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

is standard normal. Consequently, the test that rejects \( H_0 \) whenever \( |\hat{D}_n| > z_{\alpha/2} \) is of the asymptotic size \( \alpha \), where \( z_{\alpha} \) 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 \( \sigma \)-finite measure \( G \), we assume the following:
The density $h$ for some $\Sigma_w$ there exist a positive continuous function $E$ 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 $v(z) = E(Y|Z = z)$ satisfying $\int v^2dG < \infty$, where $G$ is a $\sigma$-finite measure on $\mathbb{R}^d$.

(e2) $0 < \sigma^2_e = Ee^2 < \infty$, $E|r(X)||^2 < \infty$, where $\| \cdot \|$ denotes the usual Euclidean norm. The function $\delta^2(z) = E[(\theta_0^T r(X) - \theta_0^T Q(Z))^2|Z = z]$ 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 $\int |\phi_u(t)|dt < \infty$.

(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 $I_{\delta_0}$ of $\mathbb{R}^d$, where

\[
I_{\delta_0} = \{ y \in \mathbb{R}^d : \max_{1 \leq j \leq d} |y_j - z_j| \leq \delta_0, y = (y_1, \ldots, y_d)' \}
\]

\[
z = (z_1z_d)', z \in I.
\]

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

(q) $\Sigma_0 = \int Q(z)Q'(z)dG(z)$ 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 $\|t\| > T_0$,

\[
\|t\|^{-\alpha} \left| \int (r(z - x) - r(z)) \exp(-it'x)f_u(x)dx \right| \phi_u(t) \leq J(z),
\]

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

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

About the kernel functions, we assume:

(\ell) 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, $\int \|v\|^2L(v)dv < \infty$ and $\int \|t\|^\alpha |\phi_L(t)|dt < \infty$ with $\alpha$ 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 $\alpha$ 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,

\[
\left\| \frac{\int (r(z - x) - r(z)) \exp(-it'x)f_u(x)dx}{\phi_u(t)} \right\| \leq \int x \exp(-it'x)f_u(x)dx \cdot \exp\left(\frac{t'\Sigma_ut}{2}\right)
\]

\[
= \left| \frac{\partial \phi_u(t)}{\partial t} \right| \cdot \exp(t'\Sigma_ut/2) \leq c\|t\|,
\]

where the constant $c$ depends only on $\Sigma_u$. 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 \( \sigma_u^2 \). In this case, \( \phi_u(t) = 1/(1 + \sigma_u^2 t^2/2) \) and

\[
\left| \int \frac{f_u(x)(z) - r(z)}{f_u(x)} \exp(-ix) f_u(x) dx \right| = \left| \int (-2ix + x^2) \exp(-ix) f_u(x) dx \right| \phi_u(t) \\
\leq 2|z| \left| \frac{\partial \phi_u(t)}{\partial t} \right| \phi_u(t) + \left| \frac{\partial^2 \phi_u(t)}{\partial t^2} \right| \phi_u(t) \\
\leq 2|z| \frac{\sigma_u^2 |t|}{1 + \sigma_u^2 t^2/2} + \frac{\sigma_u^2}{1 + \sigma_u^2 t^2/2} \\
+ \frac{2\sigma_u^4 t^2}{(1 + \sigma_u^2 t^2/2)^2}.
\]

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 \( d = q = 1, r(x) = e^x \), and \( u \sim N(0, \sigma_u^2) \). Then

\[
\left| \int (r(z) - r(z)) \exp(-ix) f_u(x) dx \right| = \left| \int (e^{z-x} - e^z) \exp(-ix) f_u(x) dx \right| \\
\leq e^z \left[ \int e^z e^{ix} f_u(x) dx \right| + |\phi_u(t)| \right] \leq ce^z |\phi_u(t)|,
\]

where \( c \) is some positive number depending only on \( \sigma_u^2 \). Hence, (m1) holds for \( \alpha = 0 \) and, \( J(z) = ce^z \).

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 \( d \geq 1 \), and (f1), (u), (m1), (h1) hold. Then for any \( z \in \mathbb{R}^d \),

\[
\left\| E \hat{R}_{n_2}(z) - R(z) \right\|^2 \leq cw_1^4 I^2(z), \quad E \left\| \hat{R}_{n_2}(z) - E \hat{R}_{n_2}(z) \right\|^2 \leq \frac{c}{n_2 w_1^4} (J^2(z) w_1^{-2\alpha} + \|r(z)\|^2),
\]

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,

\[
E \left\| \hat{R}_{n_2}(z) - R(z) \right\|^2 \leq cw_1^4 I^2(z) + \frac{c}{n_2 w_1^4} (J^2(z) w_1^{-2\alpha} + \|r(z)\|^2).
\]

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

\[
E \left\| \hat{R}_{n_2}(z) - R(z) \right\|^2 \leq cn_2^{-\alpha - 4/5} (I^2(z) + J^2(z) + \|r(z)\|^2).
\]
In the following, we will write

\[ T(z) = I^2(z) + J^2(z) + \|r(z)\|^2. \]  

(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

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

for any positive integer \( k \) and compact set \( I \).

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

\[
\int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z) \hat{Q}_{n_2}(Z_i) \cdot \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z) \hat{Q}'_{n_2}(Z_i) d\hat{\psi}_h(z) \cdot \hat{\theta}_n = \int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z) Y_i \cdot \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z) \hat{Q}_{n_2}(Z_i) d\hat{\psi}_h(z). \tag{3.3}
\]

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:

\[
\int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z) \hat{Q}_{n_2}(Z_i) \cdot \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z) \hat{Q}'_{n_2}(Z_i) d\hat{\psi}_h(z) \cdot (\hat{\theta}_n - \theta_0) = \int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z) (Y_i - \theta_0^* \hat{Q}_{n_2}(Z_i)) \cdot \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z) \hat{Q}_{n_2}(Z_i) d\hat{\psi}_h(z). \tag{3.4}
\]

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

\[
\Sigma = \int \tau^2(z) Q(z) Q'(z) g^2(z)/f(z) dz,
\]

\( \Sigma_0 \) is as in condition (q), and \( \tau^2(z) = \sigma_e^2 + \delta^2(z) \), where \( \sigma_e^2 \), 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 $\hat{D}_n \Longrightarrow N(0, 1)$, where $\hat{D}_n$ 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 $\sqrt{n}(\hat{θ}_n - θ_0)$ can attain its minimum if $G$ is chosen to be $f_{ZH_2}(z)$. As far as the MD test statistic $M_n(\hat{θ}_n)$ 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 $\hat{I}_n$ in (2.5) by any other consistent estimator of $I$ does not affect the validity of Theorem 3.2, where

$$
I = 2 \int (\sigma^2_ε(z))^2 g(z) dψ(z) \cdot \int \left( \int K(u)K(u+v)du \right)^2 dv,
$$

(3.5)

$\sigma^2_ε(z) = \sigma_ε^2 + \delta^2(z)$, $\delta^2(z)$ is as in condition (e2), so we can choose some other consistent estimator of $I$, such as

$$
\hat{I}_n = C \int \left( \sum_{i=1}^{n_1} \frac{K_{h1}(z)(Y_i - \hat{θ}_n Q_{n2}(Z_i))^2}{n_1 f_{ZH_2}(z)} \right)^2 g(z) d\hat{ψ}_{h2}(z),
$$

(3.6)

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

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 $\inf_θ \int [H(z) - θ Q(z)]^2 dG(z) > 0$. Then, for the MD estimator $\hat{θ}_n$ defined in (2.4), $|\hat{D}_n| → \infty$ in probability.

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

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

(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 $\hat{D}_n →_d N(Γ^{-1/2} D, 1)$, where

$$
D = \int V^2(z)dG(z) + \int V(z)Q'(z)dG(z) \cdot \int Q(z)Q'(z)dG(z) \cdot \int V(z)Q(z)dG(z)
$$

and $Γ$ is as in (3.5).
4. Monte Carlo simulation

This section contains the results of four simulations corresponding to the following cases:
Case 1: \( d = q = 1 \) and \( m_\theta \) linear, the measurement error \( \epsilon \) is chosen to be normal and \( u \) double exponential; Case 2: \( d = q = 1 \) and \( m_\theta \) linear, the measurement error \( \epsilon \) and \( u \) are chosen to be normal; Case 3: \( d = 1, q = 2 \) and \( m_\theta \) a polynomial, the measurement error \( \epsilon \) is chosen to be normal and \( u \) double exponential; Case 4: \( d = q = 2 \), and \( m_\theta \) linear, the measurement error \( \epsilon \) 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 \( \{W_{jk}\}_{k,j=1}^{n_j} \) to denote the \( j \)th subsample \( S_j \) from \( W \) with sample size \( n_j \). So the full sample is \( S_1 \cup S_2 \). Finally, to make the simulation less time consuming, \( \hat{I}_n \) defined in (3.6) will be used in the test statistic instead of \( \hat{I}_n \). So the value of the test statistic is calculated by \( \hat{D}_n = n_1 h_1^{1/2} \hat{I}_n^{-1/2} (M_n(\hat{\theta}_n) - \hat{C}_n). \)

**Case 1.** In this case, \( \{X_{jk}\}_{k,j=1}^{n_j} \) are obtained as a random sample from the uniform distribution on \([-1, 1]\), \( \{\varepsilon_{jk}\}_{k,j=1}^{n_j} \) are obtained as a random sample from the normal distribution \( \mathcal{N}(0, 0.1^2) \) and \( \{u_{jk}\}_{k,j=1}^{n_j} \) 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

\[
Y_{jk} = X_{jk} + \varepsilon_{jk}, \quad Z_{jk} = X_{jk} + u_{jk}, \quad k = 1, 2, \ldots, n_j, \quad j = 1, 2.
\]

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

\[
K(z) = K^*(z) = \frac{3}{4}(1 - z^2)I(|z| \leq 1), \quad h_1 = an_1^{-1/3}, \quad h_2 = bn_1^{-1/5}(\log n_1)^{1/5},
\]

with some choices for \( a \) and \( b \). For the chosen kernel function (4.1), the constant \( C \) in \( \hat{I}_n \) 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, \( \hat{Q}_{n_2}(z) = y \hat{R}_{n_2}(z)/\hat{f}_{Z_{w_2}}(z) \) where

\[
\hat{R}_{n_2}(z) = \int x \hat{f}_{X_{w_1}}(x)f_u(z-x)dx, \quad \hat{f}_{Z_{w_2}} = \int \hat{f}_{X_{w_2}}(x)f_u(z-x)dx.
\]

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.

| Mean and MSE of \( \hat{\theta}_n \) |
|-----------------|----------------|
| \((50, 134)\)   | \((100, 317)\) |
| Mean            | 1.0103         | 1.0095         |
| MSE             | 0.0014         | 0.0007         |
| \((200, 753)\)  | \((300, 1250)\) | \((500, 2366)\) |
| Mean            | 1.0102         | 1.005         | 1.0098         |
| MSE             | 0.0004         | 0.0003         | 0.0002         |
Table 4.2
Levels and powers of the minimum distance test

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

To assess the level and power behavior of the $\hat{D}_n$ 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: $Y = X I(X \geq 0.2) + \varepsilon$.

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 $\hat{D}_n$ 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 $\#(\lvert \hat{D}_n \rvert \geq 1.96)/1000$.

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 $\hat{\theta}_n$

<table>
<thead>
<tr>
<th></th>
<th>(50, 941)</th>
<th>(100, 3164)</th>
<th>(200, 10643)</th>
<th>(300, 21638)</th>
<th>(500, 52902)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.0051</td>
<td>1.0078</td>
<td>1.0085</td>
<td>1.0101</td>
<td>1.0169</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0013</td>
<td>0.0007</td>
<td>0.0004</td>
<td>0.0003</td>
<td>0.0004</td>
</tr>
</tbody>
</table>

Table 4.4
Levels and powers of the minimum distance test

<table>
<thead>
<tr>
<th>Model</th>
<th>(50, 941)</th>
<th>(100, 3164)</th>
<th>(200, 10643)</th>
<th>(300, 21638)</th>
<th>(500, 52902)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 0</td>
<td>0.018</td>
<td>0.022</td>
<td>0.029</td>
<td>0.035</td>
<td>0.049</td>
</tr>
<tr>
<td>Model 1</td>
<td>0.918</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.993</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

*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 $\varepsilon \sim \mathcal{N}(0, 0.1^2)$. 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 $n_2 = [n_1]^{0.7}, b > 7/4$. In the simulation, we choose $b = 7/4 + 0.0001$. The bandwidths are chosen to be

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

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

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

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 $\hat{D}_n$ 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: $Y = X I(X \geq 0.2) + \varepsilon.$

Table 4.4 reports the simulation results pertaining to $\hat{D}_n$. 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 $\hat{\theta}_n$

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

Table 4.6
Levels and powers of the minimum distance test

<table>
<thead>
<tr>
<th>Model</th>
<th>(50, 134)</th>
<th>(100, 317)</th>
<th>(200, 753)</th>
<th>(300, 1250)</th>
<th>(500, 2366)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 0</td>
<td>0.001</td>
<td>0.009</td>
<td>0.019</td>
<td>0.029</td>
<td>0.046</td>
</tr>
<tr>
<td>Model 1</td>
<td>0.297</td>
<td>0.815</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.528</td>
<td>0.965</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.996</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

$\theta_1 = 1, \theta_2 = 2$. It is easy to see that $\hat{R}_{n2}(z)$ takes the form

$$\hat{R}_{n2}(z) = \left( \int x \hat{f}_{XW_1}(x) f_u(z - x) dx, \int x^2 \hat{f}_{XW_1}(x) f_u(z - x) dx \right)'$$

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 $\hat{D}_n$ 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 $\hat{D}_n$. 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 $m_0(X) = \theta_1 X_1 + \theta_2 X_2$. $X_1$ 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

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

$$h_1 = n_1^{-1/5}, h_2 = n_1^{-1/6} (\log n_1)^{1/6}.$$ For the chosen kernel function (4.2), the constant $C$ in $\hat{\Gamma}_n$ 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 $\hat{\theta}_n$

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

Table 4.8
Levels and powers of the minimum distance test

<table>
<thead>
<tr>
<th>Model</th>
<th>(50, 354)</th>
<th>(100, 1001)</th>
<th>(200, 2830)</th>
<th>(300, 5200)</th>
<th>(500, 11188)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 0</td>
<td>0.002</td>
<td>0.012</td>
<td>0.018</td>
<td>0.016</td>
<td>0.038</td>
</tr>
<tr>
<td>Model 1</td>
<td>0.908</td>
<td>0.998</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.992</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.935</td>
<td>0.996</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

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 $\hat{D}_n$ 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 $\hat{D}_n$. 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
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 $\hat{Q}_n(z)$ 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 $n_2 = \lceil n_1^b \rceil$, 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 \( n_1 = n_2 \)

<table>
<thead>
<tr>
<th>Model</th>
<th>(50, 50)</th>
<th>(100, 100)</th>
<th>(200, 200)</th>
<th>(300, 300)</th>
<th>(500, 500)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 0</td>
<td>0.013</td>
<td>0.023</td>
<td>0.027</td>
<td>0.035</td>
<td>0.047</td>
</tr>
<tr>
<td>Model 1</td>
<td>0.931</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Model 2</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.984</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 5.4
Normal, same sample

<table>
<thead>
<tr>
<th>Model</th>
<th>50</th>
<th>100</th>
<th>200</th>
<th>300</th>
<th>500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 0</td>
<td>0.017</td>
<td>0.019</td>
<td>0.036</td>
<td>0.036</td>
<td>0.051</td>
</tr>
<tr>
<td>Model 1</td>
<td>0.954</td>
<td>0.998</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.992</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 5.5
Double exponential, same sample, \( d = 2 \)

<table>
<thead>
<tr>
<th>Model</th>
<th>50</th>
<th>100</th>
<th>200</th>
<th>300</th>
<th>500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 0</td>
<td>0.000</td>
<td>0.004</td>
<td>0.010</td>
<td>0.018</td>
<td>0.041</td>
</tr>
<tr>
<td>Model 1</td>
<td>0.628</td>
<td>0.996</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.994</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.844</td>
<td>0.998</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

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 \( x \in \mathbb{R}^d \), \( E\hat{f}_{Xw_1}(x) = \int L(v)f_X(x - vw_1)dv \). By assumption (f1), there exists a vector \( a(x, v) \) such that \( f_X(x - vw_1) \)
From assumption (m1), by subtracting $r_D$, By changing the variable, and using the fact that the variance is bounded above by the second moment. Let $D$ random vectors. A routine calculation shows that $\|E \hat{R}_n(z) - E \hat{R}_n(z)\|^2 \leq \frac{1}{n^2 w_1^2} E \left\| \int r(x) L(x) f_u(z - x) dx \right\|^2$

has a Taylor expansion up to the second order, $f_X(x - vw_1) = f_X(x) - w_1 v' \hat{f}_X(x) + w_1^2 v' \hat{f}_X(a(x, v))) v/2$, where $\hat{f}$ and $\hat{f}'$ are the first- and second-order derivatives of $f$ with respect to its argument. Hence $E \hat{R}_n(z) = \int \int r(x) L(v) f_X(x - v w_1) f_u(z - x) dv dx$

$= \int \int r(x) L(v) f_X(x) f_u(z - x) dv dx$

$- w_1 \int \int r(x) L(v) v' \hat{f}_X(x) f_u(z - x) dv dx$

$+ \frac{1}{2} \int \int r(x) L(v) w_1^2 v' \hat{f}_X(a(v, x)) v f_u(z - x) dv dx$.

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

Note that $\hat{R}_n(z) - E \hat{R}_n(z)$ is an average of independently and identically distributed centered random vectors. A routine calculation shows that $\|E \hat{R}_n(z) - E \hat{R}_n(z)\|^2 \leq \frac{1}{n^2 w_1^2} E \left\| \int r(x) L(x) f_u(z - x) dx \right\|^2$ by using the fact that the variance is bounded above by the second moment. Let $D(t, z) = \int r(x) f_u(z - x) \exp(-it'x) dx$. By the definition of the deconvolution kernel $L_b$, it follows that $\frac{1}{w_1^2} E \left\| \int r(x) L(w_1 ((x - z)/w_1)) f_u(z - x) dx \right\|^2$

$= \int \int \frac{D(t, z) D(s, z) \phi_L(t w_1) \phi_L(s w_1) \phi_X(t + s) \phi_u(t + x)}{(2\pi)^2 \phi_u(t) \phi_u(s)} ds dt$.

By changing the variable, $D(t, z) = \exp(-it'z) \int r(z - x) f_u(x) \exp(it'x) dx$. Adding and subtracting $r(z)$ from $r(z - x)$ in the integrand, we obtain $D(t, z) = \exp(-it'z) \phi_u(z) \left[ r(z) + \frac{\int (r(z - x) - r(z)) f_u(x) \exp(it'x) dx}{\phi_u(t)} \right]$.

From assumption (m1), $\|D(t, z)\|$ is bounded above by $|\phi_u(t)| \cdot [\|r(z)\| + J(z)\|t\|^q]$ for all $z \in \mathbb{R}^d$. Hence $E \|\hat{R}_n(z) - E \hat{R}_n(z)\|^2$ is bounded above by
\[
\frac{c}{n_2} \int \int \left| \phi_L(tw_1)\phi_L(sw_1)\phi_u(t + s) \right| dr ds \\
+ \frac{cJ(z)}{n_2} \int \int \left( \|t\|^\alpha + \|s\|^\rho \right) \left| \phi_L(tw_1)\phi_L(sw_1)\phi_u(t + s) \right| dr ds \\
+ \frac{cJ^2(z)}{n_2} \int \int \|t\|^\alpha \|s\|^\rho \phi_L(tw_1)\phi_L(sw_1)\phi_u(t + s) dr ds.
\]

Note that for any \( m, p = 0 \) or \( \alpha \), from assumption (\( \ell \)), we have

\[
\int \int \|t\|^p \|s\|^m \phi_L(tw_1)\phi_L(sw_1)\phi_u(t + s) dr ds
\leq w_1^{-p-m-2d} \int \int \|t\|^p \|s\|^m \phi_L(t)\phi_L(s)\phi_u((t + s)/w_1) dr ds
\leq c w_1^{-p-m-2d} \int \int \|s\|^m \phi_L(s) \phi_u((t + s)/w_1) dr ds
= c w_1^{-p-m-d} \int \|s\|^m \phi_L(s) ds \cdot \int |\phi_u(t)| dt = c w_1^{-p-m-d}.
\]

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

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

\[
U_{n_1}(z) = \frac{1}{n_1} \sum_{i=1}^{n_1} K_{hi}(z)(Y_i - \theta_0^i Q(Z_i)), \\
D_n(z) = \frac{1}{n_1} \sum_{i=1}^{n_1} K_{hi}(z)(\hat{Q}_{n_2}(Z_i) - Q(Z_i)), \\
\mu_{n_1}(z) = \frac{1}{n_1} \sum_{i=1}^{n_1} K_{hi}(z)Q(Z_i), \quad \Delta_{n_1}(z) = \frac{1}{f_{zh_2}^2(z)} - \frac{1}{f_{zh}^2(z)}.
\]

It suffices to show that the matrix before \( \hat{\theta}_n - \theta_0 \) on the left-hand side of (3.4) converges to \( \Sigma_0 \) in probability, and \( \sqrt{n_1} \) times the right-hand side of (3.4) is asymptotically normal with mean vector 0 and covariance matrix \( \Sigma \).

Consider the second claim first. Adding and subtracting \( \theta_0^i Q(Z_i) \) from \( Y_i - \theta_0^i \hat{Q}_{n_2}(Z_i) \) 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 \( 1/f_{zh_2}^2(z) \) by \( 1/f_{zh_2}^2(z) - 1/f_{zh}^2(z) + 1/f_{zh}^2(z) = \Delta_{n_1}(z) + 1/f_{zh}^2(z) \), \( \sqrt{n_1} \) times the right-hand side of (3.4) can be written as the sum of the following eight terms.

\[
S_{n1} = \sqrt{n_1} \int U_{n_1}(z) D_n(z) \Delta_{n_1}(z) dG(z), \quad S_{n2} = \sqrt{n_1} \int U_{n_1}(z) D_n(z) d\psi(z), \\
S_{n3} = \sqrt{n_1} \int U_{n_1}(z) \mu_{n_1}(z) \Delta_{n_1}(z) dG(z), \quad S_{n4} = \sqrt{n_1} \int U_{n_1}(z) \mu_{n_1}(z) d\psi(z), \\
S_{n5} = -\sqrt{n_1} \int D_n(z) D_n'(z) \Delta_{n_1}(z) dG(z) \theta_0, \quad S_{n6} = -\sqrt{n_1} \int D_n(z) D_n'(z) d\psi(z) \theta_0, \\
S_{n7} = -\sqrt{n_1} \int D_n(z) \mu_{n_1}'(z) \Delta_{n_1}(z) dG(z) \theta_0, \quad S_{n8} = -\sqrt{n_1} \int D_n(z) \mu_{n_1}'(z) d\psi(z) \theta_0.
\]

Among these terms, \( S_{n4} \) is asymptotically normal with mean vector 0 and covariance matrix \( \Sigma \). 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 \( \hat{m}_{\theta_0}(X_i) \) there replaced by \( \theta_0^T Q(Z_i) \) 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 \( U_{n1}(z) = O_p(1/\sqrt{n_1 h_1^d}) \), 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 \( K_{h_i} \) 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 \( I \), so if we define

\[
I_{h_1} = \{ y \in \mathbb{R}^d : |y_j - z_j| \leq h_1, j = 1, \ldots, d, y = (y_1, \ldots, y_d)' \}.
\]

then \( I_{h_1} \) is a compact set in \( \mathbb{R}^d \), and \( K_{h_i} \) is 0 if \( Z_i \not\in I_{h_1} \). Hence, without loss of generality, we can assume all \( Z_i \in I_{h_1} \). Since \( f_Z \) is bounded from below on the compact set \( I_{h_0} \), by assumption (f2) and \( I_{h_1} \subset I_{h_0} \) for \( n_1 \) large enough, from assumption (w2), Lemma 3.2, we obtain

\[
\sup_{z \in I_{h_1}} \frac{|f_Z(z)|}{f_{Zw_2}(z)} = o \left( (\log n_2) \left( \frac{\log n_2}{n_2} \right)^{\frac{2}{pi^2}} \right) \quad \text{a.s., sup}_{z \in I_{h_1}} \frac{|f_Z(z)|}{f_{Zw_2}(z)} = o_p(1).
\]

(6.3)

Secondly, we have the following inequality,

\[
||\hat{Q}_{n2}(Z_i) - Q(Z_i)|| \leq ||\hat{R}_{n2}(Z_i) - R(Z_i)|| \cdot \frac{f_Z(Z_i)}{f_{Zw_2}(Z_i)} + \frac{f_Z(Z_i)}{f_{Zw_2}(Z_i)} - 1 \cdot ||Q(Z_i)||.
\]

Recall the definition of \( S_{n8} \). We have

\[
||S_{n8}|| \leq \sqrt{n_1} ||\theta_0|| \int_1^{n_1} \sum_{i=1}^{n_1} K_{h_1}(z)||\hat{Q}_{n2}(Z_i) - Q(Z_i)|| \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1}(z) ||Q(Z_i)|| d\psi(z).
\]

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

\[
\sqrt{n_1} \cdot O_p(1) \cdot A_{n11} + \sqrt{n_1} \cdot o \left( (\log n_2) \left( \frac{\log n_2}{n_2} \right)^{\frac{2}{pi^2}} \right) \cdot A_{n12},
\]

(6.4)

where

\[
A_{n11} = \int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1}(z)||\hat{R}_{n2}(Z_i) - R(Z_i)|| \cdot \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1}(z) ||Q(Z_i)|| d\psi(z)
\]

\[
A_{n12} = \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1}(z) ||Q(Z_i)|| \right]^2 d\psi(z).
\]

By the Cauchy–Schwarz inequality, \( A_{n11}^2 \) is bounded above by

\[
\int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1}(z)||\hat{R}_{n2}(Z_i) - R(Z_i)|| \right]^2 d\psi(z) \cdot \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1}(z) ||Q(Z_i)|| \right]^2 d\psi(z).
\]
Note that
\[
E \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z) \| \hat{R}_{n_2}(Z_i) - R(Z_i) \| \right]^2 d\psi(z)
= \int E \left( \frac{1}{n_1} \sum_{i,j=1}^{n_1} K_{h_1i}(z) K_{h_1j}(z) E_{S_1} \| \hat{R}_{n_2}(Z_i) - R(Z_i) \| \| \hat{R}_{n_2}(Z_j) - R(Z_j) \| \right) d\psi(z).
\]

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 \( cn_2^{-4/(d+2\alpha+4)} T^{1/2}(Z_i) T^{1/2}(Z_j) \), where \( T \) is defined in (3.2). So
\[
E \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z) \| \hat{R}_{n_2}(Z_i) - R(Z_i) \| \right]^2 d\psi(z) \leq cn_2^{-\frac{4}{d+2\alpha+4}}
\times \int E \left( \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z) T^{1/2}(Z_i) \right)^2 d\psi(z).
\]

Using the similar method as in K–N, together with the assumptions (m1) and (m2), we can show that
\[
\int \left( \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z) T^{1/2}(Z_i) \right)^2 d\psi(z) = O_p(1)
= \int \left( \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z) \| Q(Z_i) \| \right)^2 d\psi(z).
\]

Finally, from (6.4), we obtain \( \| S_{n8} \| \leq \sqrt{n_1} \cdot O_p(n_2^{-2/(d+2\alpha+4)}) + \sqrt{n_1} \cdot o_p((\log n_2)^2(\log n_2)^{2/(d+2\alpha+4)}) \) 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 \( \Sigma_0 \) 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.

\[
T_{n1} = \int D_n(z) D_n'(z) \Delta_{n_1}(z) dG(z), \quad T_{n2} = \int D_n(z) \mu_{n_1}'(z) \Delta_{n_1}(z) dG(z),
T_{n3} = \int \mu_{n_1}(z) D_n'(z) \Delta_{n_1}(z) dG(z), \quad T_{n4} = \int \mu_{n_1}(z) \mu_{n_1}'(z) \Delta_{n_1}(z) dG(z),
T_{n5} = \int D_n(z) D_n'(z) d\psi(z), \quad T_{n6} = \int D_n(z) \mu_{n_1}'(z) d\psi(z),
T_{n7} = \int \mu_{n_1}(z) D_n'(z) d\psi(z), \quad T_{n8} = \int \mu_{n_1}(z) \mu_{n_1}'(z) d\psi(z).
\]

Notice the connection between \( T_{n1} \) and \( S_{n5} \), \( T_{n2} \) and \( S_{n7} \), \( T_{n3} \) and \( S_{n5} \) and \( S_{n6} \), \( T_{n4} \) and \( S_{n6} \), \( T_{n5} \) 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 $\Sigma_0$ in probability, thereby proving the theorem. □

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

$$\xi_i = Y_i - \theta_0' Q(Z_i), \quad \zeta_i = Y_i - \theta_0' \hat{Q}_{n2}(Z_i),$$

$$\tilde{C}_n = n_1^{-2} \sum_{i=1}^{n_1} \int K_{h1i}(z) \xi_i^2 d\psi(z), \quad \tilde{M}_n(\theta_0) = \int \left[ n_1^{-1} \sum_{i=1}^{n_1} K_{h1i}(z) \xi_i \right]^2 d\psi(z),$$

$$\Gamma = 2 \int (\tau^2(z))^2 g(z) d\psi(z) \cdot \int \left[ \int K(u) K(u + v) du \right]^2 dv,$$

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

**Lemma 6.1.** If $H_2$ (c1), (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 $\xi_i$ by $\xi_i + \theta_0' (Q(Z_i) - \hat{Q}_{n2}(Z_i))$ 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.

$$B_{n1} = \frac{1}{n_1^2} \sum_{i \neq j}^n \int K_{h1i}(z) K_{h1j}(z) \xi_i \xi_j d\psi(z),$$

$$B_{n2} = \frac{1}{n_1^2} \sum_{i \neq j}^n \int K_{h1i}(z) K_{h1j}(z) \xi_i \theta_0' (Q(Z_j) - \hat{Q}_{n2}(Z_j)) d\psi(z),$$

$$B_{n3} = \frac{1}{n_1^2} \sum_{i \neq j}^n \int K_{h1i}(z) K_{h1j}(z) \xi_j \theta_0' (Q(Z_i) - \hat{Q}_{n2}(Z_i)) d\psi(z),$$

and $B_{n4} = n_1^{-2} \sum_{i \neq j}^n \int K_{h1i}(z) K_{h1j}(z) \theta_0' (Q(Z_i) - \hat{Q}_{n2}(Z_i)) \theta_0' (Q(Z_j) - \hat{Q}_{n2}(Z_j)) d\psi(z)$.

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 $C_{nj} = \sum_{i \neq j}^n \int K_{h1i}(z) K_{h1j}(z) \xi_i d\psi(z)$, $B_{n2}$ is bounded above by the sum $B_{n21} + B_{n22}$, where

$$B_{n21} = O_p(1) \cdot \frac{1}{n_1^2} \sum_{j=1}^{n_1} \| \hat{R}_{n2}(Z_j) - R(Z_j) \| \cdot |C_{nj}|,$$

$$B_{n22} = o \left( \log n_2 \left( \frac{\log n_2}{n_2} \right)^{2} \right) \cdot \frac{1}{n_1^2} \sum_{j=1}^{n_1} \| Q(Z_j) \| \cdot |C_{nj}|.$$
On the one hand, by the conditional expectation argument and inequality (3.1), we have

\[
E \frac{1}{n_1} \sum_{j=1}^{n_1} [\| \hat{R}_{n_2}(Z) - R(Z) \| \cdot |C_{nij}|] = E \frac{1}{n_1} \sum_{j=1}^{n_1} [E_t(\| \hat{R}_{n_2}(Z) - R(Z) \| \cdot |C_{nij}|)] \\
\leq cn_2^{-2/(d+2\alpha+4)} E \left[ \frac{1}{n_1} \sum_{j=1}^{n_1} T^{1/2}(Z) \cdot |C_{nij}| \right] \\
= cn_2^{-2/(d+2\alpha+4)} \frac{1}{n_1} E[T^{1/2}(Z) \cdot |C_{n11}|].
\]

Now, consider the asymptotic behavior of \( E[T^{1/2}(Z) \cdot |C_{n11}|] \). Instead of considering the expectation, we investigate the second moment. It is easy to see that \( ET(Z_1)C_{n11}^2 \) equals

\[
ET(Z_1) \sum_{i \neq 1} \sum_{j \neq 1} \int \int K_{h_i}(z) K_{h_j}(y) K_{h_1}(y) \xi_i \xi_j dy \psi(y) \\
= (n_1 - 1) \int \int E(K_{h1}(z) K_{h2}(y) \xi_2^2) \cdot E(K_{h1}(z) K_{h1}(y) T(Z_1)) dy \psi(y).
\]

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

\[
E(K_{h1}(z) K_{h1}(y) \xi_2^2) = E(K_{h1}(z) K_{h2}(y)(\sigma_e^2 + \delta^2(Z_2))) \\
= \frac{1}{h_1^d} \int K(v) K\left(\frac{y-z}{h_1} - v\right) (\sigma_e^2 + \delta^2(z - h_1 v)) f_Z(z - h_1 v) dv.
\]

Similarly, we can show that

\[
E(K_{h1}(z) K_{h1}(y) T(Z_1)) = \frac{1}{h_1^d} \int K(v) K\left(\frac{y-z}{h_1} - v\right) T(z - h_1 v) f_Z(z - h_1 v) dv.
\]

Putting back these two expectations in (6.5), and changing variables \( y = z + h_1 u \), then by the continuity of \( f_Z, \delta^2(z), g(z), \) and \( T(z) \), we obtain \( ET(Z_1)C_{n1}^2 = (n_1 - 1)h_1^{-d} \). Therefore,

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

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 \( n_1^{-2} \sum_{j=1}^{n_1} [\| Q(Z) \| \cdot |C_{nij}|] \) is of the order \( O_p(n_1^{-1/2}h_1^{-d/2}) \). Thus,

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

Hence

\[
n_1 h_1^{d/2}|B_{n2}| = O_p \left(n_1^{\frac{1}{2} - \frac{d+2\alpha+4}{d+2\alpha+4}} \right) + O_p \left(n_1^{\frac{1}{2} - \frac{2b}{d+2\alpha+4}} \log n_1 \right) = o_p(1),
\]

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

By exactly the same method as above, we can show that \( n_1 h_1^{d/2} B_{n3} = o_p(1) \).
It remains to show that \( n_1 h_1^{d/2} B_n = o_p(1) \). Note that

\[
|B_{n4}| \leq \frac{1}{n^2} \sum_{i \neq j}^{n_1} \int K_{h_1}(z) K_{h_1}(z) \| \theta_0 \|^2 \cdot \| \hat{Q}_{n2}(Z_i) - Q(Z_i) \| \times \| \hat{Q}_{n2}(Z_j) - Q(Z_j) \| d\psi(z).
\]

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

\[
O_p(1) \cdot B_{n41} + o_p \left( \left( \log_k n_2 \left( \frac{\log n_2}{n_2} \right) \right)^{\frac{2}{\pi+4}} \right) (B_{n42} + B_{n43})
\]

\[
+ o_p \left( \left( \log_k n_2 \left( \frac{\log n_2}{n_2} \right) \right)^{\frac{4}{\pi+4}} \right) B_{n44},
\]

where

\[
B_{n41} = \frac{1}{n_1^2} \sum_{i \neq j}^{n_1} \int K_{h_1}(z) K_{h_1}(z) \cdot \| \hat{R}_{n2}(Z_i) - R(Z_i) \| \cdot \| \hat{R}_{n2}(Z_j) - R(Z_j) \| d\psi(z),
\]

\[
B_{n42} = \frac{1}{n_1^2} \sum_{i \neq j}^{n_1} \int K_{h_1}(z) K_{h_1}(z) \cdot \| \hat{R}_{n2}(Z_i) - R(Z_i) \| \cdot \| Q(Z) \| d\psi(z),
\]

\[
B_{n43} = \frac{1}{n_1^2} \sum_{i \neq j}^{n_1} \int K_{h_1}(z) K_{h_1}(z) \cdot \| \hat{R}_{n2}(Z_j) - R(Z_j) \| \cdot \| Q(Z) \| d\psi(z),
\]

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

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

\[
E \cdot B_{n41} \leq c n_2^{-4/(d+2\alpha+4)} \int E[ K_{h_1}(z) T^{1/2}(Z_i) ]^2 d\psi(z) = O(n_2^{-4/(d+2\alpha+4)}).
\]

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

\[
B_{n42} = O_p(n_2^{-2/(d+2\alpha+4)}), \quad B_{n43} = O_p(n_2^{-2/(d+2\alpha+4)}), \quad B_{n44} = O_p(1).
\]

Therefore, for \( l = 2, 3 \),

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

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

\[
n_1 h_1^{d/2} \cdot O_p \left( \left( \log_k n_2 \left( \frac{\log n_2}{n_2} \right) \right)^{\frac{4}{\pi+4}} \right) B_{n44} = o_p(n_1^{1 - \frac{4b}{\pi+4} h_1^{d/2} (\log_k n_1) (\log n_1)^{\frac{4}{\pi+4}}})
\]
which is also of the order \( o_p(1) \). Finally, from the above and (6.6), we prove \( n_1 h_1^{d/2} B_n^{4} = o_p(1) \), 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 \( n_1^{-1} \sum_{i=1}^{n_1} K_{h_1 i}(z) \theta'_0 \hat{Q}_n(z_i) \) 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

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

\[
W_{n2} = \int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) \zeta_i \cdot \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) (\theta_0 - \hat{\theta}_n)' \hat{Q}_n(z_i) d\hat{\psi}_h(z),
\]

and \( \zeta_i = Y_i - \theta_0' \hat{Q}_n(z_i) \). It is easy to see that

\[
W_{n1} \leq 2 \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) (\theta_0 - \hat{\theta}_n)' (\hat{Q}_n(z_i) - Q(z_i)) \right]^2 d\hat{\psi}_h(z)
\]

\[
+ 2 \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) (\theta_0 - \hat{\theta}_n)' Q(z_i) \right]^2 d\hat{\psi}_h(z).
\]

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

\[
\| \hat{\theta}_n - \theta_0 \|^2 \cdot \sup_{z \in I} \left| \frac{f_z(z)}{f_{zw_2}(z)} \right| \cdot \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) \| \hat{Q}_n(z_i) - Q(z_i) \| \right]^2 d\psi(z).
\]

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 \( I_{h_1} \), 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

\[
\| \hat{\theta}_n - \theta_0 \|^2 \cdot \sup_{z \in I} \left| \frac{f_z(z)}{f_{zw_2}(z)} \right| \cdot \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) \| Q(z_i) \| \right]^2 d\psi(z).
\]

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

\[
W_{n2} = \int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) \zeta_i \cdot \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1 i}(z) \hat{Q}'_n(z_i) d\hat{\psi}_h(z) \cdot (\theta_0 - \hat{\theta}_n).
\]

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

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

\[
W_{n21} = 2\| \hat{\theta}_n - \theta_0 \|^2 \int \left[ n_1^{-1} \sum_{i=1}^{n_1} K_{h_i}(z) \| \hat{Q}_{n2}(Z_i) - Q(Z_i) \| \right]^2 d\psi_{h_2}(z),
\]

\[
W_{n22} = 2\| \hat{\theta}_n - \theta_0 \|^2 \int \left[ n_1^{-1} \sum_{i=1}^{n_1} K_{h_i}(z) \| Q(Z_i) \| \right]^2 d\psi_{h_2}(z).
\]

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. \( \square \)

**Lemma 6.3.** If \( H_0 \), (e1), (e2), (u), (f1), (f2), (m1), (m2), (\ell), (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 \( \xi_i \) and \( U_m(z) \). Note that

\[
n_1 h_1^{d/2} |M_n(\theta_0) - \tilde{M}_n(\theta_0)| \leq n_1 h_1^{d/2} \sup_{z \in I} \left| \frac{f_z^2(z)}{f_{w_2}^2(z)} \right| - 1 \int \left[ n_1^{-1} \sum_{i=1}^{n_1} K_{h_i}(z) \xi_i \right]^2 d\psi(z).
\]

Replace \( \xi_i \) by \( \xi_i + \theta'(0)(Q(Z_i) - \hat{Q}_{n2}(Z_i)) \), the integral part of the above inequality can be bounded above by the sum

\[
2 \int U_{w_2}^2(z) d\psi(z) + 2 \int \left[ n_1^{-1} \sum_{i=1}^{n_1} K_{h_i}(z) \theta'(0)(Q(Z_i) - \hat{Q}_{n2}(Z_i)) \right]^2 d\psi(z).
\]

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

\[
\sup_{z \in I_1} \left| \frac{f_z^2(z)}{f_{w_2}^2(z)} \right| \cdot O((n_2^{-4/(d+2\alpha+4)} + \sup_{z \in I_1} \left| \frac{f_z^2(z)}{f_{w_2}^2(z)} \right| - 1)^2 \cdot O_p(1).
\]

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

\[
O_p \left( n_1 h_1^{d/2} \cdot \frac{1}{nh_1^d} \cdot \log_k n_1 (\log n_1/n_1)^{2/(d+4)} \right)
\]

\[
+ O_p \left( n_1 h_1^{d/2} \cdot \log_k n_1 (\log n_1/n_1)^{2/(d+4)} \cdot n_1^{-4b/(d+2\alpha+4)} \right)
\]

\[
+ O_p \left( n_1 h_1^{d/2} \cdot \log_k n_1 (\log n_1/n_1)^{2/(d+4)} \cdot \log_k^2 n_1 (\log n_1)^{4/(d+4)} \cdot n_1^{-4b/(d+4)} \right).
\]

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

**Lemma 6.4.** If \( H_0 \), (e1), (e2), (e4), (u), (f1), (f2), (m1), (m2), (\ell), (n), (h1), (h2), (w1) and (w2) hold, \( n_1 h_1^{d/2} (\hat{C}_n - \tilde{C}_n) = o_p(1) \).
Recall the notation $\Delta_n(z)$ in (6.2). Adding and subtracting $\theta_0' \hat{Q}_{n2}(Z_i)$ from $Y_i$ in the integrand of $hC_n$, then expand the quadratic term, then $\hat{C}_n - \check{C}_n$ can be rewritten as the sum of $C_{nl}, l = 1, 2, 3, 4, 5,$ where

$$C_{n1} = \frac{1}{n_1^2} \sum_{i=1}^{n_1} \int K_{h_{i1}}^2(z)(Y_i - \theta_0' \hat{Q}_{n2}(Z_i))^2 \Delta_n(z) d\psi(z),$$

$$C_{n2} = \frac{2}{n_1^2} \sum_{i=1}^{n_1} \int K_{h_{i1}}^2(z)(Y_i - \theta_0' \hat{Q}_{n2}(Z_i))((\theta_0 - \hat{\theta}_n)' \hat{Q}_{n2}(Z_i)) \Delta_n(z) d\psi(z),$$

$$C_{n3} = \frac{2}{n_1^2} \sum_{i=1}^{n_1} \int K_{h_{i1}}^2(z)(Y_i - \theta_0' \hat{Q}_{n2}(Z_i))((\theta_0 - \hat{\theta}_n)' \hat{Q}_{n2}(Z_i)) d\psi(z),$$

$$C_{n4} = \frac{1}{n_1^2} \sum_{i=1}^{n_1} \int K_{h_{i1}}^2(z)(Y_i - \theta_0' \hat{Q}_{n2}(Z_i))((\theta_0 - \hat{\theta}_n)' \hat{Q}_{n2}(Z_i))^2 \Delta_n(z) d\psi(z),$$

$$C_{n5} = \frac{1}{n_1^2} \sum_{i=1}^{n_1} \int K_{h_{i1}}^2(z)(Y_i - \theta_0' \hat{Q}_{n2}(Z_i))((\theta_0 - \hat{\theta}_n)' \hat{Q}_{n2}(Z_i))^2 d\psi(z).$$

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

$$|C_{n1}| \leq 2 \sup_{z \in \mathcal{Z}} |\Delta_n(z)| \cdot \frac{1}{n_1^2} \sum_{i=1}^{n_1} \int K_{h_{i1}}^2(z) \xi_i^2 d\psi(z)$$

$$+ 2 \sup_{z \in \mathcal{Z}} |\Delta_n(z)| \cdot \frac{1}{n_1^2} \sum_{i=1}^{n_1} \int K_{h_{i1}}^2(z)((\theta_0' \hat{Q}(Z_i) - \hat{Q}_{n2}(Z_i)))^2 d\psi(z)$$

$$= C_{n11} + C_{n12}.$$ 

Since $n_1^{-2} \sum_{i=1}^{n_1} \int K_{h_{i1}}^2(z) \xi_i^2 d\psi(z) = O_p(1/n_1 h_1^d)$ by a routine expectation argument,

$$n_1 h_1^{d/2} |C_{n11}| = o_p \left( n_1 h_1^{d/2} \cdot (\log k_1)(\log n_1)^{(2/d+4)} n_1^{-(2/d+4)} \cdot (n_1 h_1)^{-1} \right) = o_p(1).$$

Second, from the compactness of $\Theta$, we have

$$\frac{1}{n_1^2} \sum_{i=1}^{n_1} \int K_{h_{i1}}^2(z)((\theta_0' \hat{Q}(Z_i) - \hat{Q}_{n2}(Z_i)))^2 d\psi(z)$$

$$\leq O(1) \cdot \frac{1}{n_1^2} \sum_{i=1}^{n_1} \int K_{h_{i1}}^2(z) \|Q(Z_i) - \hat{Q}_{n2}(Z_i)\|^2 d\psi(z).$$

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

$$O_p(n_2^{-4/(d+2\alpha+4)}) \cdot \sup_{z \in \mathcal{L}_{n1}} \left| \frac{f(z)}{f_{Zw_2}(z)} \right|^2 \cdot \frac{1}{n_1^2} \sum_{i=1}^{n_1} \int K_{h_{i1}}^2(z) T^2(Z_i) d\psi(z)$$

$$+ \sup_{z \in \mathcal{L}_{n1}} \left| \frac{f(z)}{f_{Zw_2}(z)} - 1 \right|^2 \cdot \frac{1}{n_1^2} \sum_{i=1}^{n_1} \int K_{h_{i1}}^2(z) \|Q(Z_i)\|^2 d\psi(z).$$
Because
\[ \frac{1}{n^2} \sum_{i=1}^{n_1} \int K_{h_t}^2(z) T^2(Z_i) d\psi(z) = O_p(1/n_1 h_t^d), \]
then from (h2), (w2), (h1), and Lemma 3.2, we get \( n_1 h_t^{d/2} |C_{n12}| = o_p(1) \). Hence \( n_1 h_t^{d/2} |C_{n1}| = o_p(1) \). Now we will show that \( n_1 h_t^{d/2} |C_{n3}| = o_p(1) \). Once we prove this, then \( n_1 h_t^{d/2} |C_{n2}| = o_p(1) \) is a natural consequence. In fact,
\[
C_{n3} = \frac{2}{n^2} \sum_{i=1}^{n_1} \int K_{h_t}^2(z) (\xi_i + \theta_0^T Q(Z_i) - \theta_n^0 \hat{Q}_{n2}(Z_i)) \cdot (\theta_0 - \hat{\theta}_n)^T (\hat{Q}_{n2}(Z_i) - Q(Z_i)) d\psi(z).
\]
So \(|C_{n3}|\) is bounded above by the sum \( 2(C_{n31} + C_{n32} + C_{n33} + C_{n34}) \), where
\[
C_{n31} = \frac{1}{n^2} \sum_{i=1}^{n_1} \int K_{h_t}^2(z) |\xi_i| \|\theta_0 - \hat{\theta}_n\| \|\hat{Q}_{n2}(Z_i) - Q(Z_i)\| d\psi(z),
\]
\[
C_{n32} = \frac{1}{n^2} \sum_{i=1}^{n_1} \int K_{h_t}^2(z) |\xi_i| \|\theta_0 - \hat{\theta}_n\| \|Q(Z_i)\| d\psi(z),
\]
\[
C_{n33} = \frac{1}{n^2} \sum_{i=1}^{n_1} \int K_{h_t}^2(z) \|\theta_0 - \hat{\theta}_n\| \|\hat{Q}_{n2}(Z_i) - Q(Z_i)\|^2 d\psi(z),
\]
\[
C_{n34} = \frac{1}{n^2} \sum_{i=1}^{n_1} \int K_{h_t}^2(z) \|\theta_0 - \hat{\theta}_n\| \|\hat{Q}_{n2}(Z_i) - Q(Z_i)\| \|Q(Z_i)\| d\psi(z).
\]
It is sufficient to show that \( n_1 h_t^{d/2} |C_{n3l}| = o_p(1) \) for \( l = 1, 2, 3, 4 \). Because the proofs are similar, here we only show \( n_1 h_t^{d/2} |C_{n32}| = o_p(1) \), others are omitted for the sake of brevity. In fact, note that \( n_1^{-2} \sum_{i=1}^{n_1} \int K_{h_t}^2(z) |\xi_i| \|Q(Z_i)\| d\psi(z) = O_p(1/n_1 h_t^d) \) by an expectation argument, then from \( \|\theta_0 - \hat{\theta}_n\| = O_p(n_1^{-1/2}) \) by Theorem 3.1, we have \( n_1 h_t^{d/2} |C_{n32}| = n_1 h_t^{d/2} \|\theta_0 - \hat{\theta}_n\| \cdot O_p(1/n_1 h_t^d) = O_p(n_1^{-1/2} h_t^{-d/2}) \). Since \( n_1^{-1/2} h_t^{-d/2} = n_1^{-1/2 + ad/2} \) 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 \( \xi_i \). Define
\[
\hat{\Gamma}_n = 2h_t^d n_1^{-2} \sum_{i \neq j}^{n_1} \left( \int K_{h_t}(z) K_{h_t}^2(z) \xi_i \xi_j d\psi_h(z) \right)^2.
\]
The lemma is proved by showing that
\[
\hat{\Gamma}_n - \hat{\Gamma}_n = o_p(1), \quad \hat{\Gamma}_n - \Gamma = o_p(1), \quad (6.6)
\]
where \( \Gamma \) 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 \( u_n = \hat{\theta}_n - \theta_0, r_i = \theta_0^T Q(Z_i) - \hat{\theta}_n^0 \hat{Q}_{n2}(Z_i) \). Now \( \hat{\Gamma}_n \) can be
Lemma 3.2 can be proved by using the same argument as in K–N. Now, consider bounded above by the sum of the following three terms:

\[ B_{n1} = 2h_1^d n_1^{-2} \sum_{i \neq j}^{n_1} \left[ \int K_{h,i}(z)K_{h,j}(z)\xi_i r_j d\hat{\psi}_h(z) + \int K_{h,i}(z)K_{h,j}(z)\xi_j r_i d\hat{\psi}_h(z) \right] , \]

\[ B_{n2} = 4h_1^d n_1^{-2} \sum_{i \neq j}^{n_1} \left( \int K_{h,i}(z)K_{h,j}(z)\xi_i \xi_j d\hat{\psi}_h(z) \right) \cdot \left( \int K_{h,i}(z)K_{h,j}(z)\xi_i r_j d\hat{\psi}_h(z) + \int K_{h,i}(z)K_{h,j}(z)\xi_j r_i d\hat{\psi}_h(z) \right) , \]

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

\[ h_1^d n_1^{-2} \sum_{i \neq j}^{n_1} \left[ \int K_{h,i}(z)K_{h,j}(z)|\xi_i r_j| d\hat{\psi}_h(z) \right]^2 = o_p(1) \]

(6.7)

\[ h_1^d n_1^{-2} \sum_{i \neq j}^{n_1} \left[ \int K_{h,i}(z)K_{h,j}(z)|r_i r_j| d\hat{\psi}_h(z) \right]^2 = o_p(1) , \]

\[ h_1^d n_1^{-2} \sum_{i \neq j}^{n_1} \left[ \int K_{h,i}(z)K_{h,j}(z)|\xi_i \xi_j| d\hat{\psi}_h(z) \right]^2 = O_p(1) . \]

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 \( d\hat{\psi}_h(z) \) is replaced by \( d\psi(z) \). 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

\[ h_1^d n_1^{-2} \sum_{i \neq j}^{n_1} |r_j|^2 \left[ \int K_{h,i}(z)K_{h,j}(z)\xi_i d\psi(z) \right]^2 \]  

(6.8)

Note that \( r_j = u_{n,i}(Q(Z_j) - \hat{Q}_n(Z_j)) - u_{n,0}^T Q(Z_j) - \theta_0^T (\hat{Q}_n(Z_j) - Q(Z_j)) \), so (6.8) can be bounded above by the sum of the following three terms:

\[ A_{n1} = 3h_1^d n_1^{-2} \|u_n\|^2 \sum_{i \neq j}^{n_1} \left[ \int K_{h,i}(z)K_{h,j}(z)\xi_i d\psi(z) \right]^2 , \]

\[ A_{n2} = 3h_1^d n_1^{-2} \|u_n\|^2 \sum_{i \neq j}^{n_1} \left[ \int K_{h,i}(z)K_{h,j}(z)\xi_i \| Q(Z_j) \| d\psi(z) \right]^2 , \]

\[ A_{n3} = 3h_1^d n_1^{-2} \|\theta_0\|^2 \sum_{i \neq j}^{n_1} \left[ \int K_{h,i}(z)K_{h,j}(z)\xi_i \| Q(Z_j) \| d\psi(z) \right]^2 . \]

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

\[ h_1^d n_1^{-2} \sum_{i \neq j}^{n_1} \left[ \int K_{h,i}(z)K_{h,j}(z)\xi_i \| Q(Z_j) \| d\psi(z) \right]^2 = O_p(1) . \]
which can be shown by using the same argument as in K–N. Let us consider \( A_n^3 \). Using the
inequality (6.4), Lemma 3.2 or (6.3), and the compactness of \( \Theta \), it is easy to see \( A_n^3 \) is bounded
above by the sum \( A_{n1} + A_{n2} \), where

\[
A_{n1} = O_p(1) \cdot h_1^d n_1^{-2} \sum_{i \neq j} \| \hat{R}_n^2(Z_j) - R(Z_j) \|^2 \left[ \int K_{h_1i}(z) K_{h_1j}(z) |\xi_i| d\psi(z) \right]^2
\]

\[
A_{n2} = o_p(1) \cdot h_1^d n_1^{-2} \sum_{i \neq j} \left[ \int K_{h_1i}(z) K_{h_1j}(z) |\xi_i| \| Q(Z_j) \| d\psi(z) \right]^2.
\]

Apply the conditional expectation argument to the second factor in \( A_{n1} \), using the fact (3.1) and
the elementary inequality \( a < (1 + a)^2 \), we can show

\[
E \left[ h_1^d n_1^{-2} \sum_{i \neq j} \| \hat{R}_n^2(Z_j) - R(Z_j) \|^2 \left[ \int K_{h_1i}(z) K_{h_1j}(z) |\xi_i| d\psi(z) \right]^2 \right]
\]

\[
= E \left[ h_1^d n_1^{-2} \sum_{i \neq j} (E_{Z_i} \| \hat{R}_n^2(Z_j) - R(Z_j) \|^2 \left[ \int K_{h_1i}(z) K_{h_1j}(z) |\xi_i| d\psi(z) \right]^2 \right]
\]

\[
\leq cn_2^{-\frac{d}{2} - 4 + d} E \left[ h_1^d n_1^{-2} \sum_{i \neq j} \left[ \int K_{h_1i}(z) K_{h_1j}(z) |\xi_i| \| Q(Z_j) \| d\psi(z) \right]^2 \right].
\]

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, \( h_1^d n_1^{-2} \sum_{i \neq j} \| \hat{R}_n^2(Z_j) - R(Z_j) \|^2 \left[ \int K_{h_1i}(z) K_{h_1j}(z) |\xi_i| d\psi(z) \right]^2 = o_p(1) \). This, in turn, implies that the second factor in \( A_{n1} = o_p(1) \). The same method as
in K–N also leads to \( h_1^d n_1^{-2} \sum_{i \neq j} \left[ \int K_{h_1i}(z) K_{h_1j}(z) |\xi_i| \| Q(Z_j) \| d\psi(z) \right]^2 = O_p(1) \). Hence \( A_{n2} = o_p(1) \). 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 \( H_0 : \mu(x) = m(x) \), one can verify that the right-hand side of (3.3) can be written as the sum of the following two terms

\[
A_{n1} = \int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z) m(X_i) \cdot \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z) \hat{Q}_n^2(Z_i) d\hat{\psi}_h(z),
\]

\[
A_{n2} = \int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z) \varepsilon_i \cdot \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z) \hat{Q}_n^2(Z_i) d\hat{\psi}_h(z).
\]

Adding and subtracting \( Q(Z_i) \) from \( \hat{Q}_n^2(Z_i) \), on the one hand, one has

\[
A_{n1} = \int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z) m(X_i) \cdot D_n(z) d\hat{\psi}_h(z) + \int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z) m(X_i)
\]

\[
\times \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z) Q(Z_i) d\hat{\psi}_h(z),
\]
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 \( \int H(z)Q(z)dG(z) \) in probability. On the other hand,

\[
A_{n2} = \int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z)\epsilon_i \cdot D_n(z)d\hat{\psi}_{h_2}(z) + \int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z)\epsilon_i \times \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z)Q(Z_i)d\hat{\psi}_{h_2}(z).
\]

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 \( \hat{\theta}_n \to \Sigma_0^{-1} \int H(z)Q(z)dG(z) \).

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

\[
S_{n1} = \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z)(Y_i - H(Z_i)) \right]^2 d\hat{\psi}_{h_2}(z),
\]

\[
S_{n2} = \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z)(H(Z_i) - \hat{\theta}_n \hat{Q}_n2(Z_i)) \right]^2 d\hat{\psi}_{h_2}(z),
\]

\[
S_{n3} = 2 \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z)(Y_i - H(Z_i)) \right] \times \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z)(H(Z_i) - \hat{\theta}_n \hat{Q}_n2(Z_i)) \right] d\hat{\psi}_{h_2}(z).
\]

Define

\[
C_n = \frac{1}{n_1} \sum_{i=1}^{n_1} \int K_{h_1i}^2(z)(Y_i - H(Z_i))^2d\hat{\psi}_{h_2}(z),
\]

\[
\Gamma_n = 2 \int (\sigma_n^2(z))^2 g(z)d\psi(z) \int \left[ \int K(u)K(u + v)du \right]^2 dv,
\]

\[
\sigma_n^2(z) = \sigma^2 + E[(m(X) - H(Z))^2|Z = z] + (H(z) - \theta'Q(z))^2.
\]

Similarly to the proof of Theorem 3.2, one can show that \( n_1h_1^{d/2}(S_{n1} - C_n^*) \to N(0, \Gamma^*) \) in distribution. Let \( \theta = \Sigma_0^{-1} \int H(z)Q(z)dG(z) \), 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

\[
S_{n21} = \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z)(H(Z_i) - \theta'Q(Z_i)) \right]^2 d\hat{\psi}_{h_2}(z),
\]

\[
S_{n22} = -2 \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z)(H(Z_i) - \theta'Q(Z_i)) \right] \times \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z)(\hat{\theta}_n \hat{Q}_n2(Z_i) - \theta'Q(Z_i)) \right] d\hat{\psi}_{h_2}(z),
\]
\[ S_{n23} = 2 \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1}(z)(\hat{\theta}_n \hat{Q}_{n2}(Z_i) - \theta' Q(Z_i)) \right]^2 d\hat{\psi}_h(z). \]

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

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

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

\[
C_n1 = \frac{2}{n_1^2} \sum_{i=1}^{n_1} \int K_{h_1}(z)(Y_i - H(Z_i))(H(Z_i) - \hat{\theta}_n \hat{Q}_{n2}(Z_i)) \, d\hat{\psi}_h(z),
\]

\[
C_n2 = \frac{1}{n_1^2} \sum_{i=1}^{n_1} \int K_{h_1}(z)(H(Z_i) - \hat{\theta}_n \hat{Q}_{n2}(Z_i))^2 \, d\hat{\psi}_h(z).
\]

Both \( C_n1 \) and \( C_n2 \) are \( o_p(1) \). Hence \( C_n - C_n^* = o_p(1) \).

Next, we shall show that \( \hat{I}_n = I^* + o_p(1) \). To this end, write \( e_i = Y_i - H(Z_i), \) \( w(Z_i, \hat{\theta}_n) = \hat{\theta}_n \hat{Q}_{n2}(Z_i) - H(Z_i), \) then

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

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

\[
A_{n1} = \frac{4h_1^d}{n_1^4} \sum_{i \neq j} \left( \int K_{h_1}(z)K_{h_1}(z)e_i e_j \, d\hat{\psi}_h(z) \right)^2
\]

\[
A_{n2} = \frac{2h_1^d}{n_1^4} \sum_{i \neq j} \left( \int K_{h_1}(z)K_{h_1}(z)e_i w(Z_j, \hat{\theta}_n) \, d\hat{\psi}_h(z) \right)^2
\]

\[
A_{n3} = \frac{2h_1^d}{n_1^4} \sum_{i \neq j} \left( \int K_{h_1}(z)K_{h_1}(z)w(Z_i, \hat{\theta}_n) e_j \, d\hat{\psi}_h(z) \right)^2
\]

\[
A_{n4} = \frac{2h_1^d}{n_1^4} \sum_{i \neq j} \left( \int K_{h_1}(z)K_{h_1}(z)w(Z_i, \hat{\theta}_n) w(Z_j, \hat{\theta}_n) \, d\hat{\psi}_h(z) \right)^2
\]

\[
A_{n5} = -\frac{4h_1^d}{n_1^4} \sum_{i \neq j} \left( \int K_{h_1}(z)K_{h_1}(z)e_i e_j \, d\hat{\psi}_h(z) \right) \times \int K_{h_1}(z)K_{h_1}(z)e_i w(Z_j, \hat{\theta}_n) \, d\hat{\psi}_h(z)
\]

...
\( A_{n6} = -\frac{4h_1^d}{n_1^2} \sum_{i \neq j} \left( \int K_{h_{1i}}(z) K_{h_{1j}}(z)e_i e_j d\hat{\psi}_2(z) \right) \times \int K_{h_{1i}}(z) K_{h_{1j}}(z) w(Z_i, \hat{\theta}_n) e_j d\hat{\psi}_2(z) \)

\( A_{n7} = \frac{4h_1^d}{n_1^2} \sum_{i \neq j} \left( \int K_{h_{1i}}(z) K_{h_{1j}}(z) e_i e_j d\hat{\psi}_2(z) \right) \times \int K_{h_{1i}}(z) K_{h_{1j}}(z) w(Z_i, \hat{\theta}_n) w(Z_j, \hat{\theta}_n) d\hat{\psi}_2(z) \)

\( A_{n8} = \frac{4h_1^d}{n_1^2} \sum_{i \neq j} \left( \int K_{h_{1i}}(z) K_{h_{1j}}(z) e_i w(Z_j, \hat{\theta}_n) d\hat{\psi}_2(z) \right) \times \int K_{h_{1i}}(z) K_{h_{1j}}(z) w(Z_i, \hat{\theta}_n) e_j d\hat{\psi}_2(z) \)

\( A_{n9} = -\frac{4h_1^d}{n_1^2} \sum_{i \neq j} \left( \int K_{h_{1i}}(z) K_{h_{1j}}(z) e_i w(Z_j, \hat{\theta}_n) d\hat{\psi}_2(z) \right) \times \int K_{h_{1i}}(z) K_{h_{1j}}(z) w(Z_i, \hat{\theta}_n) w(Z_j, \hat{\theta}_n) d\hat{\psi}_2(z) \)

\( A_{n10} = -\frac{4h_1^d}{n_1^2} \sum_{i \neq j} \left( \int K_{h_{1i}}(z) K_{h_{1j}}(z) w(Z_i, \hat{\theta}_n) e_j d\hat{\psi}_2(z) \right) \times \int K_{h_{1i}}(z) K_{h_{1j}}(z) w(Z_i, \hat{\theta}_n) w(Z_j, \hat{\theta}_n) d\hat{\psi}_2(z) \).

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

\[
\int_{\mathbb{R}^d} K_{h_{1i}}(z) K_{h_{1j}}(z) |e_i||e_j| d\psi(z)^2 = O_p(1), \quad (6.9)
\]

\[
\int_{\mathbb{R}^d} K_{h_{1i}}(z) K_{h_{1j}}(z) |e_i| d\psi(z)^2 = O_p(1), \quad k = 0, 1. \quad (6.10)
\]

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

\[
A_{n1} \rightarrow \Gamma^*_1 = 2 \int \sigma^2_v(z) g^2(z)/f^2(z)dz \int \left[ \int K(u + v)K(u)du \right]^2 dv
\]

in probability, where \( \sigma^2_v(z) = E[(Y - H(Z))^2|Z = z] = \sigma^2_v + E[(m(X) - H(Z))^2|Z = z] \). As for \( A_{n2} \), write \( \hat{\theta}_n - \hat{\theta}_0 \) as \( \hat{\theta}_n - \theta + \theta \) \((\hat{\theta}_n - \theta + \theta)(\hat{\theta}_n - \theta + \theta)\) and expand the integral by considering \( \hat{\theta}_n - \theta_0, \hat{\theta}_n - \theta_0 \) as a whole, respectively, \( A_{n2} \) can be written as the sum of

\[
2h_1^d n_1^{-2} \sum_{i \neq j} \left( \int K_{h_{1i}}(z) K_{h_{1j}}(z) e_i (\theta' Q(Z_j) - H(Z_j)) d\hat{\psi}_2(z) \right)^2,
\]
Lemma 6.1 can be written as the sum of the following six terms:

\begin{align}
\Gamma_2^* &= 2 \int \sigma^2_e(z) [H(z) - \theta' Q(z)]^2 g^2(z)/f_Z^2(z) \int \left[ \int K(u + v) K(u) du \right]^2 dv,
\end{align}

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 \( A_{n3} \rightarrow \Gamma_2^* \) 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

\begin{align}
A_{n4} \rightarrow \Gamma_3^* &= 2 \int [H(z) - \theta' Q(z)]^4 g^2(z)/f_Z^2(z) \int \left[ \int K(u + v) K(u) du \right]^2 dv
\end{align}

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

\begin{align}
A_{n51} &= -\frac{4h_1^d}{n_1^2} \sum_{i \neq j} \left( \int K_{h1i}(z) K_{h1j}(z) e_i e_j d\hat{\psi}_{h2}(z) \right)
\times \int K_{h1i}(z) K_{h1j}(z) e_i [\theta' Q(Z_j) - H(Z_j)] d\hat{\psi}_{h2}(z),
\end{align}

\begin{align}
A_{n52} &= -\frac{4h_1^d}{n_1^2} \sum_{i \neq j} \left( \int K_{h1i}(z) K_{h1j}(z) e_i e_j d\hat{\psi}_{h2}(z) \right)
\times \int K_{h1i}(z) K_{h1j}(z) e_i [\hat{\theta}_n^j \hat{Q}_{n2}(Z_j) - \theta' Q(Z_j)] d\hat{\psi}_{h2}(z).
\end{align}

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

\begin{align}
\tilde{A}_{n51} &= -\frac{4h_1^d}{n_1^2} \sum_{i \neq j} \left( \int K_{h1i}(z) K_{h1j}(z) e_i e_j d\hat{\psi}_{h2}(z) \right)
\times \int K_{h1i}(z) K_{h1j}(z) e_i [\theta' Q(Z_j) - H(Z_j)] d\hat{\psi}_{h2}(z).
\end{align}

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 \( \hat{\Gamma}_n \rightarrow \Gamma^* \) in probability.

Finally, we get

\begin{align}
\hat{D}_n &= n_1 h_1^{d/2} \hat{\Gamma}_n^{-1/2} (S_{n1} - C_{n1}^*) + n_1 h_1^{d/2} \hat{\Gamma}_n^{-1/2} \int [H(z) - \theta' Q(z)]^2 dG(z) + o_p(n_1 h_1^{d/2})
\end{align}

and the theorem follows immediately. \( \square \)

**Proof of Theorem 3.4.** Adding and subtracting \( \theta'_0 \hat{Q}_{n2}(Z_i) \) 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 \( \hat{Q}_{n2}(Z_i) \), respectively, and letting \( \xi_{ij} = e_i + \theta'_0 r(X_{ij}) - \theta'_0 Q(Z_j) \), 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

\begin{align}
B_{n1} &= \int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h1i}(z) \xi_{ij} \cdot \frac{1}{n_1} \sum_{j=1}^{n_1} K_{h1i}(z) Q(Z_i) d\hat{\psi}_{h2}(z),
\end{align}
\[ B_{n2} = \int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h1i}(z) \xi_i \cdot D_n(z) d\hat{\psi}_h(z), \]

\[ B_{n3} = \gamma_n \int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h1i}(z) v(X_i) \cdot \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h1i}(z) Q(Z_i) d\hat{\psi}_h(z), \]

\[ B_{n4} = \gamma_n \int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h1i}(z) v(X_i) \cdot D_n(z) d\hat{\psi}_h(z), \]

\[ B_{n5} = -\int \theta_0' D_n(z) \cdot D_n(z) d\hat{\psi}_h(z), \]

\[ B_{n6} = -\int \theta_0' D_n(z) \cdot \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h1i}(z) Q(Z_i) d\hat{\psi}_h(z). \]

Note that \( \xi_i \) 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 \( \xi_i \), one can show that \( \sqrt{n_1} B_{n2} = o_p(1) \) 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

\[ \eta_v(z) = E(K_h(z - Z)v(X)), \quad \eta_Q(z) = E(K_h(z - Z)Q(Z)), \quad (6.11) \]

then

\[ \sqrt{n_1} B_{n3} = \gamma_n \sqrt{n_1} \int \left[ \left( \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h1i}(z) v(X_i) - \eta_v(z) \right) \right. \]

\[ + \left. (\eta_v(z) - f_Z(z) V(Z)) + f_Z(z) V(Z) \right] \]

\[ \times \left[ \left( \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h1i}(z) Q(Z_i) - \eta_Q(z) \right) + (\eta_Q(z) - f_Z(z) Q(Z)) \right. \]

\[ + \left. f_Z(z) Q(Z) \right] d\hat{\psi}_h(z). \]

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

\[ \gamma_n \sqrt{n_1} \int f_Z^2(z) V(z) Q(z) d\hat{\psi}_h(z) = \gamma_n \sqrt{n_1} \int V(z) Q(z) dG(z) \]

\[ + \gamma_n \sqrt{n_1} \int V(z) Q(z) [1/f_Z^2 h(z) - 1/f_Z^2(z)] dG(z). \]

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

\[ \sqrt{n_1} B_{n3} = \sqrt{n_1} \gamma_n \int V(z) Q(z) dG(z) + o_p(1). \]
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 \( \Sigma_0 \) in probability. Thus, we showed that

\[
\sqrt{n}(\hat{\theta}_n - \theta_0) - \sqrt{n_1} Y_n \int V(z)Q(z)dG(z) \to_d N_q(0, \Sigma_0^{-1}\Sigma \Sigma_0^{-1}),
\]

where \( \Sigma_0 \) and \( \Sigma \) 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

\[
M_n(\hat{\theta}_n) = \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1}(z)[Y_i - \hat{\theta}_n Q_n(Z_i)] \right]^2 d\hat{\psi}_h(z)
\]

\[
= \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1}(z)[\theta_0 r(X_i) + \gamma_n v(X_i) + \varepsilon_i - \theta_0' Q(Z_i) + \hat{\theta}_n Q_n(Z_i)] \right]^2 d\hat{\psi}_h(z).
\]

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

\[
T_{n1} = \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1}(z)[\varepsilon_i + \theta_0' r(X_i) - \theta_0' Q(Z_i)] \right]^2 d\hat{\psi}_h(z),
\]

\[
T_{n2} = \gamma_n^2 \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1}(z) v(X_i) \right]^2 d\hat{\psi}_h(z),
\]

\[
T_{n3} = \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1}(z)[\theta_0' Q(Z_i) - \hat{\theta}_n Q_n(Z_i)] \right]^2 d\hat{\psi}_h(z),
\]

\[
T_{n4} = 2\gamma_n \int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1}(z)[\varepsilon_i + \theta_0' r(X_i) - \theta_0' Q(Z_i)] \cdot \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1}(z) v(X_i) d\hat{\psi}_h(z),
\]

\[
T_{n5} = 2 \int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1}(z)[\varepsilon_i + \theta_0' r(X_i) - \theta_0' Q(Z_i)] \cdot \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1}(z)[\theta_0' Q(Z_i) - \hat{\theta}_n Q_n(Z_i)] d\hat{\psi}_h(z),
\]

\[
T_{n6} = 2\gamma_n \int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1}(z) v(X_i) \cdot \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1}(z)[\theta_0' Q(Z_i) - \hat{\theta}_n Q_n(Z_i)] d\hat{\psi}_h(z).
\]

A simple argument leads to

\[
n_1 h_1^{d/2} T_{n2} = \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1}(z) v(X_i) \right]^2 d\hat{\psi}_h(z) \to \int [V(z)]^2 dG(z)
\]

in probability.
To deal with $T_{n3}$, note that

$$n_1 h_1^{d/2} T_{n3} = \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z)[(\hat{\theta}_n - \theta_0)'(\hat{Q}_{n2}(Z_i) - Q(Z_i))
+ Q(Z_i)] - \theta_0'Q(Z_i) \right]^2 d\hat{\psi}_2(z)$$

can be written as the sum of the following six terms

$$T_{n31} = n_1 h_1^{d/2} \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z)(\hat{\theta}_n - \theta_0)'(\hat{Q}_{n2}(Z_i) - Q(Z_i)) \right]^2 d\hat{\psi}_2(z)$$

$$T_{n32} = n_1 h_1^{d/2} \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z)(\hat{\theta}_n - \theta_0)'Q(Z_i) \right]^2 d\hat{\psi}_2(z)$$

$$T_{n33} = n_1 h_1^{d/2} \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z)\theta_0'(\hat{Q}_{n2}(Z_i) - Q(Z_i)) \right]^2 d\hat{\psi}_2(z)$$

$$T_{n34} = n_1 h_1^{d/2} \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z)(\hat{\theta}_n - \theta_0)'(\hat{Q}_{n2}(Z_i) - Q(Z_i)) \right] \times \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z)(\hat{\theta}_n - \theta_0)'Q(Z_i) \right] d\hat{\psi}_2(z)$$

$$T_{n35} = n_1 h_1^{d/2} \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z)(\hat{\theta}_n - \theta_0)'(\hat{Q}_{n2}(Z_i) - Q(Z_i)) \right] \times \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z)\theta_0'(\hat{Q}_{n2}(Z_i) - Q(Z_i)) \right] d\hat{\psi}_2(z)$$

$$T_{n36} = n_1 h_1^{d/2} \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z)(\hat{\theta}_n - \theta_0)'Q(Z_i) \right] \times \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z)\theta_0'(\hat{Q}_{n2}(Z_i) - Q(Z_i)) \right] d\hat{\psi}_2(z).$$

By (6.12), and conditional expectation arguments on $\hat{Q}_{n2}(Z_i)$, one can show that $T_{n3k} = o_p(1)$ for $k = 1, 3, 4, 5, 6$.

For $T_{n32}$, we have

$$T_{n32} = n_1 h_1^{d/2}(\hat{\theta}_n - \theta_0)' \int \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z)Q(Z_i) \right] \times \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1i}(z)Q'(Z_i) \right] d\hat{\psi}_2(z)(\hat{\theta}_n - \theta_0).$$
From (6.12), we see that

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

Therefore, we can show that

\[
n_1 h_1^{d/2} T_{n3} = \int V(z) Q'(z) dG(z) \cdot \int Q(z) Q'(z) dG(z) \cdot \int V(z) Q(z) dG(z) + o_p(1).
\]

(6.14)

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

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

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

one has

\[
T_{n4} = 2\gamma_n \int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1}(z) [\varepsilon_i + \theta_0' r(X_i) - \theta_0' Q(Z_i)] \cdot \eta_v(z) d\hat{\psi}_2(z)
\]

\[
+ 2\gamma_n \int \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1}(z) [\varepsilon_i + \theta_0' r(X_i) - \theta_0' Q(Z_i)]
\]

\[
\times \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_1}(z) v(X_i) - \eta_v(z) \right] d\hat{\psi}_2(z)
\]

\[
= n_1 h_1^{d/2} \gamma_n O_p(1/\sqrt{n_1}) + n_1 h_1^{d/2} \gamma_n O_p(1/(n_1 h_1^{d/4})) = o_p(1)
\]

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

\[
\frac{1}{n_1} \sum_{i=1}^{n_1} \int K_{h_1}(z) [\xi_i + \gamma_n v(X_i) + (\theta_0 - \hat{\theta}_n)' (Q(Z_i) - \hat{Q}_n(Z_i))
\]

\[
+ (\theta_0 - \hat{\theta}_n)' \hat{Q}_n(Z_i) + (Q(Z_i) - \hat{Q}_n(Z_i))' \theta_0] d\hat{\psi}_2(z)
\]

which, by expanding the second square term in the integrand, equals \( \hat{C}_{n1} + \hat{C}_{n2} \), where \( \hat{C}_{n1} = \frac{1}{n_1^2} \sum_{i=1}^{n_1} \int K_{h_1}(z) \xi_i^2 d\hat{\psi}_2(z) \) and \( \hat{C}_{n2} \) is the remainder. By consistency of \( \hat{\theta}_n \) and the conditional argument on \( \hat{Q}_n(Z_i) \), 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 \( Y_i - \hat{\theta}_n \hat{Q}_n(Z_i) \) and \( Y_j - \hat{\theta}_n \hat{Q}_n(Z_j) \) respectively, one obtains

\[
\hat{\Gamma}_n = \frac{2 h_1^d}{n_1} \sum_{i \neq j} \left[ K_{h_1}(z) K_{h_1}(z) \xi_i \xi_j d\hat{\psi}_2(z) \right]^2 + V_n.
\]
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 $I$ in probability. Therefore,

$$\hat{D}_n = n_1 h_1^{d/2} \hat{I}_n^{-1/2} (T_n - \hat{C}_n) + n_1 h_1^{d/2} (\hat{C}_n - \hat{C}_n) + n_1 h_1^{d/2} (T_n + T_n) + o_P(1),$$

the theorem follows by noting that the second term on the right-hand side of $\hat{D}_n$ 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

