Minimum Distance Regression Model Checking [1]

by

Hira L. Koul and Pingping Ni

Michigan State University

Abstract

This paper discusses a class of minimum distance tests for fitting a parametric regression model to a regression function when the underlying d dimensional design variable is random, $d \geq 1$, and the regression model is possibly heteroscedastic. These tests are based on certain minimized L_2 distances between a nonparametric regression function estimator and the parametric model being fitted. The paper establishes the asymptotic normality of the proposed test statistics and that of the corresponding minimum distance estimators under the fitted model. These estimators turn out to be $n^{1/2}$- consistent. Some simulations are also included.

1 Introduction

This paper discusses a minimum distance method for fitting a parametric model to the regression function. To be specific, let X, Y be random variables, with X being d-dimensional and Y 1-dimensional with $E|Y| < \infty$. Let

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

denote the regression function, and let $\{m_\theta(\cdot) : \theta \in \Theta\}, \Theta \subset \mathbb{R}^q$ be a given parametric model. The statistical problem of interest here is to test the hypothesis

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

based on the random sample $\{(X_i, Y_i) : i = 1, ..., n\}$ from the distribution of (X, Y), where $\mathcal{I}$ is a compact subset of $\mathbb{R}^d$. Moreover, assuming that the given parametric model holds, one is interested in finding the parameter in the given family that best fits the data.

Several authors have addressed the problem of regression model fitting: see, e.g., Cox, Koh, Wahba and Yandell (1988), Eubank and Hart (1992, 1993), Eubank and Spiegelman (1990), Härdle and Mammen (1993), Zheng (1996), Stute (1997), Stute, González Manteiga and Presedo Quindimil (1998), Stute, Thies and Zhu (1998), Diebolt and Zuber (1999), among others. The last four references propose tests based on a certain marked empirical process while the former cited references base tests on nonparametric regression estimators. Härdle and Mammen (1993) and Stute, González Manteiga and Presedo Quindimil (1998) recommend to use wild bootstrap method to implement their tests for fitting the linear model $m_\theta(x) = \theta'\gamma(x)$, where γ is a vector of q real valued functions on $\mathbb{R}^d$. The monograph of Hart (1997) contains a large class of tests for the case $d = 1$ and numerous additional references.

In the present paper, $d \geq 1$, the design is random, and the errors are allowed to be heteroscedastic. Moreover, the asymptotic normality of the proposed test statistic that is proved here allows one

[1]Research partly supported by the NSF Grant DMS 0071619. **August 12, 2002**

2000 IMS Subject Classification: Primary: 62G08, Secondary: 62G05

Keywords and Phrases: Kernel estimator, nonparametric regression, L_2 distance, Monte Carlo.

to bypass the bootstrap methodology. The proposed inference procedures are motivated from the minimum distance ideas of Wolfowitz (1953, 1954, 1957). The inference procedures based on various L_2-distances have proved to be successful in producing tests for fitting a distribution and/or a density function, and in producing asymptotically efficient and robust estimators of the underlying parameters in the fitted model, as is evidenced in the works of Beran (1977, 1978), Donoho and Liu (1988a, 1988b), Koul (1985), and González Manteiga (1990), among others.

In the context of density fitting problem in the one sample set up, Beran (1977, 1978) showed that the inference procedures based on the Hellinger distance have desirable properties. In the current context, this motivates one to consider the L_2 distance

$$M_n^*(\theta) = \int_{\mathcal{I}} (\hat{\mu}_h(x) - m_\theta(x))^2 dG(x), \qquad \theta \in \mathbb{R}^q,$$

and the corresponding minimum distance estimator $\alpha_n^* = \text{argmin}_{\theta\in\Theta} M_n^*(\theta)$, where $\hat{\mu}_h(x)$ is a nonparametric estimator of the regression function $\mu(x)$ based on the window width h_n:

$$\begin{aligned} \hat{\mu}_h(x) &= \frac{n^{-1}\sum_{i=1}^n K_h(x-X_i)Y_i}{\hat{f}_h(x)}, \qquad \hat{f}_h(x) = n^{-1}\sum_{i=1}^n K_h(x-X_i),\ x\in\mathbb{R}^d, \\ K_h(u) &= \frac{1}{h_n^d} K(\frac{u}{h_n}), \end{aligned}$$

K a kernel function on $[-1,1]^d$ and G is a σ-finite measure on $\mathbb{R}^d$.

But because the integrand inside the square of M_n^* is not centered, and because of the non-negligible asymptotic bias in the nonparametric estimator $\hat{\mu}_h$, the goodness-of-fit statistic $M_n^*(\alpha_n^*)$ does not have a desirable asymptotic null distribution. Moreover, the estimator α_n^*, though consistent, is not asymptotically normal. In fact it can be shown that generally the sequence $(nh^d)^{1/2}\|\alpha_n^* - \theta_0\|$ is not even tight, cf. Ni (2002). To overcome this difficulty, one may think of using

$$\begin{aligned} M_n(\theta) &= \int_{\mathcal{I}} \left[n^{-1}\sum_{i=1}^n K_h(x-X_i)\,(Y_i - m_\theta(X_i)) \right]^2 \{\hat{f}_h(x)\}^{-2} dG(x), \\ \hat{\alpha}_n &= \text{argmin}_{\theta\in\Theta} M_n(\theta). \end{aligned}$$

Under the null hypothesis H_0, the i^{th} summand inside the squared integrand of $M_n(\theta_0)$ is now conditionally centered, given the i^{th} design variable, $1 \le i \le n$. But the asymptotic bias in $n^{1/2}(\hat{\alpha}_n - \theta_0)$ and $M_n(\hat{\alpha}_n)$ caused by the nonparametric estimator $\hat{f}_h$ of f in the denominator of $\hat{\mu}_h$ still exists. An important observation of this paper is that these asymptotic biases can be made negligible if one uses an optimal window width for the estimation of the density f, different from h, and possibly a different kernel, to estimate f. This leads us to consider the following modification of the above procedures: Define

$$\begin{aligned} \hat{f}_w(x) &= n^{-1}\sum_{i=1}^n K_w^*(x-X_i),\ x\in\mathbb{R}^d, \quad w_n \sim (\log n/n)^{\frac{1}{d+4}}, \\ T_n(\theta) &= \int_{\mathcal{I}} \left[n^{-1}\sum_{i=1}^n K_h(x-X_i)\,(Y_i - m_\theta(X_i)) \right]^2 \{\hat{f}_w(x)\}^{-2} dG(x), \end{aligned}$$

where K^* is a density kernel function, possibly different from K. The proposed class, one for each G, of minimum distance tests of H_0 and estimators of θ, respectively, are

$$\inf_{\theta\in\Theta} T_n(\theta) = T_n(\hat\theta_n), \qquad \hat\theta_n := \text{argmin}_{\theta\in\Theta} T_n(\theta).$$

We also need the following entities:

$$\begin{aligned}
\hat m_n(x) &:= n^{-1}\sum_{i=1}^n K_h(x-X_i)Y_i\Big/\hat f_w(x), \quad x\in\mathbb{R}^d,\\
T_n^*(\theta) &= \int_{\mathcal{I}} (\hat m_n(x) - m_\theta(x))^2 dG(x), \quad \theta\in\mathbb{R}^q, \quad \theta_n^* := \text{argmin}_{\theta\in\Theta} T_n^*(\theta).
\end{aligned}$$

This paper first proves the consistency of θ_n^*, $\hat\theta_n$, and the asymptotic normality of $n^{1/2}(\hat\theta_n - \theta_0)$ and $nh^{d/2}\left(T_n(\hat\theta_n) - \tilde C_n\right)$ under H_0, where $\tilde C_n$ is given below at (1.1). Both limiting distributions have mean 0 while the variance of the latter is Γ. Then, sequences of estimators $\hat C_n$ and $\hat\Gamma_n$ are provided such that $\hat C_n$ is $nh^{d/2}$-consistent for $\tilde C_n$ and $\hat\Gamma_n$ is consistent for Γ so that the asymptotic null distribution of $D_n := nh^{d/2}\left(T_n(\hat\theta_n) - \hat C_n\right)/\hat\Gamma_n^{1/2}$ is standard normal. This result is similar in nature as the corollary to Theorem 8 of Beran (1977, p459). A test of H_0 can be thus based on D_n. Here,

$$\begin{aligned}
(1.1) \qquad \tilde C_n &:= n^{-2}\sum_{i=1}^n \int_{\mathcal{I}} K_h^2(x-X_i)\,\varepsilon_i^2\,\{f(x)\}^{-2} dG(x),\\
\hat C_n &:= n^{-2}\sum_{i=1}^n \int_{\mathcal{I}} K_h^2(x-X_i)\,\hat\varepsilon_i^2\,\{\hat f_w(x)\}^{-2} dG(x), \ \ \hat\varepsilon_i := Y_i - m_{\hat\theta_n}(X_i),\ 1\le i\le n,\\
\Gamma &:= 2\int_{\mathcal{I}} \sigma^4(x)\,\frac{g^2(x)}{f^2(x)}dx \int\Big(\int K(u)K(v+u)du\Big)^2 dv,\\
\hat\Gamma_n &:= h^d n^{-2}\sum_{i\ne j}\Big(\int_{\mathcal{I}} K_h(x-X_i)K_h(x-X_j)\,\hat\varepsilon_i\hat\varepsilon_j\,\{\hat f_h(x)\}^{-2} dG(x)\Big)^2,
\end{aligned}$$

where $\sigma^2(x) := E\Big\{(Y - m_{\theta_0}(x))^2\Big|X = x\Big\}$, $x\in\mathbb{R}^d$, and g is the density of G.

There are three reasons for choosing the window width w_n different from h_n in the estimator of f. The first is to obtain $nh^{d/2}$-consistent estimator $\hat C_n$ of the asymptotic centering $\tilde C_n$, i.e., to prove $nh_n^{d/2}(\hat C_n - \tilde C_n) = o_p(1)$, cf. Lemma 5.4 below. The second reason is similar in that it renders the asymptotic bias of $n^{1/2}(\hat\theta_n - \theta_0)$ to be zero, cf. Lemma 4.1(B). The third is given in Lemma 5.3.

Härdle and Mammen (1993) also consider a class of goodness-of-fit tests based on $M_n(\tilde\theta)$ (see T_n there), where the estimator $\tilde\theta$ of θ_0 and the null model $\{m_\theta\}$ are assumed to satisfy, among other conditions, the following condition.

$$m_{\tilde\theta}(x) - m_{\theta_0}(x) = (1/n)\sum_{i=1}^n <\gamma(x), l(X_i)> \varepsilon_i + o_p((n\log n)^{-1/2}),$$

uniformly in x, with γ and l being bounded functions taking values in $\mathbb{R}^k$ for some k. This assumption holds for linear models and the weighted least square estimators in nonlinear models if m_θ is "smooth in θ" with $\gamma(x) \equiv (\partial/\partial\theta)m_{\theta_0}(x)$, but not for the minimum distance estimator $\hat{\theta}_n$. Among other things, they need to have the bandwidth $h_n \propto n^{-1/(d+4)}$. This bandwidth is asymptotically optimal for the class of twice continuously differentiable regression functions, and it is also crucial in getting the rates of uniform consistency of nonparametric estimators of μ, which in turn are needed in their proofs. Under some additional assumptions, they concluded that the asymptotic null distribution of $nh_n^{d/2}(M_n(\tilde{\theta}) - C_n^*)$ is $\mathcal{N}_1(0, \tilde{V})$, where C_n^* depends on $\mu - m_{\theta_0}$, the second derivative $K^{(2)}$ of K and $h_n^{-d/2}$, and where

$$\tilde{V} = 2\int \frac{\sigma^4(x)g^2(x)}{f^2(x)}dx \int \Big(K^{(2)}(t)\Big)^2 dt.$$

They do not provide any estimators of the asymptotic mean and variance, but instead recommend using the wild bootstrap to determine the critical values. They also do not discuss any asymptotics for the minimizer $\tilde{\alpha}_n$.

In contrast our results do not require the null regression function to be twice continuously differentiable nor do the proofs in this paper need the rate for uniform consistency of $\hat{\mu}_h$ for μ. Moreover, we derive the asymptotic distributions of $n^{1/2}(\hat{\theta}_n - \theta_0)$ and D_n under H_0. This was made feasible by recognizing to use different window widths for the estimation of the numerator and denominator in the nonparametric regression function estimator.

Zheng (1996) proposed a test of H_0 based on the statistic

$$V_n := \frac{1}{n(n-1)}\sum_{i=1}^n \sum_{j=1, j\neq i}^n K_h(X_i - X_j)e_i e_j,$$

where $e_i = Y_i - m_{\hat{\theta}}(X_i)$, with $\hat{\theta}$ the least square estimator under H_0. As can be directly seen, this test statistic is not a member of the above class of minimum distance statistics. Under some regularity conditions that includes the second order differentiability of m_θ in θ, Zheng proves the asymptotic normality of $nh^{d/2}V_n$ and consistency of the corresponding test against certain fixed alternatives.

The rest of the paper is organized as follows. Section 2 states various assumptions, and section 3 contains the consistency proofs. The claimed asymptotic normality of $\hat{\theta}_n$ and $T_n(\hat{\theta}_n)$ are proved in sections 4 and 5, respectively.

Section 6 contains two simulation studies. The first investigates the Monte Carlo size and power of the MD test $D_n := nh(T_n(\hat{\theta}_n) - \tilde{C}_n)$ in the case $d = 2$, and when fitting the linear model $\theta' x$, θ, $x \in \mathbb{R}^2$. The power is computed at 4 alternatives, two error distributions (double exponential and normal) and two designs: the two coordinates of X are i.i.d. normal and bivariate normal with correlation 0.36. From tables 1 and 2, one sees that this test has very good empirical power against the four chosen alternatives, but is affected by the dependence structure of the design coordinates. The performance with regards to the empirical size is not as impressive for the sample sizes $30, 50, 100$, but is satisfactory for 200.

The second simulation study investigates a Monte Carlo size and power comparison of an MD test with the tests of An and Cheng (1991) (AC) and Stute, Thies and Zhu (1998) (STZ) for $d = 1$ when fitting a simple linear regression model against two alternatives. This simulation shows no clear cut domination of any one test over the other, although the STZ test appears to have some advantage with regards to the empirical size, while the AC test performs the worse of the three tests. See Section 6 for details.

2 Assumptions

Here we shall state the needed assumptions. About the errors, the underlying design and G we assume the following:

(e1) The random variables $\{(X_i, Y_i);\ X_i \in \mathbb{R}^d,\ Y_i \in \mathbb{R}, i = 1, \cdots, n\}$, are *i.i.d.* with the regression function $\mu(x) = E(Y|X = x)$ satisfying $\int \mu^2(x)dG(x) < \infty$, where G is a σ-finite measure on $\mathbb{R}^d$.

(e2) $E(Y - \mu(X))^2 < \infty$ and the function $\sigma^2(x) := E\{(Y - \mu(x))^2|X = x\}$ is a.s. (G) continuous on $\mathcal{I}$.

(e3) $E|Y - \mu(X)|^{2+\delta} < \infty,$ for some $\delta > 0$.

(e4) $E(Y - \mu(X))^4 < \infty$.

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

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

(g) G has a continuous Lebesgue density g.

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

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

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

(m1) For each θ, $m_\theta(x)$ is a.s. continuous 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 $\beta > 0$,

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

(m4) 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 $\epsilon > 0$, $k < \infty$,

$$\limsup_n P\Big(\sup_{1\le i\le n,\, (nh_n^d)^{1/2}\|\theta-\theta_0\|\le k} \frac{|m_\theta(X_i) - m_{\theta_0}(X_i) - (\theta - \theta_0)^T \dot{m}_{\theta_0}(X_i)|}{\|\theta - \theta_0\|} > \epsilon\Big) = 0.$$

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

$$P\left(\max_{1\le i\le n,\, (nh_n^d)^{1/2}\|\theta-\theta_0\|\le k} h_n^{-d/2}\|\dot{m}_\theta(X_i) - \dot{m}_{\theta_0}(X_i)\| \ge \epsilon\right) \le \epsilon, \qquad \forall\, n > N_\epsilon.$$

About the bandwidth h_n we shall make the following assumptions:

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

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

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

Conditions (e1), (e2), (f1), (k), (m1) - (m3), (h1) and (h2) suffice for the consistency of $\hat{\theta}_n$, while these plus (e3), (f2), (m4), (m5) and (h3) are needed for the asymptotic normality of $\hat{\theta}_n$. The asymptotic normality of $T_n(\hat{\theta}_n)$ needs (e1), (e2), (e4), and (f1) - (m5), and (h3). Of course, (h3) implies (h1) and (h2). Note that the conditions (m1) - (m5) are trivially satisfied by the model $m_\theta(x) \equiv \theta'\gamma(x)$ provided the components of γ are continuous on $\mathcal{I}$.

It is well known that under (f1), (k), (h1) and (h2), cf., Mack and Silverman (1982),

$$\begin{aligned}
(2.1) \qquad & \sup_{x\in\mathcal{I}} \left|\hat{f}_h(x) - f(x)\right| = o_p(1), \quad \sup_{x\in\mathcal{I}} \left|\hat{f}_w(x) - f(x)\right| = o_p(1), \\
& \sup_{x\in\mathcal{I}} \left|\frac{f(x)}{\hat{f}_w(x)} - 1\right| = o_p(1).
\end{aligned}$$

These conclusions are often used in the proofs below.

In the sequel, we write h for h_n, w for w_n; the true parameter θ_0 is assumed to be an inner point of Θ; $\varepsilon_i \equiv Y_i - m_{\theta_0}(X_i)$, ε a copy of ε_1; and the integrals with respect to the G-measure are understood to be over the set $\mathcal{I}$. The inequality $(a+b)^2 \le 2(a^2+b^2)$, for any real numbers a, b, is often used without mention in the proofs below. The convergence in distribution is denoted by $\longrightarrow_d$ and $\mathcal{N}_p(a, B)$ denotes the p-dimensional normal distribution with mean vector a and covariance matrix B, $p \ge 1$.

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

This section proves the consistency of θ_n^* and $\hat{\theta}_n$. To state and prove these results we need the following preliminary lemma. Let $L_2(G)$ denote a class of square integrable real valued functions on $\mathbb{R}^d$ with respect to G. Define

$$\rho(\nu_1, \nu_2) \quad := \quad \int_{\mathcal{I}} (\nu_1(x) - \nu_2(x))^2 dG(x), \quad \nu_1,\, \nu_2 \in L_2(G),$$

and the map

$$\mathcal{M}(\nu) = \mathrm{argmin}_{\theta\in\Theta}\, \rho(\nu, m_\theta), \quad \nu \in L_2(G).$$

Lemma 3.1 *Let m satisfy the 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.,

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

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

The proof of this lemma is similar to that of Theorem 1 Beran (1977), hence omitted.

In the sequel we shall often write

$$d\hat{\varphi}_w := \hat{f}_w^{-2} dG, \quad d\varphi = f^{-2} dG.$$

Moreover, for any integral $L := \int \gamma d\hat{\varphi}_w$, $\tilde{L} := \int \gamma d\varphi$. Thus, e.g., $\tilde{T}_n(\theta)$ stands for $T_n(\theta)$ with $\hat{\varphi}_w$ replaced by φ, i.e., with $\hat{f}_w$ replaced by f. We also need to define for $\theta \in \mathbb{R}^d$, $x \in \mathbb{R}^q$,

$$\begin{aligned}
(3.1) \qquad \mu_n(x,\theta) &:= n^{-1}\sum_{i=1}^n K_h(x-X_i)\, m_\theta(X_i), \quad \dot{\mu}_n(x,\theta) := n^{-1}\sum_{i=1}^n K_h(x-X_i)\dot{m}_\theta(X_i), \\
U_n(x,\theta) &:= n^{-1}\sum_{i=1}^n K_h(x-X_i)Y_i - \mu_n(x,\theta) \\
&= n^{-1}\sum_{i=1}^n K_h(x-X_i)(Y_i - m_\theta(X_i)), \qquad U_n(x) = U_n(x,\theta_0), \\
Z_n(x,\theta) &:= \mu_n(x,\theta) - \mu_n(x,\theta_0) = n^{-1}\sum_{i=1}^n K_h(x-X_i)[m_\theta(X_i) - m_{\theta_0}(X_i)], \\
\Sigma_0 &:= \int \dot{m}_{\theta_0}(x)\dot{m}_{\theta_0}^T(x)dG(x), \qquad C_{n1} := \int U_n^2(x)d\hat{\varphi}_w(x), \\
C_{n2}(\theta) &:= \int \Big[\mu_n(x,\theta) - \hat{f}_w(x)m_\theta(x)\Big]^2 d\hat{\varphi}_w(x), \quad \theta \in \mathbb{R}^q.
\end{aligned}$$

A consequence of Lemma 3.1 is the following

Corollary 3.1 *Suppose H_0, (e1), (e2), (f1), and (m1) - (m3) hold. Then, $\theta_n^* \longrightarrow \theta_0$, in probability.*

Proof. We shall use part (b) of the Lemma 3.1 with $\nu_n = \hat{m}_n$, $\nu = m_{\theta_0}$. Note that $T_n^*(\theta_0) = \rho(\hat{m}_n, m_{\theta_0})$, $\theta_n^* = \mathcal{M}(\hat{m}_n)$, and by the identifiability condition (m2), $\mathcal{M}(m_{\theta_0}) = \theta_0$ is unique. It thus suffices to prove

$$(3.2) \qquad \rho(\hat{m}_n, m_{\theta_0}) = o_p(1).$$

To prove this, substitute $m_{\theta_0}(X_i) + \varepsilon_i$ for Y_i inside the i^{th} summand of $T_n^*(\theta_0)$ and expand the quadratic integrand to obtain that $\rho(\hat{m}_n, m_{\theta_0})$ is bounded above by the sum $2[C_{n1} + C_{n2}(\theta_0)]$, where C_{n1}, C_{n2} are as in (3.1). It thus suffices to show that both of these two terms are $o_p(1)$.

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

$$(3.3) \qquad E\int U_n^2(x)d\varphi(x) = n^{-1}\int EK_h^2(x-X)\sigma^2(X)d\varphi(x) = O(1/nh^d) = o(1),$$

so that

$$(3.4) \qquad \int U_n^2(x)d\varphi(x) = O_p((nh^d)^{-1}).$$

Hence, by (2.1),

$$C_{n1} \;\leq\; \sup_{x\in\mathcal{I}} |f(x)/\hat{f}_w(x)|^2 \int U_n^2(x)d\varphi(x) = O_p((nh^d)^{-1}).$$

Next, we shall show

$$(3.5) \qquad C_{n2}(\theta_0) = o_p(1).$$

Let

$$\begin{aligned} e_h(x,\theta) &= EK_h(x-X)m_\theta(X) = \int K(u)m_\theta(x-uh)f(x-uh)du, \\ e_w^*(x,\theta) &= EK_w^*(x-X)m_\theta(x) = \int K^*(u)m_\theta(x)f(x-uw)du. \end{aligned}$$

By adding and subtracting $e_h(x,\theta)$ and $e_w^*(x,\theta)$ in the quadratic term of the integrand of C_{n2}, one obtains that

$$C_{n2}(\theta) \le 3C_{n21}(\theta) + 3C_{n22}(\theta) + 3C_{n23}(\theta), \quad \theta \in \Theta, \tag{3.6}$$

where

$$\begin{aligned} C_{n21}(\theta) &= \int [\mu_n(x,\theta) - e_h(x,\theta)]^2 \, d\hat{\varphi}_w(x), \\ C_{n22}(\theta) &= \int \left[\hat{f}_w(x)m_\theta(x) - e_w^*(x,\theta)\right]^2 d\hat{\varphi}_w(x), \\ C_{n23}(\theta) &= \int [e_h(x,\theta) - e_w^*(x,\theta)]^2 \, d\hat{\varphi}_w(x). \end{aligned} \tag{3.7}$$

By Fubini, the fact that the variance is bounded above by the second moment, and by (f1), (k) and (m1), one obtains that

$$E\tilde{C}_{n21}(\theta_0) \;\le\; n^{-1}\int EK_h^2(x-X)m_{\theta_0}^2(X)d\varphi(x) = O((nh^d)^{-1}). \tag{3.8}$$

Hence $C_{n21}(\theta_0) = O_p((nh^d)^{-1})$ follows from (2.1). Similarly, one can obtain that $C_{n22}(\theta_0) = O_p((nh^d)^{-1})$. The claim $C_{n23}(\theta_0) = o(1)$ follows from the continuity of m_{θ_0} and f. This completes the proof of (3.5), and hence that of (3.2) and the corollary. □

Before stating the next result we give a fact that is often used in the proofs below. Under (f1), (k), and (h2),

$$\begin{aligned} &\int E\left[n^{-1}\sum_{i=1}^n K_h(x-X_i)\alpha(X_i)\right]^2 d\varphi(x) \\ &= n^{-1}\int EK_h^2(x-X)\alpha^2(X)d\varphi(x) + \int [EK_h(x-X)\alpha(X)]^2 d\varphi(x) \\ &= o(1) + O(1) = O(1), \qquad \text{for any continuous function } \alpha \text{ on } \mathcal{I}. \end{aligned} \tag{3.9}$$

We now proceed to state and prove

Theorem 3.1 *Under H_0, (e1), (e2), (f1), (k), (m1) - (m3), (h1), and (h2),*

$$\hat{\theta}_n \longrightarrow \theta_0, \quad \textit{in probability.}$$

Proof. We shall again use part (b) of Lemma 3.1 with $\nu(x) \equiv m_{\theta_0}(x)$, $\nu_n(x) \equiv m_{\hat{\theta}_n}(x)$. Then by (m2), $\hat{\theta}_n = \mathcal{M}(\nu_n)$, $\theta_0 = \mathcal{M}(\nu)$, uniquely. It thus suffices to show that

$$\rho(m_{\hat{\theta}_n}, m_{\theta_0}) = o_p(1). \tag{3.10}$$

But observe that

$$\rho(m_{\hat{\theta}_n}, m_{\theta_0}) \le 2[\rho(\hat{m}_n, m_{\hat{\theta}_n}) + \rho(\hat{m}_n, m_{\theta_0})].$$

Thus, in view of (3.2), it suffices to show that

$$T_n^*(\hat{\theta}_n) \equiv \rho(\hat{m}_n, m_{\hat{\theta}_n}) = o_p(1). \tag{3.11}$$

But this will be implied by the following result:

$$\sup_\theta |T_n(\theta) - T_n^*(\theta)| = o_p(1). \tag{3.12}$$

For, (3.12) implies that

$$T_n^*(\hat{\theta}_n) = T_n(\hat{\theta}_n) + o_p(1), \qquad T_n^*(\theta_n^*) = T_n(\theta_n^*) + o_p(1),$$
$$T_n^*(\hat{\theta}_n) - T_n^*(\theta_n^*) = T_n(\hat{\theta}_n) - T_n(\theta_n^*) + o_p(1). \tag{3.13}$$

By the definitions of $\hat{\theta}_n$ and θ_n^*, for every n, the left hand side of (3.13) is nonnegative, while the first term on the right hand side is nonpositive. Hence,

$$T_n^*(\hat{\theta}_n) - T_n^*(\theta_n^*) = o_p(1).$$

This together with the fact that $T_n^*(\theta_n^*) \le T_n^*(\theta_0)$ and (3.2) then proves (3.11).

We now focus on proving (3.12). Add and subtract $\mu_n(x,\theta)/\hat{f}_w(x)$ inside the parenthesis of $T_n^*(\theta)$, expand the quadratic, and use the Cauchy-Schwarz inequality on the cross product, to obtain that the left hand side of (3.12) is bounded above by

$$\sup_\theta C_{n2}(\theta) + 2\sup_\theta \left(C_{n2}(\theta) T_n(\theta)\right)^{1/2}.$$

It thus suffices to show that

$$\sup_\theta C_{n2}(\theta) = o_p(1), \qquad \sup_\theta T_n(\theta) = O_p(1). \tag{3.14}$$

Recall (3.6) and (3.7). Using the same argument as for (3.8), and by the boundedness of m on $\mathcal{I} \times \Theta$, one obtains that

$$\sup_\theta E\tilde{C}_{n21}(\theta) = o(1) = \sup_\theta E\tilde{C}_{n22}(\theta).$$

By the continuity of m_θ and f, one also readily sees that $\tilde{C}_{n23}(\theta) = o(1)$, for each $\theta \in \Theta$. In view of an inequality like (3.6) for $\tilde{C}_{n2}$, we thus obtain that $\tilde{C}_{n2}(\theta) = o_p(1)$, for each $\theta \in \Theta$. This and (2.1) in turn imply that

$$C_{n2}(\theta) \quad \le \quad \sup_{x\in\mathcal{I}} \frac{f^2(x)}{\hat{f}_w^2(x)}\, \tilde{C}_{n2}(\theta) = o_p(1), \quad \forall\, \theta \in \Theta. \tag{3.15}$$

Finally, by (m3),

$$|C_{n2}(\theta_2) - C_{n2}(\theta_1)| \quad \le \quad 2\|\theta_2 - \theta_1\| \sup_{x\in\mathcal{I}} \frac{f^2(x)}{\hat{f}_w^2(x)} \Big[\int \Big[n^{-1}\sum_{i=1}^n K_h(x - X_i)\ell(X_i)\Big]^2 d\varphi(x) + \int [\hat{f}_h^*(x)\ell(x)]^2 d\varphi(x)\Big].$$

But (3.9) applied once with $\alpha \equiv \ell$ and once with $\alpha \equiv 1$ implies that the third factor of this bound is $O_p(1)$. This bound and (2.1) together with the compactness of Θ and (3.15) completes the proof of the first part of (3.14).

To prove the second part of (3.14), note that by adding and subtracting $m_{\theta_0}(X_i)$ to the i^{th} summand in $T_n(\theta)$, we obtain

$$T_n(\theta) \le 2\sup_{x\in\mathcal{I}}(f(x)/\hat f_w(x))^2\Big(\int U_n^2(x)d\varphi(x) + \int Z_n^2(x,\theta)d\varphi(x)\Big).$$

But, by the boundedness of m over $\mathcal{I}\times\Theta$ and by (3.9) applied with $\alpha \equiv 1$,

$$\sup_\theta \int Z_n^2(x,\theta)d\varphi(x) \le C\int\Big(\hat f_h(x)\Big)^2 d\varphi(x) = O_p(1). \tag{3.16}$$

This together with (3.4) then completes the proof of the second part of (3.14), and hence that of the Theorem 3.1. □

4 Asymptotic distribution of $\hat\theta_n$

In this section we shall prove the asymptotic normality of $n^{1/2}(\hat\theta_n - \theta_0)$. Let

$$\dot\mu_h(x) \;:=\; E\dot\mu_n(x,\theta_0) = EK_h(x-X)\dot m_{\theta_0}(X), \quad S_n := \int U_n(x)\dot\mu_h(x)d\varphi(x). \tag{4.1}$$

We shall prove the following

Theorem 4.1 *Assume that H_0, (e1), (e2), (e3), (f1), (f2), (g), (k), (m1) - (m5), and (h3) hold. Then,*

$$n^{1/2}(\hat\theta_n - \theta_0) = \Sigma_0^{-1}n^{1/2}S_n + o_p(1), \tag{4.2}$$

Consequently, $n^{1/2}(\hat\theta_n - \theta_0) \longrightarrow_d \mathcal{N}_q(0, \Sigma_0^{-1}\Sigma\Sigma_0^{-1})$, *where Σ_0 is as in (3.1) and*

$$\begin{aligned} \Sigma &= lim_{h\to 0}\int\int EK_h(x-X)K_h(y-X)\;\sigma^2(X)\dot\mu_h(x)\dot\mu_h^T(y)d\varphi(x)d\varphi(y) \\ &= \int \frac{\sigma^2(x)\dot m_{\theta_0}(x)\dot m_{\theta_0}^T(x)g^2(x)}{f(x)}dx. \end{aligned} \tag{4.3}$$

Proof. The proof consists of several steps. The first is to show that

$$nh^d\|\hat\theta_n - \theta_0\|^2 = O_p(1). \tag{4.4}$$

Recall the definition of Z_n from (3.1) and let $D_n(\theta) := \int Z_n^2(x,\theta)d\varphi(x)$. We claim

$$nh^dD_n(\hat\theta_n) = O_p(1). \tag{4.5}$$

To see this, observe that

$$\begin{aligned} nh^dT_n(\theta_0) &= nh^d\int\Big(n^{-1}\sum_{i=1}^n K_h(x-X_i)\varepsilon_i\Big)^2 d\hat\varphi_w(x) \\ &\le nh^d\int U_n^2(x)d\varphi(x) + nh^d\int U_n^2(x)d\varphi(x)\sup_{x\in\mathcal{I}}|f^2(x)/\hat f_w^2(x) - 1| \\ &= O_p(1), \end{aligned}$$

by (3.3) and (2.1). But, by definition, $T_n(\hat\theta_n) \le T_n(\theta_0)$, implying that $nh^d T_n(\hat\theta_n) = O_p(1)$. These facts together with the inequality $D_n(\theta) \le 2[T_n(\theta_0) + T_n(\hat\theta_n)]$ proves (4.5).

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

$$P\left(D_n(\hat\theta_n)/\|\hat\theta_n - \theta_0\|^2 \ge a + \inf_{\|b\|=1} b^T \Sigma_0 b\right) > 1 - a, \qquad \forall\ n > N_a, \tag{4.6}$$

where Σ_0 is as in (3.1). The claim (4.4) will then follow from this, (4.5), the positive definiteness of Σ_0, and the fact

$$nh^d D_n(\hat\theta_n) = nh^d \|\hat\theta_n - \theta_0\|^2\, [D_n(\hat\theta_n)/\|\hat\theta_n - \theta_0\|^2].$$

To prove (4.6), let

$$u_n \ := \ (\hat\theta_n - \theta_0), \qquad d_{ni} := m_{\hat\theta_n}(X_i) - m_{\theta_0}(X_i) - u_n^T \dot m_{\theta_0}(X_i), \quad 1 \le i \le n. \tag{4.7}$$

We have

$$\begin{aligned}
\frac{D_n(\hat\theta_n)}{\|\hat\theta_n - \theta_0\|^2} &= \int \left[n^{-1} \sum_{i=1}^n K_h(x - X_i) \left(\frac{d_{ni}}{\|u_n\|} \right) \right]^2 d\varphi(x) + \int \left[\frac{u_n' \,\dot\mu_n(x,\theta_0)}{\|u_n\|} \right]^2 d\varphi(x) \\
&= D_{n1} + D_{n2}, \qquad \text{say.}
\end{aligned}$$

By the assumption (m4) and the consistency of $\hat\theta_n$, one verifies by a routine argument that $D_{n1} = o_p(1)$. For the second term we notice that

$$D_{n2} \ge \inf_{\|b\|=1} \Sigma_n(b), \tag{4.8}$$

where

$$\Sigma_n(b) := \int \left[b^T \,\dot\mu_n(x,\theta_0) \right]^2 d\varphi(x), \qquad b \in \mathbb{R}^d.$$

By the usual calculations one sees that for each $b \in \mathbb{R}^d$, $\Sigma_n(b) \to b^T \Sigma_0 b$, in probability. Also, note that for any $\delta > 0$, and any two unit vectors $b,\, b_1 \in \mathbb{R}^d$, $\|b - b_1\| \le \delta$, we have

$$|\Sigma_n(b) - \Sigma_n(b_1)| \le \delta(\delta + 2) \left[\int n^{-1} \sum_{i=1}^n K_h(x - X_i)\, \|\dot m_{\theta_0}(X_i)\| d\varphi(x) \right]^2.$$

But the expected value of the r.v.'s inside the square of the second factor tends to $\int \|\dot m_{\theta_0}(x)\|$ $f(x) d\varphi(x)$, and hence this factor is $O_p(1)$. From these observations and the compactness of the set $\{b \in \mathbb{R}^d; \|b\| = 1\}$, we obtain that

$$\sup_{\|b\|=1} |\Sigma_n(b) - b^T \Sigma_0 b| = o_p(1).$$

This fact together with (4.8) implies (4.6) in a routine fashion, and also concludes the proof of (4.4).

The remaining proof is classical in nature. Recall the definitions (3.1) and (4.7), and let

$$\dot T_n(\theta) := -2 \int U_n(x,\theta)\, \dot\mu_n(x,\theta) d\hat\varphi_w(x).$$

Since θ_0 is an interior point of Θ, by the consistency, for sufficiently large n, $\hat\theta_n$ will be in the interior of Θ and $\dot T_n(\hat\theta_n) = 0$, with arbitrarily large probability. But the equation $\dot T_n(\hat\theta_n) = 0$ is equivalent to

$$\int U_n(x)\,\dot\mu_n(x,\hat\theta_n)d\hat\varphi_w(x) \;=\; \int Z_n(x,\hat\theta_n)\,\dot\mu_n(x,\hat\theta_n)d\hat\varphi_w(x). \tag{4.9}$$

In the final step of the proof we shall show that $n^{1/2}\times$ the left hand side of this equation converges in distribution to a normal r.v., while the right hand side of this equation equals $R_n(\hat\theta_n - \theta_0)$, for all $n \ge 1$, with $R_n = \Sigma_0 + o_p(1)$.

To establish the first of these two claims, rewrite this r.v. as the sum $S_n + S_{n1} + g_{n1} + g_{n2} + g_{n3} + g_{n4}$, where S_n is as in (4.1) and

$$\begin{aligned}
S_{n1} &= \int U_n(x)\dot\mu_h(x)(\hat f_w^{-2}(x) - f^{-2}(x))dG(x),\\
g_{n1} &= \int U_n(x)\,[\dot\mu_n(x,\theta_0) - \dot\mu_h(x)]\,d\varphi(x),\\
g_{n2} &= \int U_n(x)\,[\dot\mu_n(x,\theta_0) - \dot\mu_h(x)]\,(\hat f_w^{-2}(x) - f^{-2}(x))dG(x),\\
g_{n3} &= \int U_n(x)\,\Big[\dot\mu_n(x,\hat\theta_n) - \dot\mu_n(x,\theta_0)\Big]\,d\varphi(x)\\
g_{n4} &= \int U_n(x)\,\Big[\dot\mu_n(x,\hat\theta_n) - \dot\mu_n(x,\theta_0)\Big]\,(\hat f_w^{-2}(x) - f^{-2}(x))dG(x).
\end{aligned}$$

We need the following lemmas.

Lemma 4.1 *Suppose H_0, (e1), (e2), (f1), (k), (m1) - (m5), (h1) and (h2) hold.*

(A) If, additionally, (e3) and (g) hold, then, $n^{1/2}S_n \longrightarrow_d \mathcal{N}_q(0,\Sigma)$.

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

$$n^{1/2}|S_{n1}| = o_p(1). \tag{4.10}$$

Lemma 4.2 *Under H_0, (e1), (e2), (f1), (k), (m1), (m2), (m4), (m5), (h1), (h2),*

$$(a)\quad n^{1/2}g_{n1} = o_p(1), \qquad (b)\quad n^{1/2}g_{n2} = o_p(1). \tag{4.11}$$

$$(c)\quad n^{1/2}g_{n3} = o_p(1), \qquad (d)\quad n^{1/2}g_{n4} = o_p(1). \tag{4.12}$$

The proof of (4.10) is facilitated by the following lemma, which along with its proof appears as Theorem 2.2 part (2) in Bosq (1998).

Lemma 4.3 *Let $\hat f_w$ be the kernel estimate associate with a kernel K^* which satisfies a Lipschitz condition. If (f2) holds and $w_n = a_n\,(\log n/n)^{\frac{1}{d+4}}$, where $a_n \to a_0 > 0$, then*

$$(\log_k n)^{-1}\,(n/\log n)^{\frac{2}{d+4}} \sup_{x\in\mathcal{I}} |\hat f_w(x) - f(x)| \longrightarrow 0, \;\; a.s., \quad \forall\; k > 0, \; integer.$$

Proof of Lemma 4.1. For convenience, we shall give the proof here only for the case $d = 1$, i.e., when $\dot\mu_h(x)$ is one dimensional. For multidimensional case, the result can be proved by using linear combination of its components instead of $\dot\mu_h(x)$, and applying the same argument.

Let $s_{ni} := \int K_h(x - X_i)\varepsilon_i \dot\mu_h(x) d\varphi(x)$, and rewrite

$$n^{1/2} S_n = n^{-1/2} \sum_{i=1}^{n} s_{ni}.$$

Note that $\{s_{ni},\, 1 \le i \le n\}$ are i.i.d. centered r.v.'s for each n. By the L-F C.L.T., it suffices to show that as $n \to \infty$,

$$Es_{n1}^2 \to \Sigma, \tag{4.13}$$

$$E\left\{ s_{n1}^2 I(|s_{n1}| > n^{1/2}\eta) \right\} \longrightarrow 0, \qquad \forall\, \eta > 0. \tag{4.14}$$

But, by Fubini,

$$Es_{n1}^2 \;=\; \int\!\!\int EK_h(x - X)K_h(y - X)\sigma^2(X)\dot\mu_h(x)\dot\mu_h(y) d\varphi(x) d\varphi(y).$$

By the transformation $x - z = uh$, $y - z = vh$, $z = t$, taking the limit, and using the assumed continuity of σ^2, f, and g, we obtain

$$\begin{aligned}
\Sigma &= \lim_{h\to 0} \int\!\!\int\!\!\int K(u)K(v)\sigma^2(t)\dot\mu_h(x+uh)\dot\mu_h(x+vh)f(x) \\
&\qquad\qquad \times \frac{g(x+uh)g(x+vh)}{f^2(x+uh)f^2(x+vh)}\, du\, dv\, dx \\
&= \int \frac{\sigma^2(x)\dot m^2(x,\theta_0)g^2(x)}{f(x)}\, dx.
\end{aligned}$$

Hence (4.13) is proved.

To prove (4.14), note that by the Hölder inequality, the L.H.S. of (4.14) is bounded above by

$$\begin{aligned}
\lambda_d^{-\delta} n^{-\delta/2} E(s_{n1})^{2+\delta} &\le \lambda_d^{-\delta}\, n^{-\delta/2} E\left[\left(\int (K_h(x - X)\dot\mu_h(x))^{\frac{2+\delta}{2}} d\varphi(x) \right)^2 |\varepsilon|^{2+\delta} \right] \\
&= O((nh^d)^{-\delta/2}) = o(1),
\end{aligned}$$

thereby proving (4.14).

To prove (4.10), by the Cauchy-Schwarz inequality, the boundedness of $\dot\mu_h(x)$, (3.4), and by Lemma 4.3, we obtain

$$\begin{aligned}
nS_{n1}^2 &\le Cn \int (U_n(x)\dot\mu_h(x))^2 d\varphi(x) \sup_{x\in\mathcal{I}} \left| f^2(x)/\hat f_w^2(x) - 1 \right|^2 \\
&= n\, O_p((nh^d)^{-1})\, O_p((\log_k n)^2 (\log n/n)^{\frac{4}{d+4}}) \\
&= O_p\left((\log_k n)^2 (\log n)^{\frac{4}{d+4}}\, n^{ad - \frac{4}{d+4}} \right) = o_p(1), \qquad \text{by (h3)}.
\end{aligned}$$

This completes the proof of Lemma 4.1. □

Proof of Lemma 4.2. By the Cauchy-Schwarz inequality,

$$\left\| n^{1/2} g_{n1} \right\|^2 \;\le\; \left(n^{1/2} \int U_n^2(x) d\varphi(x) \right) \left(n^{1/2} \int \|\dot\mu_n(x,\theta_0) - \dot\mu_h(x)\|^2 d\varphi(x) \right).$$

By (3.3) and (h2),

$$En^{1/2}\int U_n^2(x)d\varphi(x) = O(n^{-1/2}h^{-d}) = o(1).$$

To handle the second factor, first note that $\dot\mu_n(x,\theta_0) - \dot\mu_h(x)$ is an average of centered i.i.d. r.v.'s. Using Fubini, and the fact that variance is bounded above by the second moment, we obtain that the expected value of the second factor of the above bound is bounded above by

$$n^{-1/2}\int E\|K_h(x-X)\dot m_{\theta_0}(x)\|^2\, d\varphi(x) = O(n^{-1/2}h^{-d}) = o(1). \tag{4.15}$$

This completes the proof of (4.11)(a). This together with (2.1) implies (4.11)(b).

To prove (c), similarly,

$$\left\|n^{1/2}g_{n3}\right\|^2 \;\;\le\;\; n\int U_n^2(x)d\varphi(x)\int\left\|\dot\mu_n(x,\hat\theta_n) - \dot\mu_n(x,\theta_0)\right\|^2 d\varphi(x).$$

But the second integral is bounded above by

$$\max_{1\le i\le n}\|\dot m_{\hat\theta_n}(X_i) - \dot m_{\theta_0}(X_i)\|^2\int\left(\hat f_h(x)\right)^2 d\varphi(x) = o_p(h^d)\times O_p(1),$$

by (4.4) and the assumption (m5), and by (3.9) applied with $\alpha\equiv 1$. This together with (3.3) proves (4.12)(c). The proof of (4.12)(d) uses (4.12)(c) and is similar to that of (4.11)(b), thereby completing the proof of the Lemma 4.2. □

Next, we shall show that the right hand side of (4.9) equals $R_n(\hat\theta_n - \theta_0)$, where

$$R_n = \Sigma_0 + o_p(1). \tag{4.16}$$

Again, recall the definitions (3.1) and (4.7). The right hand side of (4.9) can be written as the sum $W_{n1}+W_{n2}$, where

$$\begin{aligned}
V_n &:= \int\dot\mu_n(x,\hat\theta_n)\left[n^{-1}\sum_{i=1}^n K_h(x-X_i)\frac{d_{ni}}{\|u_n\|}\right] d\hat\varphi_w(x),\\
W_{n1} &:= \int\left[\dot\mu_n(x,\hat\theta_n)\, n^{-1}\sum_{i=1}^n K_h(x-X_i)d_{ni}\right] d\hat\varphi_w(x) = V_n u_n^T u_n,\\
W_{n2} &:= \int\dot\mu_n(x,\hat\theta_n)\dot\mu_n^T(x,\theta_0)\, d\hat\varphi_w(x)\, u_n = L_n\, u_n \quad \text{say,}
\end{aligned}$$

so that the right hand side of (4.9) equals $[V_n\, u_n^T + L_n]\, u_n$. But,

$$\begin{aligned}
\|V_n\| &\le \max_{1\le i\le n}\frac{|d_{ni}|}{\|u_n\|}V_{n1},\\
V_{n1} &:= \int\hat f_h(x)\,\|\dot\mu_n(x,\hat\theta_n)\| d\hat\varphi_w(x)\\
&\le \max_{1\le i\le n}\|\dot m_{\hat\theta_n}(X_i) - \dot m_{\theta_0}(X_i)\|\int\hat f_h(x)d\hat\varphi_w(x) + \int\hat f_h(x)\|\dot\mu_h(x,\theta_0)\|d\hat\varphi_w(x)\\
&= o_p(1) + O_p(1),
\end{aligned}$$

by (2.1), the assumption (m5), and by (4.4). This together with (m4) then implies that $\|V_n\| = o_p(1)$, and by the consistency of $\hat\theta_n$, we also have $\|V_n u_n^T\| = o_p(1)$.

Next, consider L_n. We have

$$\begin{aligned} L_n &= \int \dot\mu_n(x,\theta_0)[\dot\mu_n(x,\hat\theta_n) - \dot\mu_n(x,\theta_0)]^T d\hat\varphi_w(x) + \int \dot\mu_n(x,\theta_0)\dot\mu_n^T(x,\theta_0)d\hat\varphi_w(x) \\ &= L_{n1} + L_{n2}, \qquad \text{say.} \end{aligned}$$

But, by (2.1) and (m5), $\|L_{n1}\| = o_p(1)$, while

$$\begin{aligned} &\Big\| L_{n2} - \int \dot\mu_h(x,\theta_0)\dot\mu_h^T(x,\theta_0)d\hat\varphi_w(x)\Big\| \\ &\le \int \|\dot\mu_n(x,\theta_0) - \dot\mu_h(x,\theta_0)\|^2 d\hat\varphi_w(x) + 2\int \|\dot\mu_n(x,\theta_0) - \dot\mu_h(x,\theta_0)\|\, \|\dot\mu_h(x,\theta_0)\| d\hat\varphi_w(x). \end{aligned}$$

But, by (2.1) and (4.15), this upper bound is $o_p(1)$. Moreover, by usual calculations and using (2.1), one also obtains

$$\int \dot\mu_h(x,\theta_0)\dot\mu_h^T(x,\theta_0)d\hat\varphi_w(x) = \Sigma_0 + o_p(1).$$

This then proves the claim (4.16), thereby also completing the proof of Theorem 4.1. □

Remark 4.1 Upon choosing $g \equiv f$, one sees that

$$\Sigma = \int \sigma^2(x)\dot m_{\theta_0}(x)\dot m_{\theta_0}^T(x)f(x)dx, \qquad \Sigma_0 = \int \dot m_{\theta_0}(x)\dot m_{\theta_0}^T(x)f(x)dx.$$

It thus follows that in this case the asymptotic distribution of $n^{1/2}(\hat\theta_n - \theta_0)$ is the same as that of the least square estimator. This analogy is in flavor similar to the one observed by Beran (1977) when pointing out that the minimum Hellinger distance estimator in the context of density fitting problem is asymptotically like the maximum likelihood estimator.

Remark 4.2 *Choice of G.* Suppose $d = 1$ and $\sigma(x) \equiv \sigma$, a constant. Then the asymptotic variance of $n^{1/2}(\hat\theta_n - \theta_0)$ is

$$v := \sigma^2 \int \dot m_{\theta_0}^2 \frac{g^2}{f} \Big/ \Big(\int \dot m_{\theta_0}^2\, g\Big)^2.$$

Assuming $f = 0$ implies $g = 0$, by the Cauchy-Schwarz inequality we obtain

$$\Big(\int \dot m_{\theta_0}^2 g\Big)^2 = \Big(\int \dot m_{\theta_0} \frac{g}{\sqrt f}\, \dot m_{\theta_0}\sqrt f\Big)^2 \le \int \dot m_{\theta_0}^2 \frac{g^2}{f}\cdot \int \dot m_{\theta_0}^2\, f,$$

with equality if and only if $g \propto f$. This in turn implies that $v \ge \sigma^2/\int \dot m_{\theta_0}^2 f$, i.e., the lower bound on the asymptotic variance of $n^{1/2}(\hat\theta_n - \theta_0)$ is at that of the least square estimator. This in turn suggests that as far the estimator $\hat\theta_n$ is concerned one may use $g = \hat f_w$ in practice. Such an estimator would have the smallest asymptotic variance among this class of estimators as G varies. Its asymptotics can be derived using the methods developed here. A similar fact holds for $d > 1$.

Another interesting data dependent choice of G is obtained when the density $g = \hat f_w^2$ on $\mathcal{I}$. In other words, there now is no $\hat f_w^2$ in the denominator of the integrand in T_n and the integrating measure is simply dx. In this case the asymptotic theory is relatively simpler and

$$\Sigma = \int_{\mathcal I} \dot m_{\theta_0}(x)\dot m_{\theta_0}^T(x)f^3(x)dx, \quad \Sigma_0 = \int_{\mathcal I} \dot m_{\theta_0}(x)\dot m_{\theta_0}^T(x)f^2(x)dx.$$

Remark 4.3 Linear regression. Consider the linear regression model, where $q = d+1$, $\Theta = \mathbb{R}^{d+1}$, and $\mu_\theta(x) = \theta_1 + \theta_2' x$, with $\theta_1 \in \mathbb{R}$, $\theta_2 \in \mathbb{R}^d$. Because now the parameter space is not compact the above results are not directly applicable to this model. But, now the estimator has a closed expression and this regression function satisfies the conditions (m1) - (m5) trivially. The same techniques as above yield the following result.

With the notation in (3.1) and (4.1), in this case

$$
\begin{aligned}
\dot\mu_n(x,\theta) &\equiv \dot\mu_n(x) \equiv \begin{pmatrix} \hat f_h(x) \\ n^{-1}\sum_{i=1}^n K_h(x-X_i)X_i \end{pmatrix}, \qquad \dot\mu_h(x) \equiv \begin{pmatrix} EK_h(x-X) \\ EK_h(x-X)X \end{pmatrix} \\
\Sigma_0 &= \int \begin{pmatrix} 1 & x' \\ x & xx' \end{pmatrix} g(x)dx, \qquad \Sigma_n = \int \dot\mu_n(x)\dot\mu_n(x)' d\hat\varphi_w(x), \\
\Sigma &= \int \begin{pmatrix} 1 & x' \\ x & xx' \end{pmatrix} \frac{\sigma^2(x)g^2(x)}{f(x)} dx, \\
T_n(\theta) &= \int [U_n(x) - (\theta-\theta_0)'\dot\mu_n(x)]^2 d\hat\varphi_w(x).
\end{aligned}
$$

The positive definiteness of Σ_n and direct calculations thus yield

$$(\hat\theta_n - \theta_0) = \Sigma_n^{-1} \int \dot\mu_n(x)\, U_n(x) d\hat\varphi_w(x).$$

From the fact that $\Sigma_n \longrightarrow \Sigma_0$, in probability, parts (a) and (b) of Lemma 4.2, and from Lemma 4.1 applied to the linear case, we thus obtain that if (e1)-(e2), (f1), (f2), (k) and (h3) hold, if the regression function is a linear parametric function, and if $\int \|x\|^2 dG(x) < \infty$, then $n^{1/2}(\hat\theta_n - \theta_0) = \Sigma_0^{-1} \int U_n(x)\dot\mu_h(x) d\varphi(x) + o_p(1) \rightarrow_d \mathcal{N}_q(0, \Sigma_0^{-1}\Sigma\Sigma_0^{-1})$.

5 Asymptotic distribution of the minimized distance

This section contains a proof of the asymptotic normality of the minimized distance $T_n(\hat\theta_n)$. To state the result precisely, recall the definitions of $\tilde C_n, \hat C_n$, Γ, $\hat\Gamma_n$ from (1.1) and that $\tilde T_n$ is T_n with $\hat f_w$ replaced by f. The main result proved in this section is the following

Theorem 5.1 *Suppose H_0, (e1), (e2), (e4), (f1), (f2), (g), (h3), (k), and (m1)-(m5) hold. Then, $nh^{d/2}(T_n(\hat\theta_n) - \hat C_n) \longrightarrow_d \mathcal{N}_1(0,\Gamma)$. Moreover, $|\hat\Gamma_n \Gamma^{-1} - 1| = o_p(1)$.*

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

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

Lemma 5.1 *If H_0, (e1), (e2), (e4), (f1), (g), (h1), (h2), and (k) hold, then*

$$nh^{d/2}(\tilde T_n(\theta_0) - \tilde C_n) \longrightarrow_d \mathcal{N}_1(0,\Gamma).$$

Lemma 5.2 *Suppose H_0, (e1), (e2), (f1), (k), (m3) - (m5), (h1), and (h2) hold. Then*

$$nh^{d/2}\left| T_n(\hat\theta_n) - T_n(\theta_0) \right| = o_p(1).$$

Lemma 5.3 *Under H_0, (e1), (e2), (f1), (f2), (k), (m3) - (m5), and (h3),*

$$nh^{d/2}\left|T_n(\theta_0) - \tilde{T}_n(\theta_0)\right| = o_p(1).$$

Lemma 5.4 *Under the same conditions as in Lemma 5.3,*

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

Lemma 5.5 *Under the same conditions as in Lemma 5.2, $\hat{\Gamma}_n - \Gamma = o_p(1)$. Consequently, $\Gamma > 0$ implies $|\hat{\Gamma}_n\Gamma^{-1} - 1| = o_p(1)$.*

The proof of the Lemma 5.1 is facilitated by Theorem 1 of Hall (1984) which is reproduced here for the sake of completeness.

Theorem 5.2 *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

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

If

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

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

Proof of Lemma 5.1. Note that $\tilde{T}_n(\theta_0)$ can be written as the sum of $\tilde{C}_n$ and M_{n2}, where

$$M_{n2} \;=\; n^{-2}\sum_{i\ne j}\int K_h(x - X_i)K_h(x - X_j)\,\varepsilon_i\varepsilon_j\, d\varphi(x),$$

so that $\tilde{T}_n(\theta_0) - \tilde{C}_n = nh^{d/2}M_{n2}$. Let

$$\Gamma_n := 2h^d \int\int \big[EK_h(x - X)K_h(y - X)\sigma^2(X)\big]^2\, d\varphi(x)d\varphi(y).$$

It suffices to prove that

$$\Gamma_n \longrightarrow \Gamma, \quad \text{and} \quad \Gamma_n^{-1}nh^{d/2}M_{n2} \longrightarrow_d \mathcal{N}_1(0,1). \tag{5.1}$$

Apply Theorem 5.2 to $\tilde{X}_i = (X_i^T, \varepsilon_i)^T$ and

$$H_n(\tilde{X}_i, \tilde{X}_j) \;=\; n^{-1}h^{d/2}\int K_h(x - X_i)K_h(x - X_j)\,\varepsilon_i\varepsilon_j d\varphi(x),$$

so that

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

Observe that this $H_n(\tilde X_1, \tilde X_2)$ is symmetric, $E(H_n(\tilde X_1, \tilde X_2)|\tilde X_1) = 0$, and

$$
\begin{aligned}
&EH_n^2(\tilde X_1, \tilde X_2)\\
&= n^{-2}h^d \int\int \big[EK_h(x-X_1)K_h(y-X_1)\sigma^2(X_1)\,\big]^2 \,d\varphi(x)d\varphi(y)\\
&= (n^2h^d)^{-1}\int\int \Big[\int K(u)K(\frac{y-x}{h}+u)\sigma^2(x-uh)f(x-uh)du\,\Big]^2 \,d\varphi(x)d\varphi(y)\\
&< \infty, \qquad \text{for each } n \ge 1.
\end{aligned}
$$

Hence, in view of Theorem 5.2, we only need to show that

$$EG_n^2(\tilde X_1, \tilde X_2)\Big/\Big[EH_n^2(\tilde X_1, \tilde X_2)\Big]^2 = o(1), \tag{5.2}$$

$$n^{-1}EH_n^4(\tilde X_1, \tilde X_2)\Big/\Big[EH_n^2(\tilde X_1, \tilde X_2)\Big]^2 = o(1). \tag{5.3}$$

To prove (5.2) and (5.3), it suffices to prove the following three results:

$$EG_n^2(\tilde X_1, \tilde X_2) = O(n^{-4}h^d), \tag{5.4}$$

$$EH_n^4(\tilde X_1, \tilde X_2) = O(n^{-4}h^{-d}), \tag{5.5}$$

$$EH_n^2(\tilde X_1, \tilde X_2) = O(n^{-2}). \tag{5.6}$$

To prove (5.4), write a $t \in \mathbb{R}^{d+1}$ as $t^T = (t_1^T, t_2)$, with $t_1 \in \mathbb{R}^d$. Then, for any $t, s \in \mathbb{R}^{d+1}$,

$$
\begin{aligned}
G_n(t,s) &= n^{-2}h^d\int\int K_h(x-t_1)K_h(z-s_1)t_2s_2\\
&\qquad \times E\Big[K_h(x-X_1)K_h(z-X_1)\sigma^2(X_1)\Big]\,d\varphi(x)d\varphi(z).
\end{aligned}
$$

For the sake of brevity write $d\varphi_{xzwv} = d\varphi(x)d\varphi(z)d\varphi(w)d\varphi(v)$, and

$$
\begin{aligned}
EK_h(x-X_1)K_h(z-X_1)\sigma^2(X_1) &= \int K_h(x-t)K_h(z-t)\sigma^2(t)f(t)dt\\
&= h^{-d}\int K(u)K(\frac{z-x}{h}+u)\sigma^2(x-uh)f(x-uh)du\\
&= B_h(z-x), \qquad \text{say.}
\end{aligned}
$$

Then, by expanding square of the integrals and changing the variables, one obtains that

$$
\begin{aligned}
EG_n^2(\tilde X_1, \tilde X_2) &= n^{-4}h^{2d}\int\int\int\int B_h(x-w)B_h(z-x)B_h(z-v)B_h(v-w)\;d\varphi_{xzwv}\\
&= O(n^{-4}h^d).
\end{aligned}
$$

This proved (5.4). Similarly, one obtains

$$
\begin{aligned}
&EH_n^4(\tilde X_1, \tilde X_2)\\
&= n^{-4}h^{2d}E\Big(\int K_h(x-X_1)K_h(x-X_2)\varepsilon_1\varepsilon_2\,d\varphi(x)\,\Big)^4\\
&= n^{-4}h^{2d}\int\int\int\int \big(EK_h(x-X_1)K_h(y-X_1)K_h(s-X_1)K_h(t-X_1)\sigma^4(X_1)\big)^2\,d\varphi_{xyst}\\
&= O(n^{-4}h^{-d}),
\end{aligned}
$$

and

$$
\begin{aligned}
& EH_n^2(\tilde{X}_1, \tilde{X}_2) \\
= \;& n^{-2}h^d E\int\int K_h(x-X_1)K_h(x-X_2)K_h(y-X_1)K_h(y-X_2)\varepsilon_1^2\varepsilon_2^2 \; d\varphi(x)d\varphi(y) \\
= \;& n^{-2}h^d \int\int \big[EK_h(x-X_1)K_h(y-X_1)\sigma^2(X_1)\big]^2 \; d\varphi(x)d\varphi(y) = n^{-2}(\Gamma_n/2) \\
= \;& O(n^{-2}),
\end{aligned}
\tag{5.7}
$$

thereby verifying (5.5) and (5.6). This completes the proof of (5.1).

By (5.7),

$$
\begin{aligned}
& (1/2)n^2 EH_n^2(\tilde{X}_1, \tilde{X}_2) = \Gamma_n/4 \\
& = (1/2)h^d \int\int \Big(\int K(u)h^{-d}K(\frac{y-x}{h} + u)\sigma^2(x-uh)f(x-uh)\Big)^2 \; d\varphi(x)d\varphi(y) \\
& \longrightarrow (1/2)\int(\sigma^2(x))^2 g(x)d\varphi(x) \int(\int K(u)K(v+u)du)^2 \, dv = \Gamma/4,
\end{aligned}
$$

by the continuity of σ^2 and f. This complete the proof of Lemma 5.1. □

Remark 5.1 Let $e_n := E\int K_h^2(x-X_1)\,\varepsilon_1^2\, d\varphi(x)$. Note that $E\tilde{T}_n(\theta_0) = n^{-1}e_n$. Then, by routine calculations,

$$
\begin{aligned}
& E\Big(nh^{d/2}(\tilde{C}_n - E\tilde{T}_n(\theta_0))\Big)^2 \\
= \;& E\left(n^{-1}h^{d/2}\sum_{i=1}^n \left[\int K_h^2(x-X_i)\varepsilon_i^2 \, d\varphi(x) - e_n\right]\right)^2 \\
\le \;& n^{-1}h^d E\left(\int K_h^2(x-X_1)\varepsilon_i^2 \, d\varphi(x)\right)^2 \\
= \;& n^{-1}h^d E\left[\left(\int K_h^2(x-X_1)\, d\varphi(x)\right)^2 \varepsilon_1^4\right] = O((nh^d)^{-1}) = o(1).
\end{aligned}
$$

Combining this with Lemma 5.1, one obtains $nh^{d/2}(\tilde{T}_n(\theta_0) - E\tilde{T}_n(\theta_0)) \longrightarrow_d \mathcal{N}_1(0, \Gamma)$.

Proof of Lemma 5.2. Recall the definitions of U_n and Z_n from (3.1). Add and subtract $m_{\theta_0}(X_i)$ to the i^{th} summand inside the squared integrand of $T_n(\hat{\theta}_n)$, to obtain that

$$
\begin{aligned}
T_n(\theta_0) - T_n(\hat{\theta}_n) \;&=\; 2\int U_n(x)Z_n(x,\hat{\theta}_n)\, d\hat{\varphi}_w(x) - \int Z_n^2(x,\hat{\theta}_n)\, d\hat{\varphi}_w(x) \\
&=\; 2Q_1 - Q_2, \qquad \text{say.}
\end{aligned}
$$

It thus suffices to show that

$$
(i) \quad nh^{d/2}Q_1 = o_p(1), \qquad (ii) \quad nh^{d/2}Q_2 = o_p(1). \tag{5.8}
$$

By subtracting and adding $(\hat{\theta}_n - \theta_0)^T\dot{m}_{\theta_0}(X_i)$ to the i^{th} summand of $Z_n(x,\hat{\theta}_n)$, we can rewrite

$$
\begin{aligned}
Q_1 \;&=\; \int U_n(x)\left[n^{-1}\sum_{i=1}^n K_h(x-X_i)d_{ni}\right] d\hat{\varphi}_w(x) + (\hat{\theta}_n - \theta_0)^T \int U_n(x)\dot{\mu}_n(x,\theta_0)d\hat{\varphi}_w(x) \\
&=\; Q_{11} + Q_{12}, \qquad \text{say,}
\end{aligned}
$$

where d_{ni} are as in (4.7). By (4.4), for every $\eta > 0$, there is a $k < \infty$, $N < \infty$, such that $P(A_n) \geq 1 - \eta$, for all $n > N$, where $A_n := \{(nh^d)^{1/2}\|\hat{\theta}_n - \theta_0\| < k\}$. By the Cauchy-Schwarz inequality, (2.1), (3.4) and the fact that

$$\int \hat{f}_h^2(x)d\hat{\varphi}_w(x) = O_p(1), \tag{5.9}$$

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

$$n^{1/2}\|\hat{\theta}_n - \theta_0\|(nh^d)^{1/2} \cdot O_p((nh^d)^{-1/2}) \cdot \max_i \frac{|d_{ni}|}{\|\hat{\theta}_n - \theta_0\|}.$$

But, by (m4), on A_n, the last factor in this bound is $o_p(1)$, and in view of Theorem 4.1, this entire bound in turn is $o_p(1)$. Hence, to prove (5.8)(i), it remains to prove that $nh^{d/2}|Q_{12}| = o_p(1)$.

But Q_{12} can be rewritten as

$$\begin{aligned}&(\hat{\theta}_n - \theta_0)^T \int U_n(x)\dot{\mu}_n(x,\hat{\theta}_n)d\hat{\varphi}_w(x) - (\hat{\theta}_n - \theta_0)^T \int U_n(x)\left[\dot{\mu}_n(x,\hat{\theta}_n) - \dot{\mu}_n(x,\theta_0)\right]d\hat{\varphi}_w(x)\\ &= Q_{121} - Q_{122}, \qquad \text{say.}\end{aligned}$$

Arguing as above, on the event A_n, $(nh^{d/2}|Q_{122}|)^2$ is bounded above by

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

by (2.1), (3.4), (5.9), and assumptions (m5) and (h2).

Next, note that the integral in Q_{121} is the same as the expression in the left hand side of (4.9). Thus, it is equal to

$$\begin{aligned}&(\hat{\theta}_n - \theta_0)^T \int Z_n(x,\hat{\theta}_n)\dot{\mu}_n(x,\hat{\theta}_n)d\hat{\varphi}_w(x)\\ &= (\hat{\theta}_n - \theta_0)^T \int Z_n(x,\hat{\theta}_n)\dot{\mu}_n(x,\theta_0)d\hat{\varphi}_w(x)\\ &\quad +(\hat{\theta}_n - \theta_0)^T \int Z_n(x,\hat{\theta}_n)\left[\dot{\mu}_n(x,\hat{\theta}_n) - \dot{\mu}_n(x,\theta_0)\right]d\hat{\varphi}_w(x)\\ &= D_1 + D_2, \qquad \text{say,}\end{aligned} \tag{5.10}$$

But, by the Cauchy-Schwarz inequality, (2.1), (3.16), and (5.9), $nh^{d/2}|D_1|$ is bounded above by

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

by Theorem 4.1 and the assumption (m5) and (h2). Similarly, one shows $nh^{d/2}|D_2|$ is bounded above by

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

This completes the proof of (5.8)(i).

The proof of (5.8)(ii) similar. Details are left out for the sake of brevity. □

Proof of Lemma 5.3. Note that

$$\begin{aligned}nh^{d/2}|T_n(\theta_0) - \tilde{T}_n(\theta_0)| &\leq nh^{d/2}\int U_n^2(x)d\varphi(x)\sup_{x\in\mathcal{I}}|f^2(x)/\hat{f}_w^2(x) - 1|\\ &= nh^{d/2}O_p((nh^d)^{-1})O_p((\log_k n)(\log n/n)^{\frac{d}{d+4}}) = o_p(1),\end{aligned}$$

by (3.3) and Lemma 4.3. Hence the lemma. □

Proof of Lemma 5.4. Let

$$t_i = m_{\hat\theta_n}(X_i) - m_{\theta_0}(X_i), \quad \Delta_w(x) := f^2(x)\left(\hat f_w^{-2}(x) - f^{-2}(x)\right).$$

Then,

$$\begin{aligned}\hat C_n &= n^{-2}\sum_{i=1}^n \int K_h^2(x-X_i)(\varepsilon_i - t_i)^2 d\hat\varphi_w(x)\\ &= n^{-2}\sum_{i=1}^n \int K_h^2(x-X_i)(\varepsilon_i - t_i)^2 d\varphi(x) + n^{-2}\sum_{i=1}^n \int K_h^2(x-X_i)(\varepsilon_i - t_i)^2\Delta_w(x) d\varphi(x)\\ &= A_{n1} + A_{n2}, \qquad \text{say.}\end{aligned}$$

In order to prove the lemma it suffices to prove that

$$(5.11) \qquad (a) \quad nh^{d/2}(A_{n1} - \tilde C_n) = o_p(1), \quad \text{and} \quad (b) \quad nh^{d/2}A_{n2} = o_p(1).$$

By expanding the quadratic term in the integrand, A_{n1} can be written as the sum of $\tilde C_n$, A_{n12}, and A_{n13}, where

$$A_{n12} = n^{-2}\sum_{i=1}^n \int K_h^2(x-X_i)t_i^2 d\varphi(x), \quad A_{n13} = -2n^{-2}\sum_{i=1}^n \int K_h^2(x-X_i)\varepsilon_i t_i d\varphi(x).$$

But $|A_{n12}| \le \max_{1\le i\le n}|t_i|^2 n^{-2}\sum_{i=1}^n \int K_h^2(x-X_i)d\varphi(x)$. By (m4) and (4.4), one obtains that $\max_{i\le n}|t_i|^2 = O_p((nh^d)^{-1})$. Moreover, by the usual calculation, one obtains that

$$n^{-2}\sum_{i=1}^n \int K_h^2(x-X_i)d\varphi(x) = O_p((nh^d)^{-1}).$$

Hence,

$$|A_{n12}| = O_p((nh^d)^{-1})O_p((nh^d)^{-1}) = O_p((nh^d)^{-2}).$$

Similarly,

$$\begin{aligned}|A_{n13}| &\le 2\max_{i\le n}|t_i| n^{-2}\sum_{i=1}^n \int K_h^2(x-X_i)|\varepsilon_i| d\varphi(x)\\ &= O_p((nh^d)^{-1/2})O_p((nh^d)^{-1}) = O_p((nh^d)^{-3/2}).\end{aligned}$$

Hence

$$\begin{aligned}|nh^{d/2}(A_{n1} - \tilde C_n)| &= nh^{d/2}\left(O_p((nh^d)^{-2}) + O_p((nh^d)^{-3/2})\right)\\ &= O_p((nh^{-3d/2})^{-1}) + O_p((nh^{2d})^{-1/2}) = o_p(1).\end{aligned}$$

To prove the part (b) of (5.11), note that A_{n2} can be written as the sum of A_{n21}, A_{n22}, and A_{n23}, where

$$A_{n21} = n^{-2}\sum_{i=1}^n \int K_h^2(x-X_i)\varepsilon_i^2\Delta_w(x)d\varphi(x),$$

$$\begin{aligned} A_{n22} &= n^{-2}\sum_{i=1}^{n}\int K_h^2(x-X_i)t_i^2\Delta_w(x)d\varphi(x), \\ A_{n23} &= -2n^{-2}\sum_{i=1}^{n}\int K_h^2(x-X_i)\varepsilon_i t_i\Delta_w(x)d\varphi(x). \end{aligned}$$

By taking the expected value and the usual calculation, one obtains that

$$n^{-2}\sum_{i=1}^{n}\int K_h^2(x-X_i)\varepsilon_i^2 d\varphi(x) = O_p((nh^d)^{-1}).$$

Hence

$$\begin{aligned} |nh^{d/2}A_{n21}| &\leq \sup_{x\in\mathcal{I}}|\Delta_w(x)|n^{-2}\sum_{i=1}^{n}\int K_h^2(x-X_i)\varepsilon_i^2 d\varphi(x) \\ &= nh^{d/2}O_p(\log_k n\,(\log n/n)^{\frac{2}{d+4}})O_p((nh^d)^{-1}) \\ &= O_p(h^{-d/2}\log_k n\,(\log n/n)^{\frac{2}{d+4}}) = o_p(1), \end{aligned}$$

by Lemma 4.3 and (2.1). Similarly, one obtains that

$$\begin{aligned} |nh^{d/2}A_{n22}| &\leq \sup_{x\in\mathcal{I}}|\Delta_w(x)|\max_{1\leq i\leq n}|t_i|^2 n^{-2}\sum_{i=1}^{n}\int K_h^2(x-X_i)d\varphi(x) \\ &= nh^{d/2}O_p(\log_k n\,(\log n/n)^{\frac{2}{d+4}})O_p((nh^d)^{-1})O_p((nh^d)^{-1}) \\ &= o_p((nh^{3d/2})^{-1}) = o_p(1), \end{aligned}$$

and

$$\begin{aligned} |nh^{d/2}A_{n23}| &\leq 2\sup_{x\in\mathcal{I}}|\Delta_w(x)|\max_{i\leq n}|t_i|n^{-2}\sum_{i=1}^{n}\int K_h^2(x-X_i)|\varepsilon_i|d\varphi(x) \\ &= nh^{d/2}O_p(\log_k n\,(\log n/n)^{\frac{2}{d+4}})O_p((nh^d)^{-1/2})O_p((nh^d)^{-1}) \\ &= o_p((nh^{2d})^{-1/2}) = o_p(1), \end{aligned}$$

thereby completing the proof of the part (b) of (5.11), and hence that of the lemma. □.

Proof of Lemma 5.5. Define

$$\begin{aligned} \tilde{\Gamma}_n &:= h^d n^{-2}\sum_{i\neq j}\left(\int K_h(x-X_i)K_h(x-X_j)\varepsilon_i\varepsilon_j d\varphi(x)\right)^2 = \sum_{i\neq j}H_n^2(\tilde{X}_i,\tilde{X}_j), \\ \Delta_h(x) &:= f^2(x)(\hat{f}_h^{-2}(x)-f^2(x)). \end{aligned}$$

We shall first prove

$$\hat{\Gamma}_n - \tilde{\Gamma}_n = o_p(1), \tag{5.12}$$

$$\tilde{\Gamma}_n - \Gamma_n = o_p(1). \tag{5.13}$$

The claim of this lemma follows from these results and the fact that $\Gamma_n \longrightarrow \Gamma$.

For the sake of convenience, write $K_h(x - X_i)$ by $K_i(x)$. Now, rewrite $\hat{\Gamma}_n$ as the sum of the following terms:

$$\begin{aligned}
B_1 &= h^d n^{-2} \sum_{i \neq j} \left(\int K_i(x) K_j(x) (\varepsilon_i - t_i)(\varepsilon_j - t_j) d\varphi(x) \right)^2, \\
B_2 &= h^d n^{-2} \sum_{i \neq j} \left(\int K_i(x) K_j(x) (\varepsilon_i - t_i)(\varepsilon_j - t_j) \Delta_h(x) d\varphi(x) \right)^2, \\
B_3 &= -2 \frac{h^d}{n^2} \sum_{i \neq j} \left(\int K_i(x) K_j(x) (\varepsilon_i - t_i)(\varepsilon_j - t_j) d\varphi(x) \right) \\
&\qquad \times \left(\int K_i(x) K_j(x) (\varepsilon_i - t_i)(\varepsilon_j - t_j) \Delta_h(x) d\varphi(x) \right).
\end{aligned}$$

In order to prove (5.12), it suffices to prove that

$$B_1 - \tilde{\Gamma}_n = o_p(1), \qquad B_2 = o_p(1), \qquad \text{and} \qquad B_3 = o_p(1). \tag{5.14}$$

By taking the expected value, Fubini, and usual calculation one obtains that

$$h^d n^{-2} \sum_{i \neq j} \left(\int K_i(x) K_j(x) |\varepsilon_i| |\varepsilon_j| d\varphi(x) \right)^2 = O_p(1), \tag{5.15}$$

$$h^d n^{-2} \sum_{i \neq j} \left(\int K_i(x) K_j(x) |\varepsilon_i| d\varphi(x) \right)^2 = O_p(1), \tag{5.16}$$

$$h^d n^{-2} \sum_{i \neq j} \left(\int K_i(x) K_j(x) d\varphi(x) \right)^2 = O_p(1). \tag{5.17}$$

Furthermore,

$$\sup_{x \in \mathcal{I}} |\Delta_h(x)| = o_p(1), \qquad \text{by (2.1)}, \tag{5.18}$$

$$\max_{i \leq i \leq n} |t_i| = o_p(1). \qquad \text{by (m4) and (4.4)}. \tag{5.19}$$

Note that by expanding $(\varepsilon_i - t_i)(\varepsilon_j - t_j)$ and the quadratic terms, $|B_1 - \tilde{\Gamma}_n|$ is bounded above by the sum of B_{12} and B_{13}, where

$$\begin{aligned}
B_{12} &= h^d n^{-2} \sum_{i \neq j} \left(\int K_i(x) K_j(x) (|t_i t_j| + |\varepsilon_i t_i| + |t_i \varepsilon_j|) d\varphi(x) \right)^2, \\
B_{13} &= h^d n^{-2} \sum_{i \neq j} \left(\int K_i(x) K_j(x) |\varepsilon_i \varepsilon_j| d\varphi(x) \right) \\
&\qquad \times \left(\int K_i(x) K_j(x) (|t_i t_j| + |\varepsilon_i t_i| + |t_i \varepsilon_j|) d\varphi(x) \right).
\end{aligned}$$

But $B_{12} = o_p(1)$ by (5.16), (5.17), (5.19), and the fact that $\{t_i\}$ are free of x. It further implies that $B_{13} = o_p(1)$ by (5.15) and applying the Cauchy-Schwarz inequality to the double sum. Hence $|B_1 - \tilde{\Gamma}_n| = o_p(1)$.

Note that

$$\begin{aligned} B_2 &\leq \sup_{x\in\mathcal{I}} |\Delta_h(x)| h^d n^{-2} \sum_{i\neq j} \left(\int K_i(x)K_j(x)|\varepsilon_i - t_i||\varepsilon_j - t_j| d\varphi(x) \right)^2 \\ &= o_p(1)O_p(1) = o_p(1), \end{aligned}$$

by the inequality

$$|\varepsilon_i - t_i||\varepsilon_j - t_j| \leq |\varepsilon_i\varepsilon_j| + (|t_i t_j| + |\varepsilon_i t_i| + |t_i\varepsilon_j|),$$

and expanding the quadratic terms, and by (5.18), (5.15), and the result that B_{12} and B_{13} are both $o_p(1)$. Finally, again an application of the Cauchy-Schwarz inequality to the double sum yields $B_3 = o_p(1)$. This completes the proof of (5.14), and hence that of (5.12).

To prove (5.13), note that $\Gamma_n = E\tilde{\Gamma}_n$. Hence

$$\begin{aligned} E\left(\tilde{\Gamma}_n - \Gamma_n\right)^2 &\leq \sum_{i\neq j} EH_n^4(\tilde{X}_i,\tilde{X}_j) + c\sum_{i\neq j\neq k} EH_n^2(\tilde{X}_i,\tilde{X}_j)H_n^2(\tilde{X}_j,\tilde{X}_k) \\ &\leq (n^2 + cn^3)EH_n^4(\tilde{X}_1,\tilde{X}_2) \end{aligned}$$

for some constant c by expanding the quadratic terms and the fact that the variance is bounded above by the second moment. But by (5.5), this upper bound is $O((nh^d)^{-1}) = o(1)$. Hence (5.13) is proved, and so is the Lemma 5.5. □

Remark 5.2 *Choice of G.* The choice of G in connection with $\hat{\theta}_n$ was discussed in Remark 4.2. As far as the MD test statistic $T_n(\hat{\theta}_n)$ is concerned, the choice of G will depend on the alternatives. In a simulation study in the next section we simulated the power of the test corresponding to $g = \hat{f}_w^2$ and found this test competes well with some other tests against the selected alternatives.

Another data dependent choice would be the empirical d.f. G_n of $X_1, \cdots, X_n$. But to avoid edge effects, one may use the empirical of the middle $100(1-\alpha_n)$ percent of the data only, where α_n is another sequence of window widths. The asymptotics of such a statistic needs to be investigated separately.

6 Simulations

This section reports on two simulation studies. The first investigates the behavior of the empirical size and power of the MD test statistic $D_n := nh(T_n(\hat{\theta}_n) - \tilde{C}_n)$ at 4 alternatives in the case $d = 2$, and when fitting the linear model $\theta' x$, θ, $x \in \mathbb{R}^2$. In this simulation two types of designs are considered. In one the two coordinates of X are i.i.d. normal and in the second, bivariate normal with correlation 0.36.

In the second simulation study a comparison of the Monte Carlo levels and powers of an MD test is made with two other tests based on a partial sum process. All simulations are based on 1000 replications.

Case $d = 2$. Here the model fitted is $\theta_1 x_1 + \theta_2 x_2$ with $\theta_0 = (0.5, 0.8)^T$. In other words the model from where data is simulated has the regression function $m_{\theta_0}(x) := 0.5x_1 + 0.8x_2$. The errors have $\mathcal{N}_1(0, (0.3)^2)$ distribution.

To study the empirical size and power of the test, the following five models are chosen:

$$\begin{aligned}
&\textit{model}\,0. && Y_i = m_{\theta_0}(X_i) + \varepsilon_i,\\
&\textit{model}\,1. && Y_i = m_{\theta_0}(X_i) + 0.3(X_{1i} - 0.5)(X_{2i} - 0.2) + \varepsilon_i,\\
&\textit{model}\,2. && Y_i = m_{\theta_0}(X_i) + 0.3X_{1i}X_{2i} - 0.5 + \varepsilon_i,\\
&\textit{model}\,3. && Y_i = m_{\theta_0}(X_i) + 1.4(e^{-0.2X_{1i}^2} - e^{0.7X_{2i}^2}) + \varepsilon_i,\\
&\textit{model}\,4. && Y_i = I_{\{X_{2i}>0.2\}}X_{1i} + \varepsilon_i,
\end{aligned}$$

Design variables $\{X_i\}$ are i.i.d. bivariate normal $\mathcal{N}_2(0, \Sigma_j)$, $j = 1, 2$, with

$$\Sigma_1 = \begin{pmatrix} 0.49 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \Sigma_2 = \begin{pmatrix} 1 & 0.36 \\ 0.36 & 1 \end{pmatrix}.$$

The sample sizes used are 30, 50, 100, and 200, and the nominal level used is $\alpha = 0.05$. Data from model 0 are used to study the empirical size while from models 1 to 4 are used to study the empirical power of the test. The empirical size (power) is computed by the relative frequency of the event $(|D_n^2| > 1.96)$ where $D_n^2 := \hat{\Gamma}_n^{-1/2} nh(T_n(\hat{\theta}_n) - \hat{C}_n)$.

The bandwidths taken are $h = n^{-1/4.5}$ and $w = (\log n/n)^{1/(d+4)}$, $d = 2$. Note that the upper bound $\min\{1/2d,\, 4/(d(d+4))\}$ on the exponent a in n^{-a} of (h3) in the current case is 1/4. Hence the choice of $a = 1/4.5$. The choice of w is dictated by Lemma 4.3. The measure G is a measure with Lebesgue density $g(x) = 1$ on $[-1, 1]$, while the kernel $K(u, v) \equiv K^1(u)K^1(v)$, $K^* \equiv K^1$, with

$$K^1(u) := \frac{3}{4}(1 - u^2)I\{|u| \le 1\}. \tag{6.1}$$

Tables 1 and 2 give the empirical sizes and powers for testing model 0 against models 1 to 4 for two different design distributions. From table 1, one sees that the Monte Carlo power of the MD statistic D_n^2 is quite high against the alternative 2 for all sample sizes and the two chosen designs. Secondly, when the design vector has the covariance matrix Σ_1 (i.i.d. coordinates), one sees that the MD test performs well for samples of sizes 100 and 200 at all alternatives. For the alternatives 2 and 3, this power is above 97% even for the sample size 50. However, the Monte Carlo level seems to aproximate the asymptotic level only for the sample size of 200. From Table 2, i.e., when design coordinates have covariance matrix Σ_2 (dependent coordinates), one sees that the power and the level behavior is affected by the dependence in the design variable coordinates.

Even though the theory of the present paper is not applicable to the model 4, it was included to see the effect of the discontinuity in the regression function on the power of the MD test. The Monte Carlo power is clearly lower for $n = 30, 50$ compared to other models, but is quite good for $n = 100, 200$ at the $\mathcal{N}_1(0, \Sigma_1)$ design while the same is true at the other design for $n = 200$. Thus, not unexpectedly, the discontinuity of the regression function has an effect on this power of MD test.

Power Comparison. Here we make a comparison of the Monte Carlo level and power of one of the proposed MD tests corresponding to $g = \hat{f}_w^2$ with the two other tests based on a certain partial sum process when fitting the simple linear regression model $m_\theta(x) = \theta x$ with $\theta_0 = 0.8$. The error distribution is taken to be either $\mathcal{N}_1(0, (0.1)^2)$ or double exponential, and the design density

Table 1: *Empirical sizes and powers for testing model 0 vs. models 1 to 4 with* $X \sim \mathcal{N}_2(0, \Sigma_1)$.

	$n = 30$	n=50	n=100	n=200
model 0	0.005	0.022	0.036	0.049
model 1	0.003	0.062	0.670	0.895
model 2	0.931	0.999	1.000	1.000
model 3	0.461	0.975	1.000	1.000
model 4	0.035	0.368	0.977	1.000

Table 2: *Empirical sizes and powers for testing models 0 vs. model 1 to 4 with* $X \sim \mathcal{N}_2(0, \Sigma_2)$.

	$n = 30$	n=50	n=100	n=200
model 0	0.002	0.012	0.030	0.040
model 1	0.001	0.007	0.024	0.108
model 2	0.848	0.999	1.000	1.000
model 3	0.033	0.220	0.828	0.999
model 4	0.007	0.079	0.569	0.983

is $\mathcal{N}_1(0, (1/6)^2)$. The sample sizes used are $n = 50$, 100, 200, and 500, $\alpha = 0.05$. The three models chosen are:

$$\begin{aligned}
\mathit{model}\,1.\, Y_i &= m_{\theta_0}(X_i) + \varepsilon_i, \\
\mathit{model}\,2.\, Y_i &= m_{\theta_0}(X_i) - 1.2\, exp(-X_i^2)X_i + 0.1 + \varepsilon_i, \\
\mathit{model}\,3.\, Y_i &= m_{\theta_0}(X_i) + 0.5(X_i - 0.5)^2 - 0.3(X_i - 0.5)^3 + \varepsilon_i.
\end{aligned}$$

The three tests are those of An and Cheng (AC) (1991), Stute, Thies and Zhu (STZ) (1998), and the MD test. The data from model 1 are used to study the empirical size, and from models 2 and 3 are used to study the empirical power of these tests.

1. **STZ test**. Let θ_{lse} denote the least square estimator, $\hat{\varepsilon}_i := Y_i - \theta_{lse} X_i$, $\sigma_n^2 := n^{-1} \sum_{i=1}^n \hat{\varepsilon}_i^2$, and let

$$\begin{aligned}
A_n(x) &:= \frac{1}{n} \sum_{i=1}^n X_i^2 I(X_i \geq x), \qquad G_n(x) := \frac{1}{n} \sum_{i=1}^n I(X_i \leq x). \\
S_n(x) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big[I(X_i \leq x) - \frac{1}{n} \sum_{j=1}^n \frac{X_j X_i}{A_n(X_j)} I(X_j \leq X_i \wedge x) \Big] \hat{\varepsilon}_i.
\end{aligned}$$

The test statistic is

$$T_n^1 = \sup_{x \leq x_0} \frac{|S_n(x)|}{\sigma_n \sqrt{G_n(x_0)}},$$

where $x_0 = 99^{th}$ percentile of the sample $X_1, ..., X_n$. The limiting null distribution of T_n^1 is the same as that of $\sup_{0<t<1} |B(t)|$, where B is the standard Brownian motion. The 95^{th} percentile of this distribution is approximately equal to 2.2414.

2. **AC test**: This test statistic is $\hat{K}_n = \sup_t |\hat{K}_n(t)|$, where

$$\hat{K}_n(t) = \frac{1}{\sqrt{m\hat{\sigma}_N^2}} \sum_{i=1}^{m} \hat{\varepsilon}_i I(X_i \le t),$$

where $m = m(n)$ is a subsequence of n such that $m/n \approx 0.75$. The limiting null distribution of this statistic is the same as that of the above STZ statistic.

3. **MD test**: The K, K^* are as before. The statistic T_n when $g = \hat{f}_w^2$ is

$$T_n(\theta) \quad := \quad \int_{-1}^{1} [\frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i)(Y_i - m_\theta(X_i))]^2 dx, \quad \hat{\theta}_n := \mathrm{argmin}_\theta \, T_n(\theta).$$

Let $\varepsilon_{ni} \equiv Y_i - \hat{\theta}_n X_i$,

$$\mathcal{D}_n \quad = \quad \frac{nh^{1/2}}{\hat{\Gamma}_n^{1/2}} \Big[\int_{-1}^{1} \left(\frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i)\varepsilon_{ni} \right)^2 dx - \frac{1}{n^2} \sum_{i=1}^{n} \int_{-1}^{1} K_h^2(x - X_i)\, \varepsilon_{ni}^2 dx \Big].$$

where

$$\hat{\Gamma}_n \quad = \quad 4 \int_{-1}^{1} \left(\frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i)\varepsilon_{ni}^2 \right)^2 dx \int \left(\int K * K(v) \right)^2 dv.$$

Here $K = K^1$ and $h = n^{-1/4}$. The results of this simulation are shown in Tables 3 to 7.

The Tables 3 and 4 give the empirical sizes and powers of the three tests for testing model 1 against model 2, and for the error distributions double exponential and $N(0, (0.1)^2)$, respectively. Table 5 gives a similar data when testing model 1 against model 3 with the error distribution $N(0, (0.1)^2)$.

From Table 3, one sees that when testing model 1 vs. model 2 with the double exponential errors, the STZ test appears to have better power for $n = 50, 100, 200$, while for $n = 500$, MD test appears to have slight advantage. Note that the empirical size of these two tests is also comparable. The AC test seems to be worse of the three in this case.

From Table 4 one sees that when testing model 1 vs. model 2 with $N(0, (0.1)^2)$ errors, the power of all three tests is the same for $n = 100, 200, 500$, while the STZ test has an edge with regards to the size. From Table 5, one observes that the empirical power of the MD test is better than or equal to that of AC and STZ tests when testing model 1 vs. model 3 with $N(0, (0.1)^2)$ for $n = 100, 200, 500$. Moreover, in this case the AC test has better power than the STZ test for $n = 100, 200, 500$.

In summary, this simulation shows none of these tests dominate the other, though the STZ test seems to approximate the asymptotic size somewhat better than the MD test. The AC test appears to perform worse of the three with regards to the Monte Carlo size.

Tables 6 and 7 list the mean and standard deviation of $\hat{\theta}_n$ under H_0 with double exponential and $\mathcal{N}_1(0, (0.1)^2)$ errors, respectively. From these tables one can see that the bias in $\hat{\theta}_n$ is relatively smaller at the double exponential errors than at the normal errors for $n = 50, 100$.

Table 3: *Tests for model 1 v.s. model 2, double exponential errors*

.

	$n = 50$		n=100		n=200		n=500	
tests	size	power	size	power	size	power	size	power
AC	0.059	0.027	0.308	0.209	0.132	0.789	0.128	0.934
STZ	0.072	0.137	0.064	0.446	0.054	0.837	0.051	0.990
MD	0.012	0.111	0.043	0.406	0.045	0.824	0.050	0.999

Table 4: *Tests for model 1 against model 2, the* $N(0,(0.1)^2)$ *errors*

.

	$n = 50$		n=100		n=200		n=500	
tests	size	power	size	power	size	power	size	power
AC	0.010	0.867	0.022	0.999	0.025	1.000	0.034	1.000
STZ	0.029	0.998	0.036	1.000	0.042	1.000	0.049	1.000
MD	0.011	0.659	0.019	1.000	0.023	1.000	0.044	1.000

Table 5: *Tests for model 1 against model 3,* $N(0,(0.1)^2)$ *errors*

.

	$n = 50$		n=100		n=200		n=500	
tests	size	power	size	power	size	power	size	power
AC	0.019	0.372	0.019	0.947	0.025	1.000	0.029	1.000
STZ	0.013	0.554	0.071	0.777	0.059	0.871	0.046	1.000
MD	0.011	0.421	0.021	0.982	0.035	1.000	0.049	1.000

Table 6: *Mean and s.d.(θ_n) under model 1, double exponential errors*

.

sample size	n=50	n=100	n=200	n=500
mean	0.82	0.809	0.807	0.802
stdev	0.0963	0.0777	0.0533	0.0339

Table 7: *Mean and s.d(θ_n) under model 1, $N(0,(0.1)^2)$ errors*

.

sample size	n=50	n=100	n=200	n=500
mean	0.845	0.821	0.813	0.807
stdev	0.0957	0.0682	0.0475	0.0306

Acknowledgment: Authors wish to thank the two referees and Winfried Stute for careful reading of the paper and for some constructive comments.

7 Bibliography

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

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

Beran, R.J. (1978). An efficient and adaptive robust estimator of location. *Ann. Statist.*, **6**, 292-313.

Bosq, D. (1998). *Nonparametric statistics for stochastic processes.* Second Edition, Springer.

Cox, D., Koh, E., Wahba, G. and Yandell, B.S. (1988). Testing the parametric null model hypothesis in semiparametric partial and generalized spline models. *Ann. Statist.*, **16**, 113-119.

Diebolt, J. and Zuber, J. (1999). Goodness-of-fit tests for nonlinear heteroscedastic regression models. *Statist. Probab. Lett.* **42(1)**, 53–60.

Donoho, D.L. and Liu, R.C. (1988a). The "automatic" robustness of minimum distance functionals. *Ann. Statist.*, **16**, 552-586.

Donoho, D.L. and Liu, R.C. (1988b). Pathologies of some minimum distance estimators. *Ann. Statist.*, **16**, 587-608.

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

Eubank, R.L. and Hart, J.D. (1993). Commonality of CUSUM, von Neumann and smooth-ing based goodness-of-fit tests. *Biometrika*, **80**, 89-98.

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.

González Manteiga, W. (1990). Asymptotic normality of generalized functional estimators dependent on covariates. *J. Statist. Planning & Inference*, **24**, 377-390.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Mult. Analysis.*, **14**, 1-16.

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

Hart, J.D. (1997). *Nonparametric smoothing and lack-of-fit tests.* Springer, New York, N.Y.

Koul, H.L. (1985). Minimum distance estimation in multiple linear regression model. *Sankhyā, Ser. A.* **47, Part 1**, 57-74.

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

Ni, P. (2002). Minimum distance regression and autoregressive model fitting. Ph.D. thesis. Department of Statistics & Probability, Michigan State University.

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

Stute, W., González Manteiga, W. and Presedo Quindimil, M. (1998). Bootstrap approximations in model checks for regression. *J. Amer. Statist. Assoc.* **93**, 141–149

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

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

Wolfowitz, J. (1954). Estimation by the minimum distance method in nonparametric stochastic difference equations. *Ann. Math. Statist.*, **25**, 203–217.

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

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

Hira L. Koul
Department of Statistics
and Probability
Michigan State University
East Lansing, MI 48824-1027
U. S. A.

Pingping Ni
Department of Statistics
and Probability
Michigan State University
East Lansing, MI 48824-1027
U. S. A.