TESTING OF REGRESSION FUNCTIONS WHEN RESPONSES ARE MISSING AT RANDOM

By

Xiaoyu Li

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Statistics

2012

ABSTRACT

TESTING OF REGRESSION FUNCTIONS WHEN RESPONSES ARE MISSING AT RANDOM

By

Xiaoyu Li

This thesis consists two chapters. The first chapter proposes a class of minimum distance tests for fitting a parametric regression model to a regression function when some responses are missing at random. These tests are based on a class of minimum integrated square distances between a kernel type estimator of a regression function and the parametric regression function being fitted. The estimators of the regression function are based on two completed data sets constructed by imputation and inverse probability weighting methods. The corresponding test statistics are shown to have asymptotic normal distributions under null hypothesis. Some simulation results are also presented.

The second chapter considers the problem of testing the equality of two nonparametric regression curves against a one-sided alternatives based on two samples with possibly distinct design and error densities, when responses are missing at random. This chapter proposes a class of tests using imputation and covariate matching. The asymptotic distributions of these test statistics are shown to be Gaussian under null hypothesis and a class of local nonparametric alternatives. The consistency of these tests against a large class of fixed alternatives is also established. This chapter also includes a simulation study, which assesses the finite sample behavior of a member of this class of tests.

ACKNOWLEDGMENTS

I would like to sincerely and gratefully thank my advisor Professor Hira L. Koul for his excellent guidance and great patience during the past five years. He sets up a career model for me with his great enthusiasm towards science, serious attitude, hard working, and extraordinary kindness to students. His love of statistics and mathematics encourages me to keep working in my research.

I also wish to thank Professor Vidyadhar S. Mandrekar, Yimin Xiao, David Todem for serving on my dissertation committee. Special thanks to Professor Lijian Yang and Yimin Xiao for their help in my graduate study and life. Many thanks to Professor Vidyadhar S. Mandrekar for his teaching and encouragement, and to Professor James Stapleton for his help since the first day I came to Michigan State University. I am grateful to the Department of Statistics and Probability for offering assistantship to support me to complete graduate studies.

Finally, I would like to thank my family for their love which enables me to complete this work and pursue my career goal.

This research is supported by the NSF DMS Grant 0704130, under the P.I.: Professor Hira L. Koul.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

Minimum Distance Regression Model Checking when Responses are Missing At Random

1.1 Introduction

In this chapter, we discuss a class of minimum distance tests for fitting a parametric model to the regression function based on imputation and inverse probability weighting method, when responses are missing at random. To be specific, let X be an explanatory variable of dimension d with $d \geq 1$, Y be a response variable of dimension 1 with $E|Y| < \infty$, δ be an indicator for whether the response is missing or observed, i.e. $\delta = 1$ if Y is observed, and $\delta = 0$ if Y is missing. The missing mechanism of the data is missing at random, in which δ and Y are conditionally independent, given X, i.e. $P(\delta = 1|Y, X) = P(\delta = 1|X)$, a.s.; see

Little and Rubin (1987). Let

$$\mu(x) = E(Y|X = x), \quad x \in \mathbb{R}^d,$$

denote the regression function. Consider the regression model

$$Y = \mu(X) + \varepsilon \tag{1.1}$$

with response missing at random. Let $\{m_\theta(\cdot) : \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^q$, be a given parametric model and $\mathcal{I}$ be a compact subset of $\mathbb{R}^d$. The problem of interest is to test the hypothesis

$$\begin{aligned} &\mathrm{H}_0 : \mu(\mathrm{x}) = \mathrm{m}_{\theta_0}(\mathrm{x}) \qquad \text{for some } \theta_0 \in \Theta, \text{ and for all } \mathrm{x} \in \mathcal{I}, \\ &\mathrm{H}_1 : \mathrm{H}_0 \text{ is not true,} \end{aligned}$$

based on the random sample $\{(X_i, \delta_i Y_i) : \; i = 1, 2, \cdots, n\}$ from the distribution of $(X, \delta Y)$ in model (1.1). One is also interested in finding the parameter $\theta \in \Theta$ that best fits the data under the null hypothesis.

Regression model checking when data are completely observed is a classical problem in statistics. Many interesting results are available, see, e.g., Eubank and Spiegelman (1990), Eubank and Hart (1992, 1993), Härdle and Mammen (1993), Zheng (1996), Hart (1997), Stute et al. (1998), Koul and Ni (2004), Koul and Song (2009), Koul (2011), among others. Hart (1997) summarized numerous testing procedures. Koul and Ni (2004) (K-N) proposed a class of tests based on certain minimized L_2 distances between a nonparametric estimator of the regression function and the parametric model being fitted. They proved asymptotic normality of the minimum distance estimators and the proposed test statistics under the fitted model, and consistency of the proposed tests against a class of fixed alternatives. Koul and Song (2009) extended this minimum distance methodology to the regression model with Berkson measurement errors. They also obtained the asymptotic power of the proposed tests against a class of local alternatives. Koul (2011) implemented the minimum distance methodology on classical regression model with design non-random and uniform on $[0, 1]$. Sun and Wang (2009) considered the model checking problem when data are missing at random. They constructed complete data sets by imputation and inverse probability weighting methods, and proposed two score-type and two empirical process based test statistics. The asymptotic behaviors of these test statistics were investigated under the null hypothesis and local alternatives.

In this chapter we focus on adapting the minimum distance testing method of K-N to missing data at random setup when the data are completed by the imputation and inverse probability weighting methods. To describe the testing procedure, we need to estimate $\mu(x)$. Since, under H_0, μ is parametric, we only need to estimate θ_0 at $\sqrt{n}$-consistent rate. Let $\hat{\alpha}_n$ be such an estimator of θ_0 based on the random sample. A suggested choice of $\hat{\alpha}_n$ is given in Theorem 1.4.1, Section 1.4 below. Let $\tilde{K}$ be a symmetric kernel function on $[-1, 1]^d$, $b = b_n$ be a bandwidth sequence of positive numbers, $\tilde{K}_b(y) := b_n^{-d}\tilde{K}(y/b_n)$, $y \in \mathbb{R}^d$; and let for $x \in \mathbb{R}^d$,

$$\Delta(x) = P(\delta = 1|X = x), \qquad \text{and} \qquad \hat{\Delta}_n(x) = \frac{\sum_{i=1}^n \delta_i \tilde{K}_b(x - X_i)}{\sum_{i=1}^n \tilde{K}_b(x - X_i)}.$$

Note that $\hat{\Delta}_n(x)$ is the Nadaraya-Watson kernel estimator of $\Delta(x)$. We construct two complete data sets $\{(X_i, \hat{Y}_{ij}), i = 1, \cdots, n\}$, $j = 1, 2$, by imputation and inverse probability weighting methods, respectively, where

$$\hat{Y}_{i1} = \delta_i Y_i + (1-\delta_i) m_{\hat{\alpha}_n}(X_i), \qquad i = 1, \cdots, n; \tag{1.2}$$

$$\hat{Y}_{i2} = \frac{\delta_i}{\hat{\Delta}_n(X_i)} Y_i + \Big(1 - \frac{\delta_i}{\hat{\Delta}_n(X_i)}\Big) m_{\hat{\alpha}_n}(X_i), \qquad i = 1, \cdots, n. \tag{1.3}$$

To proceed further, let K and K^* be kernel functions on $[-1,1]^d$, $h = h_n$ and $w = w_n$ be window width sequences of positive numbers, G be a σ-finite measure on $\mathbb{R}^d$ with Lebesgue density g. Assume the design variable X has a uniformly continuous Lebesgue density f that is bounded from below on $\mathcal{I}$. Define

$$\hat{f}_h(x) = n^{-1} \sum_{i=1}^n K_h(x - X_i), \qquad \hat{f}_w(x) = n^{-1} \sum_{i=1}^n K^*_w(x - X_i), \quad x \in \mathbb{R}^d,$$

where $h_n \sim n^{-a}$ with $0 < a < \min(1/(2d), 4/(d(d+4)))$, and $w_n \sim (\log n/n)^{1/(d+4)}$. Adaptive versions of the L_2 distances proposed in K-N in the current setup are

$$\hat{T}_{nj}(\theta) = \int_{\mathcal{I}} \Big[n^{-1} \sum_{i=1}^n K_h(x - X_i)(\hat{Y}_{ij} - m_\theta(X_i)) \Big]^2 \{\hat{f}_w(x)\}^{-2} dG(x), \quad \theta \in \mathbb{R}^q,$$

and the corresponding minimum distance estimators are

$$\hat{\theta}_{nj} := \arg\min_{\theta \in \Theta} \hat{T}_{nj}(\theta), \qquad j = 1, 2.$$

The proposed tests of H_0 are to be based on $\hat{T}_{nj}(\hat{\theta}_{nj})$, $j = 1, 2$.

To proceed further, we need more notation. Let

$$\hat{\varepsilon}_{ij} := \hat{Y}_{ij} - m_{\hat{\theta}_{nj}}(X_i), \quad j = 1, 2, \tag{1.4}$$

$$\hat{C}_{nj} := n^{-2} \sum_{i=1}^{n} \int_{\mathcal{I}} K_h^2(x - X_i)\hat{\varepsilon}_{ij}^2 \{\hat{f}_w(x)\}^{-2} dG(x),$$

$$\hat{\Gamma}_{nj} := 2h^d n^{-2} \sum_{i \neq k} \Big(\int_{\mathcal{I}} K_h(x - X_i) K_h(x - X_k) \hat{\varepsilon}_{ij} \hat{\varepsilon}_{kj} \{\hat{f}_h(x)\}^{-2} dG(x) \Big)^2,$$

$$\hat{\mathcal{D}}_{nj} := nh^{d/2}(\hat{T}_{nj}(\hat{\theta}_{nj}) - \hat{C}_{nj})/\hat{\Gamma}_{nj}^{1/2}, \quad j = 1, 2.$$

For each $j = 1, 2$, the proposed test rejects H_0 whenever $|\hat{\mathcal{D}}_{nj}|$ is large. Asymptotic normality of $n^{1/2}(\hat{\theta}_{nj} - \theta_0)$ and $\hat{\mathcal{D}}_{nj}$, $j = 1, 2$, under H_0 are established in Section 1.4 and Section 1.5, respectively. Consistency of $\hat{\theta}_{nj}$, $j = 1, 2$, under H_0 is given in Section 1.3. Assumptions and preliminary lemmas needed to prove all these results are stated in Section 1.2, while Section 1.6 is devoted to simulation studies.

In the sequel, we write h for h_n, w for w_n, and b for b_n; the integrals with respect to the G-measure are understood to be over the set $\mathcal{I}$; all limits are taken as $n \to \infty$, unless specified otherwise; for any two sequences of real numbers a_n and b_n, notation $a_n \sim b_n$ means that $a_n/b_n \to 1$; the convergence in probability is denoted by $\to_p$, in distribution, by $\to_d$, and almost surely, by $\to_{a.s.}$; the r-dimension normal distribution with mean vector a and covariance matrix B is denoted by $\mathcal{N}_r(a, B)$, and $\mathcal{N}(a, B) = \mathcal{N}_1(a, B)$. Denoted by Φ the standard normal cumulative distribution function, and z_α the $(1 - \alpha)$-quantile.

1.2 Assumptions

Here we shall state the needed assumptions.

(e1) $(X_i, \delta_i Y_i); X_i \in \mathbb{R}^d, Y_i \in \mathbb{R}, \delta_i = 0$ or $1, i = 1, 2, \cdots, n,$ are i.i.d. random vectors with $\delta = 1$, if Y is observed, and $\delta = 0$, if Y is missing; δ and ε are conditionally independent, given X.

(e2) $E(\varepsilon|X = x) = 0$, $E\varepsilon^2 < \infty$. The function $\sigma^2(x) := E(\varepsilon^2|X = x)$ is a.e. in (G) continuous on $\mathcal{I}$, and $\Delta(x) := E(\delta|X) = P(\delta = 1|X = x)$ is positive and Lipschitz-continuous of order 1 on an open interval containing $\mathcal{I}$.

(e3) $E|\varepsilon|^{2+\delta_0} < \infty$, for some $\delta_0 > 0$.

(e4) $E\varepsilon^4 < \infty$.

(f1) The design variable X has a uniformly continuous Lebesgue density f that is bounded from below on an open interval containing $\mathcal{I}$.

(f2) The density f is twice continuously differentiable with a compact support.

(g) G is a σ-finite measure on $\mathbb{R}^d$ and has a continuous Lebesgue density g.

(k1) The kernels K and K^* are positive symmetric square integrable densities on $[-1, 1]^d$. In addition, K^* satisfies Lipschitz-continuity of order 1.

(k2) The kernel $\tilde{K}$ is positive symmetric square integrable density on $[-1, 1]^d$, satisfying Lipschitz-continuity of order γ, $\gamma > 0$. $\tilde{K}(u)$ attains its maximum at $u = 0$.

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

(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(G), then $\theta_1 = \theta_2$.

(m3) For some positive continuous function ℓ on $\mathcal{I}$ and for some $\beta > 0$,

$$|m_{\theta_2}(x) - m_{\theta_1}(x)| \leq \|\theta_2 - \theta_1\|^\beta \ell(x), \quad \forall \theta_2, \theta_1 \in \Theta,\ x \in \mathcal{I}.$$

(m4) The true parameter θ_0 is an inner point of Θ. For every x, $m_\theta(x)$ is differentiable in θ in a neighborhood of θ_0 with the vector of derivatives $\dot{m}_\theta(x)$, such that for every

$\varepsilon > 0$, $k < \infty$,

$$\limsup_n P\Big(\sup_{1\le i\le n, \sqrt{nh^d}\|\theta-\theta_0\|\le k} \frac{m_\theta(X_i) - m_{\theta_0}(X_i) - (\theta-\theta_0)^T \dot m_{\theta_0}(X_i)}{\|\theta-\theta_0\|} > \varepsilon\Big)$$

$= 0$.

(m5) The vector function $x \mapsto \dot m_{\theta_0}(x)$ is continuous in $x \in \mathcal{I}$ and for every $\varepsilon > 0$, there is an $N_\varepsilon < \infty$ such that for every $0 < k < \infty$,

$$P\Big(\max_{1\le i\le n, (nh^d)^{1/2}\|\theta-\theta_0\|\le k} h^{-d/2}\|\dot m_\theta(X_i) - \dot m_{\theta_0}(X_i)\| \ge \varepsilon\Big) \le \varepsilon, \quad \forall n > N_\varepsilon.$$

(m6) $n^{-1}\sum_{i=1}^n \delta_i \dot m_{\theta_0}(X_i)\dot m_{\theta_0}^T(X_i)$, $n \ge q$, and $E[\delta \dot m_{\theta_0}(X)\dot m_{\theta_0}^T(X)]$ are positive definite.

(a) The estimator $\hat\alpha_n$ is $\sqrt{n}$-consistent for θ_0 under H_0.

(b1) $nb_n^d \to \infty$, $nb_n^{d+1} \to 0$.

(b2) $b_n \sim n^{-r}$, where $1/(d+1) < r < 1/d$.

(h1) $h_n \to 0$.

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

(h3) $h_n \sim n^{-a}$, where $0 < a < \min(1/(2d), 4/(d(d+4)))$.

(h4) $h_n \sim n^{-a}$, where $0 < a < 1/d - r$, with r in (b2).

(w) $w_n = a_n(\log n/n)^{1/(d+4)}$, $a_n \to a_0 > 0$.

Note that (h3) implies (h1) and (h2), (h4) implies (h3), and (b2) implies (b1). Among these assumptions, (e3), (e4), (f1), (f2), (g), (k1), (m1)-(m5), (h1)-(h3), (w), and part of (e1) and (e2), are similar as in K-N when no data are missing; conditions on δ and Δ in (e1) and (e2) are for the missing data at random setup; (m6) and (a) are used for the imputation method, while (k2), (a), (b1), (b2), and (h4) are for the inverse probability weighting method. An example of r in (b2) and a in (h4) is $r = (2d+1)/(2d(d+1))$, $a = 1/(2d(d+1))$.

We need the following notation in the proofs later. For $i = 1, \cdots, n$, $j = 1, 2$, $x \in \mathbb{R}^d$, define

$$
\begin{aligned}
&\varepsilon_{i1}^* := \delta_i \varepsilon_i, \qquad \varepsilon_{i2}^* := \frac{\delta_i}{\Delta(X_i)} \varepsilon_i, && (1.5)\\
&\tilde{Y}_{i1} := m_{\theta_0}(X_i) + \varepsilon_{i1}^*, \qquad \tilde{Y}_{i2} := m_{\theta_0}(X_i) + \varepsilon_{i2}^*,\\
&K_{hi}(x) := K_h(x - X_i), \quad K_{wi}^* := K_w^*(x - X_i), \quad \tilde{K}_{bi}(x) := \tilde{K}_b(x - X_i),\\
&\hat{f}_h(x) := n^{-1} \sum_{i=1}^n K_{hi}(x), \quad \hat{f}_w(x) := n^{-1} \sum_{i=1}^n K_{wi}^*(x), \quad \hat{f}_b(x) := n^{-1} \sum_{i=1}^n \tilde{K}_{bi}(x),\\
&d\psi(x) := \{f(x)\}^{-2} dG(x), \quad d\hat{\psi}_h(x) := \{\hat{f}_h(x)\}^{-2} dG(x), \quad d\hat{\psi}_w(x) := \{\hat{f}_w(x)\}^{-2} dG(x),\\
&\mu_n(x, \theta) := n^{-1} \sum_{i=1}^n K_{hi}(x) m_\theta(X_i), \quad \dot{\mu}_n(x, \theta) := n^{-1} \sum_{i=1}^n K_{hi}(x) \dot{m}_\theta(X_i),\\
&Z_n(x, \theta) := n^{-1} \sum_{i=1}^n K_{hi}(x)(m_\theta(X_i) - m_{\theta_0}(X_i)),\\
&\dot{\mu}_h(x) := E\dot{\mu}_n(x, \theta_0) = EK_h(x - X)\dot{m}_{\theta_0}(X),\\
&\dot{\mu}_{n\delta}(x, \theta) := n^{-1} \sum_{i=1}^n K_{hi}(x)(1 - \delta_i)\dot{m}_\theta(X_i),\\
&\dot{\mu}_{h\delta}(x) := E\dot{\mu}_{n\delta}(x, \theta_0) = EK_h(x - X)(1 - \delta)\dot{m}_{\theta_0}(X),\\
&U_{nj}(x, \theta) := n^{-1} \sum_{i=1}^n K_{hi}(x)(\tilde{Y}_{ij} - m_\theta(X_i)),\\
&\hat{U}_{nj}(x, \theta) := n^{-1} \sum_{i=1}^n K_{hi}(x)(\hat{Y}_{ij} - m_\theta(X_i)),\\
&U_{nj}(x) := U_{nj}(x, \theta_0) = n^{-1} \sum_{i=1}^n K_{hi}(x)\varepsilon_{ij}^*,\\
&T_{nj}(\theta) := \int \Big[n^{-1} \sum_{i=1}^n K_{hi}(x)(\tilde{Y}_{ij} - m_\theta(X_i))\Big]^2 d\psi(x), \quad \theta \in \mathbb{R}^q,
\end{aligned}
$$

$$\tilde{T}_{nj}(\theta) := \int \Big[n^{-1}\sum_{i=1}^{n} K_{hi}(x)(\tilde{Y}_{ij} - m_\theta(X_i))\Big]^2 d\hat{\psi}_w(x), \quad \theta \in \mathbb{R}^q,$$
$$\tilde{\theta}_{nj} := \arg\min_{\theta\in\Theta} \tilde{T}_{nj}(\theta), \qquad \tilde{\varepsilon}_{ij} := \tilde{Y}_{ij} - m_{\tilde{\theta}_{nj}}(X_i),$$
$$\tilde{A}_{nj} := \int \Big[n^{-1}\sum_{i=1}^{n} K_{hi}(x)(\hat{Y}_{ij} - \tilde{Y}_{ij})\Big]^2 d\psi(x),$$
$$C_{nj} := n^{-2}\sum_{i=1}^{n}\int K_{hi}^2(x)(\tilde{Y}_{ij} - m_{\theta_0}(X_i))^2 d\psi(x),$$
$$\hat{r}_n(x) := \frac{1}{\hat{\Delta}_n(x)} - \frac{1}{\Delta(x)}, \qquad \hat{r}_{ni} := \hat{r}_n(X_i),$$
$$u_n := \hat{\alpha}_n - \theta_0, \qquad d_{ni} := m_{\hat{\alpha}_n}(X_i) - m_{\theta_0}(X_i) - u_n^T \dot{m}_{\theta_0}(X_i).$$

The following lemmas are found useful in proofs later. Lemma 1.2.1 is facilitated by Mack and Silverman (1982), and Lemma 1.2.3 is derived by Theorem 3 of Collomb and Härdle (1986).

Lemma 1.2.1. *Under the conditions (f1), (k1), (h1), and (h2), the following hold.*

$$\sup_{x\in\mathcal{I}} |\hat{f}_h(x) - f(x)| = o_p(1), \tag{1.6}$$
$$\sup_{x\in\mathcal{I}} |\hat{f}_w(x) - f(x)| = o_p(1), \tag{1.7}$$
$$\sup_{x\in\mathcal{I}} \left|\frac{f(x)}{\hat{f}_w(x)} - 1\right| = o_p(1). \tag{1.8}$$

Lemma 1.2.2. (Theorem 2.2 part (2), Bosq (1998)) *Under the assumptions (f2), (k1), and (w), we have for* $\forall k > 0$*, and* $k \in \mathbb{N}$*,*

$$(\log_k n)^{-1}(n/\log n)^{2/(d+4)} \sup_{x\in\mathcal{I}} \left|\hat{f}_w(x) - f(x)\right| \to 0, \quad a.s. \tag{1.9}$$

Lemma 1.2.3. *Suppose (e2), (f2), (k2), and (b1) hold, then*

$$\sup_{x\in\mathcal{I}} |\hat{f}_b(x) - f(x)| = o_p(1), \tag{1.10}$$

$$\sup_{x\in\mathcal{I}} |\hat{\Delta}_n(x) - \Delta(x)| = o_p(1), \tag{1.11}$$

$$\sup_{x\in\mathcal{I}} \Big|\frac{1}{\hat{\Delta}_n(x)} - \frac{1}{\Delta(x)}\Big| = o_p(1), \tag{1.12}$$

$$n^{1/2} b^{d/2} (\log n)^{-1/2} \sup_{x\in\mathcal{I}} \Big|\frac{1}{\hat{\Delta}_n(x)} - \frac{1}{\Delta(x)}\Big| = O_p(1). \tag{1.13}$$

1.3 Consistency of the minimum distance estimators

In this section we prove the consistency of the minimum distance estimators $\hat{\theta}_{nj}$, $j = 1, 2$, under H_0. To state the results, we need Lemma 3.1 in K-N as a preliminary reproduced here for the sake of completeness. Let $L_2(G)$ denote a class of square integrable real valued functions on $\mathbb{R}^d$ with respect to G. Define

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

and the map

$$\mathcal{M}(\nu) := \arg\min_{\theta\in\Theta} \rho(\nu, m_\theta), \quad \nu \in L_2(G).$$

Lemma 1.3.1. (Koul and Ni (2004)) *Let m satisfy conditions (m1)-(m3). Then the following hold.*

(a) *$\mathcal{M}(\nu)$ always exists, $\forall \nu \in L_2(G)$.*

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

$$\rho(\nu_n, \nu) \to 0 \quad \textit{implies} \quad \mathcal{M}(\nu_n) \to \mathcal{M}(\nu) \quad \textit{as } n \to \infty.$$

(c) *$\mathcal{M}(m_\theta(\cdot)) = \theta$, uniquely for $\forall \theta \in \Theta$.*

We now proceed to state and prove the consistency of $\hat{\theta}_{nj}$, $j = 1, 2$.

Theorem 1.3.1. *Under* H_0*, (e1), (e2), (f1), (k1), (m1)-(m4), (a), (h1), and (h2),*

$$\hat{\theta}_{nj} \to_p \theta_0, \quad j = 1, 2.$$

Proof. The basic idea of the proof is the same as in K-N, Theorem 3.1; Only details with respect to $\hat{Y}_{ij} - \tilde{Y}_{ij}$, $i = 1, \cdots, n$, are different. By part (c) in Lemma 1.3.1, one has $\hat{\theta}_{nj} = \mathcal{M}(m_{\hat{\theta}_{nj}})$, $j = 1, 2$, and $\theta_0 = \mathcal{M}(m_{\theta_0})$. Then it suffices to prove $\rho(m_{\hat{\theta}_{nj}}, m_{\theta_0}) = o_p(1)$, $j = 1, 2$, by part (b1) in Lemma 1.3.1. Define

$$\begin{aligned}
&\tilde{m}_{nj}(x) := n^{-1}\sum_{i=1}^{n} K_{hi}(x)\tilde{Y}_{ij}/\hat{f}_w(x), \quad \hat{m}_{nj}(x) := n^{-1}\sum_{i=1}^{n} K_{hi}(x)\hat{Y}_{ij}/\hat{f}_w(x), \\
&\hat{R}_{nj}(\theta) = \int [\hat{m}_{nj}(x) - m_\theta(x)]^2 dG(x), \quad \theta \in \mathbb{R}^q, \qquad \hat{\beta}_{nj} := \arg\min_{\theta\in\Theta} \hat{R}_{nj}(\theta), \\
&C_n(\theta) := \int [\mu_n(x, \theta) - \hat{f}_w(x) m_\theta]^2 d\hat{\psi}_w(x).
\end{aligned}$$

By the fact that

$$\rho(m_{\hat{\theta}_{nj}}, m_{\theta_0}) \le 2[\rho(m_{\hat{\theta}_{nj}}, \hat{m}_{nj}) + \rho(\hat{m}_{nj}, m_{\theta_0})] = 2[\hat{R}_{nj}(\hat{\theta}_{nj}) + \hat{R}_{nj}(\theta_0)],$$

it suffices to show

$$\hat{R}_{nj}(\theta_0) = o_p(1), \quad j = 1, 2, \tag{1.14}$$

$$\hat{R}_{nj}(\hat{\theta}_{nj}) = o_p(1), \quad j = 1, 2. \tag{1.15}$$

If we can prove (1.14) and the following result

$$\sup_{\theta\in\Theta} |\hat{T}_{nj}(\theta) - \hat{R}_{nj}(\theta)| = o_p(1), \quad j = 1, 2, \tag{1.16}$$

we can obtain (1.15). This is because the definition of $\hat{\beta}_{nj}$ and (1.14) lead to the result $\hat{R}_{nj}(\hat{\beta}_{nj}) = o_p(1)$, which together with (1.16) leads to $\hat{T}_{nj}(\hat{\beta}_{nj}) = o_p(1)$; by the definition of $\hat{\theta}_{nj}$, one has $\hat{T}_{nj}(\hat{\theta}_{nj}) = o_p(1)$; this result and (1.16) bring the claim (1.15). Therefore, we only need to prove (1.14) and (1.16).

Recall $\tilde{A}_{nj}$ from (1.5). To prove (1.14), note that

$$\begin{aligned}
\hat{R}_{nj}(\theta_0) &= \int [\hat{f}_w(x)(\hat{m}_{nj}(x) - \tilde{m}_{nj}(x)) + U_{nj}(x) \\
&\qquad + \mu_n(x, \theta_0) - \hat{f}_w(x) m_{\theta_0}(x)]^2 d\hat{\psi}_w(x) \\
&\leq 3 \int [\hat{m}_{nj}(x) - \tilde{m}_{nj}(x)]^2 dG(x) + 3 \int U_{nj}^2(x) d\hat{\psi}_w(x) \\
&\quad + 3 \int [\mu_n(x, \theta_0) - \hat{f}_w(x) m_{\theta_0}(x)]^2 d\hat{\psi}_w(x) \\
&\leq 3(1 + \sup_{x\in\mathcal{I}} |f^2(x)/\hat{f}_w^2(x) - 1|)\tilde{A}_{nj} + 3\tilde{T}_{nj}(\theta_0) + 3C_n(\theta_0), \quad j = 1, 2,
\end{aligned}$$

By Fubini, the continuity of f, σ^2, and Δ, assured by (e2) and (f1), and by (k1) and (h2), we have

$$E\int U_{n1}^2(x)d\psi(x) = n^{-1}\int EK_h^2(x-X)\Delta(X)\sigma^2(X)d\psi(x) = O((nh^d)^{-1}),$$
$$E\int U_{n2}^2(x)d\psi(x) = n^{-1}\int EK_h^2(x-X)\{\Delta(X)\}^{-1}\sigma^2(X)d\psi(x) = O((nh^d)^{-1}),$$

so that $T_{nj}(\theta_0) = \int U_{nj}^2(x)d\psi(x) = O_p((nh^d)^{-1})$, $j = 1, 2$. Together by (1.8), we have

$$\tilde{T}_{nj}(\theta_0) \le \sup_{x\in\mathcal{I}} |f(x)/\hat{f}_w(x)|^2 T_{nj}(\theta_0) = O_p((nh^d)^{-1}), \quad j = 1, 2.$$

The claim $C_n(\theta_0) = o_p(1)$ can be derived by the same argument as that of proving (3.5) in K-N. Note that for $i = 1, \cdots, n$,

$$\begin{aligned}\hat{Y}_{i1} - \tilde{Y}_{i1} &= (1-\delta_i)(m_{\hat{\alpha}_n}(X_i) - m_{\theta_0}(X_i)), \\ \hat{Y}_{i2} - \tilde{Y}_{i2} &= \hat{r}_{ni}\delta_i\varepsilon_i + \Big(1 - \frac{\delta_i}{\Delta(X_i)}\Big)(m_{\hat{\alpha}_n}(X_i) - m_{\theta_0}(X_i)) \\ &\quad -\hat{r}_{ni}\delta_i(m_{\hat{\alpha}_n}(X_i) - m_{\theta_0}(X_i)).\end{aligned} \tag{1.17}$$

Recall u_n and d_{ni} from (1.5). By calculation in (3.9) in K-N, (m4), and (a), we have

$$\begin{aligned}\tilde{A}_{n1} &\le 2\|u_n\|^2 \max_{1\le i\le n} \frac{|d_{ni}|^2}{\|u_n\|^2} \int \hat{f}_h^2(x)d\psi(x) \\ &\quad +2\|u_n\|^2 \int \Big[n^{-1}\sum_{i=1}^n K_{hi}(x)(1-\delta_i)\|\dot{m}_{\theta_0}(X_i)\|\Big]^2 d\psi(x) \\ &= o_p(1), \qquad\qquad (1.18)\\ \tilde{A}_{n2} &\le 3\int \Big[n^{-1}\sum_{i=1}^n K_{hi}(x)\hat{r}_{ni}\delta_i\varepsilon_i\Big]^2 d\psi(x) \\ &\quad +3\int \Big[n^{-1}\sum_{i=1}^n K_{hi}(x)\Big(1-\frac{\delta_i}{\Delta(X_i)}\Big)(m_{\hat{\alpha}_n}(X_i) - m_{\theta_0}(X_i))\Big]^2 d\psi(x)\end{aligned}$$

$$
\begin{aligned}
&+3\int\Big[n^{-1}\sum_{i=1}^{n}K_{hi}(x)\hat r_{ni}\delta_i(m_{\hat\alpha_n}(X_i)-m_{\theta_0}(X_i))\Big]^2 d\psi(x)\\
\le\;& 3\int\Big[n^{-1}\sum_{i=1}^{n}K_{hi}(x)\hat r_{ni}\delta_i\varepsilon_i\Big]^2 d\psi(x)\\
&+6\|u_n\|^2\max_{1\le i\le n}\frac{d_{ni}^2}{\|u_n\|^2}\int\Big[n^{-1}\sum_{i=1}^{n}K_{hi}(x)\Big(1+\frac{\delta_i}{\Delta(X_i)}\Big)\Big]^2 d\psi(x)\\
&+6\|u_n\|^2\int\Big[n^{-1}\sum_{i=1}^{n}K_{hi}(x)\Big(1-\frac{\delta_i}{\Delta(X_i)}\Big)\|\dot m_{\theta_0}(X_i)\|\Big]^2 d\psi(x)\\
&+6\|u_n\|^2\max_{1\le i\le n}\frac{d_{ni}^2}{\|u_n\|^2}\sup_{1\le i\le n}\hat r_{ni}^2\int\Big[n^{-1}\sum_{i=1}^{n}K_{hi}(x)\delta_i\Big]^2 d\psi(x)\\
&+6\|u_n\|^2\sup_{1\le i\le n}\hat r_{ni}^2\int\Big[n^{-1}\sum_{i=1}^{n}K_{hi}(x)\delta_i\|\dot m_{\theta_0}(X_i)\|\Big]^2 d\psi(x)\\
=\;& o_p(1), \qquad (1.19)
\end{aligned}
$$

Therefore, together with (1.8) and (1.12), we obtain (1.14). To prove (1.16), write

$$
\begin{aligned}
&\hat T_{nj}(\theta)-\hat R_{nj}(\theta)\\
&=\int\Big[\hat m_{nj}(x)-\frac{\mu_n(x,\theta)}{\hat f_w(x)}\Big]^2 dG(x)-\int[\hat m_{nj}(x)-m_\theta(x)]^2 dG(x)\\
&=-\int\Big[\frac{\mu_n(x,\theta)}{\hat f_w(x)}-m_\theta(x)\Big]^2 dG(x)\\
&\quad-2\int\Big[\hat m_{nj}(x)-\frac{\mu_n(x,\theta)}{\hat f_w(x)}\Big]\Big[\frac{\mu_n(x,\theta)}{\hat f_w(x)}-m_\theta(x)\Big]dG(x), \quad j=1,2.
\end{aligned}
$$

By Cauchy-Schwarz (C-S) inequality, we have

$$
\sup_{\theta\in\Theta}|\hat T_{nj}(\theta)-\hat R_{nj}(\theta)|\le\sup_{\theta\in\Theta}C_n(\theta)+2\sup_{\theta\in\Theta}\hat T_{nj}^{1/2}(\theta)C_n^{1/2}(\theta), \quad j=1,2.
$$

Hence it suffices to prove

$$\sup_{\theta\in\Theta} C_n(\theta) = o_p(1), \qquad \sup_{\theta\in\Theta} \hat{T}_{nj}(\theta) = O_p(1), \quad j = 1, 2. \tag{1.20}$$

One can prove the first claim in (1.20) by the same argument as in proving (3.14) in K-N. To prove the second part of (1.20), note that by adding and subtracting $\tilde{Y}_{ij}$ to the i-th summand in $\hat{T}_{nj}(\theta)$, we obtain

$$\begin{aligned}
\hat{T}_{nj}(\theta) &\leq 2(1 + \sup_{x\in\mathcal{I}} |f^2(x)/\hat{f}_w^2(x) - 1|)\Big(\tilde{A}_{nj} + \int [U_{nj}(x) - Z_n(x,\theta)]^2 d\psi(x)\Big) \\
&\leq 2(1 + \sup_{x\in\mathcal{I}} |f^2(x)/\hat{f}_w^2(x) - 1|) \\
&\quad \times\Big(\tilde{A}_{nj} + 2\int U_{nj}^2(x)d\psi(x) + 2\int Z_n^2(x,\theta)d\psi(x)\Big).
\end{aligned}$$

From (3.16) in K-N, one obtains $\sup_{\theta\in\Theta}\int Z_n^2(x,\theta)d\psi(x) = O_p(1)$. By (1.8) and $A_{nj} = o_p(1)$, $\int U_{nj}^2(x)d\psi(x) = o_p(1)$ in the argument above, we have $\sup_{\theta\in\Theta}\hat{T}_{nj}(\theta) = O_p(1)$, $j = 1, 2$. Thus the proof of the theorem is complete.

1.4 Asymptotic distribution of the minimum distance estimators under H_0

This section states and proves the asymptotic normality of $\hat{\theta}_{nj}$, $j = 1, 2$. To proceed further, we need the following notation. Let

$$
\begin{aligned}
\Sigma_0 &:= \int \dot{m}_{\theta_0}(x)\dot{m}^T_{\theta_0}(x)g(x)dx, \qquad (1.21)\\
\Sigma_0^* &:= \int (1-\Delta(x))\dot{m}_{\theta_0}(x)\dot{m}^T_{\theta_0}(x)g(x)dx,\\
\Sigma_1 &:= \int \sigma^2(x)\Delta(x)\dot{m}_{\theta_0}(x)\dot{m}^T_{\theta_0}(x)g^2(x)(f(x))^{-1}dx\\
&\quad \Sigma_1^* + 2\Big(\int \sigma^2(x)\Delta(x)\dot{m}_{\theta_0}(x)\dot{m}^T_{\theta_0}(x)g(x)dx\Big)\tilde{\Sigma}_0^{-1}\Sigma_0^*\\
&\quad +\Sigma_0^*\tilde{\Sigma}_0^{-1}\Big(\int \sigma^2(x)\Delta(x)\dot{m}_{\theta_0}(x)\dot{m}^T_{\theta_0}(x)f(x)dx\Big)\tilde{\Sigma}_0^{-1}\Sigma_0^*,\\
\Sigma_2 &:= \int \sigma^2(x)\dot{m}_{\theta_0}(x)\dot{m}^T_{\theta_0}(x)g^2(x)(\Delta(x)f(x))^{-1}dx,\\
\tilde{\Sigma}_0 &:= \int \Delta(x)\dot{m}_{\theta_0}(x)\dot{m}^T_{\theta_0}(x)f(x)dx, \qquad \tilde{\Sigma}_n := n^{-1}\sum_{i=1}^n \delta_i\dot{m}_{\theta_0}(X_i)\dot{m}^T_{\theta_0}(X_i),\\
\tilde{S}_n &:= n^{-1}\sum_{i=1}^n \delta_i\varepsilon_i\dot{m}_{\theta_0}(X_i), \qquad S_{nj} := \int U_{nj}(x)\dot{\mu}_h(x)d\psi(x), \quad j=1,2.
\end{aligned}
$$

Theorem 1.4.1. *Suppose* H_0*, (e1), (e2), (e3), (f1), (f2), (g), (k1), (m1)-(m5), (a), and (h3) hold. Then,*

$$n^{1/2}(\hat{\theta}_{n1} - \theta_0) = \Sigma_0^{-1}n^{1/2}\{S_{n1} + \Sigma_0^*(\hat{\alpha}_n - \theta_0)\} + o_p(1), \qquad (1.22)$$

where $\hat{\alpha}_n$ *is in (1.2), and*

$$n^{1/2}(\hat{\theta}_{n1} - \theta_0) = O_p(1). \qquad (1.23)$$

If under H_0, $m_{\theta_0}(x)$ *is a linear function of* θ_0, *i.e.* $m_{\theta_0}(x) = \theta_0^T l(x)$, *for all* $x \in \mathcal{I}$, *where* $l(x)$ *satisfies (m1)-(m3) and (m6), we can take* $\hat{\alpha}_n = \tilde{\Sigma}_n^{-1}\{n^{-1}\sum_{i=1}^n \delta_i Y_i \dot{m}_{\theta_0}(X_i)\}$, *which is the least square estimator and satisfies condition (a), and we obtain*

$$n^{1/2}(\hat{\theta}_{n1} - \theta_0) = \Sigma_0^{-1} n^{1/2}\{S_{n1} + \Sigma_0^* \tilde{\Sigma}_n^{-1} \tilde{S}_n\} + o_p(1). \tag{1.24}$$

If (k2), (b2), and (h4) hold, one has

$$n^{1/2}(\hat{\theta}_{n2} - \theta_0) = \Sigma_0^{-1} n^{1/2} S_{n2} + o_p(1). \tag{1.25}$$

Consequently, (1.24) and (1.25) lead to

$$n^{1/2}(\hat{\theta}_{nj} - \theta_0) \to_d \mathcal{N}_q(0, \Sigma_0^{-1}\Sigma_j\Sigma_0^{-1}), \quad j = 1, 2. \tag{1.26}$$

Here Σ_0, Σ_0^*, $\tilde{\Sigma}_0$, $\tilde{\Sigma}_n$, $\tilde{S}_n$, S_{nj}, *and* Σ_j, $j = 1, 2$, *are in (1.21).*

Proof. We prove the theorem in two steps, following the routine to prove Theorem 4.1 in K-N.

Step 1. The first step is to show that

$$nh^d \|\hat{\theta}_{nj} - \theta_0\|^2 = O_p(1), \quad j = 1, 2. \tag{1.27}$$

Let $D_n(\theta) := \int Z_n^2(x, \theta) d\psi(x)$. Note that

$$nh^d D_n(\hat{\theta}_{nj}) = nh^d \|\hat{\theta}_{nj} - \theta_0\|^2 \frac{D_n(\hat{\theta}_{nj})}{\|\hat{\theta}_{nj} - \theta_0\|^2}, \quad j = 1, 2.$$

It suffices to prove

$$nh^d D_n(\hat{\theta}_{nj}) = O_p(1), \quad j = 1, 2, \tag{1.28}$$

because the rest follows the a similar argument used in proving (4.4) in K-N, if the corresponding $\hat{\theta}_n$ is changed to $\hat{\theta}_{nj}$, $j = 1, 2$. Observe that

$$\begin{aligned}
& nh^d D_n(\hat{\theta}_{nj}) \\
= \;& nh^d \int [\hat{U}_{nj}(x, \hat{\theta}_{nj}) - \hat{U}_{nj}(x, \theta_0)]^2 d\psi(x) \\
\le \;& 2nh^d (1 + \sup_{x \in \mathcal{I}} |\hat{f}_w^2(x)/f^2(x) - 1|) \\
& \times \Big\{ \int \hat{U}_{nj}^2(x, \hat{\theta}_{nj}) d\hat{\psi}_w(x) + \int \hat{U}_{nj}^2(x, \theta_0) d\hat{\psi}_w(x) \Big\} \\
\le \;& 4nh^d (1 + \sup_{x \in \mathcal{I}} |\hat{f}_w^2(x)/f^2(x) - 1|) \hat{T}_{nj}(\theta_0) \\
\le \;& 8nh^d (1 + \sup_{x \in \mathcal{I}} |\hat{f}_w^2(x)/f^2(x) - 1|)(1 + \sup_{x \in \mathcal{I}} |f^2(x)/\hat{f}_w^2(x) - 1|) \{T_{nj}(\theta_0) + \tilde{A}_{nj}\}.
\end{aligned}$$

By (1.7), (1.8), and $T_{nj}(\theta_0) = O_p((nh^d)^{-1})$, $j = 1, 2$, it suffices to prove $nh^d \tilde{A}_{nj} = O_p(1)$, $j = 1, 2$. This result hold for $j = 1$ because of (1.18). When $j = 2$, by (a), (1.12), and calculation in (1.19), it suffices to show the following results:

$$nh^d \int \Big[n^{-1} \sum_{i=1}^n K_{hi}(x) \hat{r}_{ni} \delta_i \varepsilon_i \Big]^2 d\psi(x) = O_p(1). \tag{1.29}$$

To prove (1.29), we have

$$nh^d E \int \Big[n^{-1} \sum_{i=1}^n K_{hi}(x) \hat{r}_{ni} \delta_i \varepsilon_i \Big]^2 d\psi(x) \;\; = \;\; n^{-1} h^d \sum_{i=1}^n E\Big(\hat{r}_{ni} \delta_i^2 \varepsilon_i^2 \int K_{hi}^2(x) d\psi(x) \Big)$$

$$
\begin{aligned}
&= h^d E\Big(\delta\hat{r}_n^2(X)\varepsilon^2\int K_h^2(x-X)d\psi(x)\Big)\\
&= h^{-d}E\int\int K^2((x-z)/h)\{f(x)\}^{-2}g(x)f(z)\{\Delta(z)\}^{-1}\sigma^2(z)\\
&\quad\times\Big[\frac{(\Delta(z)-1)\tilde{K}(0)+\sum_{i=2}^n(\Delta(z)-\delta_i)\tilde{K}_{bi}(z)}{\tilde{K}(0)+\sum_{i=2}^n\delta_i\tilde{K}_{bi}(z)}\Big]^2 dzdx\\
&= \int\int K^2(u)\{f(z+uh)\}^{-2}g(z+uh)f(z)\{\Delta(z)\}^{-1}\sigma^2(z)\\
&\quad\times E\Big[\frac{(\Delta(z)-1)\tilde{K}(0)+\sum_{i=2}^n(\Delta(z)-\delta_i)\tilde{K}_{bi}(z)}{\tilde{K}(0)+\sum_{i=2}^n\delta_i\tilde{K}_{bi}(z))}\Big]^2 dzdu,
\end{aligned}
$$

where the last equality is derived by Fubini's theorem. Let

$$
B_n(z) := E\frac{(\Delta(z)-1)^2\tilde{K}^2(0)+\sum_{i=1}^n(\Delta(z)-\delta_i)^2\tilde{K}_{bi}^2(z)}{[\tilde{K}(0)+\sum_{i=1}^n\delta_i\tilde{K}_{bi}(z))]^2},\quad z\in\mathbb{R}^d.
$$

Let $\mathcal{I}_b$ be the b_n-neighborhood of compact set $\mathcal{I}$. By (e2), (f1), and (k1), it is sufficient to show $\sup_{z\in\mathcal{I}_b}B_n(z)=O(1)$. Let $I_n(z):=\tilde{K}(0)+\sum_{i=1}^n\delta_i\tilde{K}_{bi}(z)$, $z\in\mathcal{I}_b$, $n\geq 1$, and $I_0(z)\equiv\tilde{K}(0)$. For any $z\in\mathcal{I}_b$, write $B_n(z)=B_{n1}(z)+B_{2n}(z)+2B_{n3}(z)-2B_{n4}(z)$, where

$$
\begin{aligned}
B_{n1}(z) &:= E\Big\{[I_n(z)]^{-2}(\Delta(z)-1)^2\tilde{K}^2(0)\Big\},\\
B_{n2}(z) &:= E\Big\{[I_n(z)]^{-2}\sum_{i=1}^n(\Delta(z)-\delta_i)^2\tilde{K}_{bi}^2(z)\Big\},\\
B_{n3}(z) &:= E\Big\{[I_n(z)]^{-2}\sum_{1\leq i<j\leq n}(\Delta(z)-\delta_i)(\Delta(z)-\delta_j)\tilde{K}_{bi}(z)\tilde{K}_{bj}(z)\Big\},\\
B_{n4}(z) &:= E\Big\{[I_n(z)]^{-2}(1-\Delta(z))\tilde{K}(0)\sum_{i=1}^n(\Delta(z)-\delta_i)\tilde{K}_{bi}(z)\Big\}.
\end{aligned}
$$

Observe that

$$
B_{n1}(z) \;\leq\; (\Delta(z)-1)^2(\tilde{K}(0))^2E[I_n(z)]^{-2},
$$

$$
\begin{aligned}
B_{n2}(z) &\leq nE(\Delta(z)-\delta_n)^2\tilde{K}_{bn}^2(z)E[I_{n-1}(z)]^{-2} \\
&\leq nb^d E[I_{n-1}(z)]^{-2}\int\{\Delta^2(z)-2\Delta(z)\Delta(z-bv)+\Delta(z-bv)\}\tilde{K}^2(v)dv,
\end{aligned}
$$

hence it is vital to analyze $E[I_n(z)]^{-2}$, $z \in \mathcal{I}_b$. To proceed further, we shall calculate the marginal probability mass function of δ and conditional probability density function of X given δ based on the joint distribution of (X,δ). Let $f_{X,\delta}$ be the joint p.d.f. of X and δ, f_X be the marginal p.d.f. of X, f_δ be the marginal p.m.f. of δ, $f_{X|\delta}$ be the conditional p.d.f. of X given δ, $f_{\delta|X}$ be the conditional p.m.f. of δ given X. For $k \in \{0,1\}$ and $x \in \mathbb{R}^d$, by (e2) and (f1),

$$
f_{\delta|X}(k|x) = \Delta^k(x)(1-\Delta(x))^{1-k}, \qquad f_X(x) = f(x),
$$

thus

$$
f_{X|\delta}(x|k) = \frac{f_{X,\delta}(x,k)}{f_\delta(k)} = \frac{f_{\delta|X}(k|x)f_X(x)}{f_\delta(k)} = \frac{\Delta^k(x)(1-\Delta(x))^{1-k}f(x)}{\int \Delta^k(x)(1-\Delta(x))^{1-k}f(x)dx}.
$$

Let

$$
\begin{aligned}
&\underline{X}_n := (X_1, X_2, \cdots, X_n), \qquad \underline{\delta}_n := (\delta_1, \delta_2, \cdots, \delta_n), \qquad p := \int \Delta(x)f(x)dx, \\
&\rho_1(x) := f_{X|\delta}(x|1) = \frac{\Delta(x)f(x)}{p}, \qquad \rho_0(x) := f_{X|\delta}(x|0) = \frac{(1-\Delta(x))f(x)}{1-p}, \\
&\rho_1^*(z,b) := \sup_{u\in[-1,1]^d} \rho_1(z-bu), \quad \rho_{1*}(z,b) := \inf_{u\in[-1,1]^d} \rho_1(z-bu), \qquad z \in \mathcal{I}_b.
\end{aligned}
$$

Write the conditional expectation $E[\cdot|\underline{X}_n, \underline{\delta}_n]$ as $E_n[\cdot]$. For $z \in \mathcal{I}_b$, if $E[I_n(z)]^{-2}$ can be bounded by an expression of $E[I_{n-1}(z)]^{-2}$, then $E[I_{n-k}(z)]^{-2}$ can be bounded by the expression of $E[I_{n-k-1}(z)]^{-2}$, $k = 0, 1, \cdots, n-1$, and we can finally obtain a bound of $E[I_n(z)]^{-2}$. Note that

$$
\begin{aligned}
& E[I_n(z)]^{-2} \\
= \;& E\Big(E\big\{[I_{n-1}(z) + \delta_n \tilde{K}_{bn}(z)]^{-2} | \underline{X}_{n-1}, \underline{\delta}_{n-1}, X_n\big\}\Big) \\
= \;& E\Big(E\big\{[I_{n-1}(z)]^{-2} | \underline{X}_{n-1}, \underline{\delta}_{n-1}, X_n\big\}(1 - \Delta(X_n)) \\
& \qquad + E\big\{[I_{n-1}(z) + \tilde{K}_{bn}(z)]^{-2} | \underline{X}_{n-1}, \underline{\delta}_{n-1}, X_n\big\}\Delta(X_n)\Big) \\
= \;& E\Big(\int E_{n-1}[I_{n-1}(z)]^{-2}(1 - \Delta(x)) f(x) dx \\
& \qquad + b^d \int E_{n-1}[I_{n-1}(z) + \tilde{K}(u)]^{-2} \Delta(z - bu) f(z - bu) du\Big) \\
= \;& E\Big((1-p)\int E_{n-1}[I_{n-1}(z)]^{-2} \rho_0(x) dx \\
& \qquad + pb^d \int E_{n-1}[I_{n-1}(z) + \tilde{K}(u)]^{-2} \rho_1(z - bu) du\Big) \\
= \;& E\Big((1-p)[I_{n-1}(z)]^{-2} + pb^d \int_{[-1,1]^d} [I_{n-1} + \tilde{K}(u)]^{-2} \rho_1(z - bu) du \\
& \qquad + p[I_{n-1}(z)]^{-2}\Big\{1 - b^d \int_{[-1,1]^d} \rho_1(z - bu) du\Big\}\Big) \\
= \;& \Big\{1 - pb^d \int_{[-1,1]^d} \rho_1(z - bu) du\Big\} E[I_{n-1}(z)]^{-2} \\
& + pb^d E \int_{[-1,1]^d} [I_{n-1} + \tilde{K}(u)]^{-2} \rho_1(z - bu) du \\
\le \;& \{1 - pb^d(2^d \rho_{1*}(z,b))\} E[I_{n-1}(z)]^{-2} + pb^d \rho_1^*(z,b) E \int_{[-1,1]^d} [I_{n-1} + \tilde{K}(u)]^{-2} du \\
\le \;& \{1 - p(2b)^d \rho_{1*}(z,b)\} E[I_{n-1}(z)]^{-2} + p(2b)^d \rho_1^*(z,b) E[I_{n-1}(z) + c_0]^{-2}, \qquad (1.30)
\end{aligned}
$$

where $c_0 = \min\{2^{-(d+2)}, 2^{-(d+2)/2}(\int \tilde{K}^2(u)du)^{1/2}\}$. To obtain the last inequality above, we used the following fact. For any $a \ge \tilde{K}(0) > 0$,

$$
\begin{aligned}
&\int_{[-1,1]^d}(a+c_0)^{-2}du - \int_{[-1,1]^d}(a+\tilde{K}(u))^{-2}du \\
&= \int_{[-1,1]^d}\frac{\tilde{K}^2(u)+2a\tilde{K}(u)-c_0^2-2ac_0}{(a+c_0)^2(a+\tilde{K}(u))^2}du \\
&= (a+c_0)^{-2}\Big\{\int_{[-1,1]^d}\frac{\tilde{K}^2(u)-c_0^2}{(a+\tilde{K}(u))^2}du + 2a\int_{[-1,1]^d}\frac{\tilde{K}(u)-c_0}{(a+\tilde{K}(u))^2}du\Big\} \\
&\geq (a+c_0)^{-2}\Big\{\Big(\int_{[-1,1]^d}\frac{\tilde{K}^2(u)}{(2a)^2}du - a^{-2}\int_{[-1,1]^d}\frac{c_0^2}{a^2}du\Big) \\
&\qquad\qquad +2a\Big(\int_{[-1,1]^d}\frac{\tilde{K}(u)}{(2a)^2}du - \int_{[-1,1]^d}\frac{c_0}{a^2}du\Big)\Big\} \\
&\geq (a+c_0)^{-2}\Big\{(2a)^{-2}\Big(\int\tilde{K}^2(u)du - 2^{d+2}c_0^2\Big) + (2a)^{-1}\Big(1-2^{d+2}c_0\Big)\Big\} \\
&\geq 0,
\end{aligned}
$$

thus,

$$
\begin{aligned}
&E\Big(\int_{[-1,1]^d}[I_{n-1}+\tilde{K}(u)]^{-2}du\Big) \\
&= E\Big(E_{n-1}\Big\{\int_{[-1,1]^d}[I_{n-1}+\tilde{K}(u)]^{-2}du\Big\}\Big) \\
&\leq E\int_{[-1,1]^d}[I_{n-1}+c_0]^{-2}du \ = 2^dE[I_{n-1}+c_0]^{-2}.
\end{aligned}
$$

By a similar argument used in proving (1.30), we have $k, j = 0, 1, \cdots, n$,

$$
\begin{aligned}
E[I_{n-k}(z)+jc_0]^{-2} \ &\leq \ \{1-p(2b)^d\rho_{1*}(z,b)\}E[I_{n-k-1}(z)+jc_0]^{-2} \\
&\qquad +p(2b)^d\rho_1^*(z,b)E[I_{n-k-1}(z)+(j+1)c_0]^{-2}. \qquad (1.31)
\end{aligned}
$$

Therefore, by (1.30) and (1.31), the following hold:

$$
\begin{aligned}
&E[I_n(z)]^{-2} \\
&\le \{1-p(2b)^d\rho_{1*}(z,b)\}\Big(\{1-p(2b)^d\rho_{1*}(z,b)\}E[I_{n-2}(z)]^{-2} \\
&\qquad\qquad +p(2b)^d\rho_1^*(z,b)E[I_{n-2}(z)+c_0]^{-2}\Big) \\
&\quad +\{p(2b)^d\rho_1^*(z,b)\}\Big(\{1-p(2b)^d\rho_{1*}(z,b)\}E[I_{n-2}(z)+c_0]^{-2} \\
&\qquad\qquad +p(2b)^d\rho_1^*(z,b)E[I_{n-2}(z)+2c_0]^{-2}\Big) \\
&= \sum_{k=0}^{2}\binom{2}{k}\{1-p(2b)^d\rho_{1*}(z,b)\}^{2-k}\{p(2b)^d\rho_1^*(z,b)\}^k E[I_{n-2}(z)+kc_0]^{-2} \\
&\le \{1-p(2b)^d\rho_{1*}(z,b)\}^2\Big(\{1-p(2b)^d\rho_{1*}(z,b)\}E[I_{n-3}(z)]^{-2} \\
&\qquad\qquad +p(2b)^d\rho_1^*(z,b)E[I_{n-3}(z)+c_0]^{-2}\Big) \\
&\quad +2\{1-p(2b)^d\rho_{1*}(z,b)\}\{p(2b)^d\rho_1^*(z,b)\}\Big(\{1-p(2b)^d\rho_{1*}(z,b)\} \\
&\qquad\qquad \times E[I_{n-3}(z)+c_0]^{-2}+p(2b)^d\rho_1^*(z,b)E[I_{n-3}(z)+2c_0]^{-2}\Big) \\
&\quad +\{p(2b)^d\rho_1^*(z,b)\}^2\Big(\{1-p(2b)^d\rho_{1*}(z,b)\}E[I_{n-3}(z)+2c_0]^{-2} \\
&\qquad\qquad +p(2b)^d\rho_1^*(z,b)E[I_{n-3}(z)+3c_0]^{-2}\Big) \\
&= \sum_{k=0}^{3}\binom{3}{k}\{1-p(2b)^d\rho_{1*}(z,b)\}^{3-k}\{p(2b)^d\rho_1^*(z,b)\}^k E[I_{n-3}(z)+kc_0]^{-2} \\
&\le \cdots \\
&\le \sum_{k=0}^{n}\binom{n}{k}\{1-p(2b)^d\rho_{1*}(z,b)\}^{n-k}\{p(2b)^d\rho_1^*(z,b)\}^k E[I_0(z)+kc_0]^{-2} \\
&\le \{1-p(2b)^d\rho_{1*}(z,b)\}^n[\tilde{K}(0)]^{-2} \\
&\quad +c_0^{-2}\sum_{k=1}^{n}\binom{n}{k}k^{-2}\{1-p(2b)^d\rho_{1*}(z,b)\}^{n-k}\{p(2b)^d\rho_1^*(z,b)\}^k. \qquad (1.32)
\end{aligned}
$$

By (e2) and (f1), for large enough n, $f(x)$ and $\Delta(x)$ are bounded and bounded below from zero, and Lipschitz-continuous on $\mathcal{I}_b$. Let ℓ_f and ℓ_Δ denote the Lipschitz constants of f and

Δ, respectively. Define

$$c_1 := \min_{z\in\mathcal{I}_b} \rho_{1*}(z,b) > 0, \qquad c_2 := (\ell_f \sup_{z\in\mathcal{I}_{2b}} \Delta(z) + \ell_\Delta \sup_{z\in\mathcal{I}_{2b}} f(z)),$$
$$\tilde{p}(z,b) := \frac{p(2b)^d \rho_1^*(z,b)}{1 + p(2b)^d(\rho_1^*(z,b) - \rho_{1*}(z,b))}.$$

By (1.30) and the fact that $\sup_{z\in\mathcal{I}_b}(\rho_1^*(z,b) - \rho_{1*}(z,b)) \le 2bd^{1/2}c_2$, we have

$$\begin{aligned} E[I_n(z)]^{-2} \;\le\; & \{1 - p(2b)^d c_1\}^n[\tilde{K}(0)]^{-2} + c_0^{-2}\{1 + p(2b)^{d+1}d^{1/2}c_2\}^n \\ & \times \sum_{k=1}^{n} \binom{n}{k} k^{-2}\{1 - \tilde{p}(z,b)\}^{n-k}\{\tilde{p}(z,b)\}^k. \end{aligned} \tag{1.33}$$

Hence,

$$\begin{aligned} nb^d E[I_n(z)]^{-2} \;\le\; & nb^d[\tilde{K}(0)]^{-2}\Big(\{1 - p(2b)^d c_1\}^{-(p(2b)^d c_1)^{-1}}\Big)^{-n(2b)^d pc_1} \\ & + c_0^{-2}\Big(\{1 + p(2b)^{d+1}d^{1/2}c_2\}^{(p(2b)^{d+1}d^{1/2}c_2)^{-1}}\Big)^{n(2b)^{d+1}pd^{1/2}c_2} \\ & \times nb^d \sum_{k=1}^{n} \binom{n}{k} k^{-2}\{1 - \tilde{p}(z,b)\}^{n-k}\{\tilde{p}(z,b)\}^k. \end{aligned}$$

Note that $n!(k!)^{-1}((n-k)!)^{-1}\{1 - \tilde{p}(z,b)\}^{n-k}\{\tilde{p}(z,b)\}^k$ is the probability mass function of the Binomial(n, $\tilde{p}(z,b)$) distribution. Recall the Chernoff's bound for a r.v. $\zeta \sim B(n, p_0)$, and a constant $\eta \in (0,1)$,

$$P(\zeta < (1-\eta)np_0) < \exp(-np_0\eta^2/2).$$

Using this bound, with $\eta = 1/2$, we obtain that for any $z \in \mathcal{I}_b$,

$$
\begin{aligned}
& nb^d \sum_{k=1}^{n} \binom{n}{k} k^{-2}\{1-\tilde{p}(z,b)\}^{n-k}\{\tilde{p}(z,b)\}^k \\
&= nb^d \Big(\sum_{k=1}^{\lfloor n\tilde{p}(z,b)/2 \rfloor} + \sum_{k=\lfloor n\tilde{p}(z,b)/2\rfloor+1}^{n} \Big) \binom{n}{k} k^{-2}\{1-\tilde{p}(z,b)\}^{n-k}\{\tilde{p}(z,b)\}^k \\
&\leq nb^d \sum_{k=1}^{\lfloor n\tilde{p}(z,b)/2 \rfloor} \binom{n}{k} \{1-\tilde{p}(z,b)\}^{n-k}\{\tilde{p}(z,b)\}^k + nb^d\{n\tilde{p}(z,b)/2\}^{-2} \\
&\leq nb^d \exp(-n\tilde{p}(z,b)/8) + nb^d\{n\tilde{p}(z,b)/2\}^{-2} \\
&= nb^d \exp(-nb^d 2^{d-3} pc_1(1+p(2b)^{d+1}d^{1/2}c_2)^{-1}) \\
&\quad +(nb^d)^{-1}4^{1-d}p^{-2}c_1^{-2}(1+p(2b)^{d+1}d^{1/2}c_2)^2 \ = O((nb^d)^{-1}),
\end{aligned}
$$

by condition (b1). Together with the fact that

$$
\begin{aligned}
&\{1-p(2b)^d c_1\}^{-(p(2b)^d c_1)^{-1}} \to \exp(1), \\
&\{1+p(2b)^{d+1}d^{1/2}c_2\}^{(p(2b)^{d+1}d^{1/2}c_2)^{-1}} \to \exp(1),
\end{aligned}
$$

we have

$$
\begin{aligned}
&nb^d E[I_n(z)]^{-2} = O((nb^d)^{-1}), \quad z \in \mathcal{I}_b, \\
&\sup_{z\in\mathcal{I}_b} nb^d E[I_n(z)]^{-2} = O((nb^d)^{-1}).
\end{aligned}
$$

Hence

$$
\sup_{z\in\mathcal{I}_b} B_{n1}(z) = O((nb^d)^{-2}), \qquad \sup_{z\in\mathcal{I}_b} B_{n2}(z) = O((nb^d)^{-1}).
$$

Observe that

$$
\begin{aligned}
&B_{n3}(z)\\
&= n(n-1)E\Big(\frac{(\Delta(z)-\delta_{n-1})(\Delta(z)-\delta_n)\tilde{K}_{b(n-1)}(z)\tilde{K}_{bn}(z)}{[I_{n-2}(z)+\delta_{n-1}\tilde{K}_{b(n-1)}(z)+\delta_n\tilde{K}_{bn}(z)]^2}\Big)\\
&= n(n-1)b^{2d}E\int\int\Big\{\frac{\tilde{K}(u)\tilde{K}(v)}{[I_{n-2}(z)]^2}\Delta^2(z)(1-\Delta(z-bu))(1-\Delta(z-bv))\\
&\qquad -2\frac{\tilde{K}(u)\tilde{K}(v)}{[I_{n-2}(z)+\tilde{K}(u)]^2}\Delta(z)(1-\Delta(z))\\
&\qquad\qquad \times\Delta(z-bu)(1-\Delta(z-bv))\\
&\qquad +\frac{\tilde{K}(u)\tilde{K}(v)}{[I_{n-2}(z)+\tilde{K}(u)+\tilde{K}(v)]^2}(\Delta(z)-1)^2\\
&\qquad\qquad \times\Delta(z-bu)\Delta(z-bv)\Big\}\\
&\qquad\qquad\qquad \times f(z-bu)f(z-bv)dudv,
\end{aligned}
$$

thus we have

$$
\begin{aligned}
&|B_{n3}(z)|\\
&\le n(n-1)b^{2d}E\int\int\Big\{\frac{\tilde{K}(u)\tilde{K}(v)}{[I_{n-2}(z)]^2}[\Delta^2(z)(1-\Delta(z-bu))(1-\Delta(z-bv))\\
&\qquad -2\Delta(z)(1-\Delta(z))\Delta(z-bu)(1-\Delta(z-bv))\\
&\qquad +(\Delta(z)-1)^2\Delta(z-bu)\Delta(z-bv)]\\
&\qquad +2\Big(\frac{\tilde{K}(u)\tilde{K}(v)}{[I_{n-2}(z)]^2}-\frac{\tilde{K}(u)\tilde{K}(v)}{[I_{n-2}(z)+\tilde{K}(u)]^2}\Big)\\
&\qquad\qquad \times\Delta(z)(1-\Delta(z))\Delta(z-bu)(1-\Delta(z-bv))\Big\}\\
&\qquad\qquad\qquad \times f(z-bu)f(z-bv)dudv\\
&\le n(n-1)b^{2d}E\int\int\Big\{\frac{\tilde{K}(u)\tilde{K}(v)}{[I_{n-2}(z)]^2}[\,|\Delta(z)-\Delta(z-bu)||\Delta(z)-\Delta(z-bv)|\\
&\qquad +\Delta(z)(1-\Delta(z))|\Delta(z-bv)-\Delta(z-bu)|\,]
\end{aligned}
$$

$$+4\frac{\tilde{K}^2(u)\tilde{K}(v)}{[I_{n-2}(z)]^3}\Delta(z)(1-\Delta(z))\Delta(z-bu)(1-\Delta(z-bv))\Big\}$$
$$\times f(z-bu)f(z-bv)dudv$$

$$\begin{aligned}
\leq\ & nb^{d+2}(nb^d)E[I_{n-2}(z)]^{-2}\ell_\Delta^2 \\
& \times\int\int\tilde{K}(u)\tilde{K}(v)\|u\|\,\|v\|f(z-bu)f(z-bv)dudv \\
& +nb^{d+1}(nb^d)E[I_{n-2}(z)]^{-2}\Delta(z)(1-\Delta(z))\ell_\Delta \\
& \times\int\int\tilde{K}(u)\tilde{K}(v)\|u-v\|f(z-bu)f(z-bv)dudv \\
& +4(nb^d)^2E[I_{n-2}(z)]^{-3}\Delta(z)(1-\Delta(z)) \\
& \times\int\int\tilde{K}^2(u)\tilde{K}(v)\Delta(z-bu)(1-\Delta(z-bv))f(z-bu)f(z-bv)dudv.
\end{aligned}$$

By a similar argument used in proving (1.33), for $z\in\mathcal{I}_b$ and $j=3,4,\cdots$, one has

$$\begin{aligned}
E[I_n(z)]^{-j}\ \leq\ & \{1-p(2b)^dc_1\}^n[\tilde{K}(0)]^{-j} \\
& +c_0^{-j}\{1+p(2b)^{d+1}d^{1/2}c_2\}^n\sum_{k=1}^n\binom{n}{k}k^{-j}\{1-\tilde{p}(z,b)\}^{n-k}\{\tilde{p}(z,b)\}^k,
\end{aligned}$$

hence by (b1) and Chernoff's bound, we obtain that

$$\begin{aligned}
& n^2b^{2d}E[I_n(z)]^{-3} \\
& \leq\ n^2b^{2d}[\tilde{K}(0)]^{-3}\{1-p(2b)^dc_1\}^n \\
& \quad +c_0^{-3}\{1+p(2b)^{d+1}d^{1/2}c_2\}^n\times n^2b^{2d}\sum_{k=1}^n\binom{n}{k}k^{-3}\{1-\tilde{p}(z,b)\}^{n-k}\{\tilde{p}(z,b)\}^k \\
& \leq\ n^2b^{2d}[\tilde{K}(0)]^{-3}\Big(\{1-p(2b)^dc_1\}^{-(p(2b)^dc_1)^{-1}}\Big)^{-n(2b)^dpc_1} \\
& \quad +c_0^{-3}\Big(\{1+p(2b)^{d+1}d^{1/2}c_2\}^{(p(2b)^{d+1}d^{1/2}c_2)^{-1}}\Big)^{n(2b)^{d+1}pd^{1/2}c_2} \\
& \qquad \times\Big(n^2b^{2d}\{n\tilde{p}(z,b)/2\}^{-3}+n^2b^{2d}\exp(-n\tilde{p}(z,b)/8)\Big)
\end{aligned}$$

$$\begin{aligned}
&\sim\ n^2b^{2d}[\tilde K(0)]^{-3}\exp(-n(2b)^dpc_1)\\
&\quad +c_0^{-3}\exp(n(2b)^{d+1}pd^{1/2}c_2)\Big((nb^d)^{-1}8^{1-d}p^{-3}c_1^{-3}(1+p(2b)^{d+1}d^{1/2}c_2)^3\\
&\qquad\qquad +n^2b^{2d}\exp(-nb^d2^{d-3}pc_1(1+p(2b)^{d+1}d^{1/2}c_2)^{-1})\Big)\\
&=\ O((nb^d)^{-1}),
\end{aligned}$$

for any $z\in\mathcal{I}_b$, and $\sup_{z\in\mathcal{I}_b}|B_{n3}(z)|=O((nb^d)^{-1})$. With the fact that

$$\begin{aligned}
B_{n4}(z) &= 2n(1-\Delta(z))\tilde K(0)E\frac{(\Delta(z)-\delta_n)\tilde K_{bn}(z)}{[I_n(z)]^2}\\
&= 2nb^d(1-\Delta(z))\tilde K(0)\\
&\quad \times E\int\tilde K(u)f(z-bu)\Big\{\frac{\Delta(z)(1-\Delta(z-bu))}{[I_{n-1}(z)]^2}\\
&\qquad -\frac{(1-\Delta(z))\Delta(z-bu)}{[I_{n-1}(z)+\tilde K(u)]^2}\Big\}du,
\end{aligned}$$

we have

$$\begin{aligned}
&|B_{n4}(z)|\\
&\le\ 2nb^d(1-\Delta(z))\tilde K(0)\\
&\quad \times E\int\Big\{\frac{\tilde K(u)}{[I_{n-1}(z)]^2}|\Delta(z)-\Delta(z-bu)|\\
&\qquad +\Big(\frac{\tilde K(u)}{[I_{n-1}(z)]^2}-\frac{\tilde K(u)}{[I_{n-1}(z)+\tilde K(u)]^2}\Big)(1-\Delta(z))\Delta(z-bu)\Big\}f(z-bu)du\\
&=\ 2nb^{d+1}(1-\Delta(z))\tilde K(0)E[I_{n-1}(z)]^{-2}\ell_\Delta\int\tilde K(u)\|u\|f(z-bu)du\\
&\quad +4nb^d(1-\Delta(z))^2\tilde K(0)E[I_{n-1}(z)]^{-3}\int\tilde K^2(u)\Delta(z-bu)f(z-bu)du\Big\}\\
&=\ O(b(nb^d)^{-1})+O((nb^d)^{-2})\ =O((nb^d)^{-2}),
\end{aligned}$$

and $\sup_{z\in\mathcal{I}_b}|B_{n4}(z)|=O((nb^d)^{-2})$. Thus we have $\sup_{z\in\mathcal{I}_b}|B_n(z)|=O((nb^d)^{-1})$, and

$$nh^d \int \Big[n^{-1} \sum_{i=1}^{n} K_{hi}(x)\hat{r}_{ni}\delta_i\varepsilon_i\Big]^2 d\psi(x) = O_p((nb^d)^{-1}).$$

Moreover, one obtains

$$\int \Big[n^{-1} \sum_{i=1}^{n} K_{hi}(x)\hat{r}_{ni}\delta_i\varepsilon_i\Big]^2 d\psi(x) = O_p((nh^d)^{-1}(nb^d)^{-1})) = o_p(n^{-1}). \tag{1.34}$$

This completes the proof of (1.29), and hence we obtain (1.27).

Step 2. In this part, we shall prove (1.22)-(1.26) in two steps, (2.a) and (2.b).

(2.a) We will prove (1.22), (1.24) and (1.25) by similar arguments used in proving the asymptotic normality of the minimum distance estimator when data is complete in K-N. Let

$$\dot{\hat{T}}_{nj}(\theta) := -2 \int \hat{U}_{nj}(x,\theta)\dot{\mu}_n(x,\theta)d\hat{\psi}_w(x), \quad j = 1, 2,$$

be the derivative of $\hat{T}_{nj}(\theta)$ with respect to θ. Since θ_0 is an interior point of Θ by condition (m4), and $\hat{\theta}_{nj}$ is consistent for θ_{nj} by Theorem 1.3.1, $\hat{\theta}_{nj}$ will be in the interior of Θ and $\dot{\hat{T}}_{nj}(\hat{\theta}_{nj}) = 0$ with arbitrarily large probability for all sufficient large n. The equation $\dot{\hat{T}}_{nj}(\theta) = 0$ is equivalent to

$$\int (\hat{U}_{nj}(x,\hat{\theta}_{nj}) - U_{nj}(x,\hat{\theta}_{nj}))\dot{\mu}_n(x,\hat{\theta}_{nj})d\hat{\psi}_w(x) + \int U_{nj}(x)\dot{\mu}_n(x,\hat{\theta}_{nj})d\hat{\psi}_w(x)$$
$$= \int Z_n(x,\hat{\theta}_{nj})\dot{\mu}_n(x,\hat{\theta}_{nj})d\hat{\psi}_w(x), \qquad j = 1, 2. \tag{1.35}$$

By similar proof as that of (4.16) in K-N, the right-hand side of (1.35) equals $R_n(\hat{\theta}_{nj} - \theta_0)$ for all $n \geq 1$, with $R_n = \Sigma_0 + o_p(1)$; while for the second term on the left-hand side, one has

$\int U_{nj}(x)\dot\mu_n(x,\hat\theta_{nj})d\hat\psi_w(x) = S_{nj} + o_p(n^{-1/2})$ by similar proofs as those of Lemma 4.1 and Lemma 4.2 in K-N, with U_n and ε_i replaced by U_{nj} and ε^*_{ij} in (1.5), respectively. Recall u_n and d_{ni} from (1.5). For the first term on the left-hand side with $j = 1$, note that

$$\begin{aligned}
&\int (\hat U_{n1}(x,\hat\theta_{n1}) - U_{n1}(x,\hat\theta_{n1}))\dot\mu_n(x,\hat\theta_{n1})d\hat\psi_w(x) \\
= \quad & \|u_n\| \int \Big[n^{-1}\sum_{i=1}^n K_{hi}(x)(1-\delta_i)\frac{d_{ni}}{\|u_n\|}\Big]\dot\mu_n(x,\hat\theta_{n1})d\hat\psi_w(x) \\
& +u_n^T \int \Big[n^{-1}\sum_{i=1}^n K_{hi}(x)(1-\delta_i)\dot m_{\theta_0}(X_i)\Big]\dot\mu_n(x,\hat\theta_{n1})d\hat\psi_w(x) := J_{n1} + J_{n2}.
\end{aligned}$$

By (m4), (m5), (a), and result (1.8), we obtain

$$\begin{aligned}
n^{1/2}\|J_{n1}\| \;&\le\; n^{1/2}\|u_n\| \max_{1\le i\le n}\frac{|d_{ni}|}{\|u_n\|}\int \hat f_h(x)\|\dot\mu_n(x,\hat\theta_{n1})\|d\hat\psi_w(x) \\
&\le\; n^{1/2}\|u_n\| \max_{1\le i\le n}\frac{|d_{ni}|}{\|u_n\|}\Big\{\int \hat f_h(x)\|\dot\mu_n(x,\theta_0)\|d\hat\psi_w(x) \\
&\qquad + \max_{1\le i\le n}\|\dot m_{\hat\theta_{n1}}(X_i) - \dot m_{\theta_0}(X_i)\|\int \hat f_h(x)d\hat\psi_w(x)\Big\} \;= o_p(1).
\end{aligned}$$

Moreover, observe that

$$\begin{aligned}
n^{1/2}J_{n2}^T \;=\; & n^{1/2}u_n^T\int \dot\mu_{h\delta}(x)\dot\mu_h^T(x)d\hat\psi_w(x) \\
& +n^{1/2}u_n^T\int \dot\mu_{h\delta}(x)\{\dot\mu_n^T(x,\theta_0) - \dot\mu_h^T(x)\}d\hat\psi_w(x) \\
& +n^{1/2}u_n^T\int \dot\mu_{h\delta}(x)\{\dot\mu_n^T(x,\hat\theta_{n1}) - \dot\mu_n^T(x,\theta_0)\}d\hat\psi_w(x) \\
& +n^{1/2}u_n^T\int \{\dot\mu_{n\delta}(x,\theta_0) - \dot\mu_{h\delta}(x)\}\dot\mu_h^T(x)d\hat\psi_w(x) \\
& +n^{1/2}u_n^T\int \{\dot\mu_{n\delta}(x,\theta_0) - \dot\mu_{h\delta}(x)\}\{\dot\mu_n^T(x,\theta_0) - \dot\mu_h^T(x)\}d\hat\psi_w(x) \\
& +n^{1/2}u_n^T\int \{\dot\mu_{n\delta}(x,\theta_0) - \dot\mu_{h\delta}(x)\}\{\dot\mu_n^T(x,\hat\theta_{n1}) - \dot\mu_n^T(x,\theta_0)\}d\hat\psi_w(x).
\end{aligned}$$

On the right-hand side of last equality, the last five terms are $o_p(1)$, because of (m5), (a), (1.8), C-S inequality and the fact that

$$\begin{aligned}
& E\int\{\dot\mu_n(x,\theta_0)-\dot\mu_h(x)\}\{\dot\mu_n^T(x,\theta_0)-\dot\mu_h^T(x)\}d\psi(x)\\
&= \int Var(\dot\mu_n(x,\theta_0))d\psi(x)=O_p((nh^d)^{-1}),\\
& \int\{\dot\mu_n(x,\hat\theta_{n1})-\dot\mu_n(x,\theta_0)\}\{\dot\mu_n^T(x,\hat\theta_{n1})-\dot\mu_n^T(x,\theta_0)\}d\psi(x)\\
&= \int \hat f_h^2(x)d\psi(x)\max_{1\le i\le n}(\dot m_{\hat\theta_{n1}}(X_i)-\dot m_{\theta_0}(X_i))(\dot m_{\hat\theta_{n1}}^T(X_i)-\dot m_{\theta_0}^T(X_i))=o_p(h^d),\\
& E\int\{\dot\mu_{n\delta}(x,\theta_0)-\dot\mu_{h\delta}(x)\}\{\dot\mu_{n\delta}^T(x,\theta_0)-\dot\mu_\delta h^T(x)\}d\psi(x)\\
&= \int Var(\dot\mu_{n\delta}(x,\theta_0))d\psi(x)=O_p((nh^d)^{-1}).
\end{aligned}$$

For the first term, by (m4), (m5), (a), (1.8), and C-S inequality, one has

$$\int \dot\mu_h(x)\dot\mu_{h\delta}^T(x)d\hat\psi_w(x)=\Sigma_0^*+o_p(1).$$

Hence (1.22) holds. If under H_0, $m_{\theta_0}(x)$ is a linear function of θ_0, and $\hat\alpha_n$ is the least square estimator, we have $u_n=\tilde\Sigma_n^{-1}\tilde S_n$ and result (1.24).

To prove (1.25), it suffices to show that when $j=2$, the first term in the left-hand side of (1.35) multiplied by $n^{1/2}$ is $o_p(1)$. Note that by C-S inequality,

$$\begin{aligned}
& \Big\|\int(\hat U_{n2}(x,\hat\theta_{n2})-U_{n2}(x,\hat\theta_{n2}))\dot\mu_n(x,\hat\theta_{n2})d\hat\psi_w(x)\Big\|^2\\
&\le \Big(1+\sup_{x\in\mathcal{I}}|f^2(x)/\hat f_w^2(x)-1|\Big)^2\tilde A_{n2}\int\|\dot\mu_n(x,\hat\theta_{n2})\|^2d\psi(x).
\end{aligned}$$

By the fact that $\int\|\dot\mu_n(x,\hat\theta_{n2})\|^2d\psi(x)=O_p(1)$, and $\sup_{x\in\mathcal{I}}|f^2(x)/\hat f_w^2(x)-1|=o_p(1)$ derived by (1.8), and it suffices to prove $\tilde A_{n2}=o_p(n^{-1})$, which in turn follows (a), (1.12), (1.19), and (1.34).

(2.b) We shall prove (1.26) in this step. Based on (1.24) and (1.25), it suffices to prove that

$$n^{1/2}\{S_{n1} + \Sigma_0^* \tilde{\Sigma}_n^{-1} \tilde{S}_n\} \to_d \mathcal{N}_q(0, \Sigma_1), \tag{1.36}$$

$$n^{1/2} S_{n2} \to_d \mathcal{N}_q(0, \Sigma_2). \tag{1.37}$$

The proof of (1.37) is similar as that of Lemma 4.1 (a) in K-N, if ε_i, σ^2 and Σ there are replaced by $\delta_i \varepsilon_i / \Delta(X_i)$, σ^2/Δ, and Σ_2 in (1.5), respectively. To prove (1.36), note that $n^{1/2}\tilde{S}_n = O_p(1)$ by the Central Limit Theorem, and $\tilde{\Sigma}_n^{-1} = \tilde{\Sigma}_0^{-1} + o_p(1)$ by Law of Large Numbers and routine calculations. Thus we have

$$\begin{aligned} n^{1/2}\{S_{n1} + \Sigma_0^* \tilde{\Sigma}_n^{-1} \tilde{S}_n\} &= n^{1/2}\{S_{n1} + \Sigma_0^* \tilde{\Sigma}_0^{-1} \tilde{S}_n\} + \Sigma_0^*(\tilde{\Sigma}_n^{-1} - \tilde{\Sigma}_0^{-1})(n^{1/2}\tilde{S}_n) \\ &= n^{1/2}\{S_{n1} + \Sigma_0^* \tilde{\Sigma}_0^{-1} \tilde{S}_n\} + o_p(1), \end{aligned}$$

and it suffices to show $n^{1/2}\{S_{n1} + \Sigma_0^* \tilde{\Sigma}_0^{-1} \tilde{S}_n\} \to_d \mathcal{N}_q(0, \Sigma_1)$. Write

$$\begin{aligned} & n^{1/2}\{S_{n1} + \Sigma_0^* \tilde{\Sigma}_0^{-1} s_{n1}\} \\ &\quad = n^{-1/2} \sum_{i=1}^n \Big(\int K_{hi}(x) \dot{\mu}_h(x) d\psi(x) + \Sigma_0^* \tilde{\Sigma}_0^{-1} \dot{m}_{\theta_0}(X_i) \Big) \delta_i \varepsilon_i \\ &\quad = n^{-1/2} \sum_{i=1}^n s_{ni}, \quad \text{say.} \end{aligned}$$

Note that by (e1) and (e2), $\{s_{ni}, i = 1, \cdots, n\}$ are i.i.d. centered r.v.'s for each n. By the Lindeberg-Feller C.L.T., it suffices to prove that as $n \to \infty$,

$$E s_{n1}^2 \to \Sigma_1, \tag{1.38}$$

$$E\{s_{n1}^2 I(|s_{n1}| > n^{1/2}\eta)\} \to 0 \quad \forall \eta > 0. \tag{1.39}$$

By the continuity of σ^2, Δ, f, and g, we obtain

$$\begin{aligned}
E s_{n1}^2 &= E\Big[\Big(\int K_h(x-X)\dot{\mu}_h(x)d\psi(x) + \Sigma_0^*\tilde{\Sigma}_0^{-1}\dot{m}_{\theta_0}(X)\Big)^2 \Delta(X)\sigma^2(X)\Big] \\
&= E\int\int K_h(x-X)K_h(y-X)\sigma^2(X)\Delta(X)\dot{\mu}_h(x)\dot{\mu}_h^T(y)d\psi(x)d\psi(y) \\
&\quad + E\int K_h(x-X)\sigma^2(X)\Delta(X)\dot{\mu}_h(x)\dot{m}_{\theta_0}^T(X)d\psi(x)\tilde{\Sigma}_0^{-1}\Sigma_0^* \\
&\quad + \Sigma_0^*\tilde{\Sigma}_0^{-1}E\int K_h(x-X)\sigma^2(X)\Delta(X)\dot{m}_{\theta_0}(X)\dot{\mu}_h^T(x)d\psi(x) \\
&\quad + \Sigma_0^*\tilde{\Sigma}_0^{-1}E[\dot{m}_{\theta_0}(X)\dot{m}_{\theta_0}^T(X)\sigma^2(X)\Delta(X)]\tilde{\Sigma}_0^{-1}\Sigma_0^* \\
&\to \int \sigma^2(x)\Delta(x)\dot{m}_{\theta_0}(x)\dot{m}_{\theta_0}^T(x)(f(x))^{-1}g^2(x)dx \\
&\quad + 2\Big(\int \sigma^2(x)\Delta(x)\dot{m}_{\theta_0}(x)\dot{m}_{\theta_0}^T(x)g(x)dx\Big)\tilde{\Sigma}_0^{-1}\Sigma_0^* \\
&\quad + \Sigma_0^*\tilde{\Sigma}_0^{-1}\Big(\int \sigma^2(x)\Delta(x)\dot{m}_{\theta_0}(x)\dot{m}_{\theta_0}^T(x)f(x)dx\Big)\tilde{\Sigma}_0^{-1}\Sigma_0^* \; = \Sigma_1,
\end{aligned}$$

Hence (1.38) is proved. Note that by the Hölder's inequality, the L.H.S. of (1.39) with $\eta = \delta_0$ in (e3) is bounded by

$$\begin{aligned}
&C n^{-\delta_0/2} E s_{n1}^{2+\delta_0} \\
&\quad = C n^{-\delta_0/2} E\Big[\Big\{\int K_h(x-X)\dot{\mu}_h(x)d\psi(x) + \Sigma_0^*\tilde{\Sigma}_0^{-1}\dot{m}_{\theta_0}(X)\Big\}^{2+\delta_0}|\delta\varepsilon|^{2+\delta_0}\Big] \\
&\quad \le C n^{-\delta_0/2} E\Big[\Big\{2\int K_h(x-X)\dot{\mu}_h(x)d\psi(x)\Big\}^{2+\delta_0}|\delta\varepsilon|^{2+\delta_0}\Big] \\
&\qquad + C n^{-\delta_0/2} E[\{2\Sigma_0^*\tilde{\Sigma}_0^{-1}\dot{m}_{\theta_0}(X)\}^{2+\delta_0}|\delta\varepsilon|^{2+\delta_0}]
\end{aligned}$$

$$
\begin{aligned}
&\le Cn^{-\delta_0/2}2^{2+\delta_0}E\Big[\Big\{\int (K_h(x-X)\dot{\mu}_h(x))^{\frac{2+\delta_0}{2}}d\psi(x)\Big\}^2\Big\{\int d\psi(x)\Big\}^{\delta_0}|\delta\varepsilon|^{2+\delta_0}\Big] \\
&\quad +Cn^{-\delta_0/2}E[\{2\Sigma_0^*\tilde{\Sigma}_0^{-1}\dot{m}_{\theta_0}(X)\}^{2+\delta_0}|\delta\varepsilon|^{2+\delta_0}] \\
&= O_p((nh^d)^{-\delta_0/2}).
\end{aligned}
$$

Therefore the proof is complete.

Remark 1.4.1. (Choice of G). Assuming $f=0$ implies $g=0$. When $q=1$ and $\sigma^2(x)\equiv\sigma^2$, a constant, the asymptotic variance of $\hat{\theta}_{n1}$ satisfies

$$
\begin{aligned}
v_1 := {} & \sigma^2\tilde{\Sigma}_0^{-1}+\sigma^2\Sigma_0^{-2}\Big[\int \Delta(x)\dot{m}_{\theta_0}^2(x)(f(x))^{-1}g^2(x)dx \\
& -\Big(\int \Delta(x)\dot{m}_{\theta_0}^2(x)f(x)dx\Big)^{-1}\Big(\int \Delta(x)\dot{m}_{\theta_0}^2(x)g(x)dx\Big)^2\Big] \\
\ge {} & \sigma^2\tilde{\Sigma}_0^{-1},
\end{aligned}
$$

because, by C-S inequality,

$$
\begin{aligned}
&\Big(\int \Delta(x)\dot{m}_{\theta_0}^2(x)g(x)dx\Big)^2 \\
&= \Big(\int \Delta^{1/2}(x)\dot{m}_{\theta_0}(x)f^{1/2}(x)\Delta^{1/2}(x)\dot{m}_{\theta_0}(x)f^{-1/2}(x)g(x)dx\Big)^2 \\
&\le \int \Delta(x)\dot{m}_{\theta_0}^2(x)f(x)dx\int \Delta(x)\dot{m}_{\theta_0}^2(x)(f(x))^{-1}g^2(x)dx,
\end{aligned}
$$

with equality if and only if $g\propto f$; and the asymptotic variance of $\hat{\theta}_{n2}$ satisfies

$$
\begin{aligned}
v_2 := {} & \sigma^2\int (\Delta(x))^{-1}\dot{m}_{\theta_0}^2(x)g^2(x)(f(x))^{-1}dx\Big(\int \dot{m}_{\theta_0}^2(x)g(x)dx\Big)^{-2} \\
\ge {} & \sigma^2\Big(\int \Delta(x)\dot{m}_{\theta_0}^2(x)f(x)dx\Big)^{-1} \\
= {} & \sigma^2\tilde{\Sigma}_0^{-1},
\end{aligned}
$$

because

$$
\begin{aligned}
&\Big(\int \dot{m}_{\theta_0}^2(x)g(x)dx\Big)^2 \\
&= \Big(\int (\Delta(x))^{-1/2}\dot{m}_{\theta_0}(x)g(x)(f(x))^{-1/2}(\Delta(x))^{1/2}\dot{m}_{\theta_0}(x)(f(x))^{1/2}dx\Big)^2 \\
&\leq \int (\Delta(x))^{-1}\dot{m}_{\theta_0}^2(x)g^2(x)(f(x))^{-1}dx \int \Delta(x)\dot{m}_{\theta_0}^2(x)f(x)dx,
\end{aligned}
$$

with equality if and only if $g \propto f\Delta$. This implies that both lower bounds on the asymptotic variances of $\hat{\theta}_{nj}$, $j = 1, 2$, are at that of the least square estimator's when the regression function is linear.

1.5 Asymptotic distribution of the test statistics under H_0

In this section we shall discuss the asymptotic null distribution of $\hat{\mathcal{D}}_{nj}$ in Theorem 1.5.1.

Theorem 1.5.1. *Assume that* H_0*, (e1), (e2), (e3), (e4), (f1), (f2), (g), (k1), (m1)-(m5), (a), and (h3) hold. Then,*

$$\hat{\mathcal{D}}_{n1} \rightarrow_d \mathcal{N}(0, 1).$$

If, in addition, (k2), (b2), and (h4) hold, then,

$$\hat{\mathcal{D}}_{n2} \rightarrow_d \mathcal{N}(0, 1).$$

Consequently, for each $j = 1, 2$, the test that rejects H_0 whenever $|\hat{\mathcal{D}}_{nj}| > z_{\alpha/2}$, is of the asymptotic size α.

The proof of Theorem 1.5.1 is facilitated by Lemma 1.5.2-1.5.7. The idea of the proof is similar to that of Theorem 5.1 in K-N. Lemma 1.5.1 is applied to prove Lemma 1.5.2.

Lemma 1.5.1. (Theorem 1 of Hall (1984)) *Let $\tilde{X}_i$, $1 \le i \le n$, be i.i.d. random vectors, and let*

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

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

$$E(H_n(\tilde{X}_1, \tilde{X}_2)|\tilde{X}_1) = 0, \quad a.s., \qquad and$$
$$EH_n^2(\tilde{X}_1, \tilde{X}_2) < \infty, \quad for\ each\ n \ge 1.$$

If

$$[EG_n^2(\tilde{X}_1, \tilde{X}_2) + n^{-1} EH_n^4(\tilde{X}_1, \tilde{X}_2)]/[EH_n^2(\tilde{X}_1, \tilde{X}_2)]^2 \to 0,$$

then, U_n is asymptotically normally distributed with mean zero and variance equal to $n^2 EH_n^2(\tilde{X}_1, \tilde{X}_2)/2$.

To proceed further, we need the following notation:

$$K_2(v) := \int K(u)K(u+v)du, \quad \|K_2\|^2 := \int K_2^2(v)dv, \tag{1.40}$$
$$\Gamma_1 := 2\|K_2\|^2 \int_{\mathcal{I}} \frac{\Delta^2(x)(\sigma^2(x))^2 g^2(x)}{f^2(x)} dx, \quad \Gamma_2 := 2\|K_2\|^2 \int_{\mathcal{I}} \frac{(\sigma^2(x))^2 g^2(x)}{\Delta^2(x) f^2(x)} dx,$$

$$\Gamma_{n1} := 2h^d \int\int [EK_h(x-X)K_h(y-X)\Delta(X)\sigma^2(X)]^2 d\psi(x)d\psi(y),$$
$$\Gamma_{n2} := 2h^d \int\int [EK_h(x-X)K_h(y-X)(\Delta(X))^{-1}\sigma^2(X)]^2 d\psi(x)d\psi(y).$$

Recall the definitions of $\hat{C}_{nj}$, $\hat{\Gamma}_{nj}$, $\hat{\mathcal{D}}_{nj}$ from (1.4), and $\tilde{T}_{nj}$, T_{nj}, C_{nj} from (1.5).

Lemma 1.5.2. *Suppose* H_0, *(e1), (e2), (e4), (f1), (g), (k1), (h1), and (h2) hold. Then,*

$$nh^{d/2}(T_{nj}(\theta_0) - C_{nj}) \to_d \mathcal{N}(0, \Gamma_j), \quad j = 1, 2.$$

The proof of Lemma 1.5.2 follows the same routine as that of Lemma 5.1 in K-N, but with the following changes: for $j = 1$, replace ε_i, $\sigma^2(x)$, $\sigma^4(x)$, Γ_n, and Γ in K-N by $\delta_i\varepsilon_i$, $\Delta(x)\sigma^2(x)$, $\Delta(x)\sigma^4(x)$, Γ_{n1}, and Γ_1, respectively; for $j = 2$, replace ε_i, $\sigma^2(x)$, $\sigma^4(x)$, Γ_n, and Γ in K-N by $(\Delta(X_i))^{-1}\delta_i\varepsilon_i$, $(\Delta(x))^{-1}\sigma^2(x)$, $(\Delta(x))^{-3}\sigma^4(x)$, Γ_{n2}, and Γ_2, respectively. The following results will be used in the proofs later:

$$\Gamma_{nj} \to_{a.s.} \Gamma_j, \quad j = 1, 2. \tag{1.41}$$

Remark 1.5.1. Similar as Remark 5.1 in K-N, one has

$$nh^{d/2}(T_{nj}(\theta_0) - ET_{nj}(\theta_0)) \to_d \mathcal{N}(0, \Gamma_j), \qquad j = 1, 2.$$

Lemma 1.5.3. *Under* H_0, *(e1), (e2), (f1), (f2), (k1), and (h3),*

$$nh^{d/2}|\tilde{T}_{nj}(\theta_0) - T_{nj}(\theta_0)| = o_p(1), \quad j = 1, 2.$$

The proof of Lemma 1.5.3 is similar to that of Lemma 5.3 in K-N where now $U_n(x)$ would be changed to $U_{nj}(x)$, $j = 1, 2$.

Lemma 1.5.4. *Under* H_0, *(e1), (e2), (e3), (f1), (k1), (m1)-(m5), (a), (h1), and (h2),*

$$nh^{d/2}|\tilde{T}_{n1}(\hat{\theta}_{n1}) - \tilde{T}_{n1}(\theta_0)| = o_p(1).$$

If, in addition, (k2), (b2), and (h4) hold, then,

$$nh^{d/2}|\tilde{T}_{n2}(\hat{\theta}_{n2}) - \tilde{T}_{n2}(\theta_0)| = o_p(1).$$

Proof. Observe that

$$\tilde{T}_{nj}(\theta_0) - \tilde{T}_{nj}(\hat{\theta}_{nj}) = 2\int U_{nj}(x) Z_n(x, \hat{\theta}_{nj}) d\hat{\psi}_w(x) - \int Z_n^2(x, \hat{\theta}_{nj}) d\hat{\psi}_w(x).$$

If we follow similar routine as proof of Lemma 5.2 in K-N, with $\hat{\theta}_n$ and U_n in K-N changed to $\hat{\theta}_{nj}$ and U_{nj}, respectively, we can find that it suffices to show

$$nh^{d/2}(\hat{\theta}_{nj} - \theta_0)^T \int U_{nj}(x) \dot{\mu}_n(x, \hat{\theta}_{nj}) d\hat{\psi}_w(x) = o_p(1), \quad j = 1, 2.$$

Note that the integral is the same as the second term in the left-hand side of (1.35). Thus,

$$\begin{aligned} \text{L.H.S.} &= nh^{d/2}(\hat{\theta}_{nj} - \theta_0)^T \int Z_n(x, \hat{\theta}_{nj}) \dot{\mu}_n(x, \hat{\theta}_{nj}) d\hat{\psi}_w(x) \\ &\quad - nh^{d/2}(\hat{\theta}_{nj} - \theta_0)^T \int (\hat{U}_{nj}(x, \hat{\theta}_{nj}) - U_{nj}(x, \hat{\theta}_{nj})) \dot{\mu}_n(x, \hat{\theta}_{nj}) d\hat{\psi}_w(x) \\ &:= Q_{n1j} - Q_{n2j}, \quad j = 1, 2, \quad \text{say.} \end{aligned}$$

We have $Q_{n1j} = o_p(1)$ for $j = 1, 2$, by the same argument as used in proving (5.10) in K-N.

By the proof of step (2.a) of Theorem 1.4.1, one obtains that

$$n^{1/2}\int(\hat{U}_{nj}(x,\hat{\theta}_{nj})-U_{nj}(x,\hat{\theta}_{nj}))\dot{\mu}_n(x,\hat{\theta}_{nj})d\hat{\psi}_w(x)=O_p(1),\quad j=1,2,$$

hence $Q_{n2j}=nh^{d/2}O_p(n^{-1/2})O_p(n^{-1/2})=O_p(h^{d/2})$, and the proof is completed.

Lemma 1.5.5. *If* H_0*, (e1), (e2), (e3), (f1), (k), (m1)-(m5), (a), (h1), and (h2) hold, then*

$$nh^{d/2}|\hat{T}_{n1}(\hat{\theta}_{n1})-\tilde{T}_{n1}(\hat{\theta}_{n1})|=o_p(1).$$

If, in addition, (k2), (b2), (h3), and (h4) hold, then,

$$nh^{d/2}|\hat{T}_{n2}(\hat{\theta}_{n2})-\tilde{T}_{n2}(\hat{\theta}_{n2})|=o_p(1).$$

Proof. Observe that

$$\begin{aligned}
&|\hat{T}_{nj}(\hat{\theta}_{nj})-\tilde{T}_{nj}(\hat{\theta}_{nj})|\\
&= \Big|\int[\hat{U}_{nj}(x,\theta_0)-Z_n(x,\hat{\theta}_{nj})]^2-[U_{nj}(x,\theta_0)-Z_n(x,\hat{\theta}_{nj})]^2d\hat{\psi}_w(x)\Big|\\
&\le (1+\sup_{x\in\mathcal{I}}|f^2(x)/\hat{f}_w^2(x)-1|)\Big[\int(\hat{U}_{nj}(x,\theta_0)-U_{nj}(x,\theta_0))^2d\psi(x)\\
&\qquad +2\Big\{\int(\hat{U}_{nj}(x,\theta_0)-U_{nj}(x,\theta_0))^2d\psi(x)\Big\}^{1/2}\\
&\qquad \times\Big\{\int(U_{nj}(x,\theta_0)-Z_n(x,\hat{\theta}_{nj}))^2d\psi(x)\Big\}^{1/2}\Big],
\end{aligned}$$

by the C-S inequality. The results in (1.18), (1.19), and (1.34) lead to the fact

$$\int(\hat{U}_{nj}(x,\theta_0)-U_{nj}(x,\theta_0))^2d\psi(x)=o_p(n^{-1}).$$

Let for $j = 1, 2$,

$$\hat{u}_{nj} := \hat{\theta}_{nj} - \theta_0, \qquad \hat{d}_{nij} := m_{\hat{\theta}_{nj}}(X_i) - m_{\theta_0}(X_i) - \hat{u}_{nj}^T \dot{m}_{\theta_0}(X_i). \tag{1.42}$$

Then,

$$\begin{aligned}
&\int [U_{nj}(x, \theta_0) - Z_n(x, \hat{\theta}_{nj})]^2 d\psi(x) \\
&\le 2 \int U_{nj}^2(x, \theta_0) d\psi(x) + 2 \int Z_n^2(x, \hat{\theta}_{nj}) d\psi(x) \\
&\le 2 \int U_{nj}^2(x, \theta_0) d\psi(x) + 4\|\hat{u}_{nj}\|^2 \max_{1 \le i \le n} \frac{\hat{d}_{nij}}{\|\hat{u}_{nj}\|} \int \hat{f}_h^2(x) d\psi(x) \\
&\quad +4\|\hat{u}_{nj}\|^2 \int \Big[n^{-1} \sum_{i=1}^{n} K_{hi}(x) \|\dot{m}_{\theta_0}(X_i)\| \Big]^2 d\psi(x) \\
&= O_p((nh^d)^{-1}) + o_p(n^{-1}) + O_p(n^{-1}) = O_p((nh^d)^{-1}), \qquad j = 1, 2.
\end{aligned}$$

by (m4) and Theorem 1.4.1. This completes the proof of Lemma 1.5.5.

Lemma 1.5.6. *If* H_0*, (e1), (e2), (e3), (f1), (f2), (k), (m1)-(m5), (a), and (h3) hold, then*

$$nh^{d/2} |\hat{C}_{n1} - C_{n1}| = o_p(1).$$

If, in addition, (k2), (b2), and (h4) hold, then,

$$nh^{d/2} |\hat{C}_{n2} - C_{n2}| = o_p(1).$$

Proof. Let for $j = 1, 2$, $i = 1, \cdots, n$,

$$v_w(x) := f^2(x)/\hat{f}_w^2(x) - 1, \qquad t_{ij} := m_{\hat{\theta}_{nj}}(X_i) - m_{\theta_0}(X_i), \tag{1.43}$$

$$s_i := (1-\delta_i)(m_{\hat{\alpha}_n}(X_i) - m_{\theta_0}(X_i)), \qquad a_i := \Delta(X_i)\hat{r}_{ni},$$

$$c_i := (1 - \frac{\delta_i}{\Delta(X_i)})(m_{\hat{\alpha}_n}(X_i) - m_{\theta_0}(X_i)),$$

$$q_i := -\hat{r}_{ni}\delta_i(m_{\hat{\alpha}_n}(X_i) - m_{\theta_0}(X_i)),$$

$$w_{i1} := s_i - t_{i1}, \qquad w_{i2} := c_i + q_i - t_{i2}.$$

Note that from result (1.17), for $i = 1, \cdots, n$,

$$\begin{aligned}
\hat{\varepsilon}_{i1} &= (\tilde{Y}_{i1} - m_{\theta_0}(X_i)) + (\hat{Y}_{i1} - \tilde{Y}_{i1}) - (m_{\hat{\theta}_{nj}}(X_i) - m_{\theta_0}(X_i)) \\
&= \varepsilon^*_{i1} + s_i - t_{i1}, \\
\hat{\varepsilon}_{i2} &= (\tilde{Y}_{i2} - m_{\theta_0}(X_i)) + (\hat{Y}_{i2} - \tilde{Y}_{i2}) - (m_{\hat{\theta}_{nj}}(X_i) - m_{\theta_0}(X_i)) \\
&= \varepsilon^*_{i2}(1 + a_i) + c_i + q_i - t_{i2},
\end{aligned}$$

and hence,

$$\begin{aligned}
&\hat{C}_{n1} - C_{n1} \\
&= n^{-2}\sum_{i=1}^{n}\int K^2_{hi}(x)[\hat{\varepsilon}^2_{i1} - (\varepsilon^*_{i1})^2]d\hat{\psi}_w(x) + n^{-2}\sum_{i=1}^{n}\int K^2_{hi}(x)(\varepsilon^*_{i1})^2 v_w d\psi(x) \\
&= n^{-2}\sum_{i=1}^{n}\int K^2_{hi}(x)[(\varepsilon^*_{i1} + w_{i1})^2 - (\varepsilon^*_{i1})^2]d\hat{\psi}_w(x) \\
&\quad + n^{-2}\sum_{i=1}^{n}\int K^2_{hi}(x)(\varepsilon^*_{i1})^2 v_w d\psi(x),
\end{aligned}$$

$$\begin{aligned}
&\hat{C}_{n2} - C_{n2} \\
&= n^{-2}\sum_{i=1}^{n}\int K^2_{hi}(x)[\hat{\varepsilon}^2_{i2} - (\varepsilon^*_{i2})^2]d\hat{\psi}_w(x) + n^{-2}\sum_{i=1}^{n}\int K^2_{hi}(x)(\varepsilon^*_{i2})^2 v_w d\psi(x)
\end{aligned}$$

$$= n^{-2}\sum_{i=1}^{n}\int K_{hi}^{2}(x)[(\varepsilon_{i2}^{*}(1+a_i)+w_{i2})^2-(\varepsilon_{i2}^{*})^2]d\hat{\psi}_w(x)$$
$$+n^{-2}\sum_{i=1}^{n}\int K_{hi}^{2}(x)(\varepsilon_{i2}^{*})^2 v_w d\psi(x).$$

By (a), (m4), (1.9), (1.13), (1.27), one has

$$v_n := \sup_{x\in\mathcal{I}}|v_w(x)| = O_p((\log_k n)(\log n/n)^{2/(d+4)}), \tag{1.44}$$
$$t_{nj} := \max_{1\le i\le n}|t_{ij}| = O_p((nh^d)^{-1/2}), \quad j=1,2,$$
$$s_n := \max_{1\le i\le n}|s_i| = O_p(n^{-1/2}), \qquad a_n := \max_{1\le i\le n}|a_i| = O_p((nb^d)^{-1/2}(\log n)^{1/2}),$$
$$c_n := \max_{1\le i\le n}|c_i| = O_p(n^{-1/2}), \qquad q_n := \max_{1\le i\le n}|q_i| = O_p((n^{-1/2}b^{-d/2}(\log n)^{1/2}),$$
$$w_{n1} := \max_{1\le i\le n}|w_{i1}| \le s_n + t_{n1} = O_p((nh^d)^{-1/2}),$$
$$w_{n2} := \max_{1\le i\le n}|w_{i2}| \le c_n + q_n + t_{n2} = O_p((n^{-1/2}b^{-d/2}(\log n)^{1/2}).$$

These facts together with the following facts that

$$S_{nj,2} := n^{-2}\sum_{i=1}^{n} K_{hi}^{2}(x)(\varepsilon_{ij}^{*})^2 d\psi(x) = O_p((nh^d)^{-1}),$$
$$S_{nj,1} := n^{-2}\sum_{i=1}^{n} K_{hi}^{2}(x)|\varepsilon_{ij}^{*}| d\psi(x) = O_p((nh^d)^{-1}),$$
$$S_{nj,0} := n^{-2}\sum_{i=1}^{n} K_{hi}^{2}(x) d\psi(x) = O_p((nh^d)^{-1}), \qquad j=1,2,$$

we obtain

$$nh^{d/2}\{(|\hat{C}_{n1}-C_{n1}|)\} \le nh^{d/2}[(1+v_n)\{2w_{n1}S_{n1,1}+w_{n1}^2 S_{n1,0}\}$$
$$+v_n S_{n1,2}] = o_p(1),$$

$$nh^{d/2}\{(|\hat{C}_{n2} - C_{n2}|)\} \leq nh^{d/2}[(1+v_n)\{2w_{n2}S_{n2,1}(1+a_n) + w_{n2}^2 S_{n2,0} + (2a_n + a_n^2)S_{n2,2}\} + v_n S_{n1,2}] = o_p(1),$$

by (b2) and (h4). This completes the proof.

Lemma 1.5.7. *Under* H_0, *(e1), (e2), (e3), (f1), (k1), (m1)-(m5), (a), (h1), and (h2),*

$$\hat{\Gamma}_{n1} - \Gamma_1 = o_p(1).$$

If in addition, (f2), (k2), (b2), and (h4) hold, then,

$$\hat{\Gamma}_{n2} - \Gamma_2 = o_p(1).$$

Consequently, $\Gamma_j > 0$ *implies* $|\hat{\Gamma}_{nj}\Gamma_j^{-1} - 1| = o_p(1)$, $j = 1, 2$.

Proof. The proof of Lemma 1.5.7 is similar to that of Lemma 5.5 in K-N. Recall v_w, t_{i1}, t_{i2}, s_i, a_i, c_i, q_i from (1.43), and v_n, t_{n1}, t_{n2}, s_n, a_n, c_n, q_n, w_{n1}, w_{n2} from (1.44). Let for $k = 1, 2$,

$$\tilde{\Gamma}_{nk} := 2h^d n^{-2} \sum_{i \neq j} \Big(\int K_{hi}(x) K_{hj}(x) \varepsilon_{ik}^* \varepsilon_{jk}^* d\psi(x) \Big)^2.$$

From result (1.41), it suffices to show

$$\tilde{\Gamma}_{nk} - \Gamma_{nk} = o_p(1), \qquad \hat{\Gamma}_{nk} - \tilde{\Gamma}_{nk} = o_p(1), \qquad k = 1, 2. \tag{1.45}$$

The first claim in (1.45) is proved similarly as (5.13) in K-N. For the second claim, note that

$$
\begin{aligned}
&\hat{\Gamma}_{n1} - \tilde{\Gamma}_{n1} \\
&= 2h^d n^{-2} \sum_{i \neq j} \Big(\int K_{hi}(x) K_{hj}(x) (\varepsilon^*_{i1} + w_{i1})(\varepsilon^*_{j1} + w_{j1})(1 + v_w(x)) d\psi(x) \Big)^2 \\
&\quad -2h^d n^{-2} \sum_{i \neq j} \Big(\int K_{hi}(x) K_{hj}(x) \varepsilon^*_{i1} \varepsilon^*_{j1} d\psi(x) \Big)^2 \\
&= 2h^d n^{-2} \sum_{i \neq j} \Big(\int K_{hi}(x) K_{hj}(x) (\varepsilon^*_{i1} + w_{i1})(\varepsilon^*_{j1} + w_{j1}) d\psi(x) \Big)^2 \\
&\quad +2h^d n^{-2} \sum_{i \neq j} \Big(\int K_{hi}(x) K_{hj}(x) (\varepsilon^*_{i1} + w_{i1})(\varepsilon^*_{j1} + w_{j1}) v_w(x) d\psi(x) \Big)^2 \\
&\quad +4h^d n^{-2} \sum_{i \neq j} \Big(\int K_{hi}(x) K_{hj}(x) (\varepsilon^*_{i1} + w_{i1})(\varepsilon^*_{j1} + w_{j1}) d\psi(x) \Big) \\
&\qquad\qquad \times \Big(\int K_{hi}(x) K_{hj}(x) (\varepsilon^*_{i1} + w_{i1})(\varepsilon^*_{j1} + w_{j1}) v_w(x) d\psi(x) \Big) \\
&\quad -2h^d n^{-2} \sum_{i \neq j} \Big(\int K_{hi}(x) K_{hj}(x) \varepsilon^*_{i1} \varepsilon^*_{j1} d\psi(x) \Big)^2, \\
&\hat{\Gamma}_{n2} - \tilde{\Gamma}_{n2} \\
&= 2h^d n^{-2} \sum_{i \neq j} \Big(\int K_{hi}(x) K_{hj}(x) \{\varepsilon^*_{i2}(1 + a_i) + w_{i2}\} \\
&\qquad\qquad \times \{\varepsilon^*_{j2}(1 + a_j) + w_{j2}\}(1 + v_w(x)) d\psi(x) \Big)^2 \\
&\quad -2h^d n^{-2} \sum_{i \neq j} \Big(\int K_{hi}(x) K_{hj}(x) \varepsilon^*_{i2} \varepsilon^*_{j2} d\psi(x) \Big)^2 \\
&= 2h^d n^{-2} \sum_{i \neq j} \Big(\int K_{hi}(x) K_{hj}(x) \{\varepsilon^*_{i2}(1 + a_i) + w_{i2}\} \\
&\qquad\qquad \times \{\varepsilon^*_{j2}(1 + a_j) + w_{j2}\} d\psi(x) \Big)^2 \\
&\quad +2h^d n^{-2} \sum_{i \neq j} \Big(\int K_{hi}(x) K_{hj}(x) \{\varepsilon^*_{i2}(1 + a_i) + w_{i2}\} \\
&\qquad\qquad \times \{\varepsilon^*_{j2}(1 + a_j) + w_{j2}\} v_w(x) d\psi(x) \Big)^2
\end{aligned}
$$

$$
\begin{aligned}
&+4h^d n^{-2}\sum_{i\neq j}\Big(\int K_{hi}(x)K_{hj}(x)\{\varepsilon^*_{i2}(1+a_i)+w_{i2}\}\\
&\qquad\qquad\times\{\varepsilon^*_{j2}(1+a_j)+w_{j2}\}d\psi(x)\Big)\\
&\qquad\times\Big(\int K_{hi}(x)K_{hj}(x)\{\varepsilon^*_{i2}(1+a_i)+w_{i2}\}\\
&\qquad\qquad\times\{\varepsilon^*_{j2}(1+a_j)+w_{j2}\}v_w(x)d\psi(x)\Big)\\
&-2h^d n^{-2}\sum_{i\neq j}\Big(\int K_{hi}(x)K_{hj}(x)\varepsilon^*_{i2}\varepsilon^*_{j2}d\psi(x)\Big)^2.
\end{aligned}
$$

By Fubini's theorem and taking the expected value, one obtains

$$
\begin{aligned}
W_{nk,2,2} &:= 2h^d n^{-2}\sum_{i\neq j}(\varepsilon^*_{ik})^2(\varepsilon^*_{jk})^2\Big(\int K_{hi}(x)K_{hj}(x)d\psi(x)\Big)^2 = O_p(1),\\
W_{nk,2,1} &:= 2h^d n^{-2}\sum_{i\neq j}(\varepsilon^*_{ik})^2|\varepsilon^*_{jk}|\Big(\int K_{hi}(x)K_{hj}(x)d\psi(x)\Big)^2 = O_p(1),\\
W_{nk,2,0} &:= 2h^d n^{-2}\sum_{i\neq j}(\varepsilon^*_{ik})^2\Big(\int K_{hi}(x)K_{hj}(x)d\psi(x)\Big)^2 = O_p(1),\\
W_{nk,1,1} &:= 2h^d n^{-2}\sum_{i\neq j}|\varepsilon^*_{ik}|\,|\varepsilon^*_{jk}|\Big(\int K_{hi}(x)K_{hj}(x)d\psi(x)\Big)^2 = O_p(1),\\
W_{nk,1,0} &:= 2h^d n^{-2}\sum_{i\neq j}|\varepsilon^*_{ik}|\Big(\int K_{hi}(x)K_{hj}(x)d\psi(x)\Big)^2 = O_p(1),\\
W_{nk,0,0} &:= 2h^d n^{-2}\sum_{i\neq j}\Big(\int K_{hi}(x)K_{hj}(x)d\psi(x)\Big)^2 = O_p(1), \qquad k=1,2.
\end{aligned}
$$

Hence, we have

$$
\begin{aligned}
|\hat{\Gamma}_{n1}-\tilde{\Gamma}_{n1}| &\le (1+v_n)^2\{2w_{n1}^2W_{n1,2,0}+w_{n1}^4W_{n1,0,0}+4w_{n1}W_{n1,2,1}\\
&\quad +4w_{n1}^2W_{n1,1,1}+4w_{n1}^3W_{n1,1,0}\}+(2v_n+v_n^2)W_{n1,2,2}\\
&= o_p(1),
\end{aligned}
$$

$$
\begin{aligned}
|\hat{\Gamma}_{n2} - \tilde{\Gamma}_{n2}| \;\le\; & (1+v_n)^2\{(2a_n + a_n^2)W_{n1,2,2} + 2w_{n2}^2(1+a_2)^2 W_{n2,2,0} \\
& \qquad + w_{n2}^4 W_{n2,0,0} + 4w_{n2}(1+a_n)^3 W_{n2,2,1} \\
& \qquad + 4w_{n2}^2(1+a_n)^2 W_{n2,1,1} + 4w_{n2}^3(1+a_n)W_{n2,1,0}\} \\
& + (2v_n + v_n^2)W_{n1,2,2} \\
= \;& o_p(1).
\end{aligned}
$$

Therefore the second claim of (1.45) is proved, and so is Lemma 1.5.7.

1.6 Simulations

In this section two simulation studies are reported. The first investigates behavior of the empirical size and power of the test $I(|\hat{\mathcal{D}}_{n1}| > 1.96)$ with $g(x) \equiv 1$ on $[-1,1]^2$ at 4 alternatives under different designs and data missing probabilities. The second lists the mean and standard deviation of the minimum distance parameter estimator $\hat{\theta}_{n1}$. In both studies, $d = 2$, and the completed data set are constructed using imputation method. All simulations are based on 1000 replications.

In the first study, we compare the empirical size and power of the test at 4 alternatives, on 2 designs X, and 3 data missing probabilities $\Delta(X)$. More precisely, the design variables $X_i = (X_{1i}, X_{2i})^T$, $i = 1, \cdots, n$, are i.i.d bivariate normal $\mathcal{N}(0, V_k)$, $k = 1, 2$, with

$$
V_1 = \begin{pmatrix} 0.36 & 0 \\ 0 & 1 \end{pmatrix}, \qquad V_2 = \begin{pmatrix} 1 & 0.64 \\ 0.64 & 1 \end{pmatrix}. \tag{1.46}
$$

The three choices of $\Delta(x)$, $x = (x_1, x_2)^T$, are as follows:

$$\Delta_1(x) = (1 + e^{-0.8-0.5x_1-0.5x_2})^{-1}, \tag{1.47}$$

$$\Delta_2(x) = (1 + e^{-0.2-0.3x_1-0.3x_2})^{-1},$$

$$\Delta_3 \equiv 1, \quad \text{the complete data.}$$

These choices are similar to those in Sun and Wang (2009). They use the data missing probabilities $\{1+\exp(-0.3-0.3x)\}^{-1}$, $\{1+\exp(-1.0-0.8x)\}^{-1}$, and $1-0.4\exp(-5(x-0.4)^2)$ when $d = 1$. The error distribution is $\mathcal{N}(0, (0.3)^2)$. The regression function under the null hypothesis is $\mu(x) = \theta_0^T l(x)$, where $\theta_0 = (0.5, 0.8)^T$, $l(x) = x = (x_1, x_2)^T$. The regression models are as follows:

Model 0. $\delta_i Y_i = \delta_i \mu(X_i) + \delta_i \varepsilon_i,$

Model 1. $\delta_i Y_i = \delta_i \mu(X_i) + 0.5\delta_i (X_{1i} - 0.2)(X_{2i} - 0.4) + \delta_i \varepsilon_i,$

Model 2. $\delta_i Y_i = \delta_i \mu(X_i) + 0.5\delta_i (X_{1i} X_{2i} - 1) + \delta_i \varepsilon_i,$

Model 3. $\delta_i Y_i = \delta_i \mu(X_i) + 2\delta_i \{\exp(-0.4X_{1i}^2) - \exp(0.6X_{2i}^2)\} + \delta_i \varepsilon_i,$

Model 4. $\delta_i Y_i = \delta_i X_{1i} I(X_{2i} > 0.2) + \delta_i \varepsilon_i,$

The nominal level is $\alpha = 0.05$. The sample sizes considered are $n = 50, 100, 200$. The first 2 tables describe empirical sizes and powers in models 0-4. Model 0 is the null model while model 1-4 are the alternatives. These empirical levels and powers are computed by the relative frequency of the event $\{|\hat{\mathcal{D}}_{n1}| > 1.96\}$ in corresponding models. Bandwidths $h = n^{-1/4.5}$ and $w = (\log n/n)^{1/6}$ are chosen because of (h3) and (1.9). The kernels are

$K(u,v) \equiv K^1(u)K^1(v)$ and $K^* \equiv K$, with $K^1(u) := \frac{3}{4}(1-u^2)I(|u| \leq 1)$.

Table 1.1: Empirical sizes and powers for model 0 vs. models 1-4 with $X \sim \mathcal{N}(0, V_1)$ and $\varepsilon \sim \mathcal{N}(0, (.3)^2)$

n	n=50			n=100			n=200		
Δ	Δ_1	Δ_2	Δ_3	Δ_1	Δ_2	Δ_3	Δ_1	Δ_2	Δ_3
Model 0	.020	.027	.031	.029	.029	.036	.033	.034	.042
Model 1	.103	.079	.224	.278	.176	.586	.633	.513	.935
Model 2	.993	.941	1	1	.999	1	1	1	1
Model 3	.315	.203	.999	.351	.270	1	.375	.338	1
Model 4	.241	.159	.484	.671	.497	.905	.980	.920	1

Table 1.1 gives the empirical sizes and powers for testing model 0 against models 1-4 with design $X \sim \mathcal{N}(0, V_1)$, when the data are randomly missing with either of the 2 missing data probabilities or with no missing data. In the simulation, the empirical sizes of the test for model 0 keep less than 0.05. When the sample size increases, it gradually approaches the asymptotic level and becomes quite close at the sample size 200. On the other hand, the empirical powers of the test are greater than 0.05 against each alternative 1-4 for all the sample sizes we take, and become closer to 1 as the sample size increases; especially against alternative 2, the power is above 0.94 even at sample size 50. From the comparison among the 3 data missing probabilities, we observe that the level behavior is affected by the data missing probability, while the power is affected much more.

Table 1.2: Empirical sizes and powers for model 0 vs. models 1-4 with $X \sim \mathcal{N}(0, V_2)$ and $\varepsilon \sim \mathcal{N}(0, (.3)^2)$

n	n=50			n=100			n=200		
Δ	Δ_1	Δ_2	Δ_3	Δ_1	Δ_2	Δ_3	Δ_1	Δ_2	Δ_3
Model 0	.025	.027	.030	.029	.031	.036	.035	.037	.043
Model 1	.115	.103	.371	.199	.164	.677	.479	.373	.952
Model 2	.965	.831	1	.999	.991	1	1	1	1
Model 3	.237	.187	1	.272	.209	1	.274	.227	1
Model 4	.203	.144	.529	.596	.471	.927	.957	.892	1

Table 1.2 lists empirical sizes and powers with design $X \sim \mathcal{N}(0, V_2)$. In addition to

obtaining similar conclusion as the first table, we can also find that the power and the level behaviors are affected by the dependence between the design variable coordinates, although they are not affected that much. Results for model 4 in both tables show that the discontinuity of regression function has an effect on the power of the test, because the power is dramatically changed as the sample size increases.

Table 1.3: Mean and s.d. of $\hat{\theta}_{n1}$ under model 0 with $X \sim \mathcal{N}(0, V_1)$, $\varepsilon \sim \mathcal{N}(0, (.3)^2)$, and $E(\delta|X = x) = \Delta_1(x)$

n	$n = 50$	$n = 100$	$n = 200$
Mean	(.494, .804)	(.503, .800)	(.499, .800)
Std dev	(.110, .084)	(.078, .061)	(.052, .043)

The second study gives the mean and standard deviation of each component of $\hat{\theta}_{n1}$ under the null hypothesis model 0 with normal error $\varepsilon \sim \mathcal{N}(0, (0.3)^2)$ when $d = q = 2$. The variance of design and data missing probability are chosen to be V_1 in (1.46) and Δ_1 in (1.47), respectively. The regression function and parameter are the same as in the first study. Results listed in Table 1.3 show that the minimum distance estimator of the parameter is very close to the true parameter and the standard deviation is quite small.

Chapter 2

Testing for Superiority of Two Regression Curves when Responses are Missing At Random

2.1 Introduction

This chapter considers a class of tests using covariate matching for comparing the equality of two nonparametric regression curves against a one-sided alternative, when responses are missing at random. More precisely, let $(X_k,\, \delta_k Y_k)$, $k = 1, 2$, be the two groups of random variables, where X_k is a one-dimensional explanatory variable, Y_k is a one-dimensional response variable, δ_k is an indicator for whether the response is missing or observed, i.e. $\delta_k = 1$, if Y_k is observed, and $\delta_k = 0$, if Y_k is missing, $k = 1, 2$. We say Y_k is missing at random, if δ_k and Y_k are conditionally independent, given X_k, i.e. $P(\delta_k = 1|Y_k, X_k) = P(\delta_k = 1|X_k)$, a.s., $k = 1, 2$; see Little and Rubin (1987).

Now, let

$$\mu_k(x) := E(Y_k|X_k = x), \quad x \in \mathbb{R}, \quad k = 1, 2,$$

be the two regression functions so that

$$Y_k = \mu_k(X_k) + \varepsilon_k, \quad E(\varepsilon_k|X_k = x) = 0, \quad \forall\, x \in \mathbb{R}, \quad k = 1, 2.$$

Let $\mathcal{I}$ be a compact interval in $\mathbb{R}$. The problem of interest is to test the hypothesis

$$\mathrm{H}_0 : \mu_1(x) = \mu_2(x), \quad \text{for all } x \in \mathcal{I},$$

$$\mathrm{H}_1 : \mu_1(x) \geq \mu_2(x), \quad \text{for all } x \in \mathcal{I} \text{ with strict inequality for at least one } x \in \mathcal{I},$$

based on independent samples $\{(X_{k,i},\, \delta_{k,i}Y_{k,i}) : i = 1, \cdots, n_k\}$ from the distributions of $(X_k,\, \delta_k Y_k)$, $k = 1, 2$, respectively. Moreover, let ϕ be a non-negative continuous function on $\mathbb{R}$. One is interested in the asymptotic power of a given test against the local alternatives

$$\mathrm{H}_{1\mathrm{N}} : \mu_1(x) = \mu_2(x) + N^{-1/2}\phi(x), \quad N := \frac{n_1 n_2}{n_1 + n_2}, \quad \text{for all } x \in \mathcal{I}. \tag{2.1}$$

When we observe complete data, this testing problem has been addressed by many researchers. In particular, Hall et. al (1997) proposed a class of tests based on the covariate-matching, and the local averaging interpolation rule. They proved the asymptotic normality of the proposed statistics under general alternatives, allowing design and error densities to be different. They also proposed an adaptive version of their test that achieves the optimal power against a sequence of local alternatives. Koul and Schick (1997) proposed four classes of tests under the assumption of possibly distinct design but common error densities. They gave a general asymptotic optimality theory against a sequence of local alternatives. One of these classes of covariate-matched tests is shown to have desirable asymptotic power properties against several alternatives. Koul and Schick (2003) (K-S) developed this class of test further and derive their asymptotic power for the local alternatives, under the heteroscedastic setting with possibly distinct error and design densities in the two regression models. They obtained an upper bound on the asymptotic power of all tests against a given sequence of local alternatives using a semiparametric approach, and showed that a member of this class of tests achieves this upper bound.

This chapter discusses the above one-sided testing problem when responses are missing at random. We construct a complete data set by imputing kernel-type estimates for the regression functions, and investigate the asymptotic properties of the modified version for missing at random setup of the covariate-matched test statistic proposed in K-S under null hypothesis and local alternatives. The consistency of the tests based on these statistics is also discussed. To set up the analysis, let $\mathcal{U}$ be the set of all non-negative functions that are continuous on $\mathcal{I}$ and vanish off $\mathcal{I}$. Assume that X_k has a density g_k that is bounded away from zero on $\mathcal{I}$, $k = 1, 2$. Let K be a symmetric Lipschitz continuous kernel density with compact support $[-1, 1]$, $a = a_N$, $b_k = b_{k,n_k}$, $c_k = c_{k,n_k}$, and $d_k = d_{k,n_k}$, be bandwidth sequences. Let $K_h(y) := K(y/h)/h$, $y \in \mathbb{R}$, $h = a, b_k, c_k, d_k$. The estimators of regression functions and the constructed responses are, respectively,

$$\hat{\mu}_k(x) := \frac{\sum_{i=1}^{n_k} \delta_{k,i} Y_{k,i} K b_k(x - X_{k,i})}{\sum_{i=1}^{n_k} \delta_{k,i} K_{b_k}(x - X_{k,i})},$$

$$\hat{Y}_{k,i} := \delta_{k,i} Y_{k,i} + (1 - \delta_{k,i}) \hat{\mu}_k(X_{k,i}), \quad 1 \le i \le n_k, \quad k = 1, 2.$$

For each $k = 1, 2$, let $\hat{v}_k$ be a non-negative estimate of $v_k := \sqrt{u}/g_k$ which vanishes off $\mathcal{I}$. The covariate-matched statistic and the adaptive version with responses missing at random, respectively, are

$$T := \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} v_1(X_{1,i}) v_2(X_{2,j})(Y_{1,i} - Y_{2,j}) K_a(X_{1,i} - X_{2,j}),$$

and

$$\hat{T} := \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \hat{v}_1(X_{1,i}) \hat{v}_2(X_{2,j})(\hat{Y}_{1,i} - \hat{Y}_{2,j}) K_a(X_{1,i} - X_{2,j}).$$

The needed assumptions and conditions to state the main results are given in Section 2.2. Section 2.3 states the asymptotic normality of $\hat{T}$ under H_0 and $\mathrm{H}_{1\mathrm{N}}$, and the consistency of the test based on $\hat{T}$. The optimal u to maximize the asymptotic power against $\mathrm{H}_{1\mathrm{N}}$ is also discussed. Section 2.4 gives the estimates needed to construct $\hat{T}$ and the corresponding test. Simulation studies are set in Section 2.5.

2.2 Assumptions

In this section we shall state the needed assumptions. The following assumptions are similar to those in K-S. For each $k = 1, 2$,

(e1) $(X_{k,i}, \delta_{k,i} Y_{k,i}) : X_{k,i} \in \mathbb{R}, Y_{k,i} \in \mathbb{R}, \delta_{k,i} = 0$ or $1, i = 1, 2, \cdots, n_k$, are i.i.d. random vectors with $\delta_{k,i} = 1$, if $Y_{k,i}$ is observed, and $\delta_{k,i} = 0$, if $Y_{k,i}$ is missing; $\mu_k(x) = E(Y_{k,1} | X_{k,1} = x)$, $x \in \mathbb{R}$, $\varepsilon_{k,i} = Y_{k,i} - \mu_k(X_{k,i})$, $\delta_{k,i}$ and $\varepsilon_{k,i}$ are conditionally independent, given $X_{k,i}$. $\{(X_{1,i}, \delta_{1,i} Y_{1,i})\}_{i=1}^{n_1}$ and $\{(X_{2,j}, \delta_{2,j} Y_{2,j})\}_{j=1}^{n_2}$ are independent.

(e2) $E\varepsilon_{k,1}^2 < \infty$, $\sigma_k^2(x) := E(\varepsilon_{k,1}^2|X_{k,1} = x)$ and $\Delta_k(x) := E(\delta_{k,1} = 1|X_{k,1} = x)$ are continuous and positive on $\mathcal{I}$.

(e3) $\nu_k^4(x) := E(\varepsilon_{k,1}^4|X_{k,1})$, $x \in \mathbb{R}$, is bounded on an open interval containing $\mathcal{I}$.

(e4) σ_k^2 and Δ_k are twice continuously differentiable on $\mathcal{I}$.

(g1) The design variable $X_{k,1}$ has a bounded Lebesgue density g_k which is continuous and positive on $\mathcal{I}$.

(g2) The density g is twice continuously differentiable on $\mathcal{I}$.

(k) The kernel w is symmetric square integrable continuous density with compact support $[-1, 1]$. In addition, w satisfies Lipschitz-continuity of order 1.

(m) μ_1 is continuous. μ_2 is Lipschitz-continuous of order 1 with Lipschitz constant ℓ_{μ_2}.

(p) ϕ is a non-negative continuous function.

(q) ξ is a non-negative continuous function with $\xi(x) > 0$ for at least one $x \in \mathcal{I}$.

(u) $\mathcal{U}$ is the set of all non-negative functions that vanish off $\mathcal{I}$ and whose restrictions to $\mathcal{I}$ are continuous.

(w1) $a^2 N \to 0$, $aN^{\eta_1} \to \infty$, for some $\eta_1 \in (1/2, 1)$.

(w2) $b_k^2 n_k \to 0$, $b_k n_k^{\eta_2} \to \infty$, for some $\eta_2 \in (1/2, 1)$.

(w3) $c_k \to 0$, $d_k \to 0$, $(c_k + d_k)n_k^{\eta_3} \to \infty$ for some $\eta_3 \in (0, 1/2)$, $(c_k^5 + d_k^5)n_k(\log n_k)^{-1} \leq C$ for some $C < \infty$.

(z) $\{I_{k,1}, \cdots, I_{k,B_k}\}$ partitions $\mathcal{I}$ into disjoint intervals of equal length π_k, with $\pi_k \to 0$ and $n_k^{1/2}\pi_k \to \infty$.

Note that (e2) and (g1) imply that for each $k = 1, 2$, the functions g_k, σ_k^2, and Δ_k, are bounded and uniformly continuous on the compact interval $\mathcal{I}$, and bounded away from zero on $\mathcal{I}$.

Rewrite H_1 into the form:

$$\mathrm{H}_1 : \mu_1 = \mu_2 + \xi, \qquad \text{where } \xi \text{ satisfies } (q) \text{ and } \int u(x)\xi(x)dx > 0, \quad u \in \mathcal{U}. \tag{2.2}$$

To state the main results, we need the following set of additional conditions on estimators. They are motivated by Schick (1987), and proposed in K-S as Definition 2.1, Assumption 2.3, and Lemma 2.4, for the case of complete responses. These conditions are reproduced as follows, only with changes from the case of complete responses to data missing at random setup. We need these conditions not only under H_0 and $\mathrm{H}_{1\mathrm{N}}$ in (2.1), but also under H_1 in (2.2). Let

$$\begin{aligned} \underline{\mathrm{X}} &:= (X_{1,1}, \cdots, X_{1,n_1}, X_{2,1}, \cdots, X_{2,n_2}), \\ \underline{\delta} &:= (\delta_{1,1}, \cdots, \delta_{1,n_1}, \delta_{2,1}, \cdots, \delta_{2,n_2}), \\ \underline{\mathrm{Y}} &:= (Y_{1,1}, \cdots, Y_{1,n_1}, Y_{2,1}, \cdots, Y_{2,n_2}), \\ r_k(x) &= u(x)/g_k(x), \ x \in \mathcal{I}. \end{aligned} \tag{2.3}$$

and $\underline{\mathrm{Y}}_{k,j}$ be the vector obtained from $\underline{\mathrm{Y}}$ by removing $Y_{k,j}$, $j = 1, \cdots, n_k$, $k = 1, 2$.

Definition 2.2.1. The estimator $\hat{r}_k$ of r_k is said to be *consistent and cross-validated on $\mathcal{I}$ for the function r_k* (short *CCV on $\mathcal{I}$ for r_k*) if the following two conditions hold:

$$\frac{N}{n_k^2}\sum_{i=1}^{n_k}\mathbf{1}_{\mathcal{I}}(X_{k,i})E[\{\hat{r}_k(X_{k,i})-r_k(X_{k,i})\}^2|\underline{\mathrm{X}},\underline{\delta}]=o_p(1), \tag{2.4}$$

$$N\max_{1\le j\le n_k}\sup_{x\in\mathcal{I}}E[\{\hat{r}_k(x)-E[\hat{r}_k(x)|\underline{\mathrm{X}},\underline{\delta},\underline{\mathrm{Y}}_{k,j}]\}^2|\underline{\mathrm{X}},\underline{\delta}]=o_p(1). \tag{2.5}$$

We say $\tilde{r}_k$ *is a modification of* $\hat{r}_k$ if $P(\sup_{x\in\mathcal{I}}|\tilde{r}_k(x)-\hat{r}_k(x)|>0)\to 0$. We say $\hat{r}_k$ *is essentially CCV on* $\mathcal{I}$ *for* r_k if there exists a modification of $\hat{r}_k$ which is CCV on $\mathcal{I}$ for r_k.

Assumption 2.2.1. The estimate $\hat{r}_k$ is essentially CCV on $\mathcal{I}$ for r_k for $k=1,2$.

Lemma 2.2.1. *Suppose there are modifications* $\tilde{v}_k$ *of* $\hat{v}_k$ *such that, for* $k=1,2$ *and* $l=1,2$,

$$0\le\tilde{v}_k(x)\le M,\quad x\in\mathcal{I}, \tag{2.6}$$

for some finite constant M, *and such that*

$$\frac{1}{n_l}\sum_{i=1}^{n_l}E[\{\tilde{v}_k(X_{l,i})-v_k(X_{l,i})\}^2|\underline{\mathrm{X}},\underline{\delta}]=o_p(1), \tag{2.7}$$

$$N\max_{1\le i\le n_l}\sup_{x\in\mathcal{I}}E[\{\tilde{v}_k(x)-E[\tilde{v}_k(x)|\underline{\mathrm{X}},\underline{\delta},\underline{\mathrm{Y}}_{l,i}]\}^2|\underline{\mathrm{X}},\underline{\delta}]=o_p(1). \tag{2.8}$$

Then Assumption 2.2.1 holds.

The proof of Lemma 2.2.1 follows that of Lemma 2.4 in K-S, only with changes from $\underline{\mathrm{X}}$ to $(\underline{\mathrm{X}},\underline{\delta})$. Since this proof does not involve the responses $\underline{\mathrm{Y}}$ but only the designs $(\underline{\mathrm{X}},\underline{\delta})$, the above lemma holds under $\mathrm{H_0}$, $\mathrm{H_{1N}}$, and $\mathrm{H_1}$.

Remark 2.2.1. Suppose modifications $\tilde{v}_k$ of $\hat{v}_k$ exist and satisfy (2.6)-(2.8), $k=1,2$. K-S show in their proof of Lemma 2.4 that the estimators

$$\hat{r}_1(x) := \hat{v}_1(x)\frac{1}{n_2}\sum_{j=1}^{n_2}\hat{v}_2(X_{2,j})K_a(x - X_{2,j}), \quad \text{and} \tag{2.9}$$
$$\hat{r}_2(x) := \hat{v}_2(x)\frac{1}{n_1}\sum_{i=1}^{n_1}\hat{v}_1(X_{1,i})K_a(x - X_{1,i}), \quad x \in \mathbb{R},$$

are essentially CCV on $\mathcal{I}$ for r_1, and r_2, respectively, and their respective modifications can be chosen as

$$\tilde{r}_1(x) = \tilde{v}_1(x)\frac{1}{n_2}\sum_{j=1}^{n_2}\tilde{v}_2(X_{2,j})K_a(x - X_{2,j}), \tag{2.10}$$
$$\tilde{r}_2(x) = \tilde{v}_2(x)\frac{1}{n_1}\sum_{i=1}^{n_1}\tilde{v}_1(X_{1,i})K_a(x - X_{1,i}).$$

We also need the following notation and results in the proofs later. Let

$$h_k(x) := \Delta_k(x)g_k(x), \qquad \lambda_k := \inf_{x\in\mathcal{I}} h_k(x), \quad k = 1, 2. \tag{2.11}$$
$$\hat{h}_k(x) := \frac{1}{n_k}\sum_{l=1}^{n_k}\delta_{k,l}K_{b_k}(x - X_{k,l}), \qquad \hat{g}_k(x) := \frac{1}{n_k}\sum_{l=1}^{n_k}K_{b_k}(x - X_{k,l}).$$

Lemma 2.2.2. *Let $t_k = t_{n_k}$, $k = 1, 2$, be bandwidths satisfying $t_k \to 0$ and $n_k t_k^5 (\log n_k)^{-1} \leq C$ for some $C < \infty$. Assume (e2), (e4), (g1), and (g2) hold. Then the following hold.*

$$\sup_{x\in\mathcal{I}}\left|\frac{1}{n_k}\sum_{i=1}^{n_k}K_{t_k}(x - X_{k,i}) - g_k(x)\right| = o_p(1). \tag{2.12}$$
$$\sup_{x\in\mathcal{I}}\left|\frac{1}{n_k}\sum_{i=1}^{n_k}\delta_{k,i}K_{t_k}(x - X_{k,i}) - h_k(x)\right| = o_p(1). \tag{2.13}$$

$$\sup_{x\in\mathcal{I}}\Big|\frac{1}{n_k}\sum_{i=1}^{n_k}\varepsilon_{k,i}\delta_{k,i}K_{t_k}(x-X_{k,i})\Big| = o_p(1). \tag{2.14}$$

$$\sup_{x\in\mathcal{I}}\Big|\frac{1}{n_k}\sum_{i=1}^{n_k}\varepsilon_{k,i}^2\delta_{k,i}K_{t_k}(x-X_{k,i}) - E\varepsilon_{k,1}^2\delta_{k,1}K_{t_k}(x-X_{k,1})\Big| = o_p(1). \tag{2.15}$$

$$\sup_{x\in\mathcal{I}}\Big|\frac{\frac{1}{n_k}\sum_{i=1}^{n_k}\delta_{k,i}K_{t_k}(x-X_{k,i})}{\frac{1}{n_k}\sum_{i=1}^{n_k}K_{t_k}(x-X_{k,i})} - \Delta_k(x)\Big| = o_p(1). \tag{2.16}$$

This lemma is obtained from Theorem 3 of Collomb and Härdle (1986).

2.3 Asymptotic distribution of the test statistic under H_0, $\mathrm{H}_{1\mathrm{N}}$, and H_1

In this section we discuss the asymptotic distribution of $\hat{T}$ against $\mathrm{H}_{1\mathrm{N}}$ in Theorem 2.3.1. The asymptotic null distribution is included because the choice $\phi = 0$ corresponds to the null hypothesis. The asymptotic behavior of $\hat{T}$ against H_1 is given in Theorem 2.3.2, while consistency of the corresponding test against H_1 is stated in Remark 2.3.1.

K-S propose an optimal u to test H_0 against $\mathrm{H}_{1\mathrm{N}}$ when data is complete. When responses are missing at random, a similar optimal u can be derived. This result is given in Remark 2.3.1.

The following definitions are used in the theorems and remarks below.

$$q_1 := \frac{N}{n_1} = \frac{n_2}{n_1+n_2}, \qquad q_2 := \frac{N}{n_2} = \frac{n_1}{n_1+n_2}, \tag{2.17}$$

$$\psi_1(x) := \frac{\sigma_1^2(x)}{\Delta_1(x)g_1(x)}, \qquad \psi_2(x) := \frac{\sigma_2^2(x)}{\Delta_2(x)g_2(x)}, \quad x\in\mathbb{R},$$

$$\tau^2 := \int u^2(x)[q_1\psi_1(x)+q_2\psi_2(x)]dx, \qquad D := \int u(x)(\mu_1(x)-\mu_2(x))dx.$$

Theorem 2.3.1. *Assume that (e1), (e2), (e4), (g1), (g2), (k), (m), (p), (u), (w1), (w2), and Assumption 2.2.1 hold. Then under* $\mathrm{H_{1N}}$ *of (2.1),*

$$N^{1/2}\Big(\hat{T} - D - \frac{1}{n_1}\sum_{i=1}^{n_1} \frac{u(X_{1,i})}{\Delta_1(X_{1,i})g_1(X_{1,i})}\delta_{1,i}\varepsilon_{1,i} - \frac{1}{n_2}\sum_{j=1}^{n_2} \frac{u(X_{2,j})}{\Delta_2(X_{2,j})g_2(X_{2,j})}\delta_{2,j}\varepsilon_{2,j}\Big) = o_p(1),$$

as both sample sizes n_1 *and* n_2 *tend to infinity. Consequently, under* $\mathrm{H_{1N}}$,

$$N^{1/2}(\hat{T} - D) \to_d \mathcal{N}(0, \tau^2), \qquad \text{as } n_1 \wedge n_2 \to \infty.$$

Proof. Recall r from (2.3), $\hat{r}_k$ from (2.9), $\hat{g}_k$ and $\hat{h}_k$ from (2.11), $k = 1, 2$. For $x \in \mathbb{R}$, $k, m = 1, 2$, let

$$\bar{\mu}_{k,m}(x) := \frac{1}{n_k}\sum_{l=1}^{n_k} \mu_m(X_{k,l})\delta_{k,l}K_{b_k}(x - X_{k,l})/\hat{h}_k(x),$$

$$\bar{\varepsilon}_k(x) := \frac{1}{n_k}\sum_{l=1}^{n_k} \varepsilon_{k,l}\delta_{k,l}K_{b_k}(x - X_{k,l})/\hat{h}_k(x),$$

$$\bar{\phi}_k(x) := \frac{1}{n_k}\sum_{l=1}^{n_k} \phi(X_{k,l})\delta_{k,l}K_{b_k}(x - X_{k,l})/\hat{h}_k(x).$$

Suppose $\mathrm{H_{1N}}$ holds. With definitions above, write $\hat{T} = A_1 + B_1 - B_2 + C_1 - C_2 + R_1 + R_2$, where

$$A_1 := \frac{1}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \hat{v}_1(X_{1,i})\hat{v}_2(X_{2,j})\Big(\mu_2(X_{1,i}) - \mu_2(X_{2,j})\Big)K_a(X_{1,i} - X_{2,j}),$$

$$B_1 := \frac{1}{n_1}\sum_{i=1}^{n_1} \hat{r}_1(X_{1,i})(1 - \delta_{1,i})\Big(\bar{\mu}_{1,2}(X_{1,i}) - \mu_2(X_{1,i})\Big),$$

$$B_2 := \frac{1}{n_2}\sum_{j=1}^{n_2}\hat{r}_2(X_{2,j})(1-\delta_{2,j})\Big(\bar{\mu}_{2,2}(X_{2,j})-\mu_2(X_{2,j})\Big),$$

$$C_1 := \frac{1}{n_1}\sum_{i=1}^{n_1}\hat{r}_1(X_{1,i})\Big(\delta_{1,i}\varepsilon_{1,i}+(1-\delta_{1,i})\bar{\varepsilon}_1(X_{1,i})\Big),$$

$$C_2 := \frac{1}{n_2}\sum_{j=1}^{n_2}\hat{r}_2(X_{2,j})\Big(\delta_{2,j}\varepsilon_{2,j}+(1-\delta_{2,j})\bar{\varepsilon}_2(X_{2,j})\Big),$$

$$R_1 := \frac{N^{-1/2}}{n_1}\sum_{i=1}^{n_1}\hat{r}_1(X_{1,i})\phi(X_{1,i}),$$

$$R_2 := \frac{N^{-1/2}}{n_1}\sum_{i=1}^{n_1}\hat{r}_1(X_{1,i})(1-\delta_{1,i})\Big(\bar{\phi}_1(X_{1,i})-\phi(X_{1,i})\Big).$$

In the following, we shall show that

$$N^{1/2}A_1 = o_p(1); \qquad N^{1/2}B_k = o_p(1), \quad k=1,2;$$

$$N^{1/2}R_1 = n^{1/2}D + o_p(1); \tag{2.18}$$

$$N^{1/2}C_k = \frac{N^{1/2}}{n_k}\sum_{i=1}^{n_k}\frac{r_k(X_{k,i})}{\Delta_k(X_{k,i})}\delta_{k,i}\varepsilon_{k,i} + o_p(1), \quad k=1,2; \tag{2.19}$$

$$N^{1/2}R_2 = o_p(1). \tag{2.20}$$

Among them, (2.18) is derived by similar proof as that of Theorem 2.6 in K-S, while some details proof of (2.19) are also inspired by those of Theorem 2.6 in K-S. Recall the Lipschitz constant ℓ_{μ_2} of μ_2 in condition (m). By (g1), (m), (u), (w1), Assumption 2.2.1, and routine calculation, one has

$$N^{1/2}|A_1| \le N^{1/2}\ell_{\mu_2}a\frac{1}{n_1}\sum_{i=1}^{n_1}\hat{r}_1(X_{1,i}) = o_p(1).$$

From (g1), (m), (u), (w2), Assumption 2.2.1, and the fact

$$
\begin{aligned}
\Big|\bar{\mu}_{k,2}(X_{k,i}) - \mu_2(X_{k,i})\Big| &\leq \frac{\frac{1}{n_k}\sum_{l=1}^{n_k} |\mu_2(X_{k,l}) - \mu_2(X_{k,i})|\delta_{k,l}K_{b_k}(X_{k,i} - X_{k,l})}{\hat{h}_k(X_{k,i})} \\
&\leq \ell_{\mu_2} b_k, \quad k = 1, 2,
\end{aligned}
$$

one obtains $N^{1/2}|B_k| \leq N^{1/2}\ell_{\mu_2} b_k \frac{1}{n_k}\sum_{i=1}^{n_k} \hat{r}_k(X_{k,i}) = o_p(1)$, $k = 1, 2$. For each $k = 1, 2$, note that

$$
C_k = \frac{1}{n_k}\sum_{i=1}^{n_k} \varepsilon_{k,i}\Big\{\frac{1}{n_k}\sum_{l=1}^{n_k} \frac{\hat{r}_k(X_{k,l})(1-\delta_{k,l})\delta_{k,i}K_{b_k}(X_{k,i} - X_{k,l})}{\hat{h}_k(X_{k,l})} + \delta_{k,i}\hat{r}_k(X_{k,i})\Big\}.
$$

Write $C_k = C_{k,1} + C_{k,2} + C_{k,3} + C_{k,4}$, where

$$
\begin{aligned}
C_{k,1} &:= \frac{1}{n_k}\sum_{i=1}^{n_k} \varepsilon_{k,i}\Big\{\frac{1}{n_k}\sum_{l=1}^{n_k} \frac{r_k(X_{k,l})(1-\delta_{k,l})\delta_{k,i}K_{b_k}(X_{k,i} - X_{k,l})}{h_k(X_{k,l})} \\
&\qquad - \frac{\delta_{k,i}(1 - \Delta_k(X_{k,i}))r_k(X_{k,i})}{\Delta_k(X_{k,i})}\Big\}, \\
C_{k,2} &:= \frac{1}{n_k}\sum_{i=1}^{n_k} \varepsilon_{k,i}\Big\{\frac{1}{n_k}\sum_{l=1}^{n_k} \Big(\frac{\hat{r}_k(X_{k,l})}{\hat{h}_k(X_{k,l})} - \frac{r_k(X_{k,l})}{h_k(X_{k,l})}\Big) \\
&\qquad \times (1-\delta_{k,l})\delta_{k,i}K_{b_k}(X_{k,i} - X_{k,l})\Big\}, \\
C_{k,3} &:= \frac{1}{n_k}\sum_{i=1}^{n_k} \varepsilon_{k,i}\delta_{k,i}(\hat{r}_k(X_{k,i}) - r_k(X_{k,i})), \\
C_{k,4} &:= \frac{1}{n_k}\sum_{i=1}^{n_k} \frac{\varepsilon_{k,i}\delta_{k,i}r_k(X_{k,i})}{\Delta_k(X_{k,i})}.
\end{aligned}
$$

For $i, l = 1, \cdots, n_k$, $k = 1, 2$, let

$$
I_{k,i,l} := \frac{r_k(X_{k,l})(1-\delta_{k,l})\delta_{k,i}K_{b_k}(X_{k,i} - X_{k,l})}{h_k(X_{k,l})},
$$

$$J_{k,i} := \frac{\delta_{k,i}(1-\Delta_k(X_{k,i}))r_k(X_{k,i})}{\Delta_k(X_{k,i})}.$$

By (e1), (e2), (g1), (u), (w2), and routine calculation, one has $E(N^{1/2}C_{k,1}) = 0$ and

$$\begin{aligned}
&Var(N^{1/2}C_{k,1}) \\
&= \frac{N}{n_k} E\Big[\varepsilon_{k,1}^2 \Big\{ \frac{1}{n_k}\sum_{l=1}^{n_k} I_{k,1,l} - J_{k,1} \Big\}^2\Big] \\
&= \frac{N}{n_k} E\Big[\varepsilon_{k,1}^2 \Big\{ \frac{n_k-1}{n_k^2} I_{k,1,2}^2 + \frac{(n_k-1)(n_k-2)}{n_k^2} I_{k,1,2}I_{k,1,3} \\
&\qquad - \frac{2(n_k-1)}{n_k} J_{k,1}I_{k,1,2} + J_{k,1}^2 \Big\}\Big] \\
&= \frac{N}{n_k}\Big[\frac{n_k-1}{n_k^2 b_k} \int\int \Big(\sigma_k^2(x)\Delta_k(x)g_k(x)\Big)\Big(r_k^2(x+bu)(1-\Delta_k(x+bu)) \\
&\qquad \times(\Delta_k(x+bu))^{-2}(g_k(x+bu))^{-1}\Big)K^2(u)dudx \\
&\quad + \frac{(n_k-1)(n_k-2)}{n_k^2}\int\int\int \Big(\sigma_k^2(x)\Delta_k(x)g_k(x)\Big)\Big(r_k(x+bu) \\
&\qquad \times(1-\Delta_k(x+bu))(\Delta_k(x+bu))^{-1}\Big)\Big(r_k(x+bv) \\
&\qquad \times(1-\Delta_k(x+bv))(\Delta_k(x+bv))^{-1}\Big)K(u)K(v)dudvdx \\
&\quad - \frac{2(n_k-1)}{n_k}\int\int \Big(\sigma_k^2(x)r_k(x)(1-\Delta_k(x))g_k(x)\Big)\Big(r_k(x+bu) \\
&\qquad \times(1-\Delta_k(x+bu))(\Delta_k(x+bu))^{-1}\Big)K(u)dudx \\
&\quad + \int \sigma_k^2(x)r_k^2(x)(1-\Delta_k(x))^2(\Delta_k(x))^{-1}g_k(x)dx\Big] \\
&\to 0, \qquad k=1,2.
\end{aligned}$$

Hence $N^{1/2}C_{k,1} = o_p(1)$, $k = 1, 2$. Recall the modification $\tilde{r}_k$ of $\hat{r}_k$ defined in (2.10) which is CCV on $\mathcal{I}$ for r_k. Let for $i, j, m = 1, \cdots, n_k$, $k = 1, 2$,

$$\tilde{r}_{k,i}(x) := E[\tilde{r}_k(x)|\underline{\mathrm{X}}, \underline{\delta}, \underline{\mathrm{Y}}_{k,i}], \qquad \tilde{r}_{k,i,j}(x) := E[\tilde{r}_{k,i}(x)|\underline{\mathrm{X}}, \underline{\delta}, \underline{\mathrm{Y}}_{k,j}],$$

$$\hat{M}_{k,i} := \frac{1}{n_k}\sum_{l=1}^{n_k}\Big(\frac{\hat{r}_k(X_{k,l})}{\hat{h}_k(X_{k,l})} - \frac{r_k(X_{k,l})}{h_k(X_{k,l})}\Big)(1-\delta_{k,l})\delta_{k,i}K_{b_k}(X_{k,i}-X_{k,l}),$$
$$\tilde{M}_{k,i} := \frac{1}{n_k}\sum_{l=1}^{n_k}\Big(\frac{\tilde{r}_k(X_{k,l})}{\hat{h}_k(X_{k,l})} - \frac{r_k(X_{k,l})}{h_k(X_{k,l})}\Big)(1-\delta_{k,l})\delta_{k,i}K_{b_k}(X_{k,i}-X_{k,l}),$$
$$\tilde{M}_{k,i;j} := E[\tilde{M}_{k,i}|\underline{\mathrm{X}},\underline{\delta},\underline{\mathrm{Y}}_{k,j}], \qquad \tilde{M}_{k,i;j,m} := E[\tilde{M}_{k,i;j}|\underline{\mathrm{X}},\underline{\delta},\underline{\mathrm{Y}}_{k,m}].$$

Then we have

$$\begin{aligned} C_{k,2} &= \frac{1}{n_k}\sum_{i=1}^{n_k}\varepsilon_{k,i}\tilde{M}_{k,i;i} + \frac{1}{n_k}\sum_{i=1}^{n_k}\varepsilon_{k,i}(\tilde{M}_{k,i}-\tilde{M}_{k,i;i}) + \frac{1}{n_k}\sum_{i=1}^{n_k}\varepsilon_{k,i}(\hat{M}_{k,i}-\tilde{M}_{k,i}) \\ &= C_{k,2,1}+C_{k,2,2}+C_{k,2,3}, \quad \text{say.} \end{aligned}$$

For each $k=1,2$, let $Q_{k,l;i,j} := E[(\tilde{r}_{k,i}(X_{k,l}) - \tilde{r}_{k,i,j}(X_{k,l}))^2|\underline{\mathrm{X}},\underline{\delta}]$, $l,i,j=1,\cdots,n_k$. By C-S inequality, one has

$$\begin{aligned} S_{k,1}: &= \frac{N}{n_k^2}\sum_{i=1}^{n_k}\sum_{j=1}^{n_k}E[(\tilde{M}_{k(i),i}-\tilde{M}_{k(i),i,j})^2|\underline{\mathrm{X}},\underline{\delta}] \\ &= \frac{N}{n_k^2}\sum_{i=1}^{n_k}\sum_{j=1}^{n_k}E\Big[\Big\{\frac{1}{n_k}\sum_{l=1}^{n_k}\Big(\frac{\tilde{r}_{k,i}(X_{k,l})}{\hat{h}_k(X_{k,l})} - \frac{\tilde{r}_{k,i,j}(X_{k,l})}{\hat{h}_k(X_{k,l})}\Big) \\ &\qquad \times(1-\delta_{k,l})\delta_{k,i}K_{b_k}(X_{k,i}-X_{k,l})\Big\}^2\Big|\underline{\mathrm{X}},\underline{\delta}\Big] \\ &\le \frac{N}{n_k^2}\sum_{i=1}^{n_k}\sum_{j=1}^{n_k}E\Big[\frac{1}{n_k}\sum_{l=1}^{n_k}\Big(\frac{\tilde{r}_{k,i}(X_{k,l})}{\hat{h}_k(X_{k,l})} - \frac{\tilde{r}_{k,i,j}(X_{k,l})}{\hat{h}_k(X_{k,l})}\Big)^2 K_{b_k}(X_{k,i}-X_{k,l})\Big|\underline{\mathrm{X}},\underline{\delta}\Big] \\ &\qquad \times\Big\{\frac{1}{n_k}\sum_{l=1}^{n_k}(1-\delta_{k,l})\delta_{k,i}K_{b_k}(X_{k,i}-X_{k,l})\Big\} \\ &\le \sup_{x\in\mathcal{I}}\hat{g}_k(x)\frac{N}{n_k^2}\sum_{i=1}^{n_k}\sum_{j=1}^{n_k}\Big\{\frac{1}{n_k}\sum_{l=1}^{n_k}E\Big[\Big(\frac{\tilde{r}_{k,i}(X_{k,l})}{\hat{h}_k(X_{k,l})} - \frac{\tilde{r}_{k,i,j}(X_{k,l})}{\hat{h}_k(X_{k,l})}\Big)^2\Big|\underline{\mathrm{X}},\underline{\delta}\Big] \\ &\qquad \times K_{b_k}(X_{k,i}-X_{k,l})\Big\} \end{aligned}$$

$$
\begin{aligned}
&= \sup_{x\in\mathcal{I}} \hat{g}_k(x)\frac{N}{n_k^3}\sum_{l=1}^{n_k}\sum_{i=1}^{n_k}\sum_{j=1}^{n_k}\frac{Q_{k,l;i,j}}{\hat{h}_k^2(X_{k,l})}K_{b_k}(X_{k,i}-X_{k,l}) \\
&= \sup_{x\in\mathcal{I}} \hat{g}_k(x) I[\bigcap_{m=1}^{n_k}\{\hat{h}_k(X_{k,m}) \geq \frac{\lambda_k}{2}\}] \\
&\qquad \times \frac{N}{n_k^3}\sum_{l=1}^{n_k}\sum_{i=1}^{n_k}\sum_{j=1}^{n_k}\frac{Q_{k,l;i,j}}{\hat{h}_k^2(X_{k,l})}K_{b_k}(X_{k,i}-X_{k,l}) \\
&\quad + \sup_{x\in\mathcal{I}} \hat{g}_k(x) I[\bigcup_{m=1}^{n_k}\{\hat{h}_k(X_{k,m}) < \frac{\lambda_k}{2}\}] \\
&\qquad \times \frac{N}{n_k^3}\sum_{l=1}^{n_k}\sum_{i=1}^{n_k}\sum_{j=1}^{n_k}\frac{Q_{k,l;i,j}}{\hat{h}_k^2(X_{k,l})}K_{b_k}(X_{k,i}-X_{k,l}) \\
&= S_{k,1,1} + S_{k,1,2}, \quad \text{say.}
\end{aligned}
$$

By Assumption 2.2.1, (2.5), (e1), and C-S inequality, one obtains

$$
\begin{aligned}
&\sup_{1\leq i,l\leq n_k} \frac{N}{n_k}\sum_{j=1}^{n_k} Q_{k,l;i,j} \\
&= \sup_{1\leq i,l\leq n_k} \frac{N}{n_k}\sum_{j=1}^{n_k} E[(\tilde{r}_{k,i}(X_{k,l}) - \tilde{r}_{k,j,i}(X_{k,l}))^2 | \underline{\mathrm{X}}, \underline{\delta}] \\
&\leq \sup_{1\leq l\leq n_k} \frac{N}{n_k^2}\sum_{i=1}^{n_k}\sum_{j=1}^{n_k} E[E(\{\tilde{r}_k(X_{k,l}) - \tilde{r}_{k,j}(X_{k,l})\}^2 | \underline{\mathrm{X}}, \underline{\delta}, \underline{\mathrm{Y}}_{k,i}) | \underline{\mathrm{X}}, \underline{\delta}] \\
&\leq N \max_{1\leq j\leq n_k} \sup_{x\in\mathcal{I}} E[\{\tilde{r}_k(x) - \tilde{r}_{k,j}(x)\}^2 | \underline{\mathrm{X}}, \underline{\delta}] = o_p(1). \qquad (2.21)
\end{aligned}
$$

(2.12) in Remark 2.2.2 shows $\sup_{x\in\mathcal{I}} \hat{g}_k(x) = o_p(1)$. Together with (2.11) and (2.21), we have

$$
\begin{aligned}
S_{k,1,1} &\leq \frac{1}{n_k^2}\sum_{l=1}^{n_k}\sum_{i=1}^{n_k} K_{b_k}(X_{k,i}-X_{k,l})\Big\{\sup_{1\leq i,l\leq n_k}\frac{N}{n_k}\sum_{j=1}^{n_k} Q_{k,l;i,j}\Big\}\Big/(\lambda_k/2)^2 \\
&\qquad \times \sup_{x\in\mathcal{I}} \hat{g}_k(x) I[\bigcap_{m=1}^{n_k}\{\hat{h}_k(X_{k,m}) \geq \lambda_k/2\}]
\end{aligned}
$$

$$\leq \Big(\sup_{x\in\mathcal{I}} \hat{g}_k(x)\Big)^2 \Big\{ \sup_{1\leq i,l\leq n_k} \frac{N}{n_k}\sum_{j=1}^{n_k} Q_{k,l;i,j}\Big\}\Big/(\lambda_k/2)^2 = o_p(1).$$

(2.13) in Remark 2.2.2 leads to the result

$$\begin{aligned} P(\bigcup_{i=1}^{n_k}\{\hat{h}_k(X_{k,i}) < \lambda/2\}) &\leq P(\max_{1\leq i\leq n_k} |\hat{h}_k(X_{k,i}) - h_k(X_{k,i})| > \lambda/2) \\ &\leq P(\sup_{x\in\mathcal{I}} |\hat{h}_k(x) - h_k(x)| > \lambda/2) \to 0. \end{aligned}$$

Together with the fact that $\bigcup_{i=1}^{n_k}\{\hat{h}_k(X_{k,i}) < \lambda/2\} \in \sigma(\underline{\mathrm{X}}, \underline{\delta})$, one has $S_{k,1,2} = o_p(1)$. Thus, we have $S_{k,1} = o_p(1)$, $k = 1, 2$. Let

$$D_i := \varepsilon_{1,i}\tilde{M}_{1,i;i}, \qquad D_{i,j} := E[D_i|\underline{\mathrm{X}}, \underline{\delta}, \underline{\mathrm{Y}}_{1,j}], \qquad i, j = 1, \cdots, n_1.$$

Note that by (e1), $D_{i,i} = 0$, and $E[D_iD_j|\underline{\mathrm{X}}, \underline{\delta}] = E[(D_i - D_{i,j})(D_j - D_{j,i})|\underline{\mathrm{X}}, \underline{\delta}]$. From (e2), one has

$$\begin{aligned} E[(N^{1/2}C_{k,2,1})^2|\underline{\mathrm{X}}, \underline{\delta}] &= \frac{N}{n_k^2}\sum_{i=1}^{n_k}\sum_{j=1}^{n_k} E[(D_i - D_{i,j})(D_j - D_{j,i})|\underline{\mathrm{X}}, \underline{\delta}] \\ &\leq \frac{N}{n_k^2}\sum_{i=1}^{n_k}\sum_{j=1}^{n_k} E[(D_i - D_{i,j})^2|\underline{\mathrm{X}}, \underline{\delta}] \\ &= \frac{N}{n_k^2}\sum_{i=1}^{n_k}\sum_{j=1}^{n_k} E[\varepsilon_{1,i}^2(\tilde{M}_{1,i;i} - \tilde{M}_{1,i;i,j})^2|\underline{\mathrm{X}}, \underline{\delta}] \\ &\leq S_{k,1}\sup_{x\in\mathcal{I}} \sigma_k^2(x) = o_p(1), \quad k = 1, 2. \end{aligned}$$

Thus $N^{1/2}C_{k,2,1} = o_p(1)$, $k = 1, 2$. By similar routine in proving $S_{k,1} = o_p(1)$, one has $S_{k,2} := \frac{N}{n_k}\sum_{i=1}^{n_k} E[(\tilde{M}_{k,i} - \tilde{M}_{k,i;i})^2|\underline{\mathrm{X}}, \underline{\delta}] = o_p(1)$, $k = 1, 2$. This together with (e2) and C-S inequality leads to the following result:

$$(N^{1/2}C_{k,2,2})^2 \le \Big(\frac{N}{n_k}\sum_{i=1}^{n_k}(\tilde{M}_{k,i} - \tilde{M}_{k,i;i})^2\Big)\Big(\frac{1}{n_k}\sum_{j=1}^{n_k}\varepsilon_{k,j}^2\Big) = o_p(1), \quad k = 1, 2.$$

Because of $P(|N^{1/2}C_{k,2,3}| > 0) \le P(\sup_{x\in\mathcal{I}} |\tilde{r}_k(x) - \hat{r}_k(x)| > 0) \to 0$, we have $N^{1/2}C_{k,2,3} = o_p(1)$, $k = 1, 2$. Therefore one obtains $N^{1/2}C_{k,2} = o_p(1)$, $k = 1, 2$. By similar proof as that of Theorem 2.6 in K-S, $N^{1/2}C_{k,3} = o_p(1)$ can be derived. Then one has

$$N^{1/2}C_k = \frac{N^{1/2}}{n_k}\sum_{i=1}^{n_k}\frac{r_k(X_{k,i})}{\Delta_k(X_{k,i})}\delta_{k,i}\varepsilon_{k,i} + o_p(1), \quad k = 1, 2.$$

Furthermore, by Assumption 2.2.1, (2.4), (p), C-S inequality, and Law of Large Numbers, one obtains

$$\begin{aligned} N^{1/2}R_1 &= \frac{1}{n_1}\sum_{i=1}^{n_1} r_1(X_{1,i})\phi(X_{1,i}) + \frac{1}{n_1}\sum_{i=1}^{n_1}(\tilde{r}_1(X_{1,i}) - r_1(X_{1,i}))\phi(X_{1,i}) \\ &\quad + \frac{1}{n_1}\sum_{i=1}^{n_1}(\hat{r}_1(X_{1,i}) - \tilde{r}_1(X_{1,i}))\phi(X_{1,i}) = N^{1/2}D + o_p(1), \\ N^{1/2}R_2 &\le \Big(\frac{1}{n_1}\sum_{i=1}^{n_1}\hat{r}_1^2(X_{1,i})\Big)^{1/2}\Big(\frac{1}{n_1}\sum_{i=1}^{n_1}(\bar{\phi}_1(X_{1,i}) - \phi(X_{1,i}))^2\Big)^{1/2}. \end{aligned}$$

Together by the result as follows:

$$\begin{aligned} &\frac{1}{n_1}\sum_{i=1}^{n_1}(\bar{\phi}_1(X_{1,i}) - \phi(X_{1,i}))^2 \\ &= \frac{1}{n_1}\sum_{i=1}^{n_1}\Big(\frac{1}{n_1}\sum_{l=1}^{n_1}(\phi(X_{1,l}) - \phi(X_{1,i}))\delta_{1,l}K_{b_1}(X_{1,i} - X_{1,l})/\hat{h}_1(X_{1,i})\Big)^2 \\ &\le \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{n_1}\sum_{l=1}^{n_1}(\phi(X_{1,l}) - \phi(X_{1,i}))^2\delta_{1,l}K_{b_1}(X_{1,i} - X_{1,l})/\hat{h}_1(X_{1,i}) \end{aligned}$$

$$\begin{aligned}
&= \Big(\frac{1}{n_1^2}\sum_{i=1}^{n_1}\sum_{l=1}^{n_1}(\phi(X_{1,l})-\phi(X_{1,i}))^2\delta_{1,l}K_{b_1}(X_{1,i}-X_{1,l})/\hat{h}_1(X_{1,i})\Big) \\
&\quad\times\Big(I[\bigcap_{m=1}^{n_1}\{\hat{h}_1(X_{1,m})\ge\lambda_1/2\}]+I[\bigcup_{m=1}^{n_1}\{\hat{h}_1(X_{1,m})<\lambda_1/2\}]\Big) \\
&= o_p(1)+o_p(1) \ = o_p(1),
\end{aligned}$$

we have $N^{1/2}R_2 = o_p(1)$. Therefore, one obtains

$$N^{1/2}\hat{T} = N^{1/2}\Big(D + \frac{1}{n_1}\sum_{i=1}^{n_1}\frac{r_1(X_{1,i})}{\Delta_1(X_{1,i})}\delta_{1,i}\varepsilon_{1,i} - \frac{1}{n_2}\sum_{j=1}^{n_2}\frac{r_2(X_{2,j})}{\Delta_2(X_{2,j})}\delta_{2,j}\varepsilon_{2,j}\Big) + o_p(1).$$

Thus the proof is complete.

Theorem 2.3.2. *Suppose (e1), (e2), (e4), (g1), (g2), (k), (m), (p), (u), (w1), (w2), and Assumption 2.2.1 hold. Then under* H_1 *in (2.2), one has* $N^{1/2}\hat{T} \to_p \infty$.

The proof of Theorem 2.3.2 is similar to that of Theorem 2.3.1, only with difference that $N^{1/2}(R_1+R_2) \to_p \infty$ under H_1.

Remark 2.3.1. Let $\gamma := \int u(x)\phi(x)dx/\tau$. Assume that under H_0, H_1, and $\mathrm{H}_{1\mathrm{N}}$, the assumptions of Theorem 2.3.1 hold, and there exists an estimate $\hat{\tau}^2$ of τ^2 which satisfies $\hat{\tau}^2 = \tau^2 + o_p(1)$. Then, one has

$$\begin{aligned}
&N^{1/2}\hat{T}/\hat{\tau} \to_d \mathcal{N}(0,1), && \text{under } \mathrm{H}_0, \\
&N^{1/2}\hat{T}/\hat{\tau} \to_d \mathcal{N}(\gamma,1), && \text{under } \mathrm{H}_{1\mathrm{N}}, \\
&N^{1/2}\hat{T}/\hat{\tau} \to_p \infty, && \text{under } \mathrm{H}_1.
\end{aligned}$$

Consequently, the asymptotic level of the test

$$\hat{V} = I\{N^{1/2}\hat{T}/\hat{\tau} \geq z_\alpha\}, \tag{2.22}$$

is α. The asymptotic power of this test under $\mathrm{H}_{1\mathrm{N}}$ is $1 - \Phi(z_\alpha - \gamma)$. An application of the Cauchy-Schwarz (C-S) inequality shows that γ and the asymptotic power are maximized by the choice

$$u = u_\phi := \frac{\phi \mathbf{1}_{\mathcal{I}}}{q_1\psi_1 + q_2\psi_2}.$$

The maximal asymptotic power is $1 - \Phi(z_\alpha - \gamma_\phi)$, where

$$\gamma_\phi := \Big(\int \frac{\phi^2(x)\mathbf{1}_{\mathcal{I}}(x)}{q_1\psi_1(x) + q_2\psi_2(x)} dx \Big)^{1/2}$$

is the maximal value of γ. This result is similar to that of the complete responses data discussed in Remark 2.8 and Remark 2.9 of K-S. The only difference in the missing data at random structure is reflected in having $\Delta_k(x)$ appear in the denominator of ψ_k, $k = 1, 2$. The result is exactly the same as that of complete responses data when $\Delta_k \equiv 1$, $k = 1, 2$.

2.4 Some suggested estimators

In this section we shall consider estimates of v_k and τ^2. K-S give these estimates for a given u and the (unknown) optimal u when responses are complete, and discuss their properties. When responses are missing at random, similar estimates and properties are still valid. They are listed as follows for the sake of completeness.

The following discussion gives an estimate of v_k, $k = 1, 2$. Recall $\hat{g}_k$ and $\hat{h}_k$ from (2.11). When u is known, consider

$$\hat{v}_k := \sqrt{u}/\hat{g}_k, \qquad \tilde{v}_k := \sqrt{u}/(\hat{g}_k \vee \eta), \tag{2.23}$$

where η is a positive number satisfying $g_k(x) > 4\eta$ for all $x \in \mathcal{I}$, $\tilde{v}_k$ is a modification of $\hat{v}_k$ which satisfies assumption of Lemma 2.2.1. This implies that Assumption 2.2.1 holds.

When $u = u_\phi$ with a known non-negative continuous function ϕ, let c_k and d_k be bandwidths satisfy (w3), and consider

$$\begin{aligned} \hat{\mu}_{k,c}(x) &:= \frac{\sum_{j=1}^{n_k} Y_{k,j}\delta_{k,j}K_{c_k}(x - X_{k,j})}{\sum_{j=1}^{n_k} \delta_{k,j}K_{c_k}(x - X_{k,j})}, \\ \hat{\sigma}_k^2(x) &:= \frac{\sum_{j=1}^{n_k}(Y_{k,j} - \hat{\mu}_{k,c}(X_{k,j}))^2\delta_{k,j}K_{d_k}(x - X_{k,j})}{\sum_{j=1}^{n_k} \delta_{k,j}K_{d_k}(x - X_{k,j})}, \quad x \in \mathbb{R}, \\ \hat{\psi}_k &:= \frac{\hat{\sigma}_k^2}{\hat{h}_k}, \quad \hat{\mu}_\phi := \frac{\phi \mathbf{1}_\mathcal{I}}{q_1\hat{\psi}_1 + q_2\hat{\psi}_2}, \quad \hat{v}_k := \frac{\sqrt{\hat{\mu}_\phi}}{\hat{g}_k}, \quad k = 1, 2. \end{aligned} \tag{2.24}$$

Arguing as in the estimation of v_k when $u = u_\phi$ in section 3 of K-S, we can find a modification $\tilde{v}_k$ of $\hat{v}_k$, which satisfies the assumptions of Lemma 2.2.1, such that Assumption 2.2.1 holds. The following lemma gives the needed properties of $\hat{\sigma}_k^2$.

Lemma 2.4.1. *Suppose (e1), (e2), (e3), (e4), (g1), (g2), (k), (m), (p), (u), (v), and (w3) hold. Then for each $k = 1, 2$,*

$$\sup_{x \in \mathcal{I}} |\hat{\sigma}_k^2(x) - \sigma_k^2(x)| = o_p(1), \quad \textit{under } \mathrm{H}_0, \mathrm{H}_1, \textit{ and } \mathrm{H}_{1\mathrm{N}}, \tag{2.25}$$

and $\hat{\sigma}_k^2$ is essentially CCV on $\mathcal{I}$ for σ_k^2, under H_0, H_1, *and* $\mathrm{H}_{1\mathrm{N}}$.

Proof. First we give the proof of (2.25) under $\mathrm{H}_{1\mathrm{N}}$. The case $\phi = 0$ corresponds to the result under H_0. For $k = 1, 2$ and $x \in \mathcal{I}$, define

$$\hat{h}_{k,c}(x) := \frac{1}{n_k}\sum_{l=1}^{n_k} \delta_{k,l} K_{c_k}(x - X_{k,l}),$$
$$\bar{\mu}_{k,c}(x) := \frac{\frac{1}{n_k}\sum_{l=1}^{n_k} \mu_k(X_{k,l})\delta_{k,l}K_{c_k}(x - X_{k,l})}{\hat{h}_{k,c}(x)},$$
$$\bar{\varepsilon}_{k,c}(x) := \frac{\frac{1}{n_k}\sum_{l=1}^{n_k} \varepsilon_{k,l}\delta_{k,l}K_{c_k}(x - X_{k,l})}{\hat{h}_{k,c}(x)},$$

while $\hat{h}_{k,d}(x)$, $\bar{\mu}_{k,d}(x)$, and $\bar{\varepsilon}_{k,d}(x)$ can be defined similarly when the bandwidth d_k is used instead of c_k. One can write $\hat{\sigma}_k^2(x) - \sigma_k^2(x)$ into the sum of the following terms:

$$
\begin{aligned}
Z_{k,1}(x) &= \frac{\frac{1}{n_k}\sum_{j=1}^{n_k}(\mu_k(X_{k,j}) - \bar{\mu}_{k,c}(X_{k,j}))^2\delta_{k,l}K_{d_k}(x - X_{k,j})}{\hat{h}_{k,d}(x)},\\
Z_{k,2}(x) &= \frac{\frac{1}{n_k}\sum_{j=1}^{n_k}\bar{\varepsilon}_{k,c}^2(X_{k,j})\delta_{k,j}K_{d_k}(x - X_{k,j})}{\hat{h}_{k,d}(x)}\\
&\quad - \frac{\frac{2}{n_k}\sum_{j=1}^{n_k}\varepsilon_{k,j}\bar{\varepsilon}_{k,c}(X_{k,j})\delta_{k,j}K_{d_k}(x - X_{k,j})}{\hat{h}_{k,d}(x)}\\
&\quad + \Big(\frac{\frac{1}{n_k}\sum_{j=1}^{n_k}\varepsilon_{k,j}^2\delta_{k,j}K_{d_k}(x - X_{k,j})}{\hat{h}_{k,d}(x)} - \sigma_k^2(x)\Big)\\
&= Z_{k,2,1} - Z_{k,2,2} + Z_{k,2,3}, \qquad \text{say,}
\end{aligned}
$$

$$Z_{k,3}(x) = \frac{\frac{2}{n_k}\sum_{j=1}^{n_k}(\mu_k(X_{k,j}) - \bar{\mu}_{k,c}(X_{k,j}))(\varepsilon_{k,j} - \bar{\varepsilon}_{k,c}(X_{k,j}))\delta_{k,j}K_{d_k}(x - X_{k,j})}{\hat{h}_{k,d}(x)}.$$

By (m), (p), and (u), we have $\sup_{x\in\mathcal{I}} Z_{k,1}(x) \le \ell_{\mu_2}^2 c_k^2 + o_p(n^{-1}) = o_p(1)$. (2.13) and (2.14) in Lemma 2.2.2 leads to the result

$$\begin{aligned}\sup_{x\in\mathcal{I}} Z_{k,2,1}(x) &\leq \max_{1\leq j\leq n_k} \bar{\varepsilon}^2_{k,c}(X_{k,j}) \\ &\leq \sup_{x\in\mathcal{I}}\Big|\frac{h(x)}{\hat{h}_{k,c}(x)}\Big|^2 \sup_{x\in\mathcal{I}} \frac{\big|\frac{1}{n_k}\sum_{l=1}^{n_k}\varepsilon_{k,l}\delta_{k,l}K_{c_k}(x-X_{k,l})\big|^2}{h^2(x)} = o_p(1).\end{aligned}$$

From (2.13) and (2.15) in Lemma 2.2.2, one obtains

$$\begin{aligned}\sup_{x\in\mathcal{I}}|Z_{k,2,3}(x)| &\leq \sup_{x\in\mathcal{I}} \frac{\big|\frac{1}{n_k}\sum_{j=1}^{n_k}\varepsilon^2_{k,j}\delta_{k,j}K_{d_k}(x-X_{k,j}) - \sigma_k^2(x)h_k(x)\big|}{h_k(x)} \\ &\quad \times \sup_{x\in\mathcal{I}}\frac{h_k(x)}{\hat{h}_{k,d}(x)} + \sup_{x\in\mathcal{I}}\sigma_k^2(x)\sup_{x\in\mathcal{I}}\Big|\frac{h_k(x)}{\hat{h}_{k,d}(x)} - 1\Big| \\ &= o_p(1)o_p(1) + o_p(1)o_p(1) = o_p(1).\end{aligned}$$

By C-S inequality, we have $\sup_{x\in\mathcal{I}}|Z_{k,2,2}(x)| = o_p(1)$ and $\sup_{x\in\mathcal{I}}|Z_{k,3}(x)| = o_p(1)$. Therefore, $\sup_{x\in\mathcal{I}}|\hat{\sigma}_k^2(x) - \sigma_k^2(x)| = o_p(1)$ holds under $\mathrm{H}_{1\mathrm{N}}$.

Under H_1 of (2.2), the above proof remain the same except that of $Z_{1,1}(x)$. By (m), (q), (u), and compactness of $\mathcal{I}$,

$$\begin{aligned}&\sup_{x\in\mathcal{I}} Z_{1,1}(x) \\ &\quad\leq \sup_{x\in\mathcal{I},0\leq t\leq c_1}(\mu_1(x)-\mu_1(x+t))^2 \\ &\quad\leq \sup_{x\in\mathcal{I},0\leq t_1\leq c_1} 2(\mu_2(x)-\mu_2(x+t_1))^2 + \sup_{x\in\mathcal{I},0\leq t_2\leq c_1} 2(\xi(x)-\xi(x+t_2))^2 \\ &\quad\leq 2\ell^2_{\mu_2}c_1^2 + \sup_{x\in\mathcal{I},0\leq t_2\leq c_1} 2(\xi(x)-\xi(x+t_2))^2 = o_p(1).\end{aligned}$$

Therefore one has (2.25) under H_1. The rest of the results in this lemma can be proved in a routine fashion. Thus the proof is complete.

To estimate τ^2, let $\{I_{k,1}, \cdots, I_{k,B_k}\}$ and π_k be as in assumption (z). Define

$$\hat{\Delta}_k(x) := \frac{\sum_{j=1}^{n_k} \delta_{k,j} K_{c_k}(x - X_{k,j})}{\sum_{j=1}^{n_k} K_{c_k}(x - X_{k,j})}, \quad x \in \mathbb{R},$$
$$\tilde{g}_k(x) := \frac{1}{n_k \pi_k} \sum_{j=1}^{n_k} \mathbf{1}_{\{X_{k,j} \in I_{k,i}\}}, \quad x \in I_{k,i}, \quad k = 1, 2.$$

By Remark 3.2 in K-S, the function $\tilde{g}_k(x)$ is a simple bin-estimate, which is uniformly consistent for $g_k(x)$ for $x \in \mathcal{I}$ under condition (z). Recall $\hat{r}_k$ from (2.9). Because τ^2 from (2.17) can be expressed as

$$\begin{aligned}
\tau &= q_1 \int \frac{r_1^2(x)\sigma_1^2(x)g_1(x)}{\Delta_1(x)} dx + q_2 \int \frac{r_2^2(x)\sigma_2^2(x)g_2(x)}{\Delta_2(x)} dx \\
&= q_1 \int \frac{v_1^4(x)\sigma_1^2(x)g_1^3(x)}{\Delta_1(x)} dx + q_2 \int \frac{v_2^4(x)\sigma_2^2(x)g_2^3(x)}{\Delta_2(x)} dx,
\end{aligned}$$

we consider two estimators of τ^2:

$$\hat{\tau}^2 := q_1 \frac{1}{n_1} \sum_{i=1}^{n_1} \frac{\hat{r}_1^2(X_{1,i})\hat{\sigma}_1^2(X_{1,i})}{\hat{\Delta}_1(X_{1,i})} + q_2 \frac{1}{n_2} \sum_{j=2}^{n_2} \frac{\hat{r}_2^2(X_{2,j})\hat{\sigma}_2^2(X_{2,j})}{\hat{\Delta}_2(X_{2,j})},$$
$$\hat{\tau}_*^2 := q_1 \frac{1}{n_1} \sum_{i=1}^{n_1} \frac{\hat{v}_1^4(X_{1,i})\hat{\sigma}_1^2(X_{1,i})\tilde{g}_1^2(X_{1,i})}{\hat{\Delta}_1(X_{1,i})} + q_2 \frac{1}{n_2} \sum_{j=2}^{n_2} \frac{\hat{v}_2^4(X_{2,j})\hat{\sigma}_2^2(X_{2,j})\tilde{g}_2^2(X_{2,j})}{\hat{\Delta}_2(X_{2,j})}.$$

These estimators have the following properties, which can be proved in a routine fashion.

Lemma 2.4.2. Suppose the assumptions of Lemma 2.4.1, (w1), and (z) hold. Then

$$\hat{\tau}^2 = \tau^2 + o_p(1), \qquad \text{and} \qquad \hat{\tau}_*^2 = \tau^2 + o_p(1)$$

hold under H_0, H_1, and $\mathrm{H}_{1\mathrm{N}}$.

2.5 Simulations

In this section we investigate the behavior of the empirical size and power of the test $\hat{V}$ defined in (2.22) against local alternatives and fixed alternatives. To be specific, let $\mathcal{I} = [0, 1]$, Z_1 and Z_2 be independent standard normal random variables, and independent of $\{X_1, X_2, \delta_1, \delta_2\}$. Recall $\hat{u}_\phi$ defined in (2.24). Design and error distributions and functions including ϕ, ξ, u, μ_2, Δ_l, $l = 1, 2$, are chosen as follows:

$$X_1 \sim \mathcal{N}(0, (0.7)^2), \quad X_2 \sim \mathcal{N}(0, 1), \quad X_1 \text{ and } X_2 \text{ are independent;}$$

$$\varepsilon_1 = \frac{Z_1}{\sqrt{1 + X_1^2}}, \quad \varepsilon_2 = Z_2(1 + X_2^2);$$

$$\Delta_l(x) = D_l(x), \quad l = 1, 2,$$

$$\text{where} \quad D_1(x) = \{1 + \exp(-0.5 - 0.5x)\}^{-1}, \quad D_2(x) = \{1 + \exp(-2 - 2x)\}^{-1},$$

$$\text{or} \quad \Delta_l(x) \equiv 1, \quad l = 1, 2, \quad \text{for complete responses;}$$

$$\phi(x) = \phi_j(x), \quad j = 0, 1, 2, 3,$$

$$\text{where} \quad \phi_0(x) = 0, \quad \phi_1(x) = (x + 1)^2, \quad \phi_2(x) = 2e^x, \quad \phi_3(x) = 4\cos(x);$$

$$\xi(x) = \xi_j(x), \quad j = 1, 2, 3, \quad \text{where} \quad \xi_j(x) = \phi_j(x);$$

$$u(x) = u_j(x), \quad j = 1, 2, 3, \quad \text{where} \quad u_j(x) = \mathbf{1}_{[0,1]}(x)\phi_j(x),$$

$$\text{or} \quad u(x) = u_j^*(x), \quad j = 1, 2, 3, \quad \text{where} \quad u_j^*(x) = \hat{u}_{\phi_j}(x);$$

$$\mu_2(x) = \log(x^2 + 0.5).$$

The kernel is chosen to be $K(u) := \frac{3}{4}(1 - u^2)I\{|u| \le 1\}$, with bandwidths $a = \rho_1 N^{-2/3}$, $b_k = \rho_2 n_k^{-2/3}$, and $c_k = d_k = \rho_3 n_k^{-1/4}$, $k = 1, 2$, where ρ_i, $i = 1, 2, 3$, are constants. The sample sizes are chosen to be $n_1 = n_2 = 50$, 100, 200. All simulations are based on 2000 replications. The nominal level is $\alpha = 0.05$. The empirical sizes and powers are computed by the relative frequency of the event $\{N^{1/2}\hat{T}/\hat{\tau} \geq 1.645\}$.

Table 2.1: Empirical sizes of $\hat{V}$, with coefficients ρ_1, ρ_2, ρ_3, and $\Delta_l = D_l$, $l = 1, 2$.

	(ρ_1, ρ_2, ρ_3)	$n_1 = n_2 = 50$	$n_1 = n_2 = 100$	$n_1 = n_2 = 200$
u_1	(.5,.2,.8)	.066	.057	.050
u_1^*	(.5,.2,.8)	.071	.059	.052
u_2	(.8,.5,.8)	.077	.053	.051
u_2^*	(.2,.5,.8)	.066	.062	.052
u_3	(.2,.2,.5)	.072	.058	.050
u_3^*	(.2,.5,.8)	.068	.055	.049

Table 2.2: Empirical sizes of $\hat{V}$, with coefficients ρ_1, ρ_2, ρ_3, and $\Delta_l = 1$, $l = 1, 2$.

	(ρ_1, ρ_2, ρ_3)	$n_1 = n_2 = 50$	$n_1 = n_2 = 100$	$n_1 = n_2 = 200$
u_1	(.8,.5,.8)	.073	.066	.052
u_1^*	(.8,.8,.8)	.085	.079	.052
u_2	(.2,.2,.8)	.072	.065	.049
u_2^*	(.5,.8,.8)	.084	.074	.051
u_3	(.2,.2,.8)	.071	.061	.052
u_3^*	(.2,.2,.8)	.063	.073	.050

Before we calculate the empirical powers, we choose suitable coefficients (ρ_1, ρ_2, ρ_3) in bandwidths for each u and the corresponding test, in order to make the empirical size close to 0.05 when $n_1 = n_2 = 200$. To find such coefficients, we compare the empirical sizes among all choices of $\rho_i \in \{0.2, 0.5, 0.8\}$, $i = 1, 2, 3$, and pick the one which is closest to 0.05 at $n_1 = n_2 = 200$. For each u, empirical sizes with the best choice of (ρ_1, ρ_2, ρ_3) at $n = 50$ and 100 are also listed. These results of data with responses missing at random, i.e. $\Delta_l = D_l$, $l = 1, 2$, are put in Table 2.1; while results of complete data set, $\Delta_l \equiv 1$, $l = 1, 2$, are reported in Table 2.2. Notice that these choices of $\rho_i's$ are just fairly good ones among many others. There doesn't really exist best choices. The behavior of $\hat{V}$ under null hypothesis will not be affected by the choices of these coefficients for large sample sizes n_1 and n_2.

Table 2.3: Empirical powers of $\hat{V}$ with ρ_1, ρ_2, ρ_3 in Table 2.1, and $\Delta_l = D_l$, $l = 1, 2$.

ϕ	$n_1 = n_2$	$u = u_1$	$u = u_1^*$	$u = u_2$	$u = u_2^*$	$u = u_3$	$u = u_3^*$
ϕ_1	50	.268	.183	.281	.230	.242	.210
	100	.238	.230	.265	.255	.233	.228
	200	.238	.308	.269	.341	.215	.282
ϕ_2	50	.420	.268	.436	.318	.356	.308
	100	.384	.281	.388	.339	.370	.324
	200	.379	.357	.389	.399	.351	.411
ϕ_3	50	.431	.292	.472	.344	.436	.371
	100	.402	.303	.403	.368	.427	.395
	200	.410	.379	.388	.421	.444	.464

Table 2.4: Empirical powers of $\hat{V}$ with ρ_1, ρ_2, ρ_3 in Table 2.2, and $\Delta_l = 1$, $l = 1, 2$.

ϕ	$n_1 = n_2$	$u = u_1$	$u = u_1^*$	$u = u_2$	$u = u_2^*$	$u = u_3$	$u = u_3^*$
ϕ_1	50	.339	.345	.238	.325	.234	.207
	100	.295	.302	.247	.308	.215	.185
	200	.280	.303	.236	.282	.237	.201
ϕ_2	50	.519	.509	.382	.503	.360	.314
	100	.472	.503	.353	.477	.376	.310
	200	.429	.468	.373	.494	.373	.329
ϕ_3	50	.495	.541	.401	.542	.448	.405
	100	.434	.490	.380	.507	.445	.420
	200	.404	.483	.376	.503	.455	.404

Table 2.5: Empirical sizes and powers of $\hat{V}$ with $\rho_1 = \rho_2 = \rho_3 = 1$ and $\Delta_l = D_l$, $l = 1, 2$.

ϕ	$n_1 = n_2$	$u = u_1$	$u = u_1^*$	$u = u_2$	$u = u_2^*$	$u = u_3$	$u = u_3^*$
ϕ_0	50	.070	.070	.068	.080	.065	.065
	100	.055	.060	.060	.052	.058	.058
	200	.056	.056	.054	.054	.060	.059
ϕ_1	50	.280	.293	.292	.299	.281	.252
	100	.262	.288	.264	.273	.258	.241
	200	.243	.275	.244	.280	.242	.230
ϕ_2	50	.439	.459	.450	.457	.439	.412
	100	.405	.448	.419	.434	.418	.394
	200	.394	.440	.374	.458	.396	.411
ϕ_3	50	.410	.463	.417	.474	.503	.521
	100	.362	.439	.391	.489	.488	.508
	200	.332	.454	.361	.490	.476	.516

Table 2.6: Empirical sizes and powers of $\hat{V}$ with $\rho_1 = \rho_2 = \rho_3 = 1$ and $\Delta_l = 1$, $l = 1, 2$.

ϕ	$n_1 = n_2$	$u = u_1$	$u = u_1^*$	$u = u_2$	$u = u_2^*$	$u = u_3$	$u = u_3^*$
	50	.068	.073	.076	.088	.063	.078
ϕ_0	100	.061	.074	.056	.061	.052	.054
	200	.060	.059	.045	.066	.041	.062
	50	.305	.331	.315	.340	.301	.274
ϕ_1	100	.294	.307	.290	.315	.266	.288
	200	.263	.317	.281	.320	.265	.253
	50	.477	.525	.494	.525	.500	.488
ϕ_2	100	.452	.524	.456	.514	.470	.467
	200	.439	.472	.463	.509	.474	.471
	50	.449	.537	.474	.571	.553	.611
ϕ_3	100	.394	.537	.412	.557	.541	.601
	200	.364	.533	.392	.541	.535	.610

Table 2.3 and 2.4 give the empirical powers of $\hat{V}$ against $\text{H}_{1\text{N}}$ of (2.1), with respect to missing data and complete data, respectively. These empirical powers of each test with corresponding u are calculated with the coefficients (ρ_1, ρ_2, ρ_3) in bandwidths given in Table 2.1 and 2.2.

Table 2.5 and 2.6 compare the empirical powers of $\hat{V}$ with different $u's$ against $\text{H}_{1\text{N}}$, by choosing common coefficients $\rho_1 = \rho_2 = \rho_3 = 1$ in bandwidths, with respect to missing data and complete data, respectively. In each table, one can see that the empirical sizes are getting closer to 0.05 as the sample sizes increase. For each $\phi = \phi_j$, $j = 1, 2, 3$, the test $\hat{V}$ with $u = u_j^*$ has the largest, or one of several largest, empirical power among all choices of u. This is consistent with the result in Remark 2.3.1. Moreover, for each $j = 1, 2, 3$, $u = u_j^*$ has larger empirical powers than $u = u_j$ for all choices of ϕ. From comparison between two tables, one can see that empirical powers of all these tests at three sample sizes of the complete data's are larger than those of the missing data's, while their empirical sizes don't show much difference. It means that data missing probability affects the power of the test.

All of the empirical powers of $\hat{V}$ with above choices of u are 1, against H_1 in (2.2) with

$\xi = \xi_j$, $j = 1, 2, 3$, for both of the missing data and the complete data, and for all three sample sizes. This result in turn shows the consistency of $\hat{V}$.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Bosq, D. 1998. *Nonparametric Statistics for stochastic Processes*, 2nd Edition. Springer, Berlin.

[2] Collomb, G. and Härdle, W. (1986). Strong uniform convergence rates in robust nonparametric time series analysis and prediction: kernel regression estimation from dependent observations. *Stochastic Process. Appl.*, **23**, no.1, 77-89.

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

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

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

[6] Hall, P., Huber, C., Speckman, P.L., 1997. Covariate-matched one-sided tests for the difference between functional means. *J. Amer. Statist. Assoc.*, **92**, 1074-1083.

[7] Hall, P. 1984. Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.*, **14**, 1-16.

[8] Härdle, W. and Mammen, E. 1993. Comparing nonparametric versus parametric regression fits. *Ann. Statist.*, **21**, 1926-1947.

[9] Hart, J.D. 1997. *Nonparametric Smoothing and Lack-of-fit Tests*. Springer, New York.

[10] Koul, H.L. 2011. Minimum distance lack-of-fit tests for fixed design. *J. Statist. Plann. Inference*, **141**, 65-79.

[11] Koul, H.L. and Ni, P.P. 2004. Minimum distance regression model checking. *J. Statist. Plann. Inference*, **119**, 109-141.

[12] Koul, H.L. and Schick, A. (1997). Testing the equality of two regression curves. *J. Statist. Plann. Inference*, **65**, 293-314.

[13] Koul, H.L. and Schick, A. (2003). Testing the superiority among two regression curves. *J. Statist. Plann. Inference*, **117**, 15-33.

[14] Koul, H.L. and Song, W.X. 2009. Minimum distance regression model checking with Berkson measurement errors. *Ann. Statist.*, **37**, 132-156.

[15] Little, R.J.A. and Rubin, D.B. 1987. *Statistical Analysis with Missing Data*. Wiley, New York.

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

[17] Schick, A. 1987. A note on the construction of asymptotic linear estimators. *J. Statist. Plann. Inference*, **16**, 89-105. Correction (1989), **22**, 269-270.

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

[19] Sun, Z.H. and Wang, Q.H. 2009. Checking the Adequacy of a General Linear Model with Responses Missing at Random. *J. Statist. Plann. Inference*, **139**, 3588-3604.

[20] Zheng, J.X. 1996. A consistent test of functional from via nonparametric estimation technique. *J. Econometrics*, **75**, 263-289.