This is to certify that the
dissertation entitled

MINIMUM DISTANCE
MEASUREMENT ERRORS MODEL FITTING

presented by

WEIXING SONG

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Department of Statistics and Probability

Major Professor's Signature

May 12, 2006

Date

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

2/05 p:/CIRC/DateDue.indd-p.1

Minimum Distance Measurement Errors Model Fitting

By

Weixing Song

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

2006

ABSTRACT

Minimum Distance

Measurement Errors Model Fitting

By

Weixing Song

This work proposes a class of minimum distance tests for fitting a parametric regression model to a class of regression functions in the measurement error models. In the errors-in-variables model case, these tests are based on certain minimized L_2 distances between a nonparametric regression function estimator and a deconvolution kernel estimator of the regression function of the parametric model being fitted. In the Berkson model case, these tests are based on certain minimized distances between a nonparametric regression function estimator and the parametric model being fitted. The thesis establishes the asymptotic normality of the proposed test statistics under the null hypothesis and that of the corresponding minimum distance estimators in both cases. 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.

ACKNOWLEDGMENTS

I wish to express my sincere gratitude to my advisor Professor Hira L. Koul for his invaluable guidance. It would have been impossible for me to finish this dissertation without the uncountable number of hours he spent sharing his knowledge and discussing various ideas throughout the study. His general thinking of statistical problem and ways to solve the problem will help my future research.

I would also like to thank Professors Sarat Dass, R.V. Ramamoorthi and Richard Baillie for serving on my guidance committee. Many thanks to Professors Connie Page and Dennis Gilliland for their advice when I was at the consulting service. Finally, I would like to thank the Department of Statistics and Probability for offering me graduate assistantships, and the Graduate School for offering me the Dissertation Completion Fellowship so that I could complete my graduate studies at the Michigan State University.

Last but not the least, I would like to give my thanks to my mother Fuying Song and my wife Xiuqin Bai, whose patient love enabled me to complete this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

In the classical regression model, we use a set of variables, say $d-$dimensional predictor X, to explain the response Y, a one dimensional real random variable, here, both X and Y are observable. But in the real applications, the predictor X is not always observable. To deal with the statistical inference problems in this case, statisticians proposed the so called measurement errors model. In this model, a surrogate of X, say Z, is observed. Then how to investigate the statistical relationships between X and Y based on the data from Z and Y is the main issue in the measurement errors models.

Based on the stochastic structure between X and Z, the measurement errors model usually can be divided into two classes, error models which including the errors-in-variables models in which $Z = X + u$ and the error calibration models in which $Z = \alpha + \beta X + u$, and the Berkson model (or Regression calibration models) in which $X = Z + \eta$, where u, η are measurement errors. About this classification, see Carroll, Rupert and Stefanski (1995) for the details.

The measurement errors regression models have been receiving a continuing attention in the statistical literature over the last century. For some literature reviews on errors-in-variables models, see Gleser (1981), Anderson (1984), Fuller (1987), Bickel and Ritov (1987), Carroll and Hall (1988), Fan (1991a, 1991b), Fan and Truong (1993), Carroll, Rupert and Stefanski (1995), and the references therein. As for the Berkson models, see Rudemo, et al. (1989), Huwang, L. and Huang, Y.H.S. (2000), Wang (2003, 2004) for some literature reviews. Most of the existing literature has focused on the estimation problem. Model checking or lack-of-fit testing problem is not discussed thoroughly. Only some sporadic results on this topic can be found in the literature.

In the errors-in-variables model case, Fuller (1987) discusses a graphic method for lack-of-fit testing of a linear errors in variables regression model. Carroll and Spiegelman (1992) consider the graphic and numerical diagnostics for nonlinearity and heteroscedasticity in linear regression model with errors in variables. Zhu, Song and Cui (2003) considered the lack-of-fit testing in the polynomial regression with errors in variables and constructed a residual-based test of score type, but their method has two limitations. First, the predictor is one dimensional and the regression function under the null hypothesis is polynomial; second, the density function of the predictor is assumed to be known which is generally unrealistic in the real applications. Cheng and Kukush (2004) also addressed the same problem based on so-called adjusted least squares estimators. Few results on the errors in variables regression model checking without imposing strict conditions are available in the literature.

Berkson model has a relatively simpler structure than errors-in-variables model in that the density function of the predictor can be estimated by the usual kernel method. Like the errors-in-variables models, there is a vast literature on the estimation problems about the parameters, but no discussion on the model checking problem for this case.

Many interesting and profound results, on the other hand, are available for the regression model checking problem in the absence of errors in predictor, see, e.g., Eubank and Spiegelman (1990), An and Cheng (1991), Eubank and Hart (1992, 1993), Hart (1997), Stute (1996), Zheng (1996), Stute, Thies, and Zhu (1998), Khmaladze and Koul (2004), among others. For a general discussion on the model fitting in the classical regression case, a good reference is Hart (1997). Stute (1996), Stute, Thies, and Zhu (1998) constructed a test statistic based on certain marked empirical processes. Their simulation results show the testing procedure is quite satisfying, but their procedure can only be used for the one dimensional case. The recent paper of Koul and Ni (2004)(K-N) uses the minimum distance (MD) ideas developed by Wolfowitz (1953, 1954, 1957) to propose tests of lack-of-fit for the regression model without errors in variables. Their work can be used to deal with the multidimensional case. In a finite sample comparison of these tests with some other existing tests, they noted that a member of this class preserves the asymptotic level and has very high power against some alternatives and compared to some other existing lack-of-fit tests. Our work will extend this methodology to the measurement errors model set up.

To be specific, in the classical regression set up, let X, Y be random variables, with X being d-dimensional and Y one dimensional with $E|Y| < \infty$. Let $\mu(x) = E(Y|X = x)$ denote the regression function, and let $\{m_\theta(\cdot) : \theta \in \Theta\}$, $\Theta \subset R^q$, $q \geq 1$, be a given parametric model. The statistical problem of interest here is to test the following hypothesis:

$$H_0 : \mu(x) = m_{\theta_0}(x), \text{ for some } \theta_0 \in \Theta, \text{ and all } x \in \mathcal{I}, \text{ v.s. } H_1 : H_0 \text{ is not true,} \tag{1}$$

where $\mathcal{I}$ is a compact subset of R^d, $d \geq 1$, based on a random sample $(X_i, Y_i); 1 \leq i \leq n$ from the distribution of (X, Y). In the K-N paper, the design is random but observable. Let K, K^* be two possibly different density kernels on $[-1, 1]^d$. For any bandwidth sequence h, let

$$K_h(x) := \frac{1}{h^d} K\Big(\frac{x}{h}\Big), \quad K_{hi}(x) := K_h(x - X_i), \quad \hat{f}_{Xh}(x) := \frac{1}{n}\sum_{i=1}^{N} K^*_{hi}(x).$$

Note that $\hat{f}_{Xh}$ is the kernel estimator of f_X corresponding to the kernel K^*. K-N defines

$$T_n(\theta) := \int_{\mathcal{I}} \Big[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)(Y_i - m_\theta(X_i))\Big]^2 \hat{f}^{-2}_{Xw}(x)\, dG(x), \tag{2}$$

and $\tilde{\theta}_n := \text{argmin}_{\theta \in \Theta} T_n(\theta)$ where $w = w_n \sim (\log n/n)^{1/(d+4)}$, and $h = h_n$ is a bandwidth depending on the sample size n. For some crucial technical reasons, different bandwidths h and w are chosen. The integration measure G is a σ-finite measure on $\mathbb{R}^d$ which may be chosen to make the test statistic to have good power. Under the null hypothesis and some regular conditions, the consistency and asymptotic normality of $\tilde{\theta}_n$ are proved. They also showed that the asymptotic null distribution of $nh_n^{d/2}\tilde{\Gamma}_n^{-1/2}(T_n(\tilde{\theta}_n) - \tilde{C}_n)$ is standard normal, where

$$\tilde{C}_n := \frac{1}{n^2}\sum_{i=1}^{n} \int_{\mathcal{I}} K^2_{hi}(x)\varepsilon_i^2 \tilde{f}^{-2}_{w_n}(x) dG(x), \qquad \hat{\varepsilon}_i = Y_i - m_{\tilde{\theta}_n}(X_i)$$

$$\tilde{\Gamma}_n := \frac{h^d}{n^2}\sum_{i \neq j}^{n} \Big(\int_{\mathcal{I}} K_{hi}(x)K_{hj}(x)\hat{\varepsilon}_i\hat{\varepsilon}_i \tilde{f}^{-2}_{w_n}(x) dG(x)\Big)^2.$$

Thus, the test that rejects the null hypothesis whenever $nh_n^{d/2}\tilde{\Gamma}_n^{-1/2}|T_n(\tilde{\theta}_n)-\tilde{C}_n| > z_{\alpha/2}$, is of the asymptotic size α, where z_α is $(1-\alpha)$th percentile of the standard normal distribution. Unlike in other related papers, K-N do not need the null regression function to be twice continuously differentiable in the parameter vector. The asymptotic normal distribution of $\tilde{\theta}_n$ and $T_n(\tilde{\theta}_n)$ were made feasible by recognizing to use different band widths for the estimation of the numerator and denominator in the nonparametric regression function estimation. A consequence of the above asymptotic normality result is that at least for large samples one need not use any resampling method to implement these tests.

In this thesis, we will discuss, in the measurement errors setup, how to develop testing procedures for the following hypothesis:

$$H_0 : \mu(x) = m_{\theta_0}(x), \text{ for some } \theta_0 \in \Theta, \text{ and all } x, \text{ v.s. } H_1 : H_0 \text{ is not true.} \quad (3)$$

From K-N's procedure, we know that if we want to use the minimum distance method, a kernel-type regression estimator must be constructed, but this in turn implies that we must find an estimator for the density function of the predictor. This is not a problem in the classical regression case in that the predictor is observable. But in the measurement errors models case, the predictor X is not observable, to adapt K-N's minimum distance method, the above procedure needs some modification.

We now briefly describe the modification needed for the errors-in-variables model. It consists of two steps:

Step 1. Hypothesis Change: The hypothesis (3) concerns with the regression function $\mu(x)$ which depends on the true predictor, but the true predictor is not observable. By recognizing that $\nu(z) := E(Y|Z = z) = E(\mu(X)|Z = z)$, we consider the new regression model $Y = \nu(z) + \zeta$, where the error ζ is uncorrelated with Z and has mean 0. The problem of testing for H_0 can be transformed to test for $\nu(z) = \nu_{\theta_0}(z)$, where $\nu_\theta(z) := E(m_\theta(X)|Z = z)$. Since Z is observable, so we can construct a classic kernel estimator for the new regression function $\nu(z)$.

Step 2. Deconvolution Kernel Density Estimator: The minimum distance will be constructed based on the classical kernel estimator of $\nu(z)$ and a proper estimator of $\nu_\theta(z) := E(m_\theta(X)|Z = z)$ under the null hypothesis. Note that, under the null hypothesis,

$$\nu_\theta(z) = \frac{\int m_\theta(x) f_X(x) f_u(z - x)dx}{\int f_X(x) f_u(z - x)dx}.$$

To estimate this quantity for given θ, we need an estimator of f_X. In this connection the deconvolution kernel density estimators are found to be useful here. Putting the deconvolution kernel density estimator of f_X into the above expression, we construct the deconvolution kernel estimator of $\nu_\theta(z)$.

To obtain the asymptotic distribution of the test statistic, we need to consider the asymptotic behavior of the deconvolution kernel estimator of $\nu_\theta(z)$. Although we extend Stefanski and Carroll (1991)'s result to a more general case, the convergence rate of the deconvolution kernel estimator is still slower than the classical kernel estimator. This brings us some difficulty in proving the technical results. To overcome this difficulty, we adopt the sample splitting technique. The sample splitting scheme required in the proof is not so realistic in certain cases, but the simulation results show that the test statistic behaves good if we do not follow the sample splitting scheme.

In the Berkson model case, things become relatively easy. From $X = Z + u$ and the independence between Z and u, $E(Y|Z)$ is known under the null hypothesis except the parameter. After changing the hypothesis, the testing procedure can be developed in the similar way as done in the errors-in-variables model case.

This thesis is organized as follows. Chapter 1 discusses the model fitting for errors-in-variables model in which the regression function under the null hypothesis is linear in parameters. Theorem 1.3.1 gives the asymptotic distribution of the underlying parameter estimator. Theorem 1.4.1 gives the asymptotic distribution of the minimized distance under the null hypothesis. A test statistic therefore can be constructed based on this theorem. Several simulations are present in section 1.5. Some problems related to the sample allocation scheme and the results about the general errors-in-variables models are discussed the subsequent section.

Chapter 2 discusses the minimum distance model fitting in Berkson model. Corollary ?? and Theorem 2.3.1 state the consistency of the underlying parameter estimators, Theorem 2.4.1 and Theorem 2.5.1 give the asymptotic distribution of the parameter estimator and the minimized distance under the null hypothesis. A test statistic therefore can be constructed based on the Theorem 2.5.1. Simulations conducted in section 2.6 show the testing procedure is quite satisfactory.

CHAPTER 1

Minimum Distance Errors-in-Variables Model Fitting

1.1 Introduction

The findings in the classical regression case motivate one to look for tests of lack-of-fit in the presence of the errors in variables based on the above minimized distances. Since the predictor in errors in variables models are unobservable, clearly the above procedures need some modification. To be specific, in an errors in variables regression model of interest here, one observes Z_i, Y_i obeying the model

$$Y_i = \mu(X_i) + \varepsilon_i, \qquad Z_i = X_i + u_i, \qquad 1 \le i \le n, \tag{1.1}$$

where X_i's are the unobservable d-dimensional random design variables. We additionally assume that $(X_i, \varepsilon_i, u_i, Z_i, Y_i)$, $i = 1, 2, \cdots, n$, are i.i.d. copies of $(X, \varepsilon, u, Z, Y)$. The r.v.'s (X, u, ε) are assumed to be mutually independent, with u being d-dimensional, and ε being 1-dimensional r.v.'s, $E(\varepsilon) = 0$, $E(u) = 0$, and their marginal distributions having densities f_X, f_u, and f_ε, respectively. For the sake of identifiability, the density f_u is assumed to be known. This is a common and standard assumption in the literature of the errors in variables regression models. The densities f_X and f_ε need not be known. The problem of interest in this chapter is to develop tests for the hypothesis

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

in the model (1.1).

A way for constructing tests here is to first recognize that the independence of X and ε and $E(\varepsilon) = 0$ imply that $\nu(z) := E(Y|Z = z) = E(\mu(X)|Z = z)$. Thus one can consider the new regression model $Y = \nu(z) + \zeta$, where the conditional expectation $E(\zeta|Z) = 0$, hence ζ is uncorrelated with Z. The problem of testing for H_0 is now transformed to test for $\nu(z) = \nu_{\theta_0}(z)$, where $\nu_\theta(z) := \theta^T E(r(X)|Z = z)$. Note that for any z for which $f_Z(z) > 0$, we have

$$\nu(z) = \frac{\int \mu(x) f_X(x) f_u(z-x)dx}{\int f_X(x) f_u(z-x)dx}. \tag{1.3}$$

From (1.3) one sees that if f_X is known then f_Z is known and hence ν_θ is known except for θ. Let $Q(z) := E(r(X)|Z = z)$. Therefore a modification of K-N's procedure in this case is as follows. Define

$$\begin{aligned} \bar{T}_n(\theta) &:= \int \Big[\frac{1}{n f_Z(z)} \sum_{i=1}^n K_{hi}(z)(Y_i - \theta^T Q(Z_i))\Big]^2 dG(z), \quad \theta \in \mathbb{R}^q, \\ \bar{\theta}_n &:= \arg\min_{\theta \in \mathbb{R}^q} \bar{T}_n(\theta), \end{aligned}$$

Here h is a bandwidth depending only on n and $K_{hi}(z)$ is redefined as $K((z - Z_i)/h)/h^d$ for any kernel function K and bandwidth h . Then we may use $\bar{\theta}_n$ to estimate θ, and construct the test based on the $\bar{T}_n(\bar{\theta}_n)$. Unfortunately, f_X is generally not known and hence f_Z and $Q(z)$ are unknown. This makes the above procedures infeasible. To construct the test statistic, one needs estimators for f_Z and $Q(z)$. In this connection the deconvolution kernel density estimators are found to be useful here.

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

$$
\begin{aligned}
L_h(x) &:= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\exp(-\mathbf{i}t\cdot x)\frac{\phi_L(t)}{\phi_u(t/h)}dt, \qquad \mathbf{i} := (-1)^{1/2}, \quad x \in \mathbb{R}, \quad (1.4)\\
\hat{f}_{Xh}(x) &:= \frac{1}{nh^d}\sum_{i=1}^{n} L_h\Big(\frac{x-Z_i}{h}\Big), \qquad x \in \mathbb{R}^d.
\end{aligned}
$$

The above L_h is called the deconvolution kernel function, while $\hat{f}_{Xh}$ is called deconvolution kernel density estimator of f_X, cf. Masry (1993), Carroll, Ruppert and Stefanski (1995).

Note that $Q(z)$ is equal to $R(z)/f_Z(z)$, where $R(z) = \int r(x)f_X(x)f_u(z-x)dx$, and $f_Z(z) = \int f_X(x)f_u(z-x)dx$. Then one can estimate $Q(z)$ by

$$
\hat{Q}_n(z) = \hat{R}_n(z)/\tilde{f}_{Zh}(z), \tag{1.5}
$$

where $\hat{R}_n(z) = \int r(x)\hat{f}_{Xh}(x)f_u(z-x)dx$, $\quad \tilde{f}_{Zh}(z) = \int \hat{f}_{Xh}(x)f_u(z-x)dx$. At this point, it is worth mentioning that, by the definition of L_h and a direct calculation, one can show $\tilde{f}_{Zh}$ is nothing but the classical kernel estimator of f_Z with kernel L and bandwidth h. That is, $\tilde{f}_{Zh}(z) = \sum_{i=1}^n L((z-Z_i)/h)/nh^d$.

Our proposed inference procedures will be based on the analogs of T_n where $Q(z)$ is replaced by the above estimator $\hat{Q}_n$, and f_Z is replaced by a kernel estimator.

A very important question related to the above procedure is the following: Are the two hypotheses, $H_{10} : \mu(x) = \theta_0^T r(x)$, for some θ_0 and all x, and $H_{20} : \nu(z) = \theta_0^T E(r(x)|Z = z)$, for some θ_0 and all z, equivalent? The answer is negative in general, but in some special case, these two hypotheses are equivalent. See a general discussion in Section 1.6.2

The large sample behavior of the deconvolution kernel density estimators strongly depends on the smoothness of the distribution of measurement error u. Using the terms from Fan and Truong (1993), a distribution is called *ordinary smooth* if the tails of its characteristic function decay to 0 at an algebraic rate; it is called *super smooth* if its characteristic function has tails approaching 0 exponentially fast. As Masry (1993) showed, the local and global rates of convergence of the sequences of deconvolution kernel density estimators are slower than that of the classical kernel density estimators. Moreover, these convergence rates are much slower in the super smooth cases than in the ordinary smooth cases. But Stefanski and Carroll (1991) shows that in the one dimensional case with $r(x) = x$, for estimating $E(X|Z = z)$ by $\hat{Q}_n(z)$, faster rates are obtainable. For example, in the case of normal measurement error, the mean squared error rate of convergence of $\hat{f}_{Xh_3}$ to f_X is of order $(\log(n))^{-2}$, while the convergence rate of $\hat{Q}_n(z)$ to $E(X|Z = z)$ is of order $n^{-4/7}$. Even so, the convergence rate is still slower than the mean squared error convergence rate of the classic kernel estimator, which is $n^{-4/5}$ in the one dimensional case. This creates extra difficulty when considering the asymptotic behaviors of the analogs of the corresponding MD estimators and test statistics. In fact, if we base the estimators of f_X, hence $Q(z)$ and the other quantities on the same sample, 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 the remaining quantities. The sample size allocation scheme is stated in section 2.

To be precise, let

$$
\begin{aligned}
\hat{f}_{Zh_2}(z) &:= \sum_{i=1}^{n_1} K^*_{h_2 i}(z)/n_1, \quad \hat{f}_{Xw}(x) := \sum_{j=n_1+1}^{n} L_w((x - Z_j)/w)/n_2 w^d, \\
\hat{R}_{n_2}(z) &:= \int r(x) \hat{f}_{Xw_1}(x) f_u(z-x)dx, \quad \tilde{f}_{Zw_2}(z) := \int \hat{f}_{Xw_2}(x) f_u(z-x)dx, \\
\hat{Q}_{n_2}(z) &:= \hat{R}_{n_2}(z)/\tilde{f}_{Zw_2}(z),
\end{aligned}
$$

where h_1, h_2 depend on n_1, and w_1 and w_2 depend on n_2. Now define

$$
\begin{aligned}
M_n(\theta) &:= \int_{\mathcal{I}} \Big[\frac{1}{n_1 \hat{f}_{Zh_2}(z)} \sum_{i=1}^{n_1} K_{h_1 i}(z)(Y_i - \theta^T \hat{Q}_{n_2}(Z_i))\Big]^2 dG(z), \\
\hat{\theta}_n &:= \text{arginf}_{\theta \in \mathbb{R}^q} M_n(\theta).
\end{aligned} \tag{1.6}
$$

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

$$
\begin{aligned}
\hat{\zeta}_i &:= Y_i - \hat{\theta}_n^T \hat{Q}_{n_2}(Z_i), \quad \hat{C}_n := n_1^{-2} \sum_{i=1}^{n_1} \int K^2_{h_1 i}(z) \hat{\zeta}_i^2 d\hat{\psi}_{h_2}(z), \\
\hat{\Gamma}_n &:= 2h_1^d n_1^{-2} \sum_{i \neq j=1}^{n_1} \Big(\int K_{h_1 i}(z) K_{h_1 j}(z) \hat{\zeta}_i \hat{\zeta}_j d\hat{\psi}_{h_2}(z)\Big)^2, \\
d\hat{\psi}_{h_2}(z) &:= \frac{dG(z)}{\hat{f}^2_{Zh_2}(z)}.
\end{aligned} \tag{1.7}
$$

We prove that the asymptotic null distribution of the normalized test statistic $n_1 h_1^{d/2} \hat{\Gamma}_n^{-1/2}(M_n(\hat{\theta}_n) - \hat{C}_n)$ is standard normal. Consequently, the test that rejects H_0 whenever $n h_1^{d/2} \hat{\Gamma}_n^{-1/2} |M_n(\hat{\theta}_n) - \hat{C}_n| > z_{\alpha/2}$ is of the asymptotic size α.

This chapter is organized as follows. Section 2 states the needed assumptions. A multidimensional extension of Lemma A.1 in Stefanski and Carroll (1991) is also proved there, together with some other needed results. Section 4 proves the asymptotic normality of the MD estimator. The asymptotic normality of the MD test statistic is discussed in section 5. Section 6 includes some results from a finite sample simulation study.

In the sequel, c will denote the generic finite positive constant whose value depends on the context. For any vector b, b^T denotes its transpose. For any function f, we will use $\dot{f}$, $\ddot{f}$ to denote the first and the second derivative with respect to its argument. The convergence in distribution is denoted by $\Longrightarrow$, and $N_d(a, B)$ stands for the d-dimensional normal distribution with mean vector a and covariance matrix B and E_{S_1} denotes the conditional expectation given the subsample S_1. The integration with respect to the G-measure is understood to be over the compact set $\mathcal{I}$.

1.2 Assumptions

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

(e1) The random variables $\{(Z_i, Y_i) : Z_i \in \mathbb{R}^d, Y_i \in \mathbb{R}, i = 1, 2, \cdots, n\}$ from (1.1) are i.i.d. with the conditional expectation $\nu(z) = E(Y|Z = z)$ satisfying $\int \nu^2 dG < \infty$, where G is a σ-finite measure on $\mathbb{R}^d$.

(e2) $0 < \sigma_{\varepsilon}^2 = E\varepsilon^2 < \infty$, $E\|r(X)\|^2 < \infty$, and the function $\delta^2(z) = E[\theta_0^T r(X) - \theta_0^T Q(Z))^2|Z = z]$ is a.s. (G) continuous on $\mathcal{I}$.

(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 of the d-dimensional r.v. X, and its all possible first and second derivatives are continuous and bounded.

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

$$\begin{aligned} \mathcal{I}_\delta &= \{y \in \mathbb{R}^d : \max_{1 \le j \le d} |y_j - z_j| \le \delta, \\ & \qquad y = (y_1, \cdots, y_d)^T, z = (z_1, \cdots, z_d)^T, z \in \mathcal{I}\}, \end{aligned} \tag{1.8}$$

(g) G has a continuous Lebesgue density g.

About the null model we need to assume the following:

(m1) There exists a positive continuous function $J(z)$, such that as $\|t\| \to \infty$,

$$\|t\|^{-\alpha} \left\| \frac{\int (r(z-x) - r(z)) \exp(-\mathbf{i}t^T x) f_u(x) dx}{\phi_u(t)} \right\| \le J(z),$$

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

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

About the kernel functions, we assume:

(ℓ) The kernel function L is a density, symmetric around the origin, $\|t\|^\alpha |\phi_L(t)| < \infty$, for all $t \in \mathbb{R}^d$; Moreover, $\int \|v\|^2 L(v)dv < \infty$ and $\int \|t\|^\beta |\phi_L(t)|dt < \infty$ for $\beta = 0, \alpha$, with α as in (m1).

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

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

(h1) $h_1 \sim n_1^a$, where $a < \min(1/2d, 4/d(d+4))$.

(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,

$$\left\| \frac{\int (r(z-x) - r(z)) \exp(-it^T x) f_u(x)dx}{\phi_u(t)} \right\|$$

$$= \left\| \int x \exp(-it^T x) f_u(x)dx \right\| \cdot \exp\left(\frac{t^T \Sigma_u t}{2}\right) = \left\| \frac{\partial \phi_u(t)}{i\partial t} \right\| \cdot \exp(t^T \Sigma_u t/2) \le c\|t\|,$$

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

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

$$\begin{aligned} & \left|\frac{\int (r(z-x)-r(z))\exp(-\mathbf{i}tx) f_u(x)dx}{\phi_u(t)}\right| \\ = & \left|\int x\exp(-\mathbf{i}tx) f_u(x)dx\right| / |\phi_u(t)| \\ = & \left|\frac{\partial \phi_u(t)}{i\partial t}\right| / |\phi_u(t)| = \frac{c|t|}{1+\sigma_u^2 t^2/2}, \end{aligned}$$

with c now depending only on σ_u^2. Hence as $|t| \to \infty$, (m1) holds for $\alpha = 0$ and, $J(z) = c$.

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

$$\begin{aligned} & \left|\int (r(z-x)-r(z))\exp(-\mathbf{i}tx) f_u(x)dx\right| \\ = & \left|\int (e^{z-x}-e^z)\exp(-\mathbf{i}tx) f_u(x)dx\right| \\ \leq & e^z\left[\left|\int e^x e^{itx} f_u(x)dx\right| + |\phi_u(t)|\right] \leq ce^z|\phi_u(t)|, \end{aligned}$$

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

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

In the case of $r(x) = x$ and $d = 1$, Stefanski and Carroll (1991) obtain the following results:

$$\{E\hat{R}_{n_2}(z) - R(z)\}^2 \leq cw_1^4(1+z^2), \quad Var(\hat{R}_{n_2}(z)) \leq c(n_2 w_1)^{-1}(w_1^{-2\alpha} + z^2),$$

for all $z \in \mathbb{R}$, and under the assumptions (i) f_X, $\dot{f}_X$ and $\ddot{f}_X$ are continuous and bounded; (ii) $\int |\phi_u(t)|dt < \infty$; (iii) as $|t| \to \infty$, $|\dot{\phi}_u(t)/\phi_u(t)| = o(|t|^\alpha)$, for some

$\alpha \geq 0$; (iv) $n_2 \to \infty$, and $w_1 \to 0$. The kernel function L used in the deconvolution estimator is assumed to be four-times continuously differentiable, compactly supported and real valued. The following lemma is a multidimensional extension of the above results which will be frequently used in the sequel.

Lemma 1.2.1 *Suppose $d \geq 1$, and (f1), (u), (m1), (h1) hold. Then for any $z \in \mathbb{R}^d$,*

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

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

Proof. 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 $f_X(x - vw_1)$ has a Taylor expansion up to the second order, $f_X(x - vw_1) = f_X(x) - w_1 v^T \dot{f}_X(x) + w_1^2 v^T \ddot{f}_X(a(x, v))v/2$. Hence

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

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 $\int v^T L(v)dv = 0$, while the third term is bounded above by $cI(z)$ by assumption (f1), where c is a positive constant depending only on the kernel function L. Therefore, the first claim in the lemma holds.

Note that $\hat{R}_{n_2}(z) - E\hat{R}_{n_2}(z)$ is an average of i.i.d. centered random vectors. A routing calculation shows that

$$E\|\hat{R}_{n_2}(z) - E\hat{R}_{n_2}(z)\|^2 \leq \frac{1}{n_2 w_1^{2d}} E\Big\| \int r(x) L_{w_1}((x-z)/w_1) f_u(z-x)dx \Big\|^2$$

by using the fact that the variance is bounded above by the second moment. Let $D(t,z) = \int r(x) f_u(z-x)\exp(-\mathbf{i}t^T x)dx$. By the definition of the deconvolution kernel L_b, it follows that

$$\begin{aligned} &\frac{1}{w_1^{2d}} E\Big\| \int r(x) L_{w_1}((x-z)/w_1) f_u(z-x)dx \Big\|^2 \\ = &\iint \frac{D^T(t,z)D(s,z)\phi_L(tw_1)\phi_L(sw_1)\phi_X(t+s)\phi_u(t+x)}{(2\pi)^{2d}\phi_u(t)\phi_u(s)} dsdt. \end{aligned}$$

By changing variable, $D(t,z) = \exp(-\mathbf{i}t^T z)\int r(z-x) f_u(x)\exp(\mathbf{i}t^T x)dx$. Adding and subtracting $r(z)$ from $r(z-x)$ in the integrand, we obtain

$$D(t,z) = \exp(-\mathbf{i}t^T z)\phi_u(z)\Big[r(z) + \frac{\int (r(z-x) - r(z)) f_u(x)\exp(\mathbf{i}t^T x)dx}{\phi_u(t)}\Big].$$

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

$$\begin{aligned} &\frac{c\|r(z)\|^2}{n_2} \iint |\phi_L(tw_1)\phi_L(sw_1)\phi_u(t+s)|dtds \\ &+ \frac{cJ(z)\|r(z)\|}{n_2} \iint (\|t\|^\alpha + \|s\|^\alpha)|\phi_L(tw_1)\phi_L(sw_1)\phi_u(t+s)|dtds \\ &+ \frac{cJ^2(z)}{n_2} \iint \|t\|^\alpha \|s\|^\alpha |\phi_L(tw_1)\phi_L(sw_1)\phi_u(t+s)|dtds. \end{aligned} \tag{1.9}$$

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

$$\begin{aligned} &\iint \|t\|^p \|s\|^m |\phi_L(tw_1)\phi_L(sw_1)\phi_u(t+s)|dtds \\ \leq\ &w_1^{-p-m-2d} \iint \|t\|^p \|s\|^m |\phi_L(t)\phi_L(s)\phi_u((t+s)/w_1)|dtds \end{aligned}$$

$$
\begin{aligned}
&\leq cw_1^{-p-m-2d} \iint \|s\|^m |\phi_L(s)||\phi_u((t+s)/w_1)| dt ds \\
&= cw_1^{-p-m-2d} \int \|s\|^m |\phi_L(s)| \Big(\int |\phi_u((t+s)/w_1)| dt \Big) ds \\
&= cw_1^{-p-m-d} \int \|s\|^m |\phi_L(s)| ds \cdot \int |\phi_u(t)| dt = cw_1^{-p-m-d}.
\end{aligned}
$$

The second claim in the lemma follows from (1.9) by using the above inequality. □

By the usual bias and variance decomposition of mean square error, the following inequality is a direct consequence of Lemma 1.2.1,

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

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

$$E\|\hat{R}_{n_2}(z) - R(z)\|^2 \leq cn_2^{-\frac{4}{d+2\alpha+4}} (I^2(z) + J^2(z) + \|r(z)\|^2). \tag{1.10}$$

In the sequel, we will write

$$\mathcal{T}(z) := I^2(z) + J^2(z) + \|r(z)\|^2. \tag{1.11}$$

The following lemma we will be used repeatedly, which along with its proof appears as Theorem 2.2 part (2) in Bosq (1998). 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 1.2.2 *Let $\hat{f}_Z$ be the kernel estimator with a kernel K which satisfies a Lipschitz condition and bandwidth h. If f_Z is twice continuously differentiable, and the bandwidth h is chosen to be $c_n(\log(n)/n)^{1/(d+4)}$, where $c_n \to c > 0$, then*

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

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

1.3 Asymptotic normality of $\hat{\theta}_n$

Recall the definitions in (1.6). Because the null model is linear in θ, so the minimizer $\hat{\theta}_n$ has an explicit form obtained by setting the derivative of $M_n(\theta)$ with respect to θ equal to 0, which gives the equation

$$\begin{aligned}
&\int_{\mathcal{I}} \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\hat{Q}_{n_2}(Z_i)\cdot \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\hat{Q}_{n_2}^T(Z_i)d\hat{\psi}_{h_2}(z)\cdot \hat{\theta}_n \\
= &\int_{\mathcal{I}} \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)Y_i\cdot \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\hat{Q}_{n_2}(Z_i)d\hat{\psi}_{h_2}(z).
\end{aligned}$$

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

$$\begin{aligned}
&\int_{\mathcal{I}} \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\hat{Q}_{n_2}(Z_i)\cdot \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\hat{Q}_{n_2}^T(Z_i)d\hat{\psi}_{h_2}(z)\cdot (\hat{\theta}_n-\theta_0) \\
= &\int_{\mathcal{I}} \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)(Y_i-\theta_0^T\hat{Q}_{n_2}(Z_i))\cdot \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\hat{Q}_{n_2}(Z_i)d\hat{\psi}_{h_2}(z).
\end{aligned} \tag{1.12}$$

The above explicit relation between $\hat{\theta}_n-\theta_0$ and the 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. To keep the exposition concise, let

$$\begin{aligned}
U_{n_1}(z) &:= \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)(Y_i-\theta_0^T Q(Z_i)), \qquad (1.13)\\
D_n(z) &:= \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)(\hat{Q}_{n_2}(Z_i)-Q(Z_i)).
\end{aligned}$$

$$\mu_{n_1}(z) := \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)Q(Z_i), \quad \Delta_{n_1}(z) := \frac{1}{\hat f^2_{Zh_2}(z)} - \frac{1}{f^2_Z(z)}$$

The main result in this section is the following theorem:

Theorem 1.3.1 *Suppose H_0, (e1), (e2), (e3), (u), (f1), (f2), (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*

$$\Sigma_0 = \int Q(z)Q^T(z)dG(z), \qquad \Sigma = \int \frac{\tau^2(z)Q(z)Q^T(z)g^2(z)}{f(z)}dz,$$

and $\tau^2(z) = \sigma^2_\varepsilon + \delta^2(z)$, where σ^2_ε, and $\delta^2(z)$ are defined as in (e2).

Proof. It suffices to show that the matrix before $\hat\theta_n - \theta_0$ on the left hand side of (1.12) converges to Σ_0 in probability, and $\sqrt{n_1}$ times the right hand side of (1.12) is asymptotically normal with mean vector 0 and covariance matrix Σ.

Consider the second claim first. Adding and subtracting $\theta_0^T Q(Z_i)$ from $Y_i - \theta_0^T \hat Q_{n_2}(Z_i)$ 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 $1/\hat f^2_{Zh_2}(z)$ by $1/\hat f^2_{Zh_2}(z) - 1/f^2_Z(z) + 1/f^2_Z(z) := \Delta_{n_1}(z) + 1/f^2_Z(z)$, $\sqrt{n_1}$ times the right hand side of (1.12) can be written as the sum of the following eight terms:

$$
\begin{aligned}
S_{n1} &:= \sqrt{n_1}\int U_{n_1}(z)D_n(z)\Delta_{n_1}(z)dG(z), \quad S_{n2} := \sqrt{n_1}\int U_{n_1}(z)D_n(z)d\psi(z),\\
S_{n3} &:= \sqrt{n_1}\int U_{n_1}(z)\mu_{n_1}(z)\Delta_{n_1}(z)dG(z), \quad S_{n4} := \sqrt{n_1}\int U_{n_1}(z)\mu_{n_1}(z)d\psi(z),\\
S_{n5} &:= -\sqrt{n_1}\int D_n(z)D_n^T(z)\Delta_{n_1}(z)dG(z)\,\theta_0,\\
S_{n6} &:= -\sqrt{n_1}\int D_n(z)D_n^T(z)d\psi(z)\,\theta_0,\\
S_{n7} &:= -\sqrt{n_1}\int D_n(z)\mu_{n_1}^T(z)\Delta_{n_1}(z)dG(z)\,\theta_0,\\
S_{n8} &:= -\sqrt{n_1}\int D_n(z)\mu_{n_1}^T(z)d\psi(z)\,\theta_0.
\end{aligned}
$$

Among these terms, S_{n4} is asymptotically normal with mean vector 0 and covariance matrix Σ. The proof uses Lindeberg-Feller central limit theorem, and the arguments are exactly the same as in K-N with $m_{\theta_0}(X_i)$ and $\dot{m}_{\theta_0}(X_i)$ there replaced by $\theta_0^T Q(Z_i)$ and $Q(Z_i)$ here, respectively. The proof is omitted. All the other seven terms are of the order $o_P(1)$. Since the proofs are similar, only $S_{n8} = o_P(1)$ will be shown below for the sake of brevity. We note that by using a similar method as in K-N, we can show $U_{n_1}(z)$ is $O_P(1/\sqrt{n_1 h_1^d})$, which is used in proving $S_{nl} = o_P(1)$ for $l = 1, 2, 3$.

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

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

then $\mathcal{I}_{h_1}$ is a compact set in $\mathbb{R}^d$, and $K_{h_1 i} = 0$ if $Z_i \notin \mathcal{I}_{h_1}$. Hence, without loss of generality, we can assume all $Z_i \in \mathcal{I}_{h_1}$. Since f_Z is bounded from below on the compact set $\mathcal{I}_{\delta_0}$ by assumption (f2) and $\mathcal{I}_{h_1} \subset \mathcal{I}_{\delta_0}$ for n_1 large enough, so from assumption (w2), Lemma 1.2.2, we obtain

$$\sup_{z \in \mathcal{I}_{h_1}} \left| \frac{f_Z(z)}{\tilde{f}_{Zw_2}(z)} - 1 \right| = o\Big((\log_k n_2)\Big(\frac{\log n_2}{n_2}\Big)^{\frac{2}{d+4}}\Big) \quad \text{a.s.,} \tag{1.14}$$

$$\sup_{z \in \mathcal{I}_{h_1}} \left| \frac{f_Z(z)}{\tilde{f}_{Zw_2}(z)} \right| = O_P(1).$$

Secondly, we have the following inequality:

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

$$+\left|\frac{f_Z(Z_i)}{\tilde{f}_{Zw_2}(Z_i)} - 1\right| \cdot \|Q(Z_i)\|. \tag{1.15}$$

Recall the definition of S_{n8}. We have

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

From (1.15) and (1.14), this upper bound satisfies

$$\sqrt{n_1} \cdot Op(1) \cdot A_{n_1 1} + \sqrt{n_1} \cdot o\Big((\log_k n_2)\Big(\frac{\log n_2}{n_2}\Big)^{\frac{2}{d+4}}\Big) \cdot A_{n_1 2}, \tag{1.16}$$

where

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

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

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

Note that

$$E\int \Big[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\|\hat{R}_{n_2}(Z_i) - R(Z_i)\|\Big]^2 d\psi(z)$$

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

$$= \int E\Big(\frac{1}{n_1^2}\sum_{i,j=1}^{n_1} K_{h_1 i}(z)K_{h_1 j}(z)E_{S_1}(\|\hat{R}_{n_2}(Z_i) - R(Z_i)\| \cdot \|\hat{R}_{n_2}(Z_j) - R(Z_j)\|)\Big)d\psi(z).$$

By the Cauchy-Schwarz again,

$$\begin{aligned} & E_{S_1}(\|\hat{R}_{n_2}(Z_i) - R(Z_i)\|\|\hat{R}_{n_2}(Z_j) - R(Z_j)\|)\big) \\ \leq\ & (E_{S_1}\|\hat{R}_{n_2}(Z_i) - R(Z_i)\|^2)^{1/2}(E_{S_1}\|\hat{R}_{n_2}(Z_j) - R(Z_j)\|^2)^{1/2}, \end{aligned}$$

which in turn, from the independence of the subsamples S_1 and S_2, the choice of bandwidth w_1, and (1.10), is bounded above by $cn_2^{-4/(d+2\alpha+4)}\mathcal{T}^{1/2}(Z_i)\mathcal{T}^{1/2}(Z_j)$, where $\mathcal{T}$ is defined in (1.11). So

$$\begin{aligned} & E\int\Big[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\|\hat{R}_{n_2}(Z_i) - R(Z_i)\|\Big]^2 d\psi(z) \\ \leq\ & cn_2^{-\frac{4}{d+2\alpha+4}}\int E\Big(\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\mathcal{T}^{1/2}(Z_i)\Big)^2 d\psi(z). \end{aligned}$$

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

$$\begin{aligned} & \int\Big(\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\mathcal{T}^{1/2}(Z_i)\Big)^2 d\psi(z) = O_P(1), \qquad (1.17) \\ & \int\Big(\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\|Q(Z_i)\|\Big)^2 d\psi(z) = O_P(1). \end{aligned}$$

Finally, from (1.16), we obtain

$$\|S_{n8}\| \leq \sqrt{n_1}\cdot O_P(n_2^{-2/(d+2\alpha+4)}) + \sqrt{n_1}\cdot o_P\Big((\log_k n_2)\Big(\frac{\log n_2}{n_2}\Big)^{\frac{2}{d+4}}\Big)$$

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 (1.12) converges to Σ_0 in probability. Adding and subtracting $Q(Z_i)$ from $\hat Q_{n_2}(Z_i)$, this matrix can be written as the sum of the following eight terms:

$$\begin{aligned}
T_{n1} &:= \int D_n(z)D_n^T(z)\Delta_{n_1}(z)dG(z), T_{n2} := \int D_n(z)\mu_{n_1}^T(z)\Delta_{n_1}(z)dG(z),\\
T_{n3} &:= \int \mu_{n_1}(z)D_n^T(z)\Delta_{n_1}(z)dG(z), T_{n4} := \int \mu_{n_1}(z)\mu_{n_1}^T(z)\Delta_{n_1}(z)dG(z),\\
T_{n5} &:= \int D_n(z)D_n^T(z)d\psi(z), \qquad T_{n6} := \int D_n(z)\mu_{n_1}^T(z)d\psi(z),\\
T_{n7} &:= \int \mu_{n_1}(z)D_n^T(z)d\psi(z), \qquad T_{n8} := \int \mu_{n_1}(z)\mu_{n_1}^T(z)d\psi(z).
\end{aligned}$$

Notice the connection between T_{n1} and S_{n5}, T_{n2}, T_{n3} and S_{n7}, T_{n5} and S_{n6}, T_{n6}, T_{n7} and S_{n8}. By using similar argument as above, we can verify that $T_{nl} = o_P(1)$ for $l = 1,2,3,4,5,6,7$. From (1.14), and the second fact in (1.17), T_{n4} is also of the order of $o_P(1)$. Finally, employing similar method as in K-N, we can show T_{n8} converges to Σ_0 in probability. Thereby proving the theorem. □

1.4 Asymptotic normality of the minimized distance

This section contains a proof of the asymptotic normality of the minimized distance $M_n(\hat\theta_n)$. To state the result precisely, the following notations are needed:

$$\begin{aligned}
\xi_i &:= Y_i - \theta_0^T Q(Z_i), \quad \zeta_i := Y_i - \theta_0^T \hat Q_{n_2}(Z_i),\\
\tilde C_n &:= n_1^{-2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)\zeta_i^2 d\psi(z), \quad \tilde M_n(\theta_0) := \int \Big[n_1^{-1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\zeta_i\Big]^2 d\psi(z),
\end{aligned}$$

$$\Gamma := 2\int (\tau^2(z))^2 g(z)d\psi(z) \cdot \int \Big[\int K(u)K(u+v)du\Big]^2 dv.$$

where $\tau^2(z)$ is as in Theorem 1.3.1.

The main result proved in this section is the following:

Theorem 1.4.1 *Suppose If H_0, (e1), (e2), (e4), (u), (f1), (f2), (m1), (m2), (ℓ), (n), (h1), (h2) ,(w1) and (w2) hold, then $n_1 h_1^{d/2} \hat{\Gamma}_n^{-1/2}(M_n(\hat{\theta}_n) - \hat{C}_n) \Longrightarrow N(0,1)$, where $\hat{C}_n$, $\hat{\Gamma}_n$ are as in (1.7).*

The proof of this theorem is facilitated by the following five lemmas:

Lemma 1.4.1 *If H_0, (e1), (e2), (e4), (u), (f1), (f2), (m1), (m2), (ℓ), (n), (h1), (w1) and (w2) hold, then*

$$n_1 h_1^{d/2}(\tilde{M}_n(\theta_0) - \tilde{C}_n) \Longrightarrow N_d(0, \Gamma)$$

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

$$B_{n1} := \frac{1}{n_1^2}\sum_{i\neq j}^{n_1}\int K_{h_1 i}(z)K_{h_1 j}(z)\xi_i\xi_j d\psi(z),$$

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

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

and

$$B_{n4} := \frac{1}{n_1^2}\sum_{i\neq j}^{n_1}\int K_{h_1 i}(z)K_{h_1 j}(z)\theta_0^T(Q(Z_i) - \hat{Q}_{n_2}(Z_i)) \theta_0^T(Q(Z_j) - \hat{Q}_{n_2}(Z_j))d\psi(z).$$

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 (1.14) and the inequality (1.15), and let

$$C_{nij} = \sum_{i \neq j}^{n_1} \int K_{h_1 i}(z) K_{h_1 j}(z) \xi_i d\psi(z),$$

then B_{n2} is bounded above by the sum $B_{n21} + B_{n22}$, where

$$\begin{aligned} B_{n21} &:= O_P(1) \cdot \frac{1}{n_1^2} \sum_{j=1}^{n_1} [\|\hat{R}_{n_2}(Z_j) - R(Z_j)\| \cdot |C_{nij}|], \\ B_{n22} &:= o\Big((\log_k n_2)\Big(\frac{\log n_2}{n_2}\Big)^{\frac{2}{d+4}}\Big) \cdot \frac{1}{n_1^2} \sum_{j=1}^{n_1} [\|Q(Z_j)\| \cdot |C_{nij}|]. \end{aligned}$$

On the one hand, by the conditional expectation argument and inequality (1.10), we have

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

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

equals to

$$ET(Z_1)\sum_{i\neq 1}\sum_{j\neq 1}\iint K_{h_1i}(z)K_{h_11}(z)K_{h_1j}(y)K_{h_11}(y)\xi_i\xi_j d\psi(z)d\psi(y) \quad (1.18)$$
$$= (n_1-1)\iint E(K_{h_12}(z)K_{h_12}(y)\xi_2^2)\cdot E(K_{h_11}(z)K_{h_11}(y)\mathcal{T}(Z_1))d\psi(z)d\psi(y).$$

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

$$E(K_{h_12}(z)K_{h_12}(y)\xi_2^2) = (K_{h_12}(z)K_{h_12}(y)(\sigma_\varepsilon^2 + \delta^2(Z_2)))$$

$$= \frac{1}{h_1^{2d}}\int K\Big(\frac{z-u}{h_1}\Big)K\Big(\frac{y-u}{h_1}\Big)(\sigma_\varepsilon^2+\delta^2(u))f_Z(u)du$$
$$= \frac{1}{h_1^d}\int K(v)K\Big(\frac{y-z}{h_1}-v\Big)(\sigma_\varepsilon^2+\delta^2(z-h_1v))f_Z(z-h_1v)dv.$$

Similarly, we can show that

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

Putting back these two expectations in (1.18), and changing variables $y = z + h_1u$, then by the continuity of f_Z, $\delta^2(z)$, $g(z)$, and $\mathcal{T}(z)$, we obtain $ET(Z_1)C_{ni1}^2 = (n_1-1)h_1^{-d}$. Therefore,

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

This, in turn, implies $B_{n21} = O_p\Big(n_1^{-2b/(d+2\alpha+4)-1/2}h_1^{-d/2}\Big)$, by assumption (n). Similarly, one can show $n_1^{-2}\sum_{j=1}^{n_1}[\|Q(Z_j)\|\cdot|C_{nij}|]$ is of the order $O_p(n_1^{-1/2}h_1^{-d/2})$. Thus, $B_{n22} = o_p\Big((\log_k n_1)\Big(\log n_1/n_1^b\Big)^{2/(d+4)}\cdot n_1^{-1/2}h_1^{-d/2}\Big)$. Hence

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

is of the order $o_P(1)$ since $b > (d+2\alpha+4)/4$ by assumption (n).

By exactly same method as above, we can show that $n_1 h_1^{d/2} B_{n3} = o_P(1)$.

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

$$\begin{aligned}|B_{n4}| \;\le\; & \frac{1}{n^2}\sum_{i\neq j}^{n_1}\int K_{h_1 i}(z)K_{h_1 j}(z)\|\theta_0\|^2\cdot\|\hat{Q}_{n_2}(Z_i)-Q(Z_i)\|\cdot \\ & \|\hat{Q}_{n_2}(Z_j)-Q(Z_j)\|d\psi(z).\end{aligned}$$

From (1.15), the right hand side of above inequality is bounded above by the sum

$$\begin{aligned}& O_P(1)\cdot B_{n41} + o_P\Big((\log_k n_2)\Big(\frac{\log n_2}{n_2}\Big)^{\frac{2}{d+4}}\Big)(B_{n42}+B_{n43}) \qquad (1.19)\\ & + o_P\Big((\log_k^2 n_2)\Big(\frac{\log n_2}{n_2}\Big)^{\frac{4}{d+4}}\Big)B_{n44},\end{aligned}$$

where

$$\begin{aligned}B_{n41} \;:=\; & \frac{1}{n_1^2}\sum_{i\neq j}^{n_1}\int K_{h_1 i}(z)K_{h_1 j}(z)\cdot\|\hat{R}_{n_2}(Z_i)-R(Z_i)\|\cdot \\ & \|\hat{R}_{n_2}(Z_j)-R(Z_j)\|d\psi(z),\\ B_{n42} \;:=\; & \frac{1}{n_1^2}\sum_{i\neq j}^{n_1}\int K_{h_1 i}(z)K_{h_1 j}(z)\cdot\|\hat{R}_{n_2}(Z_i)-R(Z_i)\|\cdot\|Q(Z_j)\|d\psi(z),\\ B_{n43} \;:=\; & \frac{1}{n_1^2}\sum_{i\neq j}^{n_1}\int K_{h_1 i}(z)K_{h_1 j}(z)\cdot\|\hat{R}_{n_2}(Z_j)-R(Z_j)\|\cdot\|Q(Z_i)\|d\psi(z),\\ B_{n44} \;:=\; & \frac{1}{n_1^2}\sum_{i\neq j}^{n_1}\int K_{h_1 i}(z)K_{h_1 j}(z)\cdot\|Q(Z_i)\|\cdot\|Q(Z_j)\|d\psi(z).\end{aligned}$$

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

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

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

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

Therefore, for $l = 2, 3$,

$$\begin{aligned} & n_1 h_1^{d/2} o_P\Big((\log_k n_2)\Big(\frac{\log n_2}{n_2}\Big)^{\frac{2}{d+4}}\Big) B_{n4l} \\ = \;& o_P(n_1^{1-\frac{2b}{d+4}-\frac{2b}{d+2\alpha+4}} h_1^{d/2} (\log_k n_1)(\log n_1)^{\frac{2}{d+4}}) \end{aligned}$$

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

$$\begin{aligned} & n_1 h_1^{d/2} \cdot o_P\Big((\log_k^2 n_2)\Big(\frac{\log n_2}{n_2}\Big)^{\frac{4}{d+4}}\Big) B_{n44} \\ = \;& o_P(n_1^{1-\frac{4b}{d+4}} h_1^{d/2} (\log_k^2 n_1)(\log n_1)^{\frac{4}{d+4}}) \end{aligned}$$

which is also of the order $o_P(1)$. Finally, from above and (1.19), we prove $n_1 h_1^{d/2} B_{n4} = o_P(1)$. Thereby proving the lemma.

Lemma 1.4.2 *In addition to the conditions in Lemma 1.4.1, 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

$$\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)\theta_0^T \hat{Q}_{n_2}(Z_i)$$

in the squared integrand of $M_n(\hat{\theta}_n)$, we can write $M_n(\hat{\theta}_n) - M_n(\theta_0)$ as the sum $W_{n1} + 2W_{n2}$, where

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

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

and $\zeta_i = Y_i - \theta_0^T \hat{Q}_{n_2}(Z_i)$. Easy to see that

$$W_{n1} \leq 2\int \Big[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)(\theta_0 - \hat{\theta}_n)^T (\hat{Q}_{n_2}(Z_i) - Q(Z_i))\Big]^2 d\hat{\psi}_{h_2}(z) \qquad (1.20)$$
$$+2\int \Big[\frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1 i}(z)(\theta_0 - \hat{\theta}_n)^T Q(Z_i)\Big]^2 d\hat{\psi}_{h_2}(z).$$

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

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

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 $\mathcal{I}_{h_1}$, and the asymptotic behavior of $\hat{\theta}_n - \theta_0$ stated in Theorem 1.3.1, $n_1 h_1^{d/2} W_{n11} = o_P(h_1^{d/2}) = o_P(1)$. On the other hand, W_{n12} is bounded above by

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

Since the integral part is of the order $O_P(1)$, so $n_1 h_1^{d/2} W_{n12} = O_P(h_1^{d/2}) = o_P(1)$ is easily obtained. Therefore, $n_1 h_1^{d/2} W_{n1} = o_P(1)$ is proved.

Now, consider W_{n2}. Rewrite it as

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

Note that integral part of W_{n2} is same as the expression on the right hand side of (1.12), thus

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

Therefore, W_{n2} is bounded above by

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

Adding and subtracting $Q(Z_i)$ from $\hat{Q}_{n_2}(Z_i)$, it turns out that W_{n2} is further bounded above by the sum $W_{n21} + W_{n22}$, where

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

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

Arguing as in W_{n11} and W_{n12}, we can show

$$n_1 h_1^{d/2} |W_{n21}| = o_P(1), \qquad n_1 h_1^{d/2} |W_{n22}| = o_P(1).$$

Therefore, $n_1 h_1^{d/2} |W_{n2}| = o_P(1)$. Together with the result $n_1 h_1^{d/2} |W_{n1}| = o_P(1)$, the lemma is proved. □

Lemma 1.4.3 *If H_0, (e1), (e2), (u), (f1), (f2), (m1), (m2), (ℓ), (n), (h1), (h2), (w1) and (w2) hold,* $n_1 h_1^{d/2}(M_n(\theta_0) - \tilde{M}_n(\theta_0)) = o_P(1)$.

Proof. Recall the definition of ζ_i and $U_{n_1}(z)$. Note that $n_1 h_1^{d/2} |M_n(\theta_0) - \tilde{M}_n(\theta_0)|$ is bounded above by

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

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

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

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

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

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

$$\begin{aligned} & O_p\Big(n_1 h_1^{d/2} \cdot \frac{1}{nh_1^d} \cdot \log_k n_1 (\log n_1/n_1)^{2/(d+4)}\Big) \\ + \ & O_p\Big(n_1 h_1^{d/2} \cdot \log_k n_1 (\log n_1/n_1)^{2/(d+4)} \cdot n_1^{-4b/(d+2\alpha+4)}\Big) \\ + \ & O_p\Big(n_1 h_1^{d/2} \cdot \log_k n_1 (\log n_1/n_1)^{2/(d+4)} \cdot \log_k^2 n_1 (\log n_1)^{4/(d+4)} n_1^{-4b/(d+4)}\Big). \end{aligned}$$

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

Lemma 1.4.4 *If H_0, (e1), (e2), (e4), (u), (f1), (f2), (m1), (m2), (ℓ), (n), (h1), (h2), (w1) and (w2) hold,* $n_1 h_1^{d/2}(\hat{C}_n - \tilde{C}_n) = o_p(1)$.

Proof. Recall the notation $\Delta_{n_1}(z)$ in (1.13). Adding and subtracting $\theta_0^T \hat{Q}_{n_2}(Z_i)$ from Y_i in the integrand of hC_n, then expand the quadratic term, then $\hat{C}_n - \tilde{C}_n$ can be rewritten as the sum of C_{nl}, $l = 1, 2, 3, 4, 5$, where

$$
\begin{aligned}
C_{n1} &:= \frac{1}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)(Y_i - \theta_0^T \hat{Q}_{n_2}(Z_i))^2 \Delta_{n_1}(z)d\psi(z),\\
C_{n2} &:= \frac{2}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)(Y_i - \theta_0^T \hat{Q}_{n_2}(Z_i)) \cdot (\theta_0 - \hat{\theta}_n)^T \hat{Q}_{n_2}(Z_i)\Delta_{n_1}(z)d\psi(z),\\
C_{n3} &:= \frac{2}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)(Y_i - \theta_0^T \hat{Q}_{n_2}(Z_i))(\theta_0 - \hat{\theta}_n)^T \hat{Q}_{n_2}(Z_i)d\psi(z),\\
C_{n4} &:= \frac{1}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)(Y_i - \theta_0^T \hat{Q}_{n_2}(Z_i)) \cdot ((\theta_0 - \hat{\theta}_n)^T \hat{Q}_{n_2}(Z_i))^2\Delta_{n_1}(z)d\psi(z),\\
C_{n5} &:= \frac{1}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)(Y_i - \theta_0^T \hat{Q}_{n_2}(Z_i))((\theta_0 - \hat{\theta}_n)^T \hat{Q}_{n_2}(Z_i))^2 d\psi(z).
\end{aligned}
$$

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

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

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

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

$$
\begin{aligned}
&n_1 h_1^{d/2}|C_{n11}| = o_p\Big(n_1 h_1^{d/2} \cdot (\log_k n_1)(\log n_1)^{2/(d+4)} n_1^{-2/(d+4)} \cdot (n_1 h_1)^{-1}\Big)\\
&= o_p\Big(n_1^{ad/2 - 2/(d+4)} \cdot (\log_k n_1)(\log n_1)^{2/(d+4)}\Big) = o_p(1).
\end{aligned}
$$

Second, from the compactness of Θ, we have

$$\frac{1}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)(\theta_0^T(Q(Z_i)-\hat{Q}_{n_2}(Z_i)))^2 d\psi(z)$$
$$\leq O(1)\cdot\frac{1}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)\|Q(Z_i)-\hat{Q}_{n_2}(Z_i)\|^2 d\psi(z).$$

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

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

Because

$$\frac{1}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)\mathcal{T}^2(Z_i)d\psi(z)=O_P(1/n_1h_1^d),$$
$$\frac{1}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)\|Q(Z_i)\|^2 d\psi(z)=O_P(1/n_1h_1^d),$$

so, from (h2), (w2), and Lemma 1.2.2, we obtain $n_1h_1^{d/2}|C_{n12}|$ is of the order

$$O_P\Big(n_1^{-2/(d+4)-4b/(d+2\alpha+4)}h_1^{-d/2}(\log_k n_1)(\log n_1)^{2/(d+4)}\Big)$$
$$+\ O_P\Big(h_1^{-d/2}n_1^{-2/(d+4)-4b/(d+4)}(\log_k^3 n_1)(\log n_1)^{6/(d+4)}\Big)$$

which is $o_P(1)$ by assumption (h1). Hence we get $n_1h_1^{d/2}|C_{n1}|=o_P(1)$.

Now we will show that $n_1h_1^{d/2}|C_{n3}| = o_P(1)$. Once we prove this, then $n_1h_1^{d/2}|C_{n2}|=o_P(1)$ is a natural consequence. In fact,

$$C_{n3} = \frac{2}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)(\xi_i+\theta_0^TQ(Z_i)-\theta_0^T\hat{Q}_{n_2}(Z_i))$$
$$\cdot(\theta_0-\hat{\theta}_n)^T(\hat{Q}_{n_2}(Z_i)-Q(Z_i)+Q(Z_i))d\psi(z).$$

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

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

It is sufficient to show that $n_1 h_1^{d/2}|C_{n3l}| = o_P(1)$ for $l = 1, 2, 3, 4$. Because the proofs are similar, here we only show $n_1 h_1^{d/2}|C_{n32}| = o_P(1)$, others are omitted for the sake of brevity. In fact, note that

$$\frac{1}{n_1^2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)|\xi_i|\|Q(Z_i)\|d\psi(z) = O_P(1/n_1 h_1^d)$$

by a expectation argument, then from $\|\hat{\theta}_n - \theta_0\| = O_P(n_1^{-1/2})$ by Theorem 1.3.1, we have $n_1 h_1^{d/2}|C_{n32}| = n_1 h_1^{d/2}\|\hat{\theta}_n - \theta_0\| \cdot O_P(1/n_1 h_1^d) = O_P(n_1^{-1/2} h_1^{-d/2})$. Because $n_1^{-1/2} h_1^{-d/2} = n_1^{-1/2+ad/2}$ and $a < 1/2d$ by assumption (h1), so the above expression is $o_P(1)$. Similarly, we can show that the same results hold for C_{n4} and C_{n5}. Details are left out. □

Lemma 1.4.5 *Under the same conditions as in Lemma 1.4.4,* $\hat{\Gamma}_n - \Gamma = o_P(1)$.

Proof. Recall the notation for ξ_i. Define

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

The lemma is proved by showing that

$$\hat{\Gamma}_n - \tilde{\Gamma}_n = o_P(1), \quad \tilde{\Gamma}_n - \Gamma = o_P(1). \tag{1.21}$$

But the second claim can be shown using the same method as in K-N, so we only prove the first claim. Write $u_n := \hat{\theta}_n - \theta_0$, $r_i := \theta_0^T Q(Z_i) - \hat{\theta}_n^T \hat{Q}_{n_2}(Z_i)$. Now $\hat{\Gamma}_n$ can be expressed as the sum of $\tilde{\Gamma}_n$ and the following terms:

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

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

$$\begin{aligned}
h_1^d n_1^{-2} \sum_{i \neq j}^{n_1} \Big[\int K_{h_1 i}(z) K_{h_1 j}(z) |\xi_i r_j| d\hat{\psi}_{h_2}(z) \Big]^2 &= o_P(1) \qquad (1.22) \\
h_1^d n_1^{-2} \sum_{i \neq j}^{n_1} \Big[\int K_{h_1 i}(z) K_{h_1 j}(z) |r_i r_j| d\hat{\psi}_{h_2}(z) \Big]^2 &= o_P(1), \\
h_1^d n_1^{-2} \sum_{i \neq j}^{n_1} \Big[\int K_{h_1 i}(z) K_{h_1 j}(z) |\xi_i \xi_j| d\hat{\psi}_{h_2}(z) \Big]^2 &= O_P(1).
\end{aligned}$$

The third claim in (1.22)can be proved by using the same argument as in K-N. Now, consider the first claim above. From Lemma 1.2.2, we only need to show the claim is true when $d\hat{\psi}_{h_2}(z)$ is replaced by $d\psi(z)$. Since r_j has nothing to do with the integration variable, so the left hand side of the first claim after the replacing can be rewritten as

$$h_1^d n_1^{-2} \sum_{i \neq j}^{n_1} |r_j|^2 \Big[\int K_{h_1 i}(z) K_{h_1 j}(z) |\xi_i| d\psi(z) \Big]^2. \tag{1.23}$$

Note that $r_j = u_n^T(Q(Z_j) - \hat{Q}_{n_2}(Z_j)) - u_n^T Q(Z_j) - \theta_0^T(\hat{Q}_{n_2}(Z_j) - Q(Z_j))$, so (1.23) can be bounded above by the sum of the following three terms:

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

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

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

which can be shown by using the same argument as in K-N. Let's consider A_{n3}. Using the inequality (1.15), Lemma 1.2.2 or (1.14), and the compactness of Θ, it is easy to see A_{n3} is bounded above by the sum $A_{n31} + A_{n32}$, where

$$\begin{aligned}
A_{n31} &= O_P(1) \cdot h_1^d n_1^{-2} \sum_{i \neq j}^{n_1} \|\hat{R}_{n_2}(Z_j) - R(Z_j)\|^2 \cdot \Big[\int K_{h_1 i}(z) K_{h_1 j}(z) |\xi_i| d\psi(z) \Big]^2 \\
A_{n32} &= o_P(1) \cdot h_1^d n_1^{-2} \sum_{i \neq j}^{n_1} \Big[\int K_{h_1 i}(z) K_{h_1 j}(z) |\xi_i| \|Q(Z_j)\| d\psi(z) \Big]^2.
\end{aligned}$$

Apply the conditional expectation argument to the second factor in A_{n31}, using the fact (1.10) and the elementary inequality $a < (1+a)^2$, we can show

$$E\Big[h_1^d n_1^{-2}\sum_{i\neq j}^{n_1}\|\hat{R}_{n_2}(Z_j)-R(Z_j)\|^2\Big[\int K_{h_1 i}(z)K_{h_1 j}(z)|\xi_i|d\psi(z)\Big]^2\Big]$$

$$= E\Big[h_1^d n_1^{-2}\sum_{i\neq j}^{n_1}(E_{S_1}\|\hat{R}_{n_2}(Z_j)-R(Z_j)\|^2)\Big[\int K_{h_1 i}(z)K_{h_1 j}(z)|\xi_i|d\psi(z)\Big]^2\Big]$$

$$\leq cn_2^{-\frac{4}{d+2\alpha+4}}E\Big[h_1^d n_1^{-2}\sum_{i\neq j}^{n_1}\Big[\int K_{h_1 i}(z)K_{h_1 j}(z)|\xi_i||1+\mathcal{T}(Z_j)|d\psi(z)\Big]^2\Big].$$

The expectation of the right hand side of above inequality turns out to be $O(1)$ by using same argument as in K-N. So,

$$h_1^d n_1^{-2}\sum_{i\neq j}^{n_1}\|\hat{R}_{n_2}(Z_j)-R(Z_j)\|^2\Big[\int K_{h_1 i}(z)K_{h_1 j}(z)|\xi_i|d\psi(z)\Big]^2 = o_P(1).$$

This, in turn, implies that the second factor in $A_{n31} = o_P(1)$. Same method as in K-N also leads to the following fact:

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

Hence $A_{n32} = o_P(1)$. Therefore, $B_{n1} = o_P(1)$, and $B_{n2} = o_P(1)$. Thereby proving the first claim in (1.21), hence the lemma. □

We end this section by adding some remarks. First, the MD estimator and testing procedure depends on the choice of the integrating measure. In the classical regression case, K-N provides some guidelines on how to choose G. The same guidelines also apply here. For example, in the one-dimensional case, the asymptotic variance of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ can attain its minimum if G is chosen to be $\hat{f}_{Zh_2}(z)$. As far as the MD test statistic $M_n(\hat{\theta}_n)$ is concerned, the choice of G will depend on the alternatives. In the classical regression case, K-N found that the test has high power against the selected alternatives, if the density function is chosen to be the square of the density estimator of the design variables. Same phenomenon happens in our case. Secondly, since replacing $\hat{\Gamma}_n$ in Theorem 4.1 by other consistent estimator of Γ does not affect the validity of the result, so we can choose some other consistent estimator of Γ, for example,

$$\bar{\Gamma}_n = C\int\Big(\frac{\sum_{i=1}^{n_1} K_{h_1 i}(z)(Y_i - \hat{\theta}_n^T\hat{Q}_{n_2}(Z_i))^2}{n_1\hat{f}_{Zh_2}(z)}\Big)^2 g(z)d\hat{\psi}_{h_2}(z), \qquad (1.24)$$

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

1.5 Simulations

This section contains results of four simulations corresponding to the following cases: Case 1: $d = q = 1$ and m_θ linear, the measurement error ϵ is chosen to be normal and u double exponential; Case 2: $d = q = 1$ and m_θ linear, the measurement error ϵ and u are chosen to be normal; Case 3: $d = 1, q = 2$, and m_θ polynomial, the measurement error ϵ is chosen to be normal and u double exponential; Case 4: $d = q = 2$, and m_θ linear, the measurement error ϵ is chosen to be normal and u double exponential. In each case the Monte Carlo average of $\hat{\theta}_n$, MSE($\hat{\theta}_n$), empirical levels and powers of the MD test are reported. The asymptotic level is taken to be 0.05 in all cases. For any random variable W, we will use $\{W_{jk_j}\}_{k_j=1}^{n_j}$, $j = 1, 2$ to denote the j-th subsample S_j from W with sample size n_j. So the full sample is $S_1 \bigcup S_2$. Finally, to make the simulation less time consuming, $\bar{\Gamma}_n$ defined in (1.24) will be used in the test statistic in stead of $\hat{\Gamma}_n$. So the value of the test statistic is calculated by $\widehat{\mathcal{D}}_n := n_1 h_1^{1/2} \bar{\Gamma}_n^{-1/2} (M_n(\hat{\theta}_n) - \hat{C}_n)$.

Case 1 In this case, $\{X_{jk_j}\}_{k_j=1}^{n_j}$ are obtained as a random sample form the uniform distribution on $[-1,1]$, $\{\varepsilon_{jk_j}\}_{k_j=1}^{n_j}$ are obtained as a random sample from the normal distribution $\mathcal{N}(0,(0.1)^2)$, and $\{u_{jk_j}\}_{k_j=1}^{n_j}$ are obtained as a random sample from the double exponential distribution with mean 0 and variance 0.01. The parametric model is taken to be $m_\theta(X) = \theta X$, and the true parameter $\theta_0 = 1$. Then (Y_i, Z_i) are generated using the model

$$Y_{jk_j} = X_{jk_j} + \varepsilon_{jk_j}, \; Z_{jk_j} = X_{jk_j} + u_{jk_j},$$

$k_j = 1, 2, \cdots, n_j$, $j = 1, 2$. From example 2, we know that the assumption (m1) is held for $\alpha = 0$. The kernel functions K and K^* and the band widths used in all the simulations are

$$K(z) = K^*(z) = \frac{3}{4}(1 - z^2) I(|z| \le 1), \; h_1 = a n_1^{-1/3}, \; h_2 = b n_1^{-1/5} (\log n_1)^{1/5}, \tag{1.25}$$

with some choices for a and b. For the chosen kernel function (1.25), the constant C in $\bar{\Gamma}_n$ is equal to 0.7642. The kernel function used in (1.4) is chosen to be the standard normal, so that the deconvolution kernel function with bandwidth w takes the form

$$L_w(x) = \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2}x^2\Big)\Big[1 - \frac{0.005(x^2-1)}{w^2}\Big],$$

and the band width $w_1 = n_2^{-1/5}$, $w_2 = (\log(n_2)/n_2)^{1/5})$ which are chosen by the assumptions (w1) and (w2). Correspondingly, $\hat{Q}_{n_2}(z) = \hat{R}_{n_2}(z)/\tilde{f}_{Zw_2}(z)$, where

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

Table 1.1 reports the Monte Carlo mean and the MSE($\hat{\theta}_n$) under H_0 for the sample

(n_1, n_2)	(50,134)	(100,317)	(200,753)	(300,1250)	(500,2366)
Mean	1.0103	1.0095	1.0102	1.0105	1.0098
MSE	0.0014	0.0007	0.0004	0.0003	0.0002

Table 1.1: Mean and MSE of $\hat{\theta}_n$, $d = 1, q = 1$, Double Exponential

sizes n_1 = 50, 100, 200, 500, correspondingly, $n_2 = 134, 317, 753, 1250, 2366$, each repeated 1000 times. One can see there appears to be small bias in $\hat{\theta}_n$ for all chosen sample sizes and as expected, the MSE decreases as the sample size increases.

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

$$\text{Model } 0: \ Y = X + \varepsilon,$$

$$\text{Model } 1: \ Y = X + 0.3X^2 + \varepsilon,$$

$$\text{Model } 2: \ Y = X + 1.4\exp(-0.2X^2) + \varepsilon,$$

$$\text{Model } 3: \ Y = XI(X \geq 0.2) + \varepsilon.$$

To assess the effect of the choice of (a, b) that appear 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 1.2 reports the simulation results pertaining to $\widehat{\mathcal{D}}_n$ for three choices of (a, b). The simulation results for the other choices were similar to those reported here. Data from Model 0 in this table are used to study the empirical sizes, and from Models 1 to 3 are used to study the empirical powers of the test. These entities are obtained by computing $\#\{|\widehat{\mathcal{D}}_n| \geq 1.96\}/1000$.

From Table 1.2, 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 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.

Case 2: The measurement error in this case has normal distribution $\mathcal{N}(0, (0.1)^2)$. 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 $n_2 = [n_1]^b$, $b > 7/4$. In the simulation, we choose $b = 7/4 + 0.0001$. The band widths are chosen to be

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

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

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

$$\text{Model } 0: \; Y = X + \varepsilon,$$

$$\text{Model } 1: \; Y = X + 0.3X^2 + \varepsilon,$$

$$\text{Model } 2: \; Y = X + 1.4\exp(-0.2X^2) + \varepsilon,$$

$$\text{Model } 3: \; Y = XI(X \geq 0.2) + \varepsilon.$$

Table 1.4 reports the simulation results pertaining to $\widehat{\mathcal{D}}_n$. Data from Model 0 in this table are used to study the empirical sizes, and from Models 1 to 3 are used to study the empirical powers of the test.

Case 3: This simulation considers the case of $d = 1, q = 2$. Everything here is same as in Case 1 except the null model we want to test is $m_\theta(X) = \theta_1 X + \theta_2 X^2$. The true parameters are $\theta_1 = 1, \theta_2 = 2$. Easy to see that $\hat{R}_{n2}(z)$ takes the form

$$\hat{R}_{n2}(z) \;:=\; \Big(\int x\hat{f}_{Xw_1}(x)f_u(z-x)dx, \int x^2\hat{f}_{Xw_1}(x)f_u(z-x)dx\Big)^T.$$

Table 1.5 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 there appears to be small bias in $\hat{\theta}_n$ for all chosen sample sizes and as expected, the MSE decreases as the sample size increases.

		(n_1, n_2)				
	(a,b)	(50,134)	(100,317)	(200,753)	(300,1250)	(500,2366)
Model 0	(0.3,0.5)	0.003	0.008	0.009	0.020	0.041
	(0.3,0.8)	0.008	0.014	0.017	0.031	0.053
	(0.5,0.5)	0.010	0.011	0.020	0.030	0.049
	(0.8,0.8)	0.020	0.024	0.027	0.042	0.052
	(1.0,0.8)	0.024	0.028	0.026	0.039	0.050
	(1.0,1.0)	0.028	0.037	0.030	0.048	0.054
Model 1	(0.3,0.5)	0.407	0.865	0.987	0.997	1.000
	(0.3,0.8)	0.491	0.888	0.990	0.998	1.000
	(0.5,0.5)	0.704	0.975	0.999	1.000	1.000
	(0.8,0.8)	0.896	0.997	1.000	1.000	1.000
	(1.0,0.8)	0.921	0.999	1.000	1.000	1.000
	(1.0,1.0)	0.926	0.997	1.000	1.000	1.000
Model 2	(0.3,0.5)	0.898	0.972	0.999	0.999	1.000
	(0.3,0.8)	0.919	0.976	0.999	0.999	1.000
	(0.5,0.5)	0.985	0.999	0.999	1.000	1.000
	(0.8,0.8)	0.998	1.000	1.000	1.000	1.000
	(1.0,0.8)	0.999	1.000	1.000	1.000	1.000
	(1.0,1.0)	0.999	1.000	1.000	1.000	1.000
Model 3	(0.3,0.5)	0.774	0.959	0.993	0.998	1.000
	(0.3,0.8)	0.807	0.964	0.993	0.998	1.000
	(0.5,0.5)	0.933	0.966	0.999	1.000	1.000
	(0.8,0.8)	0.999	1.000	1.000	1.000	1.000
	(1.0,0.8)	0.992	1.000	1.000	1.000	1.000
	(1.0,1.0)	0.988	1.000	1.000	1.000	1.000

Table 1.2: Levels and powers of the M.D. test, $d = 1, q = 1$, Double Exponential

(n_1, n_2)	(50,941)	(100,3164)	(200,10643)	(300,21638)	(500,52902)
Mean	1.0051	1.0078	1.0085	1.0101	1.0169
MSE	0.0013	0.0007	0.0004	0.0003	0.0004

Table 1.3: Mean and MSE of $\hat{\theta}_n$, $d = 1, q = 1$, Normal

	(n_1, n_2)				
Model	(50,941)	(100,3164)	(200,10643)	(300,21638)	(500,52902)
Model 0	0.018	0.022	0.029	0.035	0.049
Model 1	0.918	0.999	1.000	1.000	1.000
Model 2	0.999	1.000	1.000	1.000	1.000
Model 3	0.993	1.000	1.000	1.000	1.000

Table 1.4: Levels and powers of the M.D. test, $d = 1, q = 1$, Normal

(n_1, n_2)	(50,134)	(100,317)	(200,753)	(300,1250)	(500,2366)
Mean of $\hat{\theta}_{n1}$	1.0169	1.0144	1.0139	1.0136	1.0128
MSE of $\hat{\theta}_{n1}$	0.0058	0.0031	0.0015	0.0011	0.0007
Mean of $\hat{\theta}_{n2}$	2.0450	2.0452	2.0463	2.0493	2.0473
MSE of $\hat{\theta}_{n2}$	0.0124	0.0076	0.0046	0.0042	0.0033

Table 1.5: Mean and MSE of $\hat{\theta}_n$, $d = 1, q = 2$, Double Exponential

	(n_1, n_2)				
Model	(50,134)	(100,317)	(200,753)	(300,1250)	(500,2366)
Model 0	0.001	0.009	0.019	0.029	0.046
Model 1	0.297	0.815	0.999	1.000	1.000
Model 2	0.528	0.965	0.999	1.000	1.000
Model 3	0.996	0.999	1.000	1.000	1.000

Table 1.6: Levels and powers of the M.D. test, $d = 1, q = 2$, Double Exponential

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

$$\text{Model } 0: \; Y = X + 2X^2 + \varepsilon,$$

$$\text{Model } 1: \; Y = X + 2X^2 + 0.3X^3 + 0.1 + \varepsilon,$$

$$\text{Model } 2: \; Y = X + 2X^2 + 1.4\exp(-0.2X^2) + e.$$

$$\text{Model } 3: \; Y = X + 2X^2 \sin(X) + \varepsilon,$$

Table 1.6 reports the simulation results pertaining to $\widehat{\mathcal{D}}_n$. Data from Model 0 in this table are used to study the empirical sizes, and from Models 1 to 3 are used to study the empirical powers of the test.

Case 4: This simulation considers the case of $d = 2, q = 2$. The null model we want to test is $m_\theta(X) = \theta_1 X_1 + \theta_2 X_2$. The true parameters are $\theta_1 = 1, \theta_2 = 2$. The kernel functions K and K^* and the band widths used in the simulation are

$$K(z_1, z_2) = K^*(z_1, z_2) = \frac{9}{16}(1 - z_1^2)(1 - z_2^2)I(|z_1| \le 1, |z_2| \le 1), \quad (1.26)$$
$$h_1 = n_1^{-1/5}, \; h_2 = n_1^{-1/6}(\log n_1)^{1/6},$$

(n_1, n_2)	(50,354)	(100,1001)	(200,2830)	(300,5200)	(500,11188)
Mean of $\hat{\theta}_{n1}$	1.0099	1.0120	1.0115	1.0094	1.0113
MSE of $\hat{\theta}_{n1}$	0.0042	0.0019	0.0011	0.0008	0.0005
Mean of $\hat{\theta}_{n2}$	2.0202	2.0220	2.0213	2.0225	2.0209
MSE of $\hat{\theta}_{n2}$	0.0042	0.0027	0.0014	0.0011	0.0008

Table 1.7: Mean and MSE of $\hat{\theta}_n$, $d = 2, q = 2$, Double Exponential

For the chosen kernel function (1.26), the constant C in $\bar{\Gamma}_n$ is equal to 0.292. The kernel function used in the (1.4) is chosen to be the bivariate standard normal, so the deconvolution kernel function with band width w takes the form

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

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

Table 1.7 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 there appears to be small bias in $\hat{\theta}_n$ for all chosen sample sizes and as expected, the MSE decreases as the sample size increases.

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

$$\text{Model } 0: \; Y = X_1 + 2X_2 + \varepsilon,$$

$$\text{Model } 1: \; Y = X_1 + 2X_2 + 0.3X_1X_2 + 0.9 + \varepsilon,$$

$$\text{Model } 2: \; Y = X_1 + 2X_2 + 1.4(\exp(-0.2X_1) - \exp(0.7X_2)) + \varepsilon.$$

Model 3: $Y = X_1 I(X_2 \geq 0.2) + \varepsilon$,

	(n_1, n_2)				
Model	(50,354)	(100,1001)	(200,2830)	(300,5200)	(500,11188)
Model 0	0.002	0.012	0.018	0.016	0.038
Model 1	0.908	0.998	1.000	1.000	1.000
Model 2	0.992	0.999	1.000	1.000	1.000
Model 3	0.935	0.996	1.000	1.000	1.000

Table 1.8: Levels and powers of the M.D. test, $d = 2, q = 2$, Double Exponential

Table 1.8 reports the simulation results pertaining to $\widehat{\mathcal{D}}_n$. Data from Model 0 in this table are used to study the empirical sizes, and from Models 1 to 3 are used to study the empirical powers of the test.

1.6 Discussion

1.6.1 Sample Size Allocation

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

Then an interesting question arises. What is the small sample behavior of the test procedure if (1) $n_1 = n_2$ and the two subsamples S_1 and S_2 are independent or (2) $n = n_1 = n_2$ and the same sample is used in the test? 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 Table 1.9-Table 1.12. 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.

	Sample size: (n_1, n_2)				
Model	(50,50)	(100,100)	(200,200)	(300,300)	(500,500)
Model 0	0.008	0.036	0.033	0.038	0.049
Model 1	0.938	1.000	1.000	1.000	1.000
Model 2	1.000	1.000	1.000	1.000	1.000
Model 3	0.990	1.000	1.000	1.000	1.000

Table 1.9: $n_1 = n_2$, $d = 1, q = 1$, Double exponential

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

Table 1.10: Same sample, $d = 1, q = 1$, Double exponential

	Sample size: (n_1, n_2)				
Model	(50,50)	(100,100)	(200,200)	(300,300)	(500,500)
Model 0	0.013	0.023	0.027	0.035	0.047
Model 1	0.931	0.999	1.000	1.000	1.000
Model 2	1.000	1.000	1.000	1.000	1.000
Model 3	0.984	1.000	1.000	1.000	1.000

Table 1.11: $n_1 = n_2$, $d = 1, q = 1$, Normal

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

Table 1.12: Same sample, $d = 1, q = 1$, Normal

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

Table 1.13: Same sample, $d = 2, q = 2$, Double Exponential

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 1.13. The test procedure is conservative for small sample sizes, but the empirical level is close to the nominal level 0.05 when sample size reaches 500. This phenomenon suggests us that by loosing some conditions, such as (n), even the assumptions on the choices of the bandwidths, Theorem 1.3.1 and Theorem 1.4.1 maybe still valid.

1.6.2 General Errors-in-Variables Model Fitting

In the previous sections we have so far discussed the model fitting problem in the errors-in-variables models in which the regression function is linear in θ under the null hypothesis. The separation between the parameter and the predictor enables us not only to get an explicit expression for the estimator, but also to utilize a conditional expectation argument, so that we can use Lemma 1.2.1 to get a better sample allocation scheme. If the regression function under the null hypothesis has a general form other than the form we discussed in this chapter, things become complicated.

For the sake of brevity, this section only reports the results we obtained for the general errors-in-variables model fitting.

To be specific, in the errors-in-variables model (1.1), the problem of interest is to develop tests for the following hypotheses:

$$H_0 : \mu(x) = m_{\theta_0}(x), \text{ for some } \theta_0 \in \Theta, \quad \text{v.s.} \quad H_1 : H_0 \text{ is not true}, \tag{1.27}$$

where $\{m_\theta(x) : \theta \in \Theta\}$ is a given parametric family. Just like in the special case considered in the previous sections, the problem of testing for H_0 is transformed to test for $\nu(z) = \nu_{\theta_0}(z)$, where now $\nu_\theta(z) := E(m_\theta(X)|Z = z)$. A very important question related to this hypothesis change is the following: Are the two hypotheses, $H_{10} : \mu(x) = m_{\theta_0}(x)$, for some θ_0 and all x, and $H_{20} : \nu(z) = \nu_{\theta_0}(z)$, for some

θ_0 and all z, equivalent? The answer is negative generally, because for any two measurable functions $m_1(x), m_2(x)$, $E(m_1(X)|Z = z) = E(m_2(X)|Z = z)$, for all z, need not imply $m_1(x) = m_2(x)$ for all x. In this case, if our test rejects H_{20}, then we can reject H_{10} as well, but if the test fails to reject H_{20}, then we can say nothing about H_{10}. Note that $E(m_1(X)|Z = z) = E(m_2(X)|Z = z)$ is equivalent to

$$\int m_1(x) f_X(x) f_u(z - x) dx = \int m_2(x) f_X(x) f_u(z - x) dx$$

for all z. Hence if $f_u(z - \cdot)$, as a distribution family with parameter $z \in \mathbb{R}^d$, forms a complete family, then these two hypotheses are indeed equivalent. This is the case, for example, for the normal distribution, and if $d = 1$, for double exponential distribution.

From (1.3) one sees that if f_X is known then f_Z is known and hence ν_θ is known except for θ. Therefore a modification of K-N's procedure in this case is as follows. Let

$$\begin{aligned} \bar{T}_n(\theta) &:= \int \Big[\frac{1}{n f_Z(z)} \sum_{i=1}^n K_{hi}(z) Y_i - \nu_\theta(z)\Big]^2 dG(z), \quad \theta \in \Theta, \qquad (1.28) \\ T_n(\theta) &:= \int \Big[\frac{1}{n f_Z(z)} \sum_{i=1}^n K_{hi}(z)(Y_i - \nu_\theta(Z_i))\Big]^2 dG(z), \quad \theta \in \Theta, \\ \bar{\theta}_n &:= \arg\min_{\theta\in\Theta} \bar{T}_n(\theta), \quad \theta_n := \arg\min_{\theta\in\Theta} T_n(\theta), \end{aligned}$$

Here h is a bandwidth only depending on n. Then we may use θ_n to estimate θ, and construct the test statistic through $T_n(\theta_n)$.

Unfortunately, f_X is generally not known and hence f_Z and H_θ are unknown. This makes the above procedures infeasible. To construct the test statistic, one needs estimators for f_Z and H_θ. For f_Z, one can still use the classical kernel estimator, with a possibly different kernel function K^* and a bandwidth h_2. So one only needs to find an estimator for ν_θ. Using deconvoluting kernel density estimator with bandwidth h_3 for f_X One can estimate $\nu_\theta(z)$ by

$$\begin{aligned} \hat{H}_\theta(z) &= \frac{\int m_\theta(x)\hat{f}_{Xh_3}(x)f_\eta(z-x)dx}{\tilde{f}_{Zh_3}(z)}, \\ \tilde{f}_{Zh_3}(z) &= \int \hat{f}_{Xh_3}(x)f_\eta(z-x)dx. \end{aligned}$$

Our proposed inference procedures will be based on the analogs of T_n where $\nu_\theta(z)$ in (1.28) is replaced by its estimator $\hat{\nu}_\theta(z)$.

To be precise, we assign the first $n_1 = n_1(n)$ and $n_1 < n$ observations to estimate f_Z, and use all n observations to estimate f_X. The bandwidths h_1, h_2 will depend on the sub-sample size n_1, and h_3 will still depend on the full sample size n.

Replace $\nu_\theta(z)$ in (1.28) by its estimator $\hat{\nu}_\theta(z)$ and define

$$\begin{aligned} M_n^*(\theta) &:= \int \Big[\frac{1}{n_1\hat{f}_{Zh_2}(z)}\sum_{i=1}^{n_1} K_{h_1 i}(z)Y_i - \nu_\theta(z)\Big]^2 dG(z), \\ M_n(\theta) &:= \int_{\mathcal{I}} \Big[\frac{1}{n_1\hat{f}_{Zh_2}(z)}\sum_{i=1}^{n_1} K_{h_1 i}(z)(Y_i - \hat{\nu}_\theta(Z_i))\Big]^2 dG(z), \quad \theta \in \Theta, \\ \theta_n^* &:= \text{arginf}_{\theta\in\Theta} M_n^*(\theta), \qquad \hat{\theta}_n := \text{arginf}_{\theta\in\Theta} M_n(\theta). \end{aligned}$$

Then we may use $\hat{\theta}_n$ to estimate θ, and construct the test statistic through $M_n(\hat{\theta}_n)$. We can show that θ_n^* converges to θ in probability. But as is clear θ_n^* is really not an estimator, but we need this convergence result to prove the consistency for $\hat{\theta}_n$ for θ, and the asymptotic normality of $\sqrt{n_1}(\hat{\theta}_n - \theta_0)$. Finally, let g be a density of G, and let

$$\zeta_i := Y_i - \hat{H}_{\theta_0}(Z_i), \quad \hat{\zeta}_i := Y_i - \hat{H}_{\hat{\theta}_n}(Z_i),$$

$$
\begin{aligned}
\tilde{C}_n &:= n_1^{-2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)\zeta_i^2 d\psi(z),\\
\hat{C}_n &:= n_1^{-2}\sum_{i=1}^{n_1}\int K_{h_1 i}^2(z)\hat{\zeta}_i^2 d\hat{\psi}_{h_2}(z),\\
\hat{\Gamma}_n &:= 2h_1^d n_1^{-2}\sum_{i\neq j}\Big(\int K_{h_1 i}(z)K_{h_1}(z)\hat{\zeta}_i\hat{\zeta}_j d\hat{\psi}_{h_2}(z)\Big)^2,\\
\tau^2(z) &:= \sigma_\varepsilon^2 + E((m_{\theta_0}(X) - H_{\theta_0}(z))^2|Z = z), \quad \sigma_\varepsilon^2 := \mathrm{Var}(\varepsilon),\\
\Gamma &:= 2\int(\tau^2(z))^2 g(z)d\psi(z)\cdot\int\Big(\int K(u)K(u+v)du\Big)^2 dv,\\
d\hat{\psi}_{h_2}(z) &:= \frac{dG(z)}{\hat{f}_{Zh_2}^2(z)}, \qquad d\psi(z) := \frac{dG(z)}{f_Z^2(z)}.
\end{aligned}
$$

Under appropriate sample size allocation scheme, and under the null hypothesis and other regular conditions, we can show that the asymptotic distribution of $n_1 h_1^{d/2}\hat{\Gamma}_n^{-1/2}(M_n(\hat{\theta}_n) - \hat{C}_n)$ is standard normal. But the sample allocation scheme $n_1 = n_1(n)$ is not feasible, particularly in the super smooth case. Simulation results show that, if we do not follow the sample allocation scheme, just like we did in the previous section, the test statistic behaves quite satisfactory.

CHAPTER 2

Minimum Distance Berkson Model Fitting

2.1 Introduction

Berkson model is also commonly used in the real applications. As an example, consider the herbicide study of Rudemo, et al. (1989) in which a nominal measured amount Z of herbicide was applied to a plant but the actual amount absorbed by the plant X is unobservable. As another example, from Wang (2004), an epidemiologist studies the severity of a lung disease, Y, among the residents in a city in relation to the amount of certain air pollutants, X. The amount of the air pollutants Z can be measured at certain observation stations in the city, but the actual exposure of the residents to the pollutants, X, is unobservable and may vary randomly from the Z-values. In both cases, X can be expressed as Z plus a random error. There are many similar examples in agricultural or medical studies, see e.g., Fuller (1987), Carroll, Ruppert and Stefanski (1995), among others.

All these examples can be formalized into the so called Berkson model

$$Y = \mu(X) + \varepsilon, \qquad X = Z + \eta, \tag{2.1}$$

where η and ε are random errors with $E\varepsilon = 0$, and where η is d-dimensional, and Z is the observable d-dimensional control variable. All three variables ε, η, and Z are assumed to be mutually independent.

The parametric Berkson model where the regression function is of a parametric form $\{m_\theta(x) : x \in \mathbb{R}^d, \theta \in \Theta \subset \mathbb{R}^q\}$, $q \geq 1$, has been focus of numerous authors. Fuller (1987) and Cheng and Van Ness (1999), among others, discuss the estimation in the linear Berkson measurement error models. For nonlinear models, Carroll et al. (1995) and references therein, consider the estimation problem by using regression calibration method. Huwang and Huang (2000) studies the estimation problem when $m_\theta(x)$ is a polynomial in x of a known order and shows that the least square estimators based on the first two conditional moments of Y, given Z, are consistent. Wang (2003, 2004) addresses the same problem in general nonlinear models and shows that the estimators obtained by minimizing the first two conditional moments of Y, given Z, are consistent and asymptotically normal.

But literature appears to be scant on the lack-of-fit testing problem in this important model. This paper makes an attempt in filling this void. To be precise, with (X, Y) obeying the model (2.1), the problem of interest here is to test the hypothesis

$$H_0: \quad \mu(x) = m_{\theta_0}(x), \quad \text{for some } \theta_0 \in \Theta \quad \text{and for all } x;$$

$$H_1: \quad H_0 \text{ is not true,}$$

based on a random sample (X_i, Y_i), $1 \le i \le n$, from the distribution of (X, Y).

Many interesting and profound results, on the contrary, are obtained for the regression model checking problem in the absence of errors in independent variables, see, e.g., Eubank and Spiegelman (1990), An and Cheng (1991), Hart (1997) and references therein, Stute (1997), Stute, Thies, and Zhu (1998), among others. The recent paper of Koul and Ni (2004) uses the minimum distance methodology to propose tests of lack-of-fit for the regression model without errors in variables. In a finite sample comparison of these tests with some other existing tests, they noted that a member of this class preserves the asymptotic level and has very high power against some alternatives and compared to some other existing lack-of-fit tests. This paper extends this methodology to the above Berkson model.

To be specific, Koul and Ni (2004) (K-N) considered the following tests of H_0 where the design is random and observable, and the errors are heteroscedastic. For any density kernel K, let $K_h(x) := K(x/h)/h^d$, $h > 0$, $x \in \mathbb{R}^d$. Define, as in K-N,

$$\tilde{f}_w(x) := \frac{1}{n}\sum_{j=1}^{n} K_w^*(x - X_j), \quad w = w_n \sim (\log n/n)^{1/(d+4)},$$

$$T_n(\theta) := \int_C \Big[\frac{1}{n}\sum_{j=1}^{n} K_h(x - X_j)(Y_j - m_\theta(X_j))\Big]^2 \frac{d\bar{G}(x)}{\tilde{f}_w^2(x)},$$

and $\tilde{\theta}_n := \text{argmin}_{\theta\in\Theta} T_n(\theta)$, where K, K^* are density kernel functions, possibly different, $h = h_n$ and $w = w_n$ are the window widths, depending on the sample size n, and $\bar{G}$ is a sigma finite measure on C which is a compact subset of $\mathbb{R}^d$. They proved the consistency and asymptotic normality of this estimator, and that the asymptotic null distribution, under H_0, of $\mathcal{D}_n := nh_n^{d/2}(T_n(\tilde{\theta}_n) - \tilde{C}_n)/\tilde{\Gamma}_n^{1/2}$ is standard normal, where

$$\tilde{C}_n := \frac{1}{n^2}\sum_{i=1}^{n}\int_{\mathcal{C}} K_h^2(x - X_i)\,\varepsilon_i^2 \tilde{f}_w^{-2}(x)d\bar{G}(x), \qquad \hat{\varepsilon}_i = Y_i - m_{\tilde{\theta}_n}(X_i)$$

$$\tilde{\Gamma}_n := \frac{1}{n^2h^{3d}}\sum_{i\neq j=1}^{n}\Big(\int_{\mathcal{C}} K\Big(\frac{x - X_i}{h}\Big)K\Big(\frac{x - X_j}{h}\Big)\hat{\varepsilon}_i\hat{\varepsilon}_j\tilde{f}_w^{-2}(x)d\bar{G}(x)\Big)^2.$$

The test based on $\mathcal{D}_n$ is preferable over the tests developed by Härdle and Mammen (1993), and Zheng (1996). Unlike in these and other related papers, K-N do not need the null regression function to be twice continuously differentiable in the parameter vector nor do their proofs need the rate for uniform consistency of nonparametric regression function estimators. Moreover, the asymptotic normality of $n^{1/2}(\tilde{\theta}_n - \theta)$ and $\mathcal{D}_n$ was made feasible by recognizing to use different window widths for the estimation of the numerator and denominator in the nonparametric regression function estimation. A consequence of the above asymptotic normality result is that at least for large samples one does not need to use any resampling method to implement these tests.

These findings thus motivate one to look for tests of lack-of-fit in the Berkson model based on the above minimized distances. Since the predictors in Berkson models are unobservable, clearly the above procedures need some modifications.

Let $f_\varepsilon, f_X, f_\eta, f_Z$ denote the density functions of the r.v.'s in their sub-scripts and σ_ε^2 denote the variance of ε. In linear regression models if one is interested in making inference about the coefficient parameters only, these density functions need not be known. Berkson (1950) pointed out that the ordinary least square estimators are unbiased and consistent in these models and one can simply ignore the measurement error η. But if the regression model is nonlinear or if there are other parameters in the Berkson model that need to be estimated, then extra information about these densities should be supplied to ensure the identifiability. A standard assumption in the literature is to assume that f_η is known or unknown only up to an Euclidean parameter vector, cf., Carroll, et al. (1995), Huwang and Huang (2000), Wang (2004), among others. Throughout this paper, we shall assume that f_η is known unless the regression function under the null hypothesis is linear.

To adopt K-N's procedure to the current setup, we first need to obtain a nonparametric estimator of μ. Note that in the model (2.1), $f_X(x) = \int f_Z(z) f_\eta(x-z)dz$. Let K be a kernel density,

$$\hat{f}_Z(z) = n^{-1} \sum_{i=1}^{n} K_h(z - Z_i)$$

be the kernel estimator of $f_Z(z)$, and

$$\bar{K}_h(x,z) := \int K_h(y-z) f_\eta(x-y)dy, \; x, \, z \in \mathbb{R}^d.$$

It is then natural to estimate $f_X(x)$ by

$$\hat{f}_X(x) \;:=\; \int \hat{f}_Z(z) f_\eta(x-z)dz = \frac{1}{n}\sum_{i=1}^{n} \bar{K}_h(x, Z_i), \qquad x \in \mathbb{R}^d.$$

Given the estimator $\hat{f}_X(x)$, one is then tempted to estimate the regression function $\mu(x)$ by

$$\hat{J}_n(x) := \sum_{i=1}^{n} \bar{K}_h(x, Z_i) Y_i \Big/ \sum_{i=1}^{n} \bar{K}_h(x, Z_i).$$

Unfortunately, the classical argument shows that $\hat{J}_n(x)$ is not a consistent estimator of $\mu(x)$. It in fact is consistent for $J(x) = E[H(Z)|X = x]$, where $H(z) = E[\mu(X)|Z = z]$.

We include the following simulation study to illustrate this point. Consider the model $Y = X^2 + \varepsilon$, $X = Z + \eta$, where ε and η are Gaussian r.v.'s with means zero and variances 0.01, and 0.05, respectively. The r.v. Z is the standard Gaussian. Then $J(x) = 0.0976 + 0.907x^2$. We generated 500 samples from this model, calculated $\hat{J}_n$, and then put all three graphs, $\hat{J}_n(x)$, $\mu(x) = x^2$, $J(x) = 0.0976 + 0.907x^2$ into one plot in the Figure 2.1. The curves with solid, dash-dot, dot lines are those of $\hat{J}_n$,

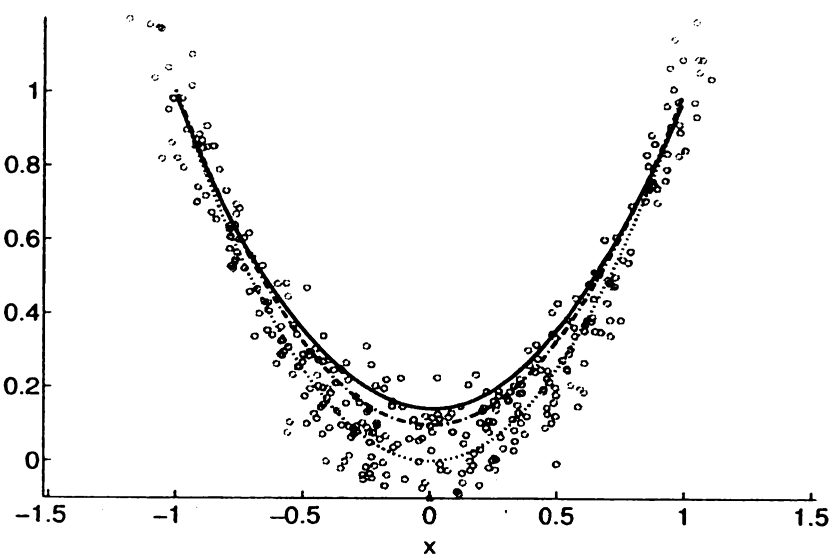

Figure 2.1: Comparison Plot

$J(x)$, and $\mu(x) = x^2$, respectively.

To overcome this difficulty, one way to proceed is as follows. Define

$$\begin{aligned} H_\theta(z) &:= E[m_\theta(X)|Z = z], \qquad J_\theta(x) = E[H_\theta(Z)|X = x], \\ \tilde{Q}_n(\theta) &= \int_C \Big[\frac{1}{n\hat{f}_X(x)} \sum_{i=1}^{n} \bar{K}_h(x, Z_i) Y_i - J_\theta(x)\Big]^2 d\bar{G}(x), \qquad (2.2) \\ Q_n(\theta) &= \int_C \Big[\frac{1}{n\hat{f}_X(x)} \sum_{i=1}^{n} \bar{K}_h(x, Z_i)[Y_i - H_\theta(Z_i)]\Big]^2 d\bar{G}(x), \end{aligned}$$

and $\widetilde{\theta}_n = \text{argmin}_{\theta\in\Theta}\widetilde{Q}_n(\theta)$, $\theta_n = \text{argmin}_{\theta\in\Theta}Q_n(\theta)$.

Under some conditions, we can show that θ_n, $\widetilde{\theta}_n$ are weakly consistent for θ, and the asymptotic null distribution of the test statistic based on the suitably standardized minimum distance $Q_n(\theta_n)$ is the same as that of a degenerate U-statistic, whose asymptotic distribution in turn is the same as that of an infinite sum of weighted centered chi square random variables. Since the kernel function in the degenerate U-statistic is complicated, the computation of the eigenvalues and the eigenfunctions is not easy and hence this test is hard to implement in practice.

An alternative way to proceed as we do here is to recognize that $E(Y|Z) = H(Z)$ and hence consider the new regression model $Y = H(Z) + \zeta$, where the error ζ is uncorrelated with Z and has mean zero. The problem of testing for H_0 is now transformed to test for $H(z) = H_{\theta_0}(z)$. Thus we do the following modification of the above K-N procedure to adjust for not observing the design variable. Let

$$\hat{f}_{Zw}(z) := \frac{1}{n}\sum_{i=1}^{n} K_w^*(z - Z_i), \quad w \sim (\log n/n)^{1/(d+4)};$$
$$\hat{H}_n(z) := \frac{\sum_{i=1}^{n} K_h(z - Z_i)Y_i}{n\hat{f}_{Zw}(z)}, \; z \in \mathbb{R}^d.$$

Note that $\hat{H}_n$ is a nonparametric estimator of the conditional expectation $H(z) = E(\mu(X)|Z = z)$. Define

$$\begin{aligned}
M_n^*(\theta) &= \int_{\mathcal{I}} \Big[\frac{1}{n\hat{f}_{Zw}(z)}\sum_{i=1}^{n} K_h(z - Z_i)Y_i - H_\theta(z)\Big]^2 dG(z),\\
M_n(\theta) &= \int_{\mathcal{I}} \Big[\frac{1}{n\hat{f}_{Zw}(z)}\sum_{i=1}^{n} K_h(z - Z_i)[Y_i - H_\theta(Z_i)]\Big]^2 dG(z),\\
\theta_n^* &= \text{argmin}_{\theta\in\Theta}M_n^*(\theta), \qquad \hat{\theta}_n = \text{argmin}_{\theta\in\Theta}M_n(\theta),
\end{aligned}$$

where G is a measure supported on a compact subset $\mathcal{I} \subset \mathbb{R}^d$. We consider M_n to be the right analog of the above T_n for the Berkson model. Let θ_0 be the true parameter under H_0. This paper proves that θ_n^* converges in probability to θ_0, under H_0. This in turn is used to prove the consistency of $\hat{\theta}_n$ for θ_0, and the asymptotic normality of $\sqrt{n}(\hat{\theta}_n - \theta_0)$, under H_0. Additionally, we prove that the asymptotic null distribution of the normalized test statistic $nh^{d/2}\hat{\Gamma}_n^{-1/2}(M_n(\hat{\theta}_n) - \hat{C}_n)$, based on the minimum distance $M_n(\hat{\theta}_n)$, is standard normal, which, unlike the first modification of (2.2), can be easily used to implement this testing procedure, at least for the large samples. Here,

$$
\begin{aligned}
d\hat{\psi}(z) &:= \frac{dG(z)}{\hat{f}^2_{Zw}(z)}, \quad z \in \mathbb{R}^d; \qquad \hat{\zeta}_i := Y_i - H_{\hat{\theta}_n}(Z_i), \quad 1 \le i \le n, \qquad (2.3)\\
\hat{C}_n &:= \frac{1}{n^2}\sum_{i=1}^{n}\int K_h^2(z - Z_i)\,\hat{\zeta}_i^2\, d\hat{\psi}(z),\\
\hat{\Gamma}_n &:= 2n^{-2}h^d \sum_{i \ne j}\Big(\int K_h(z - Z_i)K_h(z - Z_j)\,\hat{\zeta}_i\hat{\zeta}_j\, d\hat{\psi}_{h_2}(z)\Big)^2.
\end{aligned}
$$

We note that there is a typo in the definition of the $\hat{\Gamma}_n$ of K-N, there should be a factor of 2 in there also.

The paper is organized as follow. The needed assumptions are stated in the next section. Section 3 contain the proofs of consistency of θ_n^* and $\hat{\theta}_n$ while sections 4 and 5 contains the proofs of the asymptotic normality of $\hat{\theta}_n$ and that of the proposed test statistic. The simulation results in section 6 show little bias in the estimator $\hat{\theta}_n$ for all chosen sample sizes. The finite sample level approximates the nominal level well for larger sample sizes and the empirical power is high (above 0.9) for moderate to large sample sizes against the chosen alternatives.

2.2 Assumptions

Here we shall state the needed assumptions in this paper. Throughout the paper θ_0 denotes the true parameter value under H_0. About the errors, the underlying design and G we assume the following:

(e1) The random variables $\{(Z_i, Y_i) : Z_i \in \mathbb{R}^d, i = 1, 2, \cdots, n\}$ are i.i.d. with the conditional expectation $H(z) = E(Y|Z = z)$ satisfying $\int H^2(z)dG(z) < \infty$, where G is a σ-finite measure on $\mathcal{I}$.

(e2) $0 < \sigma_\varepsilon^2 < \infty$, $Em_{\theta_0}^2(X) < \infty$, and the function $\tau^2(z) = E[(m_{\theta_0}(X) - H_{\theta_0}(Z))^2|Z = z]$ is a.s. (G) continuous on $\mathcal{I}$.

(e3) $E|\varepsilon|^{2+\delta} < \infty$, $E[m_{\theta_0}(X) - H_{\theta_0}(Z)]^{2+\delta} < \infty$, for some $\delta > 0$.

(e4) $E|\varepsilon|^4 < \infty$, $E[m_{\theta_0}(X) - H_{\theta_0}(Z)]^4 < \infty$.

(f1) The density f_Z is uniformly continuous and bounded from below on $\mathcal{I}$.

(f2) The density f_Z is twice continuously differentiable.

(g) The integrating measure G has a continuous Lebesgue density g on $\mathcal{I}$.

About the kernel functions K and K^*, we shall assume the following:

(k) The kernel functions K, K^* are positive symmetric square integrable densities on $[-1, 1]^d$. In addition, K^* satisfies a Lipschitz condition.

About the parametric family of functions to be fitted we need to assume the following:

(m1) For each θ, $m_\theta(x)$ is a.s. continuous w.r.t. the Lebesgue measure.

(m2) The function $H_\theta(z)$ is identifiable w.r.t. θ. i.e., if $H_{\theta_1}(z) = H_{\theta_2}(z)$ for almost all $z(G)$, then $\theta_1 = \theta_2$.

(m3) For some positive continuous function ℓ on $\mathcal{I}$ with $E\ell(Z) < \infty$ and for some $\beta > 0$,

$$|H_{\theta_2}(z) - H_{\theta_1}(z)| \le \|\theta_2 - \theta_1\|^\beta \ell(z), \quad \forall \theta_1, \theta_2 \in \Theta, z \in \mathcal{I}.$$

(m4) For every z, $H_\theta(z)$ is differentiable in θ in a neighborhood of θ_0 with the vector of derivative $\dot{H}_\theta(z)$, such that for every $0 < k < \infty$,

$$\sup_{1 \le i \le n, \sqrt{nh_n^d}\|\theta - \theta_0\| \le k} \frac{|H_\theta(Z_i) - H_{\theta_0}(Z_i) - (\theta - \theta_0)'\dot{H}_{\theta_0}(Z_i)|}{\|\theta - \theta_0\|} = o_P(1).$$

(m5) For every $0 < k < \infty$,

$$\sup_{1 \le i \le n, \sqrt{nh_n^d}\|\theta - \theta_0\| \le k} h_n^{-d/2} \|\dot{H}_\theta(Z_i) - \dot{H}_{\theta_0}(Z_i)\| = o_P(1), \quad \forall n > N_\varepsilon.$$

(m6) $\Sigma_0 := \int \dot{H}_{\theta_0} \dot{H}'_{\theta_0} dG$ is positive definite.

About the bandwidth h_n we shall make the following assumptions:

(h1) $h_n \to 0$ as $n \to \infty$.

(h2) $nh_n^{2d} \to \infty$ as $n \to \infty$.

(h3) $h_n \sim n^{-a}$, where $a < \min(1/2d, 4/(d(d+4)))$.

The above conditions are similar to those imposed in K-N on the model m_θ. Consider the following conditions in terms of the given model.

(m2$'$) The parametric family of models $m_\theta(x)$ is identifiable w.r.t. θ, i.e., if $m_{\theta_1}(x) = m_{\theta_2}(x)$ for almost all x, then $\theta_1 = \theta_2$.

(m3$'$) For some positive continuous function L on $\mathbb{R}^d$ with $EL(X) < \infty$ and for some $\beta > 0$,

$$|m_{\theta_2}(x) - m_{\theta_1}(x)| \leq \|\theta_2 - \theta_1\|^\beta L(x), \quad \forall\, \theta_1, \theta_2 \in \Theta,\ x \in \mathbb{R}^d.$$

(m4$'$) The function $m_\theta(x)$ is differentiable in θ in a neighborhood of θ_0, with the vector of differential $\dot{m}_{\theta_0}$ such that for every $k < \infty$,

$$\sup_{x \in \mathbb{R}^d, \sqrt{nh_n^d}\|\theta - \theta_0\| \leq k} \frac{|m_\theta(x) - m_{\theta_0}(x) - (\theta - \theta_0)'\dot{m}_{\theta_0}(x)|}{\|\theta - \theta_0\|} = o_P(1).$$

(m5$'$) For every $0 < k < \infty$,

$$\sup_{x \in \mathbb{R}^d, \sqrt{nh_n^d}\|\theta - \theta_0\| \leq k} h_n^{-d/2} \|\dot{m}_\theta(x) - \dot{m}_{\theta_0}(x)\| = o_P(1), \quad \forall n > N_\varepsilon.$$

In some cases, (m2) and (m2$'$) are equivalent. For example, if the family of densities $\{f_\eta(\cdot - z);\ z \in \mathbb{R}\}$ is complete then this holds. Similarly, if $m_\theta(x) = \theta'\gamma(x)$ and $\int \gamma(x) f_\eta(x - z)dx \neq 0$, for all z, then also (m2) and (m2$'$) are equivalent.

We can also show that (m3$'$)-(m5$'$) imply (m3)-(m5), respectively. This follows because $H_\theta(z) \equiv \int m_\theta(x) f_\eta(x - z)dx$. Thus under (m3$'$),

$$|H_{\theta_2}(z) - H_{\theta_1}(z)| \leq \|\theta_2 - \theta_1\| \int L(x) f_\eta(x - z)dx, \quad \forall\, z \in \mathbb{R}^d.$$

Hence (m3) holds with $\ell(z) = \int L(x) f_\eta(x - z)dx$. Note that $E\ell(Z) = EL(X) < \infty$.

Similarly, using the fact that $\int f_\eta(x-z)dx \equiv 1$, the left hand side of (m4) is bounded above by

$$\sup_{x \in \mathbb{R}^d, \sqrt{nh_n^d}\|\theta-\theta_0\| \le k} \frac{|m_\theta(x) - m_{\theta_0}(x) - (\theta-\theta_0)'\dot{m}_{\theta_0}(x)|}{\|\theta-\theta_0\|} = o_P(1),$$

by (m4′). Similarly, (m5′) implies (m5) and (m1) implies that $H_\theta(z)$ is a.s. continuous in z (G).

The conditions (m1)-(m6) are trivially satisfied by the model $m_\theta(x) = \theta'\gamma(x)$ provided the components of $E[\gamma(X)|Z=z]$ are continuous, non-zero on $\mathcal{I}$, and the matrix $\int E[\gamma(X)\gamma'(X)|Z=z]dG(z)$ is positive definite.

The conditions (e1), (e2), (f1), (k), (m1)-(m3), (h1) and (h2) suffice for the consistency of $\hat{\theta}_n$, while these plus (e3), (f2), (m4), (m5), (m6) and (h3) are needed for the asymptotic normality of $\hat{\theta}_n$. The asymptotic normality of $M_n(\hat{\theta}_n)$ needs (e1), (e2), (e3), (e4), and (f1)-(m6), and (h3). Of course, (h3) implies (h1) and (h2).

Let $\hat{f}_{Zh}$ denote kernel density estimator of f_Z with bandwidth $h \equiv h_n$. From Mack and Silverman (1982), we obtain that under (f1), (k), (h1) and (h2),

$$\sup_{z \in \mathcal{I}} |\hat{f}_{Zh_n}(z) - f_Z(z)| = o_P(1), \quad \sup_{z \in \mathcal{I}} |\hat{f}_{Zw}(z) - f_Z(z)| = o_P(1), \quad (2.4)$$
$$\sup_{z \in \mathcal{I}} \left|\frac{f_Z(z)}{\hat{f}_{Zw}(z)} - 1\right| = o_P(1).$$

These conclusions are often used in the proofs below.

In the sequel, the true parameter θ_0 is assumed to be an inner point of Θ and $\zeta := Y - H_{\theta_0}(Z)$. The integrals with respect to the G-measure are understood to be over the compact set $\mathcal{I}$. The convergence in distribution is denoted by $\rightarrow_d$ and $\mathcal{N}_p(a, B)$ denotes the p-dimensional normal distribution with mean vector a and covariance matrix B, $p \geq 1$. We shall also need the following notation.

$$\begin{aligned} d\psi(z) &:= \frac{dG(z)}{f_Z^2(z)}, \quad \sigma_\zeta^2(z) := \mathrm{Var}_{\theta_0}(\zeta | Z = z) = \sigma_\varepsilon^2 + \tau^2(z), \ z \in \mathbb{R}^d, \quad (2.5) \\ \zeta_i &:= Y_i - H_{\theta_0}(Z_i), \ 1 \leq i \leq n; \quad \tilde{C}_n := \frac{1}{n^2} \sum_{i=1}^n \int K_h^2(z - Z_i) \zeta_i^2 d\psi(z), \\ \Gamma &:= 2 \int (\sigma_\zeta^2(z))^2 g(z) d\psi(z) \cdot \int \Big(\int K(u) K(u+v) du \Big)^2 dv. \end{aligned}$$

2.3 The Consistency of θ_n^* and $\hat{\theta}_n$

This section proves the consistency of θ_n^* and $\hat{\theta}_n$. Let $L_2(G)$ denote a class of square integrable real valued functions on $\mathbb{R}^d$ with respect to G. Define

$$\rho(\nu_1, \nu_2) := \int [\nu_1(x) - \nu_2(x)]^2 dG(x), \ \nu_1, \nu_2 \in L_2(G),$$

and the map $T(\nu) = \arg\min_{\theta \in \Theta} \rho(\nu, H_\theta), \quad \nu \in L_2(G)$.

The following lemma is found useful in the proofs here. Its proof is similar to that of Theorem 1 in Beran (1977).

Lemma 2.3.1 *Let H_θ satisfy conditions (m1)-(m3). Then the following hold.*

(a). $T(\nu)$ always exists, for $\forall \nu \in L_2(G)$.

(b). If $T(\nu)$ is unique, then T is continuous at ν in the sense that for any sequence of $\{\nu_n\} \in L_2(G)$ converging to ν in $L_2(G)$, $T(\nu_n) \to T(\nu)$, i.e.

$$\rho(\nu_n, \nu) \to 0 \text{ implies } T(\nu_n) \to T(\nu), \text{ as } n \to \infty.$$

(c). $T(H_\theta(\cdot)) = \theta$, uniquely for $\forall \theta \in \Theta$.

Recall the notation at (2.3) and (2.5). As in K-N, for any integral $J := \int r d\hat{\psi}$, the replacement of $d\hat{\psi}$ by $d\psi(z)$ is reflected by the notation $\tilde{J} := \int r d\psi$. We also need to define, for a $\theta \in \mathbb{R}^q$,

$$\begin{aligned}
\mu_n(z,\theta) &:= \frac{1}{n}\sum_{i=1}^{n} K_h(z-Z_i)H_\theta(Z_i), \qquad (2.6)\\
\dot\mu_n(z,\theta) &:= \frac{1}{n}\sum_{i=1}^{n} K_h(z-Z_i)\dot H_\theta(Z_i),\\
U_n(z,\theta) &:= \frac{1}{n}\sum_{i=1}^{n} K_h(z-Z_i)Y_i - \mu_n(z,\theta)\\
&= \frac{1}{n}\sum_{i=1}^{n} K_h(z-Z_i)[Y_i - H_\theta(Z_i)], \quad U_n(z) := U_n(z,\theta_0)\\
\mathcal{Z}_n(z,\theta) &:= \mu_n(z,\theta) - \mu_n(z,\theta_0) = \frac{1}{n}\sum_{i=1}^{n} K_h(z-Z_i)[H_\theta(Z_i) - H_{\theta_0}(Z_i)],
\end{aligned}$$

These entities are the analogs of the similar entities defined at (3.1) in K-N. The main difference is that μ_θ there is replaced by H_θ and X_i's by Z_i's. A consequence of Lemma 2.3.1 is the following

Corollary 2.3.1 *Suppose H_0, (e1), (e2), (f1), and (m1)-(m3) hold. Then $\theta_n^* \to \theta_0$, in probability.*

Proof. We shall use part (b) of the Lemma 2.3.1 with $\nu_n = \hat{H}_n(z)$, and $\nu = H_{\theta_0}(z)$. Note that $M_n^*(\theta_0) = \rho(\hat{H}_n, H_{\theta_0})$, $\theta_n^* = T(\nu_n)$, and by the identifiability condition (m2), $T(\nu) = \theta_0$ is unique. It thus suffices to prove

$$\rho(\hat{H}_n, H_{\theta_0}) = o_P(1). \qquad (2.7)$$

To show this, by plugging in $Y_i = \zeta_i + H_{\theta_0}(Z_i)$, and expanding the quadratic integrand, $\rho(\hat{H}_n, H_{\theta_0})$ is bounded above by the sum $2[C_{n1} + C_{n2}(\theta_0)]$, where

$$C_{n1} := \int U_n^2(z)d\hat\psi(z), \; C_{n2}(\theta) := \int [\mu_n(z,\theta) - \hat{f}_{Zw}(z)H_\theta(z)]^2 d\hat\psi(z), \; \theta \in \mathbb{R}^q.$$

By Fubini and the independence of Z and ϵ, we have

$$E\int U_n^2(z)d\psi(z) = n^{-1}\int EK_h^2(z - Z_1)(\sigma_\epsilon^2 + \tau^2(Z_1))d\psi(z). \tag{2.8}$$

By the uniform continuity of f_Z ensured by (f1),

$$\begin{aligned} EK_h^2(z - Z_1) &= \frac{1}{h^{2d}}\int K^2\Big(\frac{z-y}{h}\Big)f_Z(y)dy = \frac{1}{h^d}\int K^2(y)f_Z(z - yh)dy \\ &= O(1/h^d). \end{aligned}$$

Similarly, using additionally the a.s. continuity of $\tau^2(z)$, we also have

$$EK_h^2(z - Z_1)\tau^2(Z_1) = E\Big[\frac{1}{h^{2d}}\int K^2\Big(\frac{z - Z_1}{h}\Big)\tau^2(Z_1)dz\Big] = O(1/h^d).$$

These calculations imply that

$$E\int U_n^2(z)d\psi(z) = O\Big(\frac{1}{nh^d}\Big) \quad \text{and} \quad \int U_n^2(z)d\psi(z) = O_P\Big(\frac{1}{nh^d}\Big) \tag{2.9}$$

Hence by (2.4), we obtain

$$C_{n1} \le \sup_{z\in\mathcal{I}}\Big|\frac{f(z)}{\hat f_{Zw}(z)}\Big|^2\int U_n^2(z)d\psi(z) = O_P\Big(\frac{1}{nh^d}\Big) = o_P(1).$$

Let

$$\begin{aligned} e_h(z,\theta) &:= EK_h(z - Z_1)H_\theta(Z_1) = \int K(u)H_\theta(z - uh)f_Z(z - uh)du \\ e_w^*(z,\theta) &:= EK_h^*(z - Z_1)H_\theta(z) = \int K^*(u)f_Z(z - uh)du \cdot H_\theta(z). \end{aligned}$$

By adding and subtracting $e_h(z,\theta)$ and $e_w^*(z,\theta)$ in the quadratic term of the integrand in C_{n2}, and using the similar method as in K-N, one can show that $C_{n2}(\theta_0) = o_P(1)$ by (f1), (m1). This proves (2.7) and hence the corollary. □

Remark 2.3.1 Lemmas 3.1 and Corollary 3.1 are similar to those in the K-N paper. The only difference is that here we have H_θ and Z_i in place of m_θ and X_i there so that the current ζ_i are analogs of ε_i of the K-N paper. Another difference is because of the measurement error in design, (2.8) here has the extra variance term $\tau^2(Z)$, although the asymptotic order of this expectation is the same as in the no measurement error model given in (2.9) above. Thus, from now onwards, in many proofs below we shall be brief.

The proof of the following theorem is exactly similar to that of Theorem 3.1 of K-N after the above said modifications are made in there. Details are left out for the sake of brevity.

Theorem 2.3.1 *Suppose (e1), (e2), (e3), (f1), (m1)-(m3), and (h2) hold. Then under H_0, $\hat{\theta}_n \to \theta_0$, in probability.*

2.4 Asymptotic Distribution of $\hat{\theta}_n$

In this section, we shall prove the asymptotic normality of $\sqrt{n}(\hat{\theta}_n - \theta_0)$. The first step towards this goal is to show that

$$nh^d\|\hat{\theta}_n - \theta_0\|^2 = O_p(1). \tag{2.10}$$

Recall the definition of $\mathcal{Z}_n$, and let $D_n(\theta) = \int \mathcal{Z}_n^2(z,\theta)d\hat{\psi}_{h_2}(z)$. We claim that

$$nh^d D_n(\hat{\theta}_n) = O_p(1). \tag{2.11}$$

To see this, observe that $nh^d M_n(\theta_0) = nh^d \int[\frac{1}{n}\sum_{i=1}^n K_h(z - Z_i)\zeta_i]^2 d\hat{\psi}_{h_2}(z) = O_p(1)$ by (2.9) and (2.4). But, according to the definition of $\hat{\theta}_n$, one has $M_n(\hat{\theta}_n) \leq$

$M_n(\theta_0)$, so $nh^d M_n(\hat{\theta}_n) = O_p(1)$. This fact, together with the inequality $D_n(\hat{\theta}_n) \le 2M_n(\hat{\theta}_n) + 2M_n(\theta_0)$, proves (2.11).

Next, we shall show that for any $a > 0$, there exists an N_a such that

$$P\left(D_n(\hat{\theta}_n)/\|\hat{\theta}_n - \theta_0\|^2 \ge a + \inf_{\|b\|=1} b^T \Sigma_0 b \right) > 1 - a, \qquad \forall n > N_a, \tag{2.12}$$

where Σ_0 as in (m6). The claim (2.10) then will follow from (2.11), (2.12), (m6), and the fact

$$nh^d D_n(\hat{\theta}_n) = nh^d \|\hat{\theta}_n - \theta_0\|^2 \cdot [D_n(\hat{\theta}_n)/\|\hat{\theta}_n - \theta_0\|^2].$$

To prove (2.12), let

$$u_n := \hat{\theta}_n - \theta_0, \tag{2.13}$$

$$d_{ni} := H_{\hat{\theta}_n}(Z_i) - H_{\theta_0}(Z_i) - u_n' \dot{H}_{\theta_0}(Z_i), \quad 1 \le i \le n,$$

$$\Sigma_n(b) := \int \Big[b' \cdot n^{-1} \sum_{i=1}^n K_h(z - Z_i) \dot{H}_{\theta_0}(Z_i) \Big]^2 d\hat{\psi}_{h_2}(z), \quad b \in \mathbb{R}^q.$$

Note that

$$\frac{D_n(\hat{\theta}_n)}{\|\hat{\theta}_n - \theta_0\|^2} = \int \frac{\mathcal{Z}_n^2(z, \hat{\theta}_n)}{\|u_n\|^2} d\hat{\psi}_{h_2}(z) \ge D_{n1} + D_{n2} - 2D_{n1}^{1/2} D_{n2}^{1/2}, \tag{2.14}$$

where

$$D_{n1} := \int \Big[\frac{1}{n} \sum_{i=1}^n K_h(z - Z_i) \Big(\frac{d_{ni}}{\|u_n\|} \Big) \Big]^2 d\hat{\psi}_{h_2}(z),$$

$$D_{n2} := \int \Big[\frac{u_n' n^{-1} \sum_{i=1}^n K_h(z - Z_i) \dot{H}_{\theta_0}(Z_i)}{\|u_n\|} \Big]^2 d\hat{\psi}_{h_2}(z).$$

By the assumption (m4) and (2.10), one verifies that $D_{n1} = o_p(1)$. For the term D_{n2}, note that

$$D_{n2} \ge \inf_{\|b\|=1} \Sigma_n(b). \tag{2.15}$$

Decompose

$$
\begin{aligned}
\Sigma_n(b) &= \int \Big[b'\frac{1}{n}\sum_{i=1}^{n} K_h(z-Z_i)\dot{H}_{\theta_0}(Z_i)\Big]^2 d\psi(z) \\
&\quad + \int \Big[b'\frac{1}{n}\sum_{i=1}^{n} K_h(z-Z_i)\dot{H}_{\theta_0}(Z_i)\Big]^2 \cdot \Big[\frac{f_Z^2(z)}{\hat{f}_{Zw}^2(z)} - 1\Big] d\psi(z) \cdot \\
&:= \Sigma_{n1}(b) + \Sigma_{n2}(b).
\end{aligned}
$$

Note that $EK_h(z-Z)\dot{H}_{\theta_0}(Z) = \dot{H}_{\theta_0}(z)f_Z(z) + o(1)$. Hence, by the Law of Large Numbers, $\Sigma_{n1}(b) \to b'\Sigma_0 b$, for every $b \in \mathbb{R}^q$. Moreover,

$$
\Sigma_{n2}(b) \le \sup_{z\in\mathcal{I}} \Big|\frac{f_Z^2(z)}{\hat{f}_{Zw}^2(z)} - 1\Big| \cdot \Sigma_{n1}(b) = o_P(1), \quad \forall\, b \in \mathbb{R}^q.
$$

Also, note that for any $\delta > 0$, and any two unit vectors $b_1, b_2 \in \mathbb{R}^d$ and $\|b_1 - b_2\| \le \delta$, one has

$$
\begin{aligned}
&|\Sigma_{n1}(b_2) - \Sigma_{n1}(b_1)| \\
&= \Big|\int \Big[(b_2-b_1)'\frac{1}{n}\sum_{i=1}^{n} K_h(z-Z_i)\dot{H}_{\theta_0}(Z_i)\Big]^2 d\psi(z)\Big| \\
&\quad + \Big|\int \Big[(b_2-b_1)'\frac{1}{n}\sum_{i=1}^{n} K_h(z-Z_i)\dot{H}_{\theta_0}(Z_i)\Big]\Big[\frac{b_2'}{n}\sum_{i=1}^{n} K_h(z-Z_i)\dot{H}_{\theta_0}(Z_i)\Big] d\psi(z)\Big| \\
&\le \delta(\delta+1)\int \Big\|\frac{1}{n}\sum_{i=1}^{n} K_h(z-Z_i)\dot{H}_{\theta_0}(Z_i)\Big\|^2 d\psi(z).
\end{aligned}
$$

But the expected value of the random variables inside the square of the second factor tends to $\dot{H}_{\theta_0}(z)f_Z(z)$ in probability, so the second factor is $O_P(1)$. From these observations and the compactness of $\{b \in \mathbb{R}^d : \|b\| = 1\}$, one obtains $\sup_{\|b\|=1} \|\Sigma_n(b) - b'\Sigma_0 b\| = o_P(1)$. This fact, together with (2.15), implies (2.12) in a routing fashion, and also concludes the proof of (2.10). We remark here that the inequality (2.14) above corrects a typo in the K-N paper in the equation just above (4.8) there on page 120.

We shall prove the asymptotic normality of $\sqrt{n}(\hat{\theta}_n - \theta_0)$. The proof is classical in nature. Recall the definition of $M_n(\theta)$, and let $\dot{M}_n(\theta) = -2\int U_n(z,\theta)\dot{\mu}_n(z,\theta)d\hat{\psi}_{h_2}(z)$. Sice θ_0 is an interior point of Θ, by consistency, for sufficiently large n, $\hat{\theta}_n$ will be in the interior of Θ, and $\dot{M}_n(\hat{\theta}_n) = 0$, with arbitrarily large probability. But the equation $\dot{M}_n(\hat{\theta}_n) = 0$ is equivalent to

$$\int U_n(z)\dot{\mu}_n(z,\hat{\theta}_n)d\hat{\psi}_{h_2}(z) = \int Z_n(z,\hat{\theta}_n)\dot{\mu}_n(z,\hat{\theta}_n)d\hat{\psi}_{h_2}(z). \tag{2.16}$$

We shall show that $\sqrt{n}\times$ the left hand side of this equation converges in distribution to a normal random variable, while the right hand side of this equation equals $R_n(\hat{\theta}_n - \theta_0)$, for all $n \geq 1$, with $R_n = \Sigma_0 + o_p(1)$. To establish the first of these two claims, rewrite this random variable as the sum of $S_n + S_{n1} + g_{n1} + g_{n2} + g_{n3} + g_{n4}$, where

$$\begin{aligned}
S_n &:= \int U_n(z)\dot{\mu}_h(z)d\psi(z), \quad \dot{\mu}_h(z) := EK_h(z-Z)\dot{H}_{\theta_0}(Z), \\
S_{n1} &:= \int U_n(z)\dot{\mu}_h(z)(1/\hat{f}^2_{Zw}(z) - 1/f^2_Z(z))dG(z), \\
g_{n1} &:= \int U_n(z)[\dot{\mu}_n(z,\theta_0) - \dot{\mu}_h(z)]d\psi(z), \\
g_{n2} &:= \int U_n(z)[\dot{\mu}_n(z,\theta_0) - \dot{\mu}_h(z)](1/\hat{f}^2_{Zw}(z) - 1/f^2_Z(z))dG(z), \\
g_{n3} &:= \int U_n(z)[\dot{\mu}_n(z,\hat{\theta}_n) - \dot{\mu}_n(z,\theta_0)]d\psi(z), \\
g_{n4} &:= \int U_n(z)[\dot{\mu}_n(z,\hat{\theta}_n) - \dot{\mu}_n(z,\theta_0)](1/\hat{f}^2_{Zw}(z) - 1/f^2_Z(z))dG(z).
\end{aligned}$$

We need the following lemmas.

Lemma 2.4.1 *Suppose (e1), (e2), (f1), (k), (m1)-(m6), (h1),(h2) H_0 hold.*

(i) If, additionally, (e3) and (g) hold, then $\sqrt{n}S_n \to_d N(0,\Sigma)$, where

$$\Sigma = \int \frac{(\sigma^2_\epsilon + \tau^2(u)) \cdot \dot{H}_{\theta_0}(u)\dot{H}'_{\theta_0}(u)g^2(u)}{f_Z(u)}du.$$

(ii) If, additionally, (f2) and (h3) hold, then

$$\sqrt{n}|S_{n1}| = o_P(1). \tag{2.17}$$

Lemma 2.4.2 *Under H_0, (e1), (e2), (f1), (k), (m1), (m2), (m4), (m5), (h1) and (h2),*

$$n^{1/2} g_{nk} = o_P(1), \quad k = 1, 2, 3, 4.$$

The proof of (2.17) is facilitated by the following lemma, which along with its proof appears as Theorem 2.2 part (2) in Bosq (1998).

Lemma 2.4.3 *Let $\hat{f}_{Zw}(z)$ be the kernel estimator associated with a kernel K^* which satisfies a Lipschitz condition. If (f2) holds and w is chosen to be $a_n(\log n/n)^{1/(d+4)}$, where $a_n \to a_0 > 0$, then*

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

for any positive integer k.

Proof of Lemma 2.4.1. Again this proof is similar to that of Lemma 4.1 of K-N but we include details here to see how the difference in the asymptotic variance appears. For convenience, we shall give the proof here only for the case $q = 1$, i.e., when $\dot{\mu}_h(z)$ is one dimensional. For multidimensional case the result can be proved by using linear combination of its components instead of $\dot{\mu}_h(z)$, and applying the same argument.

Let $s_{ni} := \int K_h(z - Z_i)\zeta_i \dot{\mu}_h(z) d\psi(z)$. Then $\sqrt{n}S_n$ can be rewritten as $\sqrt{n}S_n = n^{-1/2}\sum_{i=1}^n s_{ni}$. Note that $s_{ni} : 1 \le i \le n$ are i.i.d. centered random variables for each n. By the Lindeberg-Feller CLT, it suffices to show that

$$Es_{n1}^2 \to \Sigma, \quad Es_{n1}^2 I[|s_{n1}| > n^{1/2}\lambda] \to 0, \qquad \text{for } \forall \lambda > 0. \tag{2.18}$$

In fact, one can show that Es_{n1}^2 is equal to

$$\iiint K(v)K(t)\sigma_\zeta^2(u)f_Z(u)\dot{\mu}_h(u+vh)\dot{\mu}_h(u+th)\frac{g(u+vh)}{f_Z^2(u+vh)}\frac{g(u+th)}{f_Z^2(u+th)}dudvdt \to \Sigma,$$

thereby proving the first claim in (2.18). To prove the second claim, note that by the Hölder inequality, $Es_{n1}^2 I[|s_{n1}| > n^{1/2}\lambda]$ is bounded above by

$$\lambda^{-\delta}n^{-\delta/2}Es_{n1}^\delta \le \lambda^{-\delta}n^{-\delta/2}E\Big(\Big[\int \Big|K_h(z-Z)\dot{\mu}_h(z)\Big|^{(2+\delta)/2}d\psi(z)\Big]^2 \cdot |\zeta|^{2+\delta}\Big).$$

By assumption (e3), this upper bound is seen to be of the order $O((nh^2)^{-\delta/2}) = o(1)$ by (h2), thereby proving the second claim in (2.18). The proof of (2.17) uses Lemma 2.4.3 and is similar to that of (4.6) of K-N, hence no details are given. □

Proof of Lemma 2.4.2. This proof is similar to that of Lemma 4.2 in K-N with obvious modifications. Details are left out for the sake of brevity.

Next, we shall show that the right hand side of (2.16) equals $R_n(\hat{\theta}_n - \theta_0)$, where $R_n = \Sigma_0 + o_p(1)$. Recall the notation at (2.13). The right hand side of (2.16) can be written as the sum $W_{n1} + W_{n2}$, where

$$\begin{aligned} W_{n1} &:= \|u_n\| \cdot \int \dot{\mu}_n(z,\hat{\theta}_n)\frac{1}{n}\sum_{i=1}^n K_h(z-Z_i)\frac{d_{ni}}{\|u_n\|}d\hat{\psi}_{h_2}(z), \\ W_{n2} &:= \int \dot{\mu}_n(z,\theta_0)\dot{\mu}_n'(z,\hat{\theta}_n)d\hat{\psi}_{h_2}(z) \cdot u_n. \end{aligned}$$

Observe that

$$n^{-1/2}\int E\|K_h(z-Z)\dot{H}_{\theta_0}(Z)\|^2 d\psi(z) = O(n^{-1/2}h^{-d}) = o(1). \qquad (2.19)$$

By (2.4), (2.19) and the assumptions (m4), (m5), we can show that $\|W_{n1}\| = o_p(\|u_n\|)$ and $W_{n2} = \Sigma_0 + o_p(1)$. This proves $R_n = \Sigma_0 + o_p(1)$.

Upon combining these results about the left hand side and the right hand side of (2.16), we obtain the following theorem.

Theorem 2.4.1 *Assume (e1)-(e3), (f1), (f2), (g), (k), (m1)-(m5), and (h3) hold. Then under* H_0,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \Sigma_0^{-1} n^{1/2} S_n + o_p(1).$$

Consequently, $\sqrt{n}(\hat{\theta}_n - \theta_0) \Longrightarrow N(0, \Sigma_0^{-1}\Sigma\Sigma_0^{-1})$, *where* Σ *and* Σ_0 *are defined in Lemma 2.4.1 and (m6) respectively.*

The above theorem shows that the asymptotic variance of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ consists of two parts. The part involving the element σ_ϵ^2 reflects the variation in the regression model, while the part involving the component τ^2 reflects the variation in the measurement error. This is the major difference between asymptotic distribution of the m.d. estimators discussed for the classical regression model in the K-N paper and for the Berkson model here.

2.5 Asymptotic Distribution of the Minimized Distance

This section contains a proof of the asymptotic distribution of the minimized distance $M_n(\hat{\theta}_n)$. Recall the notation in (2.3), the main result proved in this section is the following

Theorem 2.5.1 *Suppose (e1), (e2), (e4), (f1), (f2), (g), (k), (m1)-(m5) and (h3) hold. Then under* H_0, $nh^{d/2}(M_n(\hat{\theta}_n) - \hat{C}_n) \to_d \mathcal{N}_1(0, \Gamma)$. *Moreover* $|\hat{\Gamma}_n \Gamma^{-1} - 1| =$

$o_p(1)$.

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

Our proof of this theorem is facilitated by the following five lemmas.

Lemma 2.5.1 *Suppose (e1), (e2), (e4), (f1), (g), (k), (h1) and (h2) hold, then under* H_0,

$$nh^{d/2}(\tilde{M}_n(\theta_0) - \tilde{C}_n) \to_d N_1(0, \Gamma).$$

Lemma 2.5.2 *Suppose (e1), (e2), (f1),(k), (m3)-(m5) (h1) and (h2) hold, then under* H_0,

$$nh^{d/2}|M_n(\hat{\theta}_n) - M_n(\theta_0)| = o_p(1).$$

Lemma 2.5.3 *Suppose (e1), (e2), (f1), (f2), (k), (m3)-(m5) and (h3) hold, then under* H_0,

$$nh^{d/2}|M_n(\theta_0) - \tilde{M}_n(\theta_0)| = o_p(1).$$

Lemma 2.5.4 *Under the same conditions as in Lemma 2.5.3,*

$$nh^{d/2}|\hat{C}_n - \tilde{C}_n| = o_p(1).$$

Lemma 2.5.5 *Under the same conditions as in Lemma 2.5 2,* $\hat{\Gamma}_n - \Gamma = o_p(1)$, *Consequently, the positive definiteness of* Γ *implies* $|\hat{\Gamma}_n\Gamma^{-1} - 1| = o_p(1)$.

The proof of the Lemma 2.5.1 is facilitated by Theorem 1 of Hall (1984) which is reproduced here for the sake of completeness.

Theorem 2.5.2 *Let $\tilde{Z}_i, 1 \le i \le n$, be i.i.d. random vectors, and let*

$$U_n := \sum_{1 \le i < j \le n} H_n(\tilde{Z}_i, \tilde{Z}_j), \quad G_n(x, y) := E H_n(\tilde{Z}_1, x) H_n(\tilde{Z}_1, y),$$

where H_n is a sequence of measurable functions symmetric under permutation with

$$E[H_n(\tilde{Z}_1, \tilde{Z}_2)|\tilde{Z}_1] = 0, \quad EH_n^2(\tilde{Z}_1, \tilde{Z}_2) < \infty \quad \forall\, n \ge 1.$$

If, additionally,

$$\frac{EG_n^2(\tilde{Z}_1, \tilde{Z}_2) + n^{-1} EH_n^4(\tilde{Z}_1, \tilde{Z}_2)}{[EH_n^2(\tilde{Z}_1, \tilde{Z}_2)]^2} \longrightarrow 0, \quad \text{as } n \to \infty,$$

then U_n is asymptotically normally distributed with the mean 0 and the variance $\frac{n^2}{2} EH_n^2(\tilde{Z}_1, \tilde{Z}_2)$.

Proof of Lemma 2.5.1. Note that $\tilde{M}_n(\theta_0)$ can be written as the sum of $\tilde{C}_n$ and M_{n2}, where

$$M_{n2} := \frac{1}{n^2} \sum_{i \ne j} \int K_h(z - Z_i) K_h(z - Z_j) \zeta_i \zeta_j d\psi(z).$$

We shall prove that $nh^{d/2} M_{n2} \to_d \mathcal{N}_1(0, \Gamma)$ with the help of Theorem 2.5.2. Let $\tilde{Z}_i = (Z_i', \zeta_i)$ and $H_n(\tilde{Z}_i, \tilde{Z}_j) = n^{-1} h^{d/2} \int K_h(z - Z_i) K_h(z - Z_j) \zeta_i \zeta_j d\psi(z)$. Then,

$$nh^{d/2} M_{n2} = 2 \sum_{1 \le i < j \le n} H_n(\tilde{Z}_i, \tilde{Z}_j).$$

Observe that $H_n(\tilde{Z}_i, \tilde{Z}_j)$ is symmetric, $E[H_n(\tilde{Z}_1, \tilde{Z}_1)|\tilde{Z}_1] = 0$, and $EH_n^2(\tilde{Z}_1, \tilde{Z}_2)$ equals to

$$\frac{1}{n^2 h^d} \iint \Big[\int K(u) K\Big(\frac{y - x}{h} + u\Big) \sigma_\zeta^2(x - uh) f_Z(x - uh) du \Big]^2 d\psi(x) d\psi(y)$$

which is finite for each $n \geq 1$. Hence, to apply Theorem 2.5.2, it remains to show that

$$\frac{EG_n^2(\tilde{Z}_1, \tilde{Z}_2)}{[EH_n^2(\tilde{Z}_1, \tilde{Z}_2)]^2} \to 0, \qquad \frac{n^{-1}EH_n^4(\tilde{Z}_1, \tilde{Z}_2)}{[EH_n^2(\tilde{Z}_1, \tilde{Z}_2)]^2} \to 0. \tag{2.20}$$

But by the similar method as in K-N's paper, we can show that

$$EG_n^2(\tilde{Z}_1, \tilde{Z}_2) = O(n^{-4}h^d), \quad EH_n^4(\tilde{Z}_1, \tilde{Z}_2) = O(n^{-4}h^d). \tag{2.21}$$

$$\begin{aligned} & EH_n^2(\tilde{Z}_1, \tilde{Z}_2) \qquad (2.22) \\ = \; & \frac{h^d}{n^2} \iint \Big[\int K(u) K\Big(\frac{y-x}{h} + u\Big) \sigma_\zeta^2(x - uh) f_Z(x - uh) du \Big]^2 d\psi(x) d\psi(y) \\ = \; & O(n^{-2}). \end{aligned}$$

This verifies (2.20). By (2.22), the continuity of $\sigma_\zeta^2(z)$ and $f_Z(z)$, we obtain that $\frac{1}{2} n^2 EH_n^2(\tilde{Z}_1, \tilde{Z}_2)$ converges to

$$\begin{aligned} & \frac{1}{2} \iiiint K(u) K(w+u) K(v) K(v+w) (\sigma_\zeta^2(x))^2 f_Z^2(x) \frac{g^2(x)}{f_Z^4(x)} dx du dv dw \qquad (2.23) \\ = \; & \frac{1}{2} \int (\sigma_\zeta^2(x))^2 g(x) d\psi(x) \cdot \int \Big(\int K(u) K(w+u) du \Big)^2 dw. \end{aligned}$$

This completes the proof of Lemma 2.5.1. □

Proof of Lemma 2.5.2. Recall the definitions of $U_n(z)$ and $\mathcal{Z}_n(z, \theta)$ from (2.6). Add and subtract $H_{\theta_0}(Z_i)$ to the i-th summand inside the square integrand of $M_n(\hat{\theta}_n)$, to obtain that

$$M_n(\theta_0) - M_n(\hat{\theta}_n) = 2 \int U_n(z) \mathcal{Z}_n(z, \hat{\theta}_n) d\hat{\psi}_{h_2}(z) - \int \mathcal{Z}_n^2(z, \hat{\theta}_n) d\hat{\psi}_{h_2}(z) =: 2Q_1 - Q_2.$$

It thus suffices to show that

$$nh^{d/2} Q_1 = o_p(1), \quad nh^{d/2} Q_2 = o_p(1). \tag{2.24}$$

By subtracting and adding $(\hat{\theta}_n - \theta_0)'\dot{H}_{\theta_0}(Z_i)$ to the i-th summand of the second factor in Q_1, we can rewrite Q_1 as the sum of Q_{11} and Q_{12}, where

$$\begin{aligned} Q_{11} &:= \int U_n(z)\Big[\frac{1}{n}\sum_{i=1}^{n} K_h(z - Z_i)d_{ni}\Big]d\hat{\psi}_{h_2}(z), \\ Q_{12} &:= u_n' \int U_n(z)\dot{\mu}_n(z, \theta_0)d\hat{\psi}_{h_2}(z), \end{aligned}$$

where d_{ni} are as in (2.13). By (2.10), for for any $\eta > 0$, there exists a $k < \infty, N < \infty$, such that $P(A_n) \geq 1 - \eta$ for all $n > N$, where $A_n = \{(nh^d)^{1/2}\|\hat{\theta}_n - \theta_0\| \leq k\}$. By the Cauchy-Schwarz inequality, (2.4), (2.9), and the fact

$$\int \hat{f}_{Zw}^2(z)d\hat{\psi}_{h_2}(z) = O_p(1), \tag{2.25}$$

we obtain that on A_n, $nh^{d/2}|Q_{11}|$ is bounded above by

$$\sqrt{n}\|\hat{\theta}_n - \theta_0\| \cdot (nh^d)^{1/2} \sup_{1\leq i\leq n, (nh^d)^{1/2}\|\theta-\theta_0\|\leq k} \frac{d_{ni}}{\|u_n\|} \cdot O_p((nh^d)^{-1/2}).$$

This bound in turn is $o_p(1)$ by Theorem 2.4.1 and the assumption (m4). Hence to prove the first claim in (2.24), it remains to show that $nh^{d/2}|Q_{12}| = o_p(1)$. But Q_{12} can be written as the sum of Q_{121} and Q_{122}, where $Q_{121} = (\hat{\theta}_n - \theta_0)' \int U_n(z)\dot{\mu}_n(z, \hat{\theta}_n)d\hat{\psi}_{h_2}(z)$, $Q_{122} = -(\hat{\theta}_n - \theta_0)' \int U_n(z)[\dot{\mu}_n(z, \hat{\theta}_n) - \dot{\mu}_n(z, \theta_0)]d\hat{\psi}_{h_2}(z)$. Arguing as above, on the event A_n, $nh^{d/2}|Q_{122}|$ is bounded above by

$$n^2h^d\|\hat{\theta}_n - \theta_0\|^2 \cdot O_p((nh^d)^{-1}) \cdot \max_{1\leq i\leq n} \|\dot{H}_{\hat{\theta}_n}(Z_i) - \dot{H}_{\theta_0}(Z_i)\|^2 \cdot O_p(1) = o_p(1),$$

by (2.4), (2.9), (2.25) and assumptions (m4) and (h2). Next note that Q_{121} is the same as the expression in the left hand side of (2.16). Thus it is equal to

$$\begin{aligned} & u_n' \int \mathcal{Z}_n(z,\hat\theta_n)\,\dot\mu_n(z,\hat\theta_n)d\hat\psi_{h_2}(z) \\ = \; & u_n' \int \mathcal{Z}_n(z,\hat\theta_n)\,\dot\mu_n(z,\theta_0)d\hat\psi_{h_2}(z) \\ & + u_n' \int \mathcal{Z}_n(z,\hat\theta_n)\,[\dot\mu_n(z,\hat\theta_n) - \dot\mu_n(z,\theta_0)]d\hat\psi_{h_2}(z) \\ =: \; & D_1 + D_2. \end{aligned}$$

By Cauchy-Schwarz inequality, (2.4), (2.25), assumption (m1) and the compactness of Θ, $nh^{d/2}|D_1| \le nh^{d/2}\|\hat\theta_n - \theta_0\| O_P(1) = O_P(h^{d/2}) = o_P(1)$ by Theorem 2.4.1 and (h2). Similarly, one can show that $nh^{d/2}|D_2| \le nh^{d/2}\|\hat\theta_n - \theta_0\| o_P(1) = o_P(h^{d/2}) = o_P(1)$. This completes the proof of the first claim in (2.24).

The proof of the second claim in (2.24) is similar. Details are left out for the sake of brevity. □

Proof of Lemma 2.5.3. Note that

$$\begin{aligned} & nh^{d/2}|M_n(\theta_0) - \tilde M_n(\theta_0)| \\ = \; & nh^{d/2}\Big|\int \Big[\frac{1}{n}\sum_{i=1}^{n} K_h(z-Z_i)\zeta_i\Big]^2\Big(\frac{1}{\hat f_{Zw}^2(z)} - \frac{1}{f_Z^2(z)}\Big)dG(z)\Big| \\ \le \; & nh^{d/2}\cdot O_P((nh^d)^{-1})\cdot O_P((\log_k n)\cdot(\log n/n)^{2/(d+4)}) = o_P(1) \end{aligned}$$

by (2.9) and Lemma 2.4.3. Hence the lemma.

Proof of Lemma 2.5.4. For convenience, let $t_i := H_{\hat\theta_n}(Z_i) - H_{\theta_0}(Z_i)$, $\Delta_n(z) := \dfrac{f_Z^2(z)}{\hat f_{Zw}^2(z)} - 1$, then one obtains

$$\hat C_n = \frac{1}{n^2}\sum_{i=1}^{n}\int K_h^2(z-Z_i)\,\hat\zeta_i^{\,2}d\hat\psi_{h_2}(z) = \frac{1}{n^2}\sum_{i=1}^{n}\int K_h^2(z-Z_i)(\zeta_i - t_i)^2 d\hat\psi_{h_2}(z),$$

it can written as the sum of A_{n1} and A_{n2}, where

$$A_{n1} := \frac{1}{n^2}\sum_{i=1}^{n}\int K_h^2(z-Z_i)(\zeta_i - t_i)^2 d\psi(z)$$
$$A_{n2} := \frac{1}{n^2}\sum_{i=1}^{n}\int K_h^2(z-Z_i)(\zeta_i - t_i)^2 \Delta_n(z) d\psi(z).$$

In order to prove the lemma, it suffice to show that

$$nh^{d/2}(A_{n1} - \tilde{C}_n) = o_p(1), \quad nh^{d/2}A_{n2} = o_p(1). \tag{2.26}$$

By expanding the term $(\zeta_i - t_i)^2$ in A_{n1} and noting that $\max|t_i|^2 = O_p((nh^d)^{-1})$ by (m4) and (2.10), the first claim in (2.26) follows the similar argument as in K-N.

To prove the second claim in (2.26), note that A_{n2} can be written as

$$\begin{aligned}
A_{n2} &= \frac{1}{n^2}\sum_{i=1}^{n}\int K_h^2(z-Z_i)(\zeta_i - t_i)^2\Delta_n(z)d\psi(z) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\int K_h^2(z-Z_i)\zeta_i^2\Delta_n(z)d\psi(z) + \frac{1}{n^2}\sum_{i=1}^{n}\int K_h^2(z-Z_i)t_i^2\Delta_n(z)d\psi(z) \\
&\quad -\frac{2}{n^2}\sum_{i=1}^{n}\int K_h^2(z-Z_i)\zeta_i t_i\Delta_n(z)d\psi(z).
\end{aligned}$$

But all the three terms on the right hand side are of the order $o_p((nh^{d/2})^{-1})$. Thereby completing the proof of the second claim of (2.26), and hence that of the lemma. □

Proof of Lemma 2.5.5. Define

$$\begin{aligned}
\tilde{\Gamma}_n &:= 2n^{-2}h^d\sum_{i\neq j}\Big(\int K_h(z-Z_i)K_h(z-Z_j)\zeta_i\zeta_j d\psi(z)\Big)^2 = 2\sum_{i\neq j}H_n^2(\tilde{Z}_i,\tilde{Z}_j), \\
\Gamma_n &:= 2h^d(n-1)n^{-1}\int\int[EK_h(x-Z)K_h(y-Z)\sigma_\zeta^2(Z)]^2 d\psi(x)d\psi(y).
\end{aligned}$$

We shall prove

$$\hat{\Gamma}_n - \tilde{\Gamma}_n = o_p(1), \quad \tilde{\Gamma}_n - \Gamma_n = o_p(1), \quad \Gamma_n - \Gamma = o_p(1). \tag{2.27}$$

Note that $\hat{\Gamma}_n$ can be rewritten as the sum of the following three terms:

$$\begin{aligned} B_1 &:= 2n^{-2}h^d \sum_{i \neq j} \Big(\int K_h(z - Z_i) K_h(z - Z_j)(\zeta_i - t_i)(\zeta_j - t_j) d\psi(z) \Big)^2, \\ B_2 &:= 2n^{-2}h^d \sum_{i \neq j} \Big(\int K_h(z - Z_i) K_h(z - Z_j)(\zeta_i - t_i)(\zeta_j - t_j) \Delta_n(z) d\psi(z) \Big)^2, \\ B_3 &:= 4n^{-2}h^d \sum_{i \neq j} \Big(\int K_h(z - Z_i) K_h(z - Z_j)(\zeta_i - t_i)(\zeta_j - t_j) d\psi(z) \cdot \\ &\qquad \int K_h(z - Z_i) K_h(z - Z_j)(\zeta_i - t_i)(\zeta_j - t_j) \Delta_n(z) d\psi(z) \Big). \end{aligned}$$

So, to prove the first claim in (2.27), it suffices to show that

$$B_1 - \tilde{\Gamma}_n = o_p(1), \quad B_2 = o_p(1) = B_3. \tag{2.28}$$

By taking the expectation, Fubini and usual calculation one can obtain

$$n^{-2}h^d \sum_{i \neq j} \Big(\int K_h(z - Z_i) K_h(z - Z_j) |\zeta_i||\zeta_j| d\psi(z) \Big)^2 = O_p(1), \tag{2.29}$$

$$n^{-2}h^d \sum_{i \neq j} \Big(\int K_h(z - Z_i) K_h(z - Z_j) |\zeta_i| d\psi(z) \Big)^2 = O_p(1), \tag{2.30}$$

$$n^{-2}h^d \sum_{i \neq j} \Big(\int K_h(z - Z_i) K_h(z - Z_j) d\psi(z) \Big)^2 = O_p(1). \tag{2.31}$$

Furthermore, we also have

$$\sup_{z \in \mathcal{I}} \Delta_n(z) = o_p(1), \qquad \max_{1 \leq i \leq n} |t_i| = o_p(1) \tag{2.32}$$

by (2.4), (m4) and (2.10). By expanding $(\zeta_i - t_i)(\zeta_j - t_j)$ and the quadratic terms in B_1, we have,

$$|B_1 - \tilde{\Gamma}_n| \leq \frac{2h^d}{n^2} \sum_{i \neq j} \Big(\int K_h(z - Z_i) K_h(z - Z_j)(|t_i t_j| + |\zeta_i t_j| + |\zeta_j t_i|) d\psi(z) \Big)^2$$

$$+4n^{-2}h^d \sum_{i\neq j} \Big(\int K_h(z-Z_i)K_h(z-Z_j)|\zeta_i\zeta_j| d\psi(z)$$

$$\times \int K_h(z-Z_i)K_h(z-Z_j)(|t_it_j| + |\zeta_it_j| + |\zeta_jt_i|)d\psi(z)\Big)$$

$$:= B_{n1} + B_{n2}.$$

By (2.30), (2.31) and (2.32), one has $B_{n1} = o_P(1)$, and $B_{n2} = o_P(1)$. Hence $|B_1 - \tilde{\Gamma}_n| = o_P(1)$.

Next, consider B_2. Note that

$$B_2 \le 2\sup_{z\in\mathcal{I}} \Delta_n(z)\cdot n^{-2}h^d \sum_{i\neq j} \Big(\int K_h(z-Z_i)K_h(z-Z_j)|\zeta_i - t_i||\zeta_j - t_j| d\psi(z)\Big)^2$$

which is of the order $o_P(1)$ by the inequality $|\zeta_i - t_i|\cdot|\zeta_j - t_j| \le |\zeta_i\zeta_j| + (|t_it_j| + |\zeta_it_j| + |\zeta_jt_i|)$ and expanding the quadratic terms, and by (2.32), (2.29), and the results that $B_{12} = o_P(1)$, $B_{13} = o_P(1)$. Finally, again an application of the Cauchy-Schwarz inequality to the double sum yields $B_3 = o_P(1)$. This completes the proof of (2.28) and hence that of the first claim in (2.27).

To prove the second claim in (2.27), note that $\Gamma_n = E\tilde{\Gamma}_n$. Hence, with $C_{ij} = \int K_h(z-Z_1)K_h(z-Z_2)\zeta_i\zeta_j d\psi(z)$, one obtains

$$\begin{aligned}
&E[\tilde{\Gamma}_n - \Gamma_n]^2 \\
&= \frac{4h^{2d}}{n^4} E[\sum_{i\neq j}(C_{ij}^2 - EC_{ij}^2)]^2 \le 4n^{-4}h^{2d}\sum_{i\neq j} E(C_{ij}^4) + 4n^{-4}h^{2d}\sum_{k\neq j\neq l} EC_{kj}^2C_{kl}^2 \\
&= 4\sum_{i\neq j} EH_n^4(\tilde{Z}_i, \tilde{Z}_j) + 4\sum_{k\neq j\neq l} EH_n^2(\tilde{Z}_k, \tilde{Z}_j)H_n^2(\tilde{Z}_k, \tilde{Z}_l) \\
&\le 4(n^2 + n^3)\, EH^4(\tilde{Z}_i, \tilde{Z}_j) = O(n^{-1}h^d) = o(1),
\end{aligned}$$

by (2.21) and (h1), thereby proving the second claim in (2.27).

The third claim in (2.27) is easily obtained from the following fact,

$$\begin{aligned}\Gamma_n &= 2h^d(n-1)n^{-1}\iint [EK_h(x-Z)K_h(y-Z)\sigma_\zeta^2(Z)]^2 d\psi(x)d\psi(y)\\ &= 2n(n-1)\,EH_n^2(\tilde{Z}_1,\tilde{Z}_2) = \frac{4n(n-1)}{n^2}\cdot\frac{n^2}{2}EH_n^2(\tilde{Z}_1,\tilde{Z}_2) \to \Gamma,\end{aligned}$$

by (2.23). This completes the proof of Lemma 2.5.5. □

2.6 Simulations

This section contains results of two simulation studies corresponding to the following cases: Case 1: $d = q = 1$ and m_θ linear; Case 2: $d = q = 2$, and m_θ nonlinear. In each case the Monte Carlo average values of $\hat{\theta}_n$, MSE($\hat{\theta}_n$), empirical levels and powers of the m.d. test are reported. The asymptotic level is taken to be 0.05 in all cases.

In the first case $\{Z_i\}_{i=1}^n$ are obtained as a random sample from the uniform distribution on $[-1,1]$, $\{\epsilon_i\}_{i=1}^n$ and $\{\eta_i\}_{i=1}^n$ are obtained as two independent random samples from $\mathcal{N}_1(0,(0.1)^2)$. Then (X_i, Y_i) are generated using the model $Y_i = \mu(X_i)+\epsilon_i$, $X_i = Z_i + \eta_i$, $i = 1, 2, \cdots, n$.

The kernel function and the band widths used in the simulation are

$$K(z) = K^*(z) = \frac{3}{4}(1-z^2)I(|z| \le 1), h = an^{-1/3},\ w = bn^{-1/5}(\log n)^{1/5},$$

with some choices for a and b. The integrating measure G is taken to be the uniform measure on $[-1,1]$.

The parametric model is taken to be $m_\theta(x) = \theta x$, $x, \theta \in \mathbb{R}$, $\theta_0 = 1$. Then, $H_\theta(z) = \theta z$. In this case various calculations simplify as follows. By taking the derivative of $M_n(\theta)$ in θ and solving the equation of $\partial M_n(\theta)/\partial\theta = 0$, we obtain $\hat{\theta}_n = A_n/B_n$, where

$$A_n = \int_{-1}^{1} \Big[\sum_{i=1}^{n} K_h(z-Z_i)Y_i\Big] \cdot \Big[\sum_{i=1}^{n} K_h(z-Z_i)Z_i\Big] \cdot \Big[\sum_{i=1}^{n} K_w(z-Z_i)\Big]^{-2} dz,$$

$$B_n = \int_{-1}^{1} \Big[\sum_{i=1}^{n} K_h(z-Z_i)Z_i\Big]^2 \cdot \Big[\sum_{i=1}^{n} K_w(z-Z_i)\Big]^{-2} dz.$$

Then,

$$M_n(\hat{\theta}_n) = \int_{-1}^{1} \Big(\sum_{i=1}^{n} K_h(z-Z_i)(Y_i - \hat{\theta}_n Z_i)\Big)^2 \cdot \Big(\sum_{i=1}^{n} K_w(z-Z_i)\Big)^{-2} dz$$

$$\hat{C}_n = \int_{-1}^{1} \Big(\sum_{i=1}^{n} K_h^2(z-Z_i)(Y_i - \hat{\theta}_n Z_i)^2\Big) \cdot \Big(\sum_{i=1}^{n} K_w(z-Z_i)\Big)^{-2} dz.$$

The value of the test statistic is calculated by $\widehat{\mathcal{D}}_n := nh^{d/2}\hat{\Gamma}_n^{-1/2}(M_n(\hat{\theta}_n) - \hat{C}_n)$.

Table 2.1 reports the Monte Carlo mean and the MSE($\hat{\theta}_n$) under H_0 for the sample sizes 50, 100, 200, 500, each repeated 1000 times. One can see there appears to be little bias in $\hat{\theta}_n$ for all chosen sample sizes and as expected, the MSE decreases as the sample size increases. To assess the level and power behavior of the $\widehat{\mathcal{D}}_n$ test,

Sample Size	50	100	200	500
Mean	1.0003	0.9987	1.0006	0.9998
MSE	0.0012	0.0006	0.0003	0.0001

Table 2.1: Mean and MSE of $\hat{\theta}_n$, $d = 1, q = 1$

we chose the following four models to simulate data from. In each of these cases $X_i = Z_i + \eta_i$.

$$\text{Model } 0: \; Y_i = X_i + \epsilon_i,$$

$$\text{Model } 1: \; Y_i = X_i + 0.3X_i^2 + \epsilon_i,$$

$$\text{Model } 2: \; Y_i = X_i + 1.4\exp(-0.2X_i^2) + \epsilon_i,$$

$$\text{Model } 3: \; Y_i = X_i I(X_i \geq 0.2) + \epsilon_i.$$

To assess the effect of the choice of (a, b) that appear in the bandwidths on the level and power, we ran the simulations for numerous choices of (a, b), ranging from 0.2 to 1. Table 2.2 reports the simulation results pertaining to $\widehat{\mathcal{D}}_n$ for three choices of (a, b). The simulation results for the other choices were similar to those reported here. Data from Model 0 in this table are used to study the empirical sizes, and from Models 1 to 3 are used to study the empirical powers of the test. These entities are obtained by computing $\#\{|\widehat{\mathcal{D}}_n| \geq 1.96\}/1000$.

From Table 2.2, one sees that the empirical level is sensitive to the choice of (a, b) for moderate sample sizes ($n \leq 200$) 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 we developed 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.

Now consider the case 2 where $d = 2$, $q = 2$ and $\{m_\theta(x) = \theta_1 x_1 + \exp(\theta_2 x_2),\ \theta = (\theta_1, \theta_2)^T \in \mathbb{R}^2,\ x_1, x_2 \in \mathbb{R}$. Accordingly, here $H_\theta(z) = \theta_1 z_1 + \exp(\theta_2 z_2 + 0.005\theta_2^2)$. The true $\theta_0 = (1, 2)'$ was used in the simulations.

		Sample size			
Model	a, b	50	100	200	500
Model 0	0.3, 0.2	0.007	0.026	0.028	0.048
	0.5, 0.5	0.014	0.022	0.040	0.051
	1.0, 1.0	0.021	0.020	0.031	0.043
Model 1	0.3, 0.2	0.754	0.987	1.000	1.000
	0.5, 0.5	0.945	1.000	1.000	1.000
	1.0, 1.0	1.000	1.000	1.000	1.000
Model 2	0.3, 0.2	0.857	0.996	1.000	1.000
	0.5, 0.5	0.999	1.000	1.000	1.000
	1.0, 1.0	1.000	1.000	1.000	1.000
Model 3	0.3, 0.2	0.874	0.993	1.000	1.000
	0.5, 0.5	1.000	1.000	1.000	1.000
	1.0, 1.0	1.000	1.000	1.000	1.000

Table 2.2: Levels and powers of the M.D. test, $d = 1, q = 1$

In all models below, $\{Z_i = (Z_{1i}, Z_{2i})^T\}_{i=1}^n$ are obtained as a random sample from the uniform distribution on $[-1,1]^2$, $\{\varepsilon_i\}_{i=1}^n$ are obtained from $\mathcal{N}_1(0,(0.1)^2)$, and $\{\eta_i = (\eta_{1i}, \eta_{2i})^T\}_{i=1}^n$ are obtained from the bivariate normal distribution with mean vector 0 and the diagonal covariance matrix with both diagonal entries equal to $(0.1)^2$. We simulated data from the following four models, where $X_i = Z_i + \eta_i$.

Model 0: $Y_i = X_{1i} + \exp(2X_{2i}) + \epsilon_i$,

Model 1: $Y_i = X_{1i} + \exp(2X_{2i}) + 1.4X_{1i}^2 + 1 + \epsilon_i$,

Model 2: $Y_i = X_{1i} + \exp(2X_{2i}) + 1.4X_{1i}^2X_{2i}^2 + \epsilon_i$,

Model 3: $Y_i = X_{1i} + \exp(2X_{2i}) + 1.4(\exp(-0.2X_{1i}) + \exp(0.7X_{2i}^2)) + \epsilon_i$.

The kernel function and the bandwidths used in the simulation were taken to be

$$\begin{aligned} K(z) &= K^*(z) = \frac{9}{16}(1-z_1^2)(1-z_2^2)I(|z_1| \le 1, |z_2| \le 1), \\ h &= n^{-1/4.5}, \quad w = n^{-1/6}(\log n)^{1/6}. \end{aligned}$$

The sample sizes chosen are 50, 100, 200 and 300, each repeated 1000 times. Table 2.3 lists the means and the MSE of the estimator $\hat{\theta}_n = (\hat{\theta}_{n1}, \hat{\theta}_{n2})'$ which are obtained by minimizing $M_n(\theta)$ and employing the Newton-Raphson algorithm. As in the case 1, one sees little bias in the estimator for all chosen sample sizes.

Table 2.4 gives the empirical sizes and powers for testing Model 0 against Models 1 - 3. The entries in Table 2.4 corresponding to Model 0 are used to study the empirical size of the m.d. test, and the entries from Models 1 - 3 are used to study the empirical power of the test. From this table one sees that our m.d. test is conservative when the sample sizes are small, while the sizes do increase with the sample sizes and indeed preserve the nominal size 0.05. It also shows that the m.d. test performs well for sample sizes larger than 200 at all alternatives.

Sample Size	50	100	200	300
Mean of $\hat{\theta}_{n1}$	0.9978	0.9973	0.9974	0.9988
MSE of $\hat{\theta}_{n1}$	0.0190	0.0095	0.0053	0.0034
Mean of $\hat{\theta}_{n2}$	1.9962	1.9965	2.0013	2.0004
MSE of $\hat{\theta}_{n2}$	0.0063	0.0028	0.0014	0.0010

Table 2.3: Mean and MSE of $\hat{\theta}_n$, $d = 2, q = 2$

Sample size	50	100	200	300
Model 0	0.003	0.019	0.049	0.052
Model 1	0.158	0.843	0.979	0.996
Model 2	0.165	0.840	0.976	0.992
Model 3	0.044	0.608	0.954	0.997

Table 2.4: Levels and powers of the M.D. test, $d = 2, q = 2$

BIBLIOGRAPHY

[1] An, H.Z., Cheng, B., (1991). A Kolmogorov-Smirnov type statistic with application to test for nonlinearity in time series. Int. Statist. Rev. **59**, 287-307.

[2] Anderson,T.W. (1984). Estimating Linear Statistical Relationships. *Ann. Statist.* **12** 1-45.

[3] Beran, R.J. (1977). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.* **5** 445-463.

[4] Berkson, J. (1950). Are these two regressions? *J. Amer. Statist. Assoc.* **5** 164-180.

[5] Bickel, P.J. & Ritov,Y. (1987). Efficient Estimation in the Errors in Variables Model. *Ann. Statist.* **15**, 2, 513-540.

[6] Bosq, D. (1998). *Nonparametric statistics for stochastic processes: Estimation and Prediction*, 2nd edition. Springer Lecture Notes in Statistics, **110**. Springer-Verlag, New York, Inc.

[7] Carroll,R.J. & Hall,P. (1988). Optimal rates of convergence for deconvoluting a density. *JASA.* **83** 1184-1185.

[8] Carroll,R.J. & Spiegelman,C.H. (1992). Diagnostics for nonlinearity and heteroscedasticity in errors in variables regression. *Technometrics* **34** 186-196.

[9] Carroll, R.J., Ruppert, D. and Stefanski, L.A. (1995). *Measurement Error in Nonlinear Models*, Chapman & Hall/CRC, Boca Raton.

[10] Cheng, C. and Van Ness, J.W.(1999). Statistical regression with measurement error. Arnold, London.

[11] Cheng, C.L. and Kukush, A.G.(2004). A goodness-of- fit test for a polynomial errors-in-variables model, **56** 641-661.

[12] Elias Masry (1993). Strong consistency and rates for deconvolution of multivariate densities of stationary process. *Stochastic Processes and their Applications* **47** 53-74.

[13] Eubank,R.L., Hart, J.D., (1992). Testing the goodness of fit in regression via order selection criteria. *Ann. Statist.* **20** 1412-1425.

[14] Eubank,R.L., Hart, J.D., (1993). Commonality of CUMSUM, von Neumann and smoothing based goodness-of-fit tests. *Biometrika* **80** 89-98.

[15] Eubank,R.L., Spiegelman, C.H., (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques. *J. Amer. Statist. Assoc.* **85** 387-392.

[16] Fan,J. (1991a). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257-1272.

[17] Fan,J. (1991b). Asymptotic normality for deconvolution kernel density estimators. *Sankhyā Ser. A.* **53** 97-110.

[18] Fan,J. & Truong, K.T. (1993). Nonparametric regression with errors in variables. *Ann. Statist.* **21** 1900-1925.

[19] Fuller, W.A. (1987). *Measurement Error Models.*Wiley, New York.

[20] Gleser,L.J. (1981). Estimation in a Multivariate "Errors in Variables" Regression Model: Large Sample Results. *Ann. Statist.* **9**, **1**, 24-44.

[21] Hart, J.D. (1997). *Nonparametric smoothing and lack-of-fit tests.* Springer-Verlag, New York, Inc.

[22] Huwang, L. and Huang, Y.H.S. (2000). On errors-in-variables in polynomial regression - Berkson case. *Statist. Sinica.* **10**, 923-936.

[23] Koul, Hira L. and Pingping Ni (2004). Minimum distance regression model checking, *J. Stat. Plann. Inference* **119**, No.1, 109-141.

[24] Mack, Y.P. and Silverman, B.W. (1982). Weak and strong uniform consistency of kernel regression estimates, *Z. Wahrsch. Gebiete* **61**, 405-415.

[25] Rudemo, M., Ruppert, D. and Streibig, J. (1989). Random effect models in nonlinear regression with applications to bioassay. *Biometrics.* **45** 349-362.

[26] Stute, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25** 613-641.

[27] Stefanski, L.A., and Carroll, R.J. (1991). Deconvolution-based score tests in measurement error models. *The Annals of Statistics* **19** 249-259.

[28] Stute, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25** 613-641.

[29] Stute, W., Thies, S., Zhu, L.X. (1998). Model checks for regression: an innovation process approach. *Ann. Statist.* **26**, 1916-1934.

[30] Wang, L. (2003). Estimation of nonlinear Berkson-type measurement errors models. *Statist. Sinica.* **13**, 1201-1210.

[31] Wang, L. (2004). Estimation of nonlinear models with Berkson measurement errors.*Ann. Statist.* **32**, 6, 2559-2579.

[32] Wolfowitz, J., (1953). Estimation by the minimum distance method. *Ann. Inst. Statist. Math.*, Tokyo, **5** 9-23.

[33] Wolfowitz, J., (1954). Estimation by the minimum distance method in nonparametric stochastic difference equation. *Ann. Math. Statist.*, **25**, 203-217.

[34] Wolfowitz, J., (1957). The minimum distance method. *Ann. Math. Statist.*, **28**, 75-88.

[35] Zheng, J.X., (1996). A consistent test of functional form via estimation technique. *J. Econometrics.*, **75**, 263-289.

[36] Zhu,L.X., Song,W.X., & Cui,H.J. (2003). Testing lack-of-fit for a polynomial errors-in-variables model. *Acta Math. Appl. Sin. Engl. Ser.* **19** 353-362.