Testing the Equality of Two Linear Single-Index Models

Wei Lin and K.B.Kulasekera
Department of Mathematical Sciences
Clemson University
Clemson, SC 29634

Abstract

Comparison of two nonparametric regression models has been extensively discussed in the literature for one-dimensional covariate. Comparison problem largely remains open for completely nonparametric models for multi-dimensional covariates. We address this issue under the assumption that both models are single-index models (SIMs). We propose a test for assessing the equality of the mean functions and a test for the equality of error variances. The asymptotic normality of the test statistics are established and a simulation study is conducted to evaluate the finite-sample performance of each proposed procedure.

1 Introduction

Suppose we have two samples, following the models

$$Y_{1i} = m_1(X_{1i}) + \epsilon_{1i}, \quad i = 1, \cdots, n_1, \tag{1.1}$$

and

$$Y_{2j} = m_2(X_{2j}) + \epsilon_{2j}, \quad j = 1, \cdots, n_2, \tag{1.2}$$

where X's are iid p-dimensional random vectors, $m_1(\cdot)$ and $m_2(\cdot)$ are smooth functions, and ϵ's are independent random errors with $E(\epsilon_{ki}|X_{ki}) = 0$, $i = 1, .., n_k$, $k = 1, 2$. The comparison of two regression functions of this type has been extensively investigated in the literature for $p = 1$ (Hall and Hart 1990, King et. al. 1991, Kulasekera 1995, Kulasekera and Wang 1997, Dette and Neumeyer 2001, Neumeyer and Dette 2003 and references therein). The comparison issue remains an open problem for general p and for completely nonparametric models except for a very restrictive transformation method where one assumes the functions m_i are of the form $m_i(X) = g_i(r(X))$ for known r (Hart, 1997).

We shall address the comparison issue under the assumption that both models are Single-Index Models (SIMs) (Stoker 1986, Ichimura 1993, Hristache et al. 2001, Yin and Cook 2005). Testing whether a given model is a SIM has been addressed by Xia, et al. (2004) and Stute and Zhu (2005). Here we assume that $m_k(x) = g_k(\theta_k' x)$, $k = 1, 2$, for some index vectors θ_k and for some smooth univariate functions g_k. For identifiability purposes, θ_1 and θ_2 are assumed to be unit vectors with first nonzero elements positive. In this paper, we focus on testing

$$H_0 : \ g_1(\cdot) \equiv g_2(\cdot) \text{ and } \theta_1 = \theta_2 \quad \text{vs.} \quad H_a : \text{ Not } H_0. \tag{1.3}$$

In addition, we propose a test for the equality of the error variances, σ_1^2 and σ_2^2 for the two models.

There are two major methods of testing the equality of univariate mean functions in the literature. One is an ANOVA-type method, which compares the variability "between samples" and the variability "within samples" and rejects the null hypothesis if their difference is too large (King et.al. 1991, Dette and Neumeyer 2001, etc.) or if their ratio is too large (Yong and Bowman 1995), or simply rejects the null hypothesis for large values of the "between-sample" variability using resampling techniques (Hall and Hart 1990). The other major method is to construct a test statistic based on an appropriate norm, such as the sup norm or the L_2 norm, of the empirical process of residuals (Kulasekera 1995, Neumeyer and Dette 2003, etc.). Apart from these, Bowman and Young (1996) use a graphical method and Cabus (2000) uses a U-process method to test the equality of the two regression functions. Although the covariate is in a multivariate setting, the effective predictor dimension of a SIM is actually one. However, testing the equality of two SIMs is more complicated than testing the equality of two single-covariate models due to the fact that one has to estimate the index vectors. In this paper we shall apply an ANOVA-type method similar to that proposed by Dette and Neumeyer (2001).

Kulasekera and Lin (2005) show that, for example for (1.1) above with a Single Index mean function, a properly weighted sum of squared errors,

$$\hat{d}_1(\hat{\theta}_1) = \inf_{\alpha} \sum_{i=1}^{n_1} w_{1i} [Y_{1i} - \hat{g}_{\alpha}(\alpha' X_{1i})]^2, \tag{1.4}$$

is a CAN estimator for the error variance σ_1^2, where $\hat{g}_{\alpha}(\alpha' X_{1i})$ is a kernel estimator of the mean function $E[Y_{1i}|\alpha' X_{1i}]$ for a given index vector α and w_{1i}'s are suitable weights (see Section 2). Our motivation for the construction of the test statistic for equality of the two mean functions stems from the fact that, in the two-sample case, under H_0 above, we have the option of constructing the estimator $\hat{g}_{\alpha}(\alpha' X)$of the mean function $E[Y|\alpha' X]$ using combined samples as well as using each individual sample and, the three estimators should not be very different for large samples. Therefore, under H_0, the values of the minimized sums of squared errors in (1.4) should be close to the corresponding error variance σ_1^2 when using an estimator of the mean function based only on sample 1 or based on the combined samples. However, due to the bias introduced by the difference in the two mean functions, under H_a, resulting minimized sums of squared error will be systematically higher than the corresponding error variance when the combined sample estimator of the mean function is used instead of the first sample estimator. We can show that, if the mean functions are smooth enough

(r-time differentiable, $r \geq 2$), the difference in the two versions of the minimized sums of squares, after being properly standardized, is asymptotically normal under H_0 and, will diverge to infinity under H_a. Hence an asymptotic decision rule can be suitably devised.

We also propose a test for assessing the equality of the error variances of the two models. For this purpose, it is clearly reasonable to compare the two consistent estimators of the variances, constructed from each sample. It is an immediate consequence of Kulasekera and Lin (2005) that their difference, after being multiplied by $\sqrt{N}, N = n_1 + n_2$, is asymptotically normal when $\sigma_1^2 = \sigma_2^2$ and will diverge to infinity with rate $\sqrt{N}$ otherwise.

The remainder of the paper is organized as follows. In Section 2, we propose a test for hypotheses (1.3) and give the asymptotic results. In Section 3, we give a test for the equality of error variances. A small simulation study is provided in Section 4 and technical proofs are differed to Section 5.

2 Testing Equality of Regression Functions

2.1 Construction of Test Statistic

In this section we propose a test for hypotheses (1.3). We introduce some notation first. Let S_X be the domain of X. For every α, let c_α and $2w_\alpha$ denote the center and width of the set $\{\alpha' x \mid x \in S_X\}$. Fix a constant q close to (but less than) 1 as a width control parameter and let $q_\alpha = q \cdot w_\alpha$. For any function $L(\cdot)$ supported on $(-1, 1)$, define

$$L_{q,\alpha}(x) = L\Big(\frac{\alpha' x - c_\alpha}{q_\alpha}\Big),$$

which serves as a weight function to exclude the points on the boundary of S_X along the direction of vector α. Define index sets $\mathcal{H}'_1 = \{1, 2, \cdots, n_1\}$ and $\mathcal{H}'_2 = \{1, 2, \cdots, n_2\}$. Define $g_{k\alpha}(t) = E(Y_{ik} | \alpha' X = t)$ for the kth group, $k = 1, 2$, which can be consistently estimated by

$$\hat{g}_{k\alpha}(t) = \frac{\sum_{j \in \mathcal{H}'_k} Y_{kj} K_h(\alpha' X_{kj} - t)}{\sum_{j \in \mathcal{H}'_k} K_h(\alpha' X_{kj} - t)}. \tag{2.1}$$

A corresponding pooled-sample version, which is consistent only under H_0, is given by

$$\hat{g}_\alpha(t) = \frac{\sum_{k=1}^{2} \sum_{i=1}^{n_k} Y_{ki} K_h(\alpha' X_{ki} - t)}{\sum_{k=1}^{2} \sum_{i=1}^{n_k} K_h(\alpha' X_{ki} - t)}. \tag{2.2}$$

Now, define weighted average squared-errors for estimator $\hat{g}_{k\alpha}$ by

$$\hat{d}_k(\alpha) = \frac{\sum_{i \in \mathcal{H}'_k} \Big(Y_{ki} - \hat{g}_{k\alpha}(\alpha' X_{ki})\Big)^2 L_{q,\alpha}(X_{ki})}{\sum_{i \in \mathcal{H}'_k} L_{q,\alpha}(X_{ki})}, \quad k = 1, 2;$$

and for $\hat{g}_\alpha$ by

$$\hat{d}_{pk}(\alpha) = \frac{\sum_{i \in \mathcal{H}'_k} \Big(Y_{ki} - \hat{g}_{\alpha}(\alpha' X_{ki})\Big)^2 L_{q,\alpha}(X_{ki})}{\sum_{i \in \mathcal{H}'_k} L_{q,\alpha}(X_{ki})}, \quad k = 1, 2.$$

Let $\hat{d}_k = \inf_\alpha \hat{d}_k(\alpha)$ and $\hat{d}_{pk} = \inf_\alpha \hat{d}_{pk}(\alpha)$, $k = 1, 2$, where the inf is taken over the set D. Kulasekera and Lin (2005) has shown that $\hat{d}_k$ is a CAN estimator for the error variance σ_k^2, $k = 1, 2$. When the mean functions $m_1(\cdot)$ and $m_2(\cdot)$ are identical, the pooled estimator $\hat{g}_\alpha(t)$ and the estimators $\hat{g}_{k\alpha}(t)$'s estimate the same function $E(Y_{1i}|\alpha' X = t) = E(Y_{2i}|\alpha' X = t)$ and therefore $\hat{d}_{pk}$ is also consistent for σ_k^2, $k = 1, 2$. When $m_1(\cdot)$ and $m_2(\cdot)$ are different, each $\hat{d}_{pk}$, $k = 1, 2$, includes both the variability component due to the random error and the bias component caused by the difference between the two mean functions. Therefore, under H_a, $\hat{d}_{pk}$ tends to be larger than $\hat{d}_k$ for each k, $k = 1, 2$. Hence we propose to use the test statistic

$$T = \hat{d}_{p1} + \hat{d}_{p2} - \hat{d}_1 - \hat{d}_2,$$

and reject H_0 for large values of T. We show that a properly normalized version of T converges to the standard normal distribution under H_0 and it diverges to infinity under H_a (Theorem 2.3). This enables us to get critical points for large samples. The manner in which we create T, its asymptotic properties are comparable to those of Dette and Neumeyer (2001).

Alternatively, with $\hat{\theta}_k = \arg\inf_\alpha \hat{d}_k(\alpha)$, $k = 1, 2$, we can define a statistic T' as

$$T' = \hat{d}_{p1}(\hat{\theta}_1) + \hat{d}_{p2}(\hat{\theta}_2) - \hat{d}_1 - \hat{d}_2,$$

and reject H_0 for large values of T'. The motivation for this is as follows. Under H_0, $\hat{\theta}_k$, $k = 1, 2$, are $\sqrt{N}$-consistent for θ_k; thus $\hat{d}_{pk}(\hat{\theta}_k)$ and $\hat{d}_{pk}$ are asymptotically equivalent and it can be shown that both T and T' have the same asymptotic properties. However, since $T' \geq T$ and we reject H_0 for large values of T (T'), under H_a, T' may have a higher power compared with T in finite samples when using asymptotic critical points (see Remark 2.5 for more details).

2.2 Main Results

We first give some assumptions that are used in the sequel.

(A1) The mean functions $m_1(\cdot)$ and $m_2(\cdot)$ are r-time continuously differentiable. Under H_0, the true index vectors are in D where, for the remainder of the paper,

$$D = \{\theta = (\theta_1, \cdots, \theta_p)' \in \mathbb{R}^p \mid \|\theta\| = 1,\ \theta_1 > 0\}.$$

(A2) The covariates of both samples follow the same distribution as X. The domain S_X of X is a closed and bounded convex set. The support of $L_{q,\theta}(X)$ contains at least one interior ball with radius $w_0 > 0$. The density of X, $\alpha' X$ and $\alpha'(X_1 - X_2)$ will be denoted by $f(\cdot)$, $f_\alpha(\cdot)$ and $\phi_\alpha(\cdot)$, respectively. The function $f \in C^r(S_X)$ and there exists constants $0 < c_1 < c_2 < \infty$ such that $c_1 \leq f(x) \leq c_2, \quad \forall x \in S_X$.

(A3) The error ϵ in each sample has at least v, $v \geq 5$, moments. For $k = 1, 2$, $\mathrm{Var}(\epsilon_{kj}|X_{kj}) = \sigma_k^2$, $1 \leq j \leq n_k$.

(A4) The kernel function $K(\cdot)$ is a r-th order kernel and $K \in C^1([-1, 1])$. The constant $c_\theta = \int_{-1}^1 K(s)\phi_\theta(s)ds \neq 0$, where ϕ_θ is given in (A2).

(A5) The function $L(\cdot)$ is a bounded, symmetric non-negative function supported on $(-1, 1)$. It is Lipschitz continuous of order 1 and non-increasing in $|t|$ and $L(t) > 0$ for all $t \in (-1, 1)$.

(A6) The bandwidth h is such that $h = O(N^{-\beta})$ for some $\beta \in (0, \frac{1}{3})$ where $N = n_1 + n_2$, and $\lim_{N\to\infty} n_1/N = \Delta$.

Remark 2.1. The density of X is assumed to be bounded away from zero to avoid the sparseness problem. Most of our results hold under heteroscedastic error variances by taking the weight function $L_{q,\alpha}(\cdot)$ to be free of α. We only present our results under homogeneous error variances assumption (A3). The assumption that $c_\theta \neq 0$ in (A4) guarantees that the denominators of the kernel estimators $\hat{g}(t)$ and $\hat{g}_k(t)$, $k = 1, 2$, will be zero with diminishing probability as $N \to \infty$ when t is away from the boundary. To avoid the zero denominator problem for these kernel estimators in finite samples, we may, say, exclude the summand $(Y_i - \hat{Y}_i)^2$ in constructing $\hat{d}$'s when the denominator of $\hat{Y}_i$ is zero or close to zero (see Section 5.2 for construction and Remark 5.5 for a discussion). Our assumption for bandwidth is quite flexible, which includes the "optimal" rate $h = O(N^{-\frac{1}{5}})$ and our final choice of $h = O(N^{-\frac{1}{2r}})$ (note that $r \geq 2$).

To simplify the presentation, we introduce the following unified notation. Let $\epsilon_i = \epsilon_{1i}, i = 1, .., n_1$; $\epsilon_{n_1+j} = \epsilon_{2j}, j = 1, ..., n_2$; $Y_i = Y_{1i}, i = 1, .., n_1$; $Y_{n_1+j} = Y_{2j}, j = 1, ..., n_2$; $X_i = X_{1i}, i = 1, .., n_1$ and $X_{n_1+j} = X_{2j}, j = 1, ..., n_2$. Now we define

$$\mathcal{H}_1 = \{1, 2, \cdots, n_1\}; \quad \mathcal{H}_2 = \{n_1 + 1, 2, \cdots, N\}; \quad \text{and} \quad \mathcal{H} = \{1, 2, \cdots, N\}$$

so that the elements $1, .., n_2$ in $\mathcal{H}'_2$ correspond to the indicies $n_1 + 1, \cdots, N$ in $\mathcal{H}_2$ in that order.Then $\mathcal{H}$ will represent the indicies of the combined sample where the first n_1 elements represent the first sample and the rest the second sample. Now, let $\mathcal{S}$ and $\mathcal{G}$ be subsets of $\{1, 2, \cdots, N\}$. Define

$$\hat{g}_\alpha(t; \mathcal{G}) = \frac{\sum_{j\in\mathcal{G}} Y_j K_h(\alpha' X_j - t)}{\sum_{j\in\mathcal{G}} K_h(\alpha' X_j - t)};$$

and

$$\hat{d}(\alpha; \mathcal{S}, \mathcal{G}) = \frac{\sum_{i\in\mathcal{S}} \Big(Y_i - \hat{g}_\alpha(\alpha' X_i; \mathcal{G})\Big)^2 L_{q,\alpha}(X_i)}{\sum_{i\in\mathcal{S}} L_{q,\alpha}(X_i)}. \tag{2.3}$$

Here, using our notation above, for any $i \in \mathcal{S}$ with $i \leq n_1$, the observation is from the first sample and if $i > n_1$, the observation is from the second sample etc. Then we have, for $k = 1, 2$, $\hat{d}_k(\alpha) = \hat{d}(\alpha; \mathcal{H}_k, \mathcal{H}_k)$ and $\hat{d}_{pk}(\alpha) = \hat{d}(\alpha; \mathcal{H}_k, \mathcal{H})$. Letting $\hat{d}(\mathcal{S}, \mathcal{G}) = \inf_{\alpha\in D} \hat{d}(\alpha; \mathcal{S}, \mathcal{G})$, the test statistic T can be written as

$$T = \hat{d}(\mathcal{H}_1, \mathcal{H}) + \hat{d}(\mathcal{H}_2, \mathcal{H}) - \hat{d}(\mathcal{H}_1, \mathcal{H}_1) - \hat{d}(\mathcal{H}_2, \mathcal{H}_2). \tag{2.4}$$

and T' can be written as

$$T' = \hat{d}(\hat{\theta}_1; \mathcal{H}_1, \mathcal{H}) + \hat{d}(\hat{\theta}_2; \mathcal{H}_2, \mathcal{H}) - \hat{d}(\mathcal{H}_1, \mathcal{H}_1) - \hat{d}(\mathcal{H}_2, \mathcal{H}_2). \tag{2.5}$$

The asymptotic distribution of T (and T') under the null hypothesis is obtained using the following theorem. This gives a decomposition of $\hat{d}(\mathcal{S}, \mathcal{G})$ above when $(\mathcal{S}, \mathcal{G})$ is equal to each $(\mathcal{H}_1, \mathcal{H})$, $(\mathcal{H}_2, \mathcal{H})$, $(\mathcal{H}_1, \mathcal{H}_1)$, $(\mathcal{H}_2, \mathcal{H}_2)$.

Theorem 2.2. *Suppose assumptions (A1–A6) hold. Let* $(\mathcal{S}, \mathcal{G})$ *be any one of the four pairs* $(\mathcal{H}_1, \mathcal{H})$, $(\mathcal{H}_2, \mathcal{H})$, $(\mathcal{H}_1, \mathcal{H}_1)$, $(\mathcal{H}_2, \mathcal{H}_2)$. *Let* $n = |\mathcal{S}|, t = |\mathcal{G}|$ *and*

$$t_1 = |\mathcal{G} \cap \mathcal{H}_1|, \quad t_2 = |\mathcal{G} \cap \mathcal{H}2|, \quad u_1 = |\mathcal{S} \cap \mathcal{G} \cap \mathcal{H}_1|, \quad u_2 = |\mathcal{S} \cap \mathcal{G} \cap \mathcal{H}_2|.$$

Then the following holds either under H_a *for the terms with* $(\mathcal{S}, \mathcal{G}) = (\mathcal{H}_k, \mathcal{H}_k)$, $k = 1, 2$, *or under* H_0 *for all four pairs of* $(\mathcal{S}, \mathcal{G})$.

$$\hat{d}(\mathcal{S}, \mathcal{G}) = \frac{\sum_{i \in \mathcal{S}} \epsilon_i^2 L_{q,\theta}(X_i)}{\sum_{i \in \mathcal{S}} L_{q,\theta}(X_i)} - \frac{2}{b_0 tnh} \sum_{i \in \mathcal{S}, j \in \mathcal{G}, i \neq j} \frac{K_h(\theta' X_j - \theta' X_i)}{f_\theta(\theta' X_i)} L_{q,\theta}(X_i)\epsilon_i \epsilon_j$$
$$+ \frac{\int_{-1}^1 K^2(s)ds}{b_0 t^2 h} \big(t_1 a_1 + t_2 a_2\big) - \frac{2K(0)}{nthb_0} \big(u_1 a_1 + u_2 a_2\big) + o_p(N^{-1} h^{-\frac{1}{2}}),$$

where $b_0 = E[L_{q,\theta}(X)]$ *and* $a_k = \sigma_k^2 q_\theta \int_{-1}^1 L(s)ds$, $k = 1, 2$. *Under* H_a, *we have* $\hat{d}(\mathcal{H}_k, \mathcal{H}) = \sigma_k^2 + c_{0k} + o_p(1)$ *where*

$$c_{0k} = \inf_\alpha \frac{E\Big(\big[m_k(X) - \Delta g_{1\alpha}(\alpha' X) - (1 - \Delta) g_{2\alpha}(\alpha' X)\big]^2 L_{q,\alpha}(X)\Big)}{EL_{q,\alpha}(X)}. \tag{2.6}$$

Proof. It is an immediate consequence of Lemma 5.13 and Lemma 5.14. □

By examining each component of $\hat{d}(\mathcal{S}, \mathcal{G})$, we can obtain the asymptotic properties of the test statistics T and T' respectively. Note that by the construction of T and T' the first component (i.e. the weighted average of ϵ_i^2's) in the decomposition of terms like $\hat{d}(\mathcal{H}_1, \mathcal{H})$ will be canceled. The asymptotic distribution of T is then derived by analyzing the remaining leading components of $\hat{d}(\mathcal{H}_1, \mathcal{H})$ etc. Noting that these remaining leading terms are quadratic forms in $\epsilon_i, i = 1, ..., N$, we invoke Lemma 5.1 in Section 5.1, which deals with quadratic forms in random errors with random coefficients and diminishing diagonal elements. We can prove

Theorem 2.3. *Under Assumptions (A1–A6) with* $h = O(N^{-\frac{1}{2r}})$, *we have that under* H_0,

$$T = \frac{1}{Nh} D_N + \frac{\sigma_N}{N\sqrt{h}} Z + o_p(N^{-1} h^{-\frac{1}{2}}),$$

where Z *is a standard normal random variable,*

$$D_N = \frac{\int_{-1}^1 K^2(s)ds}{b_0} \Big[\Big(\frac{2n_1}{N} - \frac{N}{n_1}\Big) a_1 + \Big(\frac{2n_2}{N} - \frac{N}{n_2}\Big) a_2\Big] + \frac{2K(0)}{b_0} \Big(\frac{n_2}{n_1} a_1 + \frac{n_1}{n_2} a_2\Big), \tag{2.7}$$

and $\sigma_N^2 = \frac{4}{b_0^2}\big[\frac{2n_2^2}{n_1^2} a_{11} + \big(\frac{n_1}{n_2} + \frac{n_2}{n_1} + 2\big) a_{12} + \frac{2n_1^2}{n_2^2} a_{22}\big]$. *Here* $a_k = \sigma_k^2 q_\theta \int_{-1}^1 L(s)ds$, $k = 1, 2$, *and, for* $i, j = 1, 2$, $a_{ij} = \sigma_i^2 \sigma_j^2 q_\theta \int_{-1}^1 K^2(t)dt \int_{-1}^1 L^2(t)dt$. *Under* H_a, $T = c_{01} + c_{02} + o_p(1)$ *for the positive constants* c_{01}, c_{02} *given by (2.6). The asymptotic properties of* T' *are identical to* T *(except that under* H_a $T' \geq c_{01} + c_{02} + o_p(1)$*).*

Proof. See Section 5.4. □

Hence, we propose to reject H_0 when $N\sqrt{h}(T-\frac{1}{Nh}\hat{D}_N)/\hat{\sigma}_N$ is larger than the upper-α quantile of the standard normal distribution, where $\hat{D}_N$ and $\hat{\sigma}_N$ are any consistent estimators of D_N and σ_N respectively.

Remark 2.4. Note that all parameters given in the theorem can be $\sqrt{N}$-consistently estimated. In practice, the variances $\hat{d}_k$ are consistent estimators of $\sigma_k^2, k = 1, 2$ respectively and $a_i, i = 1, .., 3$ are known up to the parameter θ given the kernel functions K and L. The existing results in the literature give us the values minimizing $\hat{d}_k$ are $\sqrt{N}$-consistent for the corresponding index vectors θ_k, $k = 1, 2$ and, hence we can find consistent estimators of the a_i's.

Remark 2.5. Actually we can plug any root-n consistent estimator $\hat{\theta}_k$ of θ_k in the literature, $k = 1, 2$, into T'. The asymptotic properties of T' are identical to T and therefore the decision rule is identical. Since $T' \geq T$, under H_a the power for T' is always higher. The divergence rate is $N\sqrt{h}$ for both test statistics and therefore they are capable of detecting local alternatives of order $\left(N\sqrt{h}\right)^{-\frac{1}{2}}$. It should be noted that this divergence rate is slightly slower than what is proposed by Neumeyer and Dette (2003). However, for moderate sample sizes, our simulations show very competitive power of the proposed tests for the same alternative (see Example 4.2 in Section 4 for more detail). Our simulations also show that, not only being computationally easier (because only two minimizations are required for T' rather than four) when the sample sizes are large (say, 50 or above), the overall performance of T' is better than T (the empirical level is comparable or closer to target under H_0 and the power is 5%–20% higher than T under H_a).

3 Testing Equality of Error Variances

In this section we propose a test for the equality of the two sample variances. Namely, assume that $\sigma_k^2(x) = \sigma_k^2$, $k = 1, 2$, and we would like to test

$$H_{01} : \sigma_1^2 = \sigma_2^2 \quad \text{and} \quad H_{a1} : \sigma_1^2 \neq \sigma_2^2. \tag{3.1}$$

From Kulasekera and Lin (2005), $\hat{d}_k$ is consistent for σ_k^2, $k = 1, 2$, with

$$\sqrt{n_k}(\hat{d}_k - \sigma_k^2) \xrightarrow{\mathcal{D}} N\Big(0, \frac{E[\lambda_k(X)L_{q,\theta}^2(X)]}{\big(E[L_{q,\theta}(X)]\big)}\Big), \tag{3.2}$$

where $\lambda_k(x) = \text{Var}(\epsilon^2|X)$ for the kth sample, $k = 1, 2$. This immediately leads to the test statistic

$$T^* = \sqrt{N}\big(\hat{d}_1 - \hat{d}_2\big). \tag{3.3}$$

Since $\hat{d}_1$ and $\hat{d}_2$ are consistent for the same quantity under H_0 and are consistent for different quantities under H_a, $|\hat{d}_1 - \hat{d}_2|$ is close to zero under H_0 and it diverges to infinity under H_a. The complete result is summarized into the next theorem.

Theorem 3.1. *Let assumptions (A1–A6) hold and let T^* be given as in (3.3). Under H_0, $T^*/\tau_N \xrightarrow{\mathcal{D}} N(0,1)$, where*

$$\tau_N^2 = \frac{N}{n_1}\frac{E[\lambda_1(X)L_{q,\theta}^2(X)]}{\left(EL_{q,\theta}\right)^2} + \frac{N}{n_2}\frac{E[\lambda_2(X)L_{q,\theta}^2(X)]}{\left(EL_{q,\theta}\right)^2}.$$

Here λ_1 and λ_2 are as given above. Under H_a, T^* diverges to infinity with rate $\sqrt{N}$.

Proof. By (3.2) and the independence of $\hat{d}_1$ and $\hat{d}_2$. □

Hence we propose to reject H_{01} when $|T^*|/\hat{\tau}_N$ is larger than the upper-$\frac{\alpha}{2}$ quantile of the standard normal distribution, where $\hat{\tau}_N$ is any consistent estimator of τ_N.

4 Empirical Study

4.1 Simulation Study

To investigate the finite sample performance of the proposed procedures, we conducted an extensive simulation study. We considered dimensions $p = 2$ and $p = 3$. The covariates were taken to be independent, following two different design densities: uniform design, where each covariate is uniform between 0 and 1; and normal design, where each covariate is normal with mean zero and standard deviation 0.5. We used the quadratic kernel function $K(u) = \frac{3}{4}(1-u^2)I(|u| \le 1)$ and the $L(\cdot)$ function

$$L(u) = I(|u| \le 0.9) + 10(1-u)I(0.9 < |u| \le 1).$$

Our simulations show that a change of q does not make much of a difference and we present the results for $q = 0.95$ throughout. As an estimator of q_θ, we use $q_{\hat{\theta}} = qw_{\hat{\theta}}$ where $\hat{\theta}$ is the index vector estimator obtained using the pooled sample. We first consider the equal sample size cases (Examples 1–4), where we can use a common bandwidth h. We selected h by a simple cross-validation method, where the ith observation was deleted in estimating $E(Y|\alpha'X = \cdot)$. These estimators were then used in constructing $\hat{d}_k(\alpha)$, $k = 1, 2$. Finally we chose $(h, \tilde{\theta})$ that minimizes $\hat{d}_1(\alpha) + \hat{d}_2(\alpha)$. When the sample sizes are different for the two samples, we recommend the use of two different bandwidths, as discussed in Example 4.5. The error distributions were taken to be mean zero normals with several different (σ_1^2, σ_2^2) combinations. All the rejection rates reported are based on 2000 simulations.

Example 4.1 (Uniform Design; $p = 2$). In this example we consider evaluating the level properties of the proposed asymptotic tests for the equality of mean functions. We considered $\theta_1 = \theta_2 = (\sqrt{2}/2, \sqrt{2}/2)$ and

$$\text{(i)} \quad g_1(t) = g_2(t) = t^2;$$
$$\text{(ii)} \quad g_1(t) = g_2(t) = \sin(k\pi t), \quad k = 1, 2.$$

For each function we took $n_1 = n_2 = 25, 50, 100$ and $(\sigma_1^2, \sigma_2^2) = (0.5, 0.25), (0.25, 0.25)$. The results are presented in Table 1. We can see that for sample size 25 the rejection rates are somewhat higher than nominal levels. This is not surprising because unlike the truly univariate case, here we have to estimate the index vector θ. When the sample sizes increase to 50, the rejection rates become satisfactory. We observe that, although worse than T for sample size 25, the rejection rates for T' improve as the sample sizes get larger. In fact, when the sample sizes are 100, the rejection rates for T' is closer to the nominal levels than T in almost all cases. We believe the extra estimation in T (compared with T') has an adverse effect on its convergence.

	$g_1 = g_2$		t^2			$\sin(\pi t)$			$\sin(2\pi t)$		
(σ_1^2, σ_2^2)	α level	n	25	50	100	25	50	100	25	50	100
(0.5,0.25)	0.1	T	.166	.101	.080	.164	.108	.083	.139	.082	.061
		T'	.185	.115	.092	.191	.122	.093	.183	.101	.073
	0.05	T	.111	.058	.040	.107	.079	.037	.091	.048	.028
		T'	.129	.064	.043	.135	.070	.044	.126	.058	.037
	0.025	T	.099	.037	.021	.083	.042	.021	.065	.034	.013
		T'	.099	.042	.024	.099	.048	.026	.082	.037	.018
(0.25,0.25)	0.1	T	.140	.090	.069	.133	.096	.061	.161	.098	.065
		T'	.161	.103	.077	.165	.114	.070	.200	.120	.076
	0.05	T	.090	.044	.034	.079	.056	.031	.113	.056	.032
		T'	.102	.053	.036	.101	.070	.036	.138	.071	.035
	0.025	T	.062	.029	.022	.054	.033	.015	.082	.033	.018
		T'	.075	.036	.024	.071	.040	.018	.101	.041	.022

Table 1: Simulated Levels for Example 4.1 with sample size $n_1 = n_2 = n$.

Example 4.2 (Uniform Design; $p = 2$). We evaluate the power properties of the test of equality of regression functions by investigating the following different functions and index vectors.

(i) $g_1(t) = g_2(t) = \sin(\pi t), \quad \theta_1 = (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}), \quad \theta_2 = (\frac{\sqrt{14}}{4}, -\frac{\sqrt{2}}{4})$;

(ii) $g_1(t) = \sin(\pi t), \quad g_2(t) = \sin(\pi t) + \frac{t}{3}, \quad \theta_1 = \theta_2 = (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$;

(iii) $g_1(t) = \sin(\pi t), \quad g_2(t) = \sin(\pi t) + t, \quad \theta_1 = \theta_2 = (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$;

(iv) $g_1(t) = \sin(\pi t), \quad g_2(t) = 2\sin(\pi t), \quad \theta_1 = (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}), \quad \theta_2 = (\frac{\sqrt{14}}{4}, \frac{\sqrt{2}}{4})$;

(v) $g_1(t) = \sin(2\pi t), \quad g_2(t) = \sin(2\pi t) + t, \quad \theta_1 = \theta_2 = (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$;

(vi) $g_1(t) = \sin(2\pi t), \quad g_2(t) = 2\sin(2\pi t), \quad \theta_1 = \theta_2 = (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$.

In all cases we took $(\sigma_1^2, \sigma_2^2) = (0.5, 0.25)$. The results for cases (i)-(iv) are presented in Table 2 and the results for the last two cases are given in Table 3. From both tables, we observe that the proposed procedure can detect the difference in the two mean functions well and the power of T' is about 5–10% higher than T in most cases. From

	α	0.1			0.05			0.025		
	n	25	50	100	25	50	100	25	50	100
(i)	T	.732	.960	.998	.655	.936	.998	.583	.907	.996
	T'	.768	.966	1.0	.692	.952	.998	.629	.923	.998
(ii)	T	.302	.367	.554	.217	.268	.427	.175	.202	.344
	T'	.333	.391	.574	.246	.291	.448	.193	.222	.359
(iii)	T	.891	.998	1.0	.851	.991	1.0	.810	.982	1.0
	T'	.909	.999	1.0	.873	.993	1.0	.845	.987	1.0
(iv)	T	.932	.996	1.0	.902	.992	1.0	.870	.985	1.0
	T'	.948	.997	1.0	.921	.994	1.0	.899	.990	1.0

Table 2: Simulated powers for Example 4.2 with sample size $n_1 = n_2 = n$.

case (ii) and case (iii) we see that, for both tests, the power increases dramatically as the difference $g_2(t) - g_1(t)$ increases from $t/3$ to t.

Note that the powers of T and T' in these cases are not inflated due to the wrong size because the empirical sizes under H_0 (Example 4.1) are very close to or even lower than the nominal sizes (especially for sample sizes 50 or above) for the same $g(\cdot)$ functions.

Case (v) and case (vi) are included mainly for comparison purposes. These two cases were examined by Neumeyer and Dette (2003) (scenarios (vi) and (ix) of (4.11) in their paper). They proposed test statistics $K_N^{(1)}$ and $K_N^{(2)}$ which are sup norms of two empirical processes of residuals. In their simulation study, they took uniform density between 0 and 1 for covariate X and the same error variances as above and the empirical powers were obtained using a wild bootstrap method. Although not directly comparable due to the extra step of estimation of index vectors in SIMs, we can at least conclude that our proposed procedure, compared to the ones proposed for univariate case (denoted by $K_N^{(1)}$ and $K_N^{(2)}$ in Table 3) suffers no loss of power under the more complicated situation of a SIM.

We also ran a simulation for cases (v) and (vi) with design density $U(0, \sqrt{2}/2)$ so that $\theta' X$ matches the range $(0, 1)$ of design density in Neumeyer and Dette (2003). The results are given within parentheses in Table 3. It is noteworthy that the power of their second test $K_N^{(2)}$ does not depend on the design density of the covariate and it generally has shown better power. Hence, although the random number generation is not the same, we can argue that even with the extra estimation of index vectors, our method yields reasonable powers. The large difference in powers for case (vi) could be due to the difference in bootstrap method and direct computation of critical points, where the latter can be more powerful provided the critical values can be estimated rather accurately.

	α	0.1		0.05		0.025	
	n	25	50	25	50	25	50
(v)	T	.843	.993	.778	.984	.725	.978
		(.642)	(.898)	(.560)	(.837)	(.491)	(.786)
	$K_N^{(1)}$	.601	.883	.462	.717	.346	.734
	$K_N^{(2)}$	.745	.966	.634	.932	.523	.877
(vi)	T	.697	.930	.624	.899	.548	.862
		(.765)	(.950)	(.698)	(.934)	(.628)	(.924)
	$K_N^{(1)}$	.312	.617	.184	.465	.108	.314
	$K_N^{(2)}$	.322	.692	.160	.513	.067	.317

Table 3: Simulated powers for Example 4.2 with sample size $n_1 = n_2 = n$. The rejection rates inside the parentheses correspond to density design $U(0, \sqrt{2}/2)$. The rows for $K_N^{(1)}$ and $K_N^{(2)}$ are taken from Table 1 and Table 2 of Neumeyer and Dette (2003).

	α	0.1			0.05			0.02		
(σ_1^2, σ_2^2)	n	25	50	100	25	50	100	25	50	100
(0.25,0.25)	(i)	.128	.121	.104	.066	.070	.050	.033	.029	.019
	(ii)	.133	.123	.113	.079	.064	.061	.032	.029	.030
	(iii)	.134	.135	.116	.075	.071	.060	.034	.028	.026
(0.5,0.25)	(i)	.380	.693	.935	.258	.564	.887	.134	.383	.784
	(ii)	.408	.701	.936	.277	.581	.885	.150	.429	.795
	(iii)	.345	.645	.924	.234	.512	.863	.122	.351	.769
	(iv)	.391	.677	.940	.265	.556	.892	.145	.394	.798
	(v)	.392	.708	.939	.260	.579	.883	.150	.422	.803

Table 4: Simulated rejection rates for Example 4.3 with sample size $n_1 = n_2 = n$.

Example 4.3 (Uniform Design; $p = 2$). In this example we investigate the finite sample performance of the test for equal variances. We consider parameter combinations $(\sigma_1^2, \sigma_2^2) = (0.5, 0.25), (0.25, 0.25)$ in each of the following function/index vector settings:

(i) $g_1(t) = g_2(t) = t^2$, and $\theta_1 = \theta_2 = (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$;

(ii) $g_1(t) = g_2(t) = \sin(\pi t)$, and $\theta_1 = \theta_2 = (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$;

(iii) $g_1(t) = t^2$, $g_2(t) = \sin(\pi t)$ and $\theta_1 = (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$, $\theta_2 = (\frac{\sqrt{14}}{4}, -\frac{\sqrt{2}}{4})$.

and consider $(\sigma_1^2, \sigma_2^2) = (0.5, 0.25)$ for

(iv) $g_1(t) = \sin(\pi t)$, $g_2(t) = \sin(\pi t) + \dfrac{t}{3}$, and $\theta_1 = \theta_2 = (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$;

(v) $g_1(t) = \sin(\pi t)$, $g_2(t) = \sin(\pi t) + t$, and $\theta_1 = \theta_2 = (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$.

The results are presented in Table 4. We observe that under H_0, although slowly, in most cases the rejection rates converge to the nominal levels as sample sizes increase; and under H_a the test can detect the difference in error variances satisfactorily and the power increases rather quickly with sample sizes.

	(g_1, g_2)		$(\exp(t), \exp(t))$			$(\exp(t), \exp(t)+t)$		
(σ_1^2, σ_2^2)	α level	n	25	50	100	25	50	100
(0.5,0.25)	0.1	T	.206	.137	.082	.403	.663	.943
		T'	.248	.148	.096	.507	.742	.971
	0.05	T	.154	.090	.043	.315	.569	.911
		T'	.190	.097	.052	.411	.658	.945
	0.025	T	.111	.062	.028	.258	.491	.871
		T'	.144	.071	.033	.332	.581	.918
(0.25,0.25)	0.1	T	.176	.117	.071	.491	.811	.983
		T'	.224	.143	.091	.583	.877	.995
	0.05	T	.124	.072	.038	.414	.753	.978
		T'	.153	.085	.053	.492	.819	.988
	0.025	T	.091	.043	.023	.354	.688	.970
		T'	.118	.057	.031	.427	.758	.981

Table 5: Simulated rejection rates for Example 4.4 with sample size $n_1 = n_2 = n$.

Example 4.4 (Normal Design; $p = 3$). We also conducted an investigation of size and power properties of the tests for the equality of mean functions and the test for equal variances for dimension $p = 3$ using normal covariates design. All three covariates are generated independently from a normal distribution with mean 0 and standard

(σ_1^2, σ_2^2)	α / n	0.1 / 25	0.1 / 50	0.1 / 100	0.05 / 25	0.05 / 50	0.05 / 100	0.02 / 25	0.02 / 50	0.02 / 100
(0.25,0.25)	(i)	.127	.129	.105	.072	.061	.058	.028	.026	.024
	(ii)	.157	.134	.127	.087	.073	.075	.041	.035	.032
(0.5,0.25)	(i)	.349	.668	.935	.238	.533	.884	.142	.378	.790
	(ii)	.331	.631	.923	.225	.506	.866	.116	.350	.760

Table 6: Simulated rejection rates for Example 4.4 with sample size $n_1 = n_2 = n$.

deviation 0.5. We consider $\theta_1 = \theta_2 = (\sqrt{3}/3, \sqrt{3}/3, \sqrt{3}/3)$ and $(\sigma_1^2, \sigma_2^2) = (0.5, 0.25)$ and $(0.25, 0.25)$ in each of the following two settings:

$$\text{(i)} \quad g_1(t) = g_2(t) = \exp(t);$$
$$\text{(ii)} \quad g_1(t) = \exp(t) \quad \text{and} \quad g_2(t) = \exp(t) + t.$$

The rejection rates for testing the equality of regression functions are presented in Table 5 and the rejection rates for testing the equality of error variances are given in Table 6. We observe similar performances as in the previous three examples except that larger sample sizes are required to achieve the same nominal levels, which is not surprising since the dimension for the index vector has increased.

(σ_1^2, σ_2^2)	α level	(g_1, g_2) / (n_1, n_2)	(t^2, t^2) (25,50)	(t^2, t^2) (50,100)	$(t^2, t^2 + t)$ (25,50)	$(t^2, t^2 + t)$ (50,100)
(0.5,0.25)	0.1	T	.070	.044	.839	.987
		T'	.084	.049	.868	.991
	0.05	T	.042	.024	.776	.974
		T'	.052	.025	.815	.984
	0.025	T	.030	.014	.729	1.0
		T'	.034	.015	.764	1.0
(0.25,0.25)	0.1	T	.065	.032	.963	1.0
		T'	.085	.038	.973	1.0
	0.05	T	.043	.019	.939	1.0
		T'	.051	.022	.956	1.0
	0.025	T	.032	.012	.913	.999
		T'	.033	.014	.935	1.0

Table 7: Simulated rejection rates of the first test for Example 4.5.

Example 4.5 (Unbalanced Sample Sizes). When sample sizes are quite different, using a common bandwidth as we did in the previous examples leads to rejection rates under H_0 significantly higher than the nominal levels. To accommodate for unequal bandwidths we proceeded as follows. Since our asymptotic theory requires that the bandwidth parameter h should be proportional to $N^{-1/4}$, we first select an initial bandwidth parameter h_0 by the same cross-validation criteria as in the previous examples. Then let

$$h = \max(h_0,\, cN^{-\frac{1}{4}}); \quad \text{and} \quad h_k = h \cdot \left(\frac{N}{n_k}\right)^{\frac{1}{4}}, \quad k = 1, 2.$$

We took different c values ranging from 0.2 to 0.5. The difference in results is small, although we observed slightly better performance in some cases when c is larger in this range. We only present the results for $c = 0.3$ here. We use h_k for $\hat{d}_k$ and use h for $\hat{d}_{pk}$ and the construction of $\hat{\sigma}_N$. We use this criteria for uniform design and $p = 2$. Take $(n_1, n_2) = (25, 50), (50, 100)$, $\theta_1 = \theta_2 = (\sqrt{2}/2, \sqrt{2}/2)$, $(\sigma_1^2, \sigma_2^2) = (0.5, 0.25), (0.25, 0.25)$ and consider

$$\begin{aligned} &\text{(i)} \quad g_1(t) = g_2(t) = t^2; \\ &\text{(ii)} \quad g_1(t) = t^2, \quad \text{and} \quad g_2(t) = t^2 + t. \end{aligned}$$

The simulated rejection rates for testing (1.3) are presented in Table 7 and the results for testing (3.1) are given in Table 8.

	α	0.1		0.05		0.02	
(σ_1^2, σ_2^2)	(n_1, n_2)	(25,50)	(50,100)	(25,50)	(50,100)	(25,50)	(50,100)
(0.25,0.25)	(i)	.209	.177	.142	.120	.088	.067
	(ii)	.221	.150	.146	.100	.083	.062
(0.5,0.25)	(i)	.314	.661	.203	.528	.112	.361
	(ii)	.332	.685	.215	.540	.111	.360

Table 8: Simulated rejection rates of the second test for Example 4.5.

4.2 Real Data Example

We applied our method to a real data example to show its usage. We took the FAT data from (??). The data consists of the response variable FAT (the percentage of fat in pork bellies) and 10 predictor variables for a sample of 45 pork carcasses. It was shown that six of these covariates gave a well fitting linear model to the responses. These covariates were X_1; an average of three measures of back fat thickness, X_2; a muscling score for the carcass, X_3; an average of three measures of fat depth opposite the tenth rib, X_4; live weight of the carcass, X_5; the average measure of leanness of three cross sections of the belly, and X_6; total weight of the belly. Since the linear model assumption is reasonable for this data, we split the data set into two parts, one

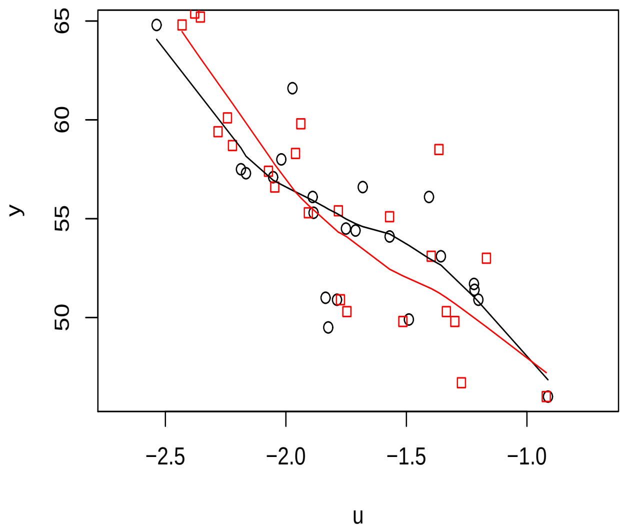

Figure 1: Plots of Y, the Fat content against $u = \hat{\theta}'X$ and LOWESS fit for group 1 (circles, black) and for group 2 (squares, red)

with the first 22 observations and the other the rest. Then we used our method to test the equality of the two regression functions and error variances. Since the two data sets actually come from the same population, we should not reject the null hypotheses in both cases. The CV method introduced in the simulation section was again used to select the bandwidths to get $h_1 = h_2 = 0.09$. The resulting test statistics values were $T = -2.51$, $T' = -0.12$ and $T^* = -0.01$. Hence we do not reject the equality of the mean functions and the error variances. Several other h values were tried and the same conclusion was reached. We give the plot of the smoothed Fat content against $u = \hat{\theta}'X$ and $\hat{\theta}'X$ for the two groups in Figure 1 where $\hat{\theta}$ is the index vector estimator for combined sample.

5 Proofs

In this section, we give a few lemmas that lead to Theorem 2.2 and Theorem 2.3. First we give a few general results in Section 5.1, followed by a sequence of technical lemmas particular to the problem in this paper in Section 5.2. Using these, we provide a decomposition of $\hat{d}(\mathcal{S}, \mathcal{G})$ and analyze each component of the decomposition in Section 5.3. Theorem 2.2 is a direct consequence of Lemma 5.14 in Section 5.3. Finally, the proof of Theorem 2.3 is given in Section 5.4. A number of lemmas in Section 5.2 are directly taken from Kulasekera and Lin (2005). They are restated here without proof.

Throughout this section, ξ is an arbitrary positive constant. In all cases, the term

N^ξ can be reduced to $(\ln N)^{\xi'}$ for a proper power ξ'. We use N^ξ for simplicity in presentation and this does not affect the results of this paper. In the sequel, unless otherwise stated, all sequences are from 1 to N (e.g. $\epsilon_i, i = 1, ..., N$).

5.1 Preparatory Lemmas

We first give a few general results that will be used in the sequel, starting with the following asymptotic result for a quadratic form. The random variable sequences used in this subsection are arbitrary sequences. We use the same notation for simplicity.

Lemma 5.1. *Let $\{(X_i, \epsilon_i), 1 \leq i \leq n\}$ be independent. Suppose $E(\epsilon_i|X_i) = 0$, $E(\epsilon_i^2|X_i) = \tau_i(X_i)$ and $E(\epsilon_i^4|X_i) = \lambda_i(X_i)$. Let $T_n = \sum_{i \neq j} w_{ij}\epsilon_i\epsilon_j$, where $w_{ij} = w(X_i, X_j)$. Suppose*

(i) *for a sequence of real numbers $h_n > 0$, $h_n = o(1)$ and $nh_n^2 \to \infty$;*

(ii) *τ_i's and λ_i's are bounded by some constant $C_1 > 0$;*

(iii) *uniformly in $u \leq 4$ and $r_1, \cdots, r_u \leq 4$,*

$$E\left|w_{i_1j_1}^{r_1} w_{i_2j_2}^{r_2} \cdots w_{i_uj_u}^{r_u}\right| = O(h_n^t); \tag{5.1}$$

where t is the number of distinct pairs in $\{(i_s, j_s)\}$ (with (i, j) and (j, i) being considered as identical) such that $i_s \neq j_s$;

(iv) *$s_n^2 = Var(T_n) = C_0 n^2 h_n + o(n^2 h_n)$ for some constant $C_0 > 0$.*

Then we have $T_n/s_n \xrightarrow{\mathcal{D}} N(0, 1)$.

Proof. Let $\tilde{w}_{ij} = w_{ij} + w_{ji}$ and $Y_i = Z_i\epsilon_i$, where $Z_1 = 0$ and $Z_i = \sum_{j=1}^{i-1} \tilde{w}_{ij}\epsilon_j$, $i = 2, \cdots, n$. Then $T_n = \sum_{i=1}^n Y_i$ and $\{(Y_i, \mathcal{F}_i)\}$ is a martingale difference sequence where $\mathcal{F}_i$ is the σ-field generated by $\{X_1, \cdots, X_i, \epsilon_1, \cdots, \epsilon_i\}$. By Theorem 1 of Heyde and Brown (1970), it suffices to show $\sum_{i=1}^n EY_i^4 = o(s_n^4)$ and $E\left(\sum_{i=1}^n Y_i^2 - s_n^2\right)^2 = o(s_n^4)$, which can be verified by straightforward computation. □

Lemma 5.2 (Bernstein's Inequality). *Let $Z_1, \cdots, Z_n$ be independent random variables with $Var(Z_i) = \sigma_i^2$ and $\sigma_1^2 + \cdots + \sigma_n^2 < \infty$. If $P(|Z_i - EZ_i| \leq b) = 1, i = 1, \cdots, n$, then*

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n (Z_i - EZ_i)\right| \geq \epsilon\right) \leq 2\exp\left\{-\frac{n^2\epsilon^2}{2\sum_{i=1}^n \sigma_i^2 + \frac{2}{3}bn\epsilon}\right\}.$$

Lemma 5.3. *Suppose for each n, $\{Z_1, \cdots, Z_n\}$ are independent conditional on X_n with $E(Z_i|X_n) = \mu(X_n)$ and $\max_{1\leq i\leq n} E(Z_i^2|X_n) \leq \sigma^2$. Then, for a sequence of functions $a_i(\cdot)$ with $\max_{1\leq i\leq n} Ea_i^2(X_n) \leq b_n$, we have*

$$\sum_{i=1}^n a_i(X_n)Z_i = \sum_{i=1}^n a_i(X_n)\mu(X_n) + O_p(\sqrt{nb_n}).$$

Proof. It suffices to observe that $E\left(\sum_{i=1}^n a_i(X_n)[Z_i - \mu(X_n)]\right)^2 \leq \sum_{i=1}^n Ea_i^2(X_n)\sigma^2$.
□

5.2 Technical Lemmas

Let $\hat{d}(\alpha; \mathcal{S}, \mathcal{G})$ be defined as in (2.3), where $\mathcal{S}$ is the index set deciding the sum of squared residuals and $\mathcal{G}$ is the index set deciding the estimator of the mean function. We shall assume that t_1, t_2, u_1, u_2 are all $O(N)$. To simplify the notation, we suppress the index sets $\mathcal{S}, \mathcal{G}$ when there is no confusion. For all $i \in \mathcal{S}$ and $j \in \mathcal{G}$ define

$$
\begin{aligned}
L_\alpha &= \sum_{i\in\mathcal{S}} L_{q,\alpha}(X_i); & K_i(\alpha) &= \sum_{j\in\mathcal{G}} K_h\big(\alpha' X_j - \alpha' X_i\big);\\
\kappa_{ij}(\alpha) &= \frac{K_h\big(\alpha' X_j - \alpha' X_i\big)}{K_i(\alpha)}; & K_{mi}(\alpha) &= \sum_{j\in\mathcal{G}} \kappa_{ij}(\alpha) m(X_j);\\
K_{ei}(\alpha) &= \sum_{j\in\mathcal{G}} \kappa_{ij}(\alpha)\epsilon_j; & l_{i\alpha} &= \frac{L_{q,\alpha}(X_i) I(|K_i(\alpha)| > a_0)}{L_\alpha};\\
\sum_{i\alpha} c_i &= \sum_{i\in\mathcal{S}} c_i l_{i,\alpha}, & & \qquad (5.2)
\end{aligned}
$$

where a_0 is an arbitrary fixed positive constant, and $\{c_i\}$ is any sequence.

Remark 5.4. Under H_a the $m(\cdot)$ function for the two populations are different. For notational simplicity, we write $m(\cdot)$ for both models which should be understood as $m(X_i) = m_k(X_i)$ for $i \in \mathcal{H}_k$, $k = 1, 2$.

Remark 5.5. By the results in Lemmas 5.6–5.9, we can add the factor $I(|K_i(\theta)| > a_0)$ to any of the terms in (5.2) and the difference is asymptotically negligible for all cases discussed in this paper. We shall use this property whenever it is necessary or sometimes for convenience.

Before partitioning $\hat{d}(\mathcal{S}, \mathcal{G})$ into several components and analyzing them, we first give a few technical results (5.6–5.11) for the quantities defined in (5.2).

Lemma 5.6 (Kulasekera and Lin, 2005). *Under Assumption (A1)-(A5), we have*

(i) $\sup_\alpha \max_{\{i\in\mathcal{S}|L_{q,\alpha}(X_i)\neq 0\}} |K_i(\alpha) - nhf_\alpha(\alpha' X_i)| = o_p(n^{\frac{1}{2}+\xi}\sqrt{h}) + O_p(nh^{r+1})$;

(ii) $\inf_\alpha \min_{\{i\in\mathcal{S}|L_{q,\alpha}(X_i)\neq 0\}} |K_i(\alpha)| \geq c_K nh + o_p(nh)$;

(iii) $\inf_\alpha L_\alpha \geq c_L N + o_p(N)$;

(iv) $\sup_\alpha \max_{1\leq i\leq n} \sum_{j=1}^n \big|K_h(\alpha' X_j - \alpha' X_i)\big| = O_p(nh)$;

(v) $\sup_\alpha \max_{\{1\leq i\leq n; L_{q,\alpha}(X_i)\neq 0\}} \sum_{j=1}^n |\kappa_{ij}(\alpha)| = O_p(1)$;

(vi) $\sup_\alpha \max_{\{1\leq j\leq n\}} \sum_{1\leq i\leq n; L_{q,\alpha}(X_i)\neq 0} |\kappa_{ij}(\alpha)| = O_p(1)$;

where c_K, c_L are some positive constants.

Remark 5.7. The upper bound in Lemma 5.6 (i) is $o_p(n^{\frac{1}{2}+\xi}\sqrt{h}) + O_p(nh^3)$ in Kulasekera and Lin (2005) because they took $r = 2$ and used a second order kernel function. A direct generalization to a general choice of r is immediate.

Lemma 5.8. *For any choice of fixed constant a_0 and a sequence $c_N = o(1)$, we have*

$$\sup_{\alpha\in\mathcal{A}} \max_{i\in\mathcal{S}} I(|K_i(\alpha)| < a_0) = o_p(N^{-t}),$$

for all $t > 0$, where $\mathcal{A} = \{\alpha \mid \|\alpha - \theta\| \le c_N\}$.

Proof. Partition the set $\mathcal{A}$ into N^s (s is decided below) small cells evenly. Take one point from each cell and let $\mathcal{A}_N$ denote the collection of them. Then for every $\alpha \in \mathcal{A}$ we can find one $\tilde{\alpha} \in \mathcal{A}_N$ such that $|\alpha - \tilde{\alpha}| = O(N^{-\frac{s}{p}})$ uniformly in α. Hence, by taking s large enough, we can make $\big||K_i(\alpha)| - |K_i(\tilde{\alpha})|\big| \le O(Nh^{-1} \cdot N^{-\frac{s}{p}}) < a_0$, uniformly in i and α. Thus, $\big|\inf_{(\alpha,i)\in\mathcal{B}} |K_i(\alpha)| - \inf_{(\alpha,i)\in\mathcal{B}_N} |K_i(\alpha)|\big| < a_0$, where $\mathcal{B} = \{(\alpha, i) \mid \alpha \in \mathcal{A}, i \in \mathcal{S}\}$ and $\mathcal{B}_N = \{(\alpha, i) \mid \alpha \in \mathcal{A}_N, i \in \mathcal{S}\}$; and the inequality $\inf_{\alpha\in\mathcal{A}} |K_i(\alpha)| < a_0$ implies that $\inf_{\alpha\in\mathcal{A}_n} K_i(\alpha) < 2a_0$. Therefore,

$$\sup_{(\alpha,i)\in\mathcal{B}} I(|K_i(\alpha)| < a_0) \le I\big(\inf_{(\alpha,i)\in\mathcal{B}} |K_i(\alpha)| < a_0\big) \le I\big(\inf_{(\alpha,i)\in\mathcal{B}_N} |K_i(\alpha)| < 2a_0\big),$$

From (5.3), $P\big(|K_i(\alpha)| < 2a_0\big) = o(N^{-t})$ for all $t > 0$ and the result follows. □

Lemma 5.9. *Under Assumption (A1)-(A5), we have, under H_0, for all $\xi > 0$,*

$$\sup_{(\alpha,i)\in\mathcal{A}} |K_{mi}(\alpha) - g_\alpha(\alpha' X_i)| = o_p\big((h^r + N^{-\frac{1}{2}} h^{\frac{1}{2}}) N^{\xi}\big).$$

where the index set $\mathcal{A} = \{(\alpha, i) \mid i \in \mathcal{S}, \quad L_{q,\alpha}(X_i) \ne 0\}$.

Proof. The proof is almost line-by-line identical to Lemma 4.8 of Kulasekera and Lin (2005) where they took $r = 2$. □

Lemma 5.10 (Kulasekera and Lin, 2005). *Let $A_{ij}(\alpha) = K_h(\alpha' X_j - \alpha' X_i)$. We have*

(i) $\sum_{i\in\mathcal{S}} \sup_{\|\alpha-\theta\|<\delta} \Big|\sum_{j\in\mathcal{G}} \big[A_{ij}(\alpha) - A_{ij}(\theta)\big]\epsilon_j\Big|^2 = o_p\big([N^2 h + N^{1+\frac{4}{v}} h^{-2}]\delta^2 N^{\xi}\big);$

(ii) $\sum_{i\in\mathcal{S}} \sup_{\alpha} \Big|\sum_{j\in\mathcal{G}} A_{ij}(\alpha)\epsilon_j\Big|^2 = o_p(N^{2+\xi} h);$

(iii) $\sup_{\|\alpha-\theta\|\le\delta} \sum_{i\in\mathcal{S}} \Big(\sum_{j\in\mathcal{G}} \big[\kappa_{ij}(\alpha) - \kappa_{ij}(\theta)\big]\epsilon_j\Big)^2 = o_p\big((h^{-1} + N^{\frac{4}{v}-1} h^{-4})\delta^2 N^{\xi}\big).$

Lemma 5.11. *Uniformly in $\{i_s, j_s\}$ and α in some neighborhood of θ, we have*

(i) $E\big|K_h(\alpha' X_{j_1} - \alpha' X_{i_1}) K_h(\alpha' X_{j_2} - \alpha' X_{i_2}) \cdots K_h(\alpha' X_{j_u} - \alpha' X_{i_u})\big| = O(h^t);$

(ii) $E\Big|\kappa_{i_1 j_1}(\alpha)\kappa_{i_2 j_2}(\alpha) \cdots \kappa_{i_u j_u}(\alpha) \cdot I(\alpha)\Big| = O\big(h^t (Nh)^{-u}\big);$

(iii) $E\Big|[\kappa_{i_1 j_1}(\alpha) - \kappa_{i_1 j_1}(\theta)] \cdots [\kappa_{i_u j_u}(\alpha) - \kappa_{i_u j_u}(\theta)] \cdot I(\alpha) I(\theta)\Big| = O\big(\delta^u (Nh)^{-u}\big),$

where $I(\alpha) = I\big(|K_{i1}(\alpha)| > a_0, \cdots, |K_{iu}(\alpha)| > a_0\big)$ and t is the number of distinct pairs in $\{(i_s, j_s)\}$ (with (i, j) and (j, i) being considered as identical) such that $i_s \ne j_s$.

Proof. For simplicity of presentation, we give the proof for $u = 2$. The proof for general u is almost identical. The first result follows immediately from direct verification. Now we show (ii). Let $Z_{ij}(\alpha) = K_h(\alpha' X_j - \alpha' X_i)$. Then $\kappa_{ij}(\alpha) = Z_{ij}(\alpha)/K_i(\alpha)$. Due to the indicator functions $\kappa_{ij}(\alpha)$'s are uniformly bounded by K_0/a_0, where $K_0 = \sup|K(t)|$. We have

$$
\begin{aligned}
&E\Big|\kappa_{i_1j_1}(\alpha)\kappa_{i_2j_2}(\alpha)I\big(|K_{i1}(\alpha)| > a_0, |K_{i2}(\alpha)| > a_0\big)\Big| \\
&\le \frac{1}{\delta_n^2}E|Z_{i_1j_1}(\alpha)Z_{i_2j_2}(\alpha)| + \frac{K_0^2}{a_0^2}[P(|K_{i_1}(\alpha)| < \delta_n) + P(|K_{i_2}(\alpha)| < \delta_n)].
\end{aligned}
$$

Note that $E[K_i(\alpha)] = K(0) + c_\alpha(N-1)h$, where $c_\alpha = \int_{-1}^{1} K(s)\phi_\alpha(s)ds$ with $\phi_\alpha(\cdot)$ being the density of $\alpha' X_1 - \alpha' X_2$. Take $\delta_N = |E[K_i(\alpha)]| - (Nh)^{\frac{2}{3}}$. By assumption (A4), $c_\theta \neq 0$ and we have $\delta_N > |c_\theta|nh/2 - (Nh)^{\frac{2}{3}}$ for all α in a small neighborhood of θ. Conditional on $\alpha' X_i$ and using Bernstein's inequality, we have

$$
\begin{aligned}
P(|K_i(\alpha)| < \delta_N) &\le P\big(|K_i(\alpha) - E[K_i(\alpha)]| > (Nh)^{\frac{2}{3}}\big) \\
&\le 2\exp\Big\{-\frac{(Nh)^{\frac{4}{3}}}{C_1 Nh + (Nh)^{\frac{2}{3}}}\Big\} = o(N^{-t}), \qquad (5.3)
\end{aligned}
$$

for all $t > 0$. The result now follows from (i).

For the last part, we have

$$
\begin{aligned}
&E\big|[\kappa_{ij}(\alpha) - \kappa_{ij}(\theta)][\kappa_{mn}(\alpha) - \kappa_{mn}(\theta)]I(\alpha)I(\theta)\big| \\
&= E\Big|\Big(\frac{Z_{ij}(\alpha) - Z_{ij}(\theta)}{K_i(\alpha)} + Z_{ij}(\theta)\frac{K_i(\theta) - K_i(\alpha)}{K_i(\alpha)K_i(\theta)}\Big) \\
&\qquad \cdot \Big(\frac{Z_{mn}(\alpha) - Z_{mn}(\theta)}{K_m(\alpha)} + Z_{mn}(\theta)\frac{K_m(\theta) - K_m(\alpha)}{K_m(\alpha)K_m(\theta)}\Big)I(\alpha)I(\theta)\Big|.
\end{aligned}
$$

Uniformly in i, j, m, n, we have

$$
\begin{aligned}
&E\Big|Z_{ij}(\theta)Z_{mn}(\theta)\frac{K_i(\theta) - K_i(\alpha)}{K_i(\alpha)K_i(\theta)}\frac{K_m(\theta) - K_n(\alpha)}{K_m(\alpha)K_n(\theta)}I(\alpha)I(\theta)\Big| \\
&\le \frac{C_1}{\delta_N^4}E\big|[K_i(\theta) - K_i(\alpha)][K_m(\theta) - K_n(\alpha)]\big| + o(N^{-t}) = \frac{C_1}{\delta_N^4}O\big((Nh\delta)^2\big) + o(N^{-t}),
\end{aligned}
$$

for all $t > 0$. Similarly we can show that the other three terms are of the same order. The proof is completed by taking the same δ_N as above. □

5.3 Decomposing $\hat{d}(\mathcal{S}, \mathcal{G})$

Now we are ready to partition $\hat{d}(\mathcal{S}, \mathcal{G})$. Let $m_\alpha(X_i) = m(X_i) - g_\alpha(\alpha' X_i)$. By (5.2) we can write $\hat{g}_\alpha(\alpha' X_i) = \sum_{j\in\mathcal{G}} \kappa_{ij}(\alpha)Y_j = K_{mi}(\alpha) + K_{ei}(\alpha)$, $i \in \mathcal{S}$; and

$$
\hat{d}(\alpha; \mathcal{S}, \mathcal{G}) = \sum_{i\alpha}\Big(\epsilon_i + m_\alpha(X_i) + g_\alpha(\alpha' X_i) - K_{mi}(\alpha) - K_{ei}(\alpha)\Big)^2 = \sum_{i=0}^{9} R_i(\alpha; \mathcal{S}, \mathcal{G});
$$

where, suppressing the dependence on $\mathcal{S}$ and $\mathcal{G}$,

$$R_0(\alpha) = \sum_{i\alpha} \epsilon_i^2; \qquad R_1(\alpha) = \sum_{i\alpha} m_\alpha^2(X_i);$$

$$R_2(\alpha) = \sum_{i\alpha} \Big(g_\alpha(\alpha' X_i) - K_{mi}(\alpha) \Big)^2; \quad R_3(\alpha) = \sum_{i\alpha} \Big(\sum_{j\in\mathcal{G}} \kappa_{ij}(\alpha)\epsilon_j \Big)^2;$$

$$R_4(\alpha) = 2\sum_{i\alpha} m_\alpha(X_i)\epsilon_i; \qquad R_5(\alpha) = 2\sum_{i\alpha} \Big(g_\alpha(\alpha' X_i) - K_{mi}(\alpha) \Big)\epsilon_i;$$

$$R_6(\alpha) = -2\sum_{i\alpha}\sum_{j\in\mathcal{G}} \kappa_{ij}(\alpha)\epsilon_i\epsilon_j; \qquad R_7(\alpha) = 2\sum_{i\alpha} m_\alpha(X_i)\Big(g_\alpha(\alpha' X_i) - K_{mi}(\alpha) \Big);$$

$$R_8(\alpha) = -2\sum_{i\alpha} m_\alpha(X_i) K_{ei}(\alpha); \qquad R_9(\alpha) = -2\sum_{i\alpha} \Big(g_\alpha(\alpha' X_i) - K_{mi}(\alpha) \Big) K_{ei}(\alpha).$$

Remark 5.12. As in Remark 5.4, we do not distinguish two $m(\cdot)$ functions here as well. It should be understood as $g_\alpha(\alpha' X_i) = E(m_k(X)|\alpha' X = \alpha' X_i)$ for $i \in \mathcal{H}_k$, $k = 1, 2$, and similarly for $m_\alpha(\cdot)$.

Lemma 5.13. *Suppose Assumptions (A1)-(A6) hold. Then under H_0 with θ being the true index vector, or under H_a with $\mathcal{G} = \mathcal{H}_k$ and $\theta = \theta_k$, $k = 1, 2$, we have that, for any $\xi > 0$,*

R0) $\sup_{\beta\in D} \sup_{\|\alpha-\beta\|\le\delta} |R_0(\alpha) - R_0(\beta)| = o_p\big(N^{-\frac{1}{2}+\xi}\delta\big)$;

R1) *there exist constants $c_1 > 0$ and $c_2 > 0$ such that, uniformly in α,*

$$c_1\|\alpha - \theta\|^2 + o(\|\alpha - \theta\|^2) \le R_1(\alpha) \le c_2\|\alpha - \theta\|^2 + o(\|\alpha - \theta\|^2);$$

R2) $\sup_{\alpha\in D} |R_2(\alpha)| = o_p\big((N^{-1}h + h^{2r})N^\xi\big)$;

R3) $\sup_{\|\alpha-\theta\|\le\delta} |R_3(\alpha) - R_3(\theta)| = O_p\big(N^{-1}h^{-1}\delta\big)$;

R4) $\sup_{\|\alpha-\theta\|\le\delta} |R_4(\alpha)| = o_p(N^{-\frac{1}{2}+\xi}\delta)$;

R5) $\sup_{\alpha\in D} |R_5(\alpha)| = o_p\big((N^{-\frac{1}{2}}h^{\frac{1}{2}} + h^r)N^{-\frac{1}{2}+\xi}\big)$;

R6) $\sup_{\|\alpha-\theta\|\le\delta} |R_6(\alpha) - R_6(\theta)| = o_p\big(h^{-1}\delta N^{-1+\xi}\big)$;

R7) $\sup_{\|\alpha-\theta\|\le\delta} |R_7(\alpha)| = o_p\big((N^{-\frac{1}{2}}h^{\frac{1}{2}} + h^r)\delta N^\xi\big)$;

R8) $\sup_{\|\alpha-\theta\|\le\delta} |R_8(\alpha)| = o_p(N^{-\frac{1}{2}+\xi}\delta)$;

R9) $\sup_{\alpha\in D} |R_9(\alpha)| = o_p\big(N^{-\frac{1}{2}+\xi}h^r + N^{-1+\xi}h\big)$.

Under H_a with $\mathcal{G} = \mathcal{H}$, the above results hold for R_0, R_3, R_6 with θ being arbitrarily fixed; the bound is $o_p\big(N^{-\frac{1}{2}+\xi}\big)$ for R_4, R_5, R_8, R_9 and, for $\mathcal{S} = \mathcal{H}_k$, $k = 1, 2$,

$$\inf_\alpha \big(R_1(\alpha) + R_2(\alpha) + R_7(\alpha)\big) = c_{0k} + o_p(1)$$

where c_{0k} is given by (2.6).

Proof. Under H_0 with θ being the true index vector, or under H_a with $\mathcal{G} = \mathcal{H}_k$, $k = 1, 2$, the proof is identical to Kulasekera and Lin (2005) except for terms R_3, R_6. Under H_a with $\mathcal{G} = \mathcal{H}$ the proof is identical to Kulasekera and Lin (2005) except for terms R_1, R_2, R_7. We will therefore provide proofs of these few cases. Let M_1, M_2 and M_3 be generic functions in the following.

First consider R_3 term. Write $R_3(\alpha) = \sum_{i=1}^{3} M_i(\alpha)$ where

$$
\begin{aligned}
M_1(\alpha) &= \sum_{i\alpha}\Big(\sum_{j\in\mathcal{G}}\big(\kappa_{ij}(\alpha) - \kappa_{ij}(\theta)\big)\epsilon_j\Big)^2 \\
M_2(\alpha) &= 2\sum_{i\alpha}\Big(\sum_{j\in\mathcal{G}}\big(\kappa_{ij}(\alpha) - \kappa_{ij}(\theta)\big)\epsilon_j\Big)\Big(\sum_{j\in\mathcal{G}}\kappa_{ij}(\alpha)\epsilon_j\Big) \\
M_3(\alpha) &= \sum_{i\alpha}\Big(\sum_{j\in\mathcal{G}}\kappa_{ij}(\theta)\epsilon_j\Big)^2 = M_{31}(\alpha) + M_{32}(\alpha) + M_{33},
\end{aligned}
$$

with

$$
\begin{aligned}
M_{31}(\alpha) &= \frac{1}{L_\alpha}\sum_{i\in\mathcal{S}}\Big(\sum_{j\in\mathcal{G}}\kappa_{ij}(\theta)\epsilon_j\Big)^2\big(L_{q,\alpha}(X_i) - L_{q,\theta}(X_i)\big), \\
M_{32}(\alpha) &= \Big(\frac{1}{L_\alpha} - \frac{1}{L_\theta}\Big)\sum_{i\in\mathcal{S}}\Big(\sum_{j\in\mathcal{G}}\kappa_{ij}(\theta)\epsilon_j\Big)^2 L_{q,\theta}(X_i), \\
M_{33} &= \frac{1}{L_\theta}\sum_{i\in\mathcal{S}}\Big(\sum_{j\in\mathcal{G}}\kappa_{ij}(\theta)\epsilon_j\Big)^2 L_{q,\theta}(X_i).
\end{aligned}
$$

Since $M_{33} = R_3(\theta)$, $|R_3(\alpha) - R_3(\theta)| \le |M_1(\alpha)| + |M_2(\alpha)| + |M_{31}(\alpha)| + |M_{32}(\alpha)|$, and the result follows by Lemma 5.6 and Lemma 5.10.

Now we consider R_6 term. Note that

$$
R_6(\alpha) - R_6(\theta) = \frac{1}{L_\alpha}[M_1(\alpha) + M_2(\alpha)] + \frac{L_\theta - L_\alpha}{L_\theta L_\alpha}M_3(\theta), \tag{5.4}
$$

where $M_3(\theta) = \sum_{i\in\mathcal{S},j\in\mathcal{G}}\kappa_{ij}(\theta)L_{q,\theta}(X_i)\epsilon_i\epsilon_j$ and

$$
\begin{aligned}
M_1(\alpha) &= \sum_{i\in\mathcal{S},j\in\mathcal{G}}[\kappa_{ij}(\alpha) - \kappa_{ij}(\theta)]L_{q,\alpha}(X_i)\epsilon_i\epsilon_j, \\
M_2(\alpha) &= \sum_{i\in\mathcal{S},j\in\mathcal{G}}\kappa_{ij}(\theta)[L_{q,\alpha}(X_i) - L_{q,\theta}(X_i)]\epsilon_i\epsilon_j.
\end{aligned}
$$

From (5.9), $M_3(\theta) = O_p(h^{-1})$. For $M_1(\alpha)$ and $M_2(\alpha)$ we apply the discretization technique again. Let the discrete set of α be denoted by $\mathcal{A}$ with size N^c (hence each cell has diameter of order $N^{-c/p}$). Note that

$$
\begin{aligned}
&M_1(\alpha) + M_2(\alpha) - M_1(\beta) - M_2(\beta) \\
&= \sum_{i\in\mathcal{S}}\epsilon_i L_{q,\alpha}(X_i)\sum_{j\in\mathcal{G}}[\kappa_{ij}(\alpha) - \kappa_{ij}(\beta)]\epsilon_j + \sum_{i\in\mathcal{S}}\epsilon_i[L_{q,\alpha}(X_i) - L_{q,\beta}(X_i)]\sum_{j\in\mathcal{G}}\kappa_{ij}(\beta)\epsilon_j.
\end{aligned}
$$

Then, by Cauchy-Schwarz inequality and Lemma 5.10 we can show

$$\begin{aligned}
&\sup_{\|\alpha-\theta\|\le\delta}\big|M_1(\alpha)+M_2(\alpha)\big| - \sup_{\alpha\in\mathcal{A}}\big|M_1(\alpha)+M_2(\alpha)\big| \\
&\quad\le \sup_{\|\alpha-\beta\|\le N^{-\frac{c}{p}}\delta}\big|M_1(\alpha)+M_2(\alpha)-M_1(\beta)-M_2(\beta)\big| \\
&\quad= o_p((h^{-\frac{1}{2}}+N^{\frac{2}{v}-\frac{1}{2}}h^{-2})N^{\frac{1}{2}-\frac{c}{p}+\xi}\delta). \qquad (5.5)
\end{aligned}$$

By Lemma 5.11 we have, $\sup_{\alpha\in\mathcal{A}} E[M_k(\alpha)]^t = O(N^t\frac{\delta^t}{(Nh)^t})$, for all $t\in\mathbb{N}$, $k=1,2$. Thus, for $k=1,2$,

$$P\big(\sup_{\alpha\in\mathcal{A}}|M_k(\alpha)|>a_N\big)\le N^c\frac{\sup_{\alpha\in\mathcal{A}}E[M_k(\alpha)]^t}{a_N^t}=O(\frac{1}{a_N}N^{\frac{c}{t}}h^{-1}\delta)^t,$$

which converges to zero provided $a_N = N^{\frac{c}{t}+\xi}h^{-1}\delta$ for any $\xi>0$. Therefore we have $\sup_{\alpha\in\mathcal{A}}|M_k(\alpha)| = o_p(N^{\frac{c}{t}+\xi}h^{-1}\delta)$, $k=1,2$. Taking c and t sufficiently large, the result follows from (5.4) and (5.5) above.

Finally we consider the term $R_1+R_2+R_7$ under H_a with $\mathcal{G}=\mathcal{H}$. We shall give the proof for the case of $\mathcal{S}=\mathcal{H}_1$. Let $g_{k\alpha}(t)=E(m_k(X)|\alpha'X=t)$, $k=1,2$. Define $K_{ik}(\alpha)=\sum_{j\in\mathcal{H}_k}K_h(\alpha'X_j-\alpha'X_i)$, $k=1,2$, and

$$K_{mik}(\alpha)=K_{ik}^{-1}(\alpha)\sum_{j\in\mathcal{H}_k}K_h(\alpha'X_j-\alpha'X_i)m_k(X_j).$$

By Lemma 5.9 we have

$$\begin{aligned}
&\inf_\alpha\big(R_1(\alpha)+R_2(\alpha)+R_7(\alpha)\big)=\inf_\alpha\sum_{i\alpha}\Big(K_{mi}(\alpha)-m_1(X_i)\Big)^2\\
&=\inf_\alpha\sum_{i\alpha}\Big(\big[K_{mi1}(\alpha)-g_{1\alpha}(\alpha'X_i)\big]\frac{K_{i1}(\alpha)}{K_i(\alpha)}+\big[K_{mi2}(\alpha)-g_{2\alpha}(\alpha'X_i)\big]\frac{K_{i2}(\alpha)}{K_i(\alpha)}\\
&\quad+g_{1\alpha}(\alpha'X_i)\frac{K_{i1}(\alpha)}{K_i(\alpha)}+g_{2\alpha}(\alpha'X_i)\frac{K_{i2}(\alpha)}{K_i(\alpha)}-m_1(X_i)\Big)^2\\
&=\inf_\alpha\sum_{i\alpha}\Big(m_1(X_i)-\Delta g_{1\alpha}(\alpha'X_i)-(1-\Delta)g_{2\alpha}(\alpha'X_i)\Big)^2+o_p(1)\\
&=c_{01}+o_p(1),
\end{aligned}$$

where c_{01} is given as in (2.6). □

Lemma 5.14. *Let $h=O(N^{-\frac{1}{2r}})$ and $\hat\theta$ minimizes $\hat d(\alpha)=\hat d(\alpha;\mathcal{S},\mathcal{G})$ over D. Let $b_0=E[L_{q,\theta}(X)]$, $n=|\mathcal{S}|, t=|\mathcal{G}|$ and*

$$t_1=|\mathcal{G}\cap\mathcal{H}_1|,\quad t_2=|\mathcal{G}\cap\mathcal{H}2|,\quad u_1=|\mathcal{S}\cap\mathcal{G}\cap\mathcal{H}_1|,\quad u_2=|\mathcal{S}\cap\mathcal{G}\cap\mathcal{H}_2|.$$

Then under Assumptions (A1)-(A6), we have, for all $\xi>0$,

(i) *under H_0, $\|\hat\theta-\theta\|=o_p\big(n^{-\frac{1}{2}+\xi}\big)$ for all $\xi>0$;*

(ii) *under* H_0, $\sum_{i=0}^{9} \tilde{R}_i(\hat{\theta}) = o_p(N^{-1+\xi})$, *where* $\tilde{R}_i(\hat{\theta}) = |R_i(\hat{\theta})|$, $i = 1, 2, 4, 5, 7, 8, 9$, *and* $\tilde{R}_i(\hat{\theta}) = |R_i(\hat{\theta}) - R_i(\theta)|$, $i = 0, 3, 6$*; the result for* $\tilde{R}_0$ *holds under* H_a *as well;*

(iii) *under* H_0, $R_3(\theta) = O_p(N^{-1}h^{-1})$ *and, more specifically,*

$$R_3(\theta) = \frac{q_\theta \int_{-1}^{1} K^2(s)ds}{b_0 t^2 h}(t_1\sigma_1^2 + t_2\sigma_2^2) \int_{-1}^{1} L(s)ds + O_p(N^{-1});$$

(iv) *under* H_0, $R_6(\theta) = O_p(N^{-1}h^{-1})$ *and, more specifically,*

$$\begin{aligned} R_6(\theta) = &-\frac{2q_\theta K(0)}{nthb_0}(u_1\sigma_1^2 + u_2\sigma_2^2)\int_{-1}^{1} L(s)ds \\ &- \frac{2}{nthb_0} \sum_{i\in\mathcal{S}, j\in\mathcal{G}, i\neq j} \frac{K_h(\theta' X_j - \theta' X_i)}{f_\theta(\theta' X_i)} L_{q,\theta}(X_i)\epsilon_i\epsilon_j + o_p(N^{-1}). \end{aligned}$$

Proof.

(i)-(ii) Since $\hat{d}(\hat{\theta})$ minimizes $\hat{d}(\alpha)$, we have $\hat{d}(\hat{\theta}) \le \hat{d}(\theta)$, i.e. $\sum_{i=0}^{9} R_i(\hat{\theta}) \le \sum_{i=0}^{9} R_i(\theta)$. Note that $R_1(\theta) = 0$ under H_0. Hence, by Lemma 5.13 we have

$$R_1(\hat{\theta}) \le \sum_{i\neq 1} \left|R_i(\hat{\theta}) - R_i(\theta)\right| \le \sum_{i\neq 1} \sup_{\|\alpha-\theta\|\le\delta} \left|R_i(\alpha) - R_i(\theta)\right| = o_p\left(h^{2r}n^{\xi} + \delta h^r N^{\xi}\right),$$

where we currently take $\delta = O(1)$. In view of Lemma 5.13(R1), a method identical to that of Lemma 4.13 of Kulasekera and Lin (2005) gives the result.

(iii) By Lemma 5.11, $E\sum_{i\in\mathcal{S}}\left(\sum_{j\in\mathcal{G}} \kappa_{ij}(\theta)\epsilon_j\right)^2 L_{q,\theta}(X_i)I(|K_i(\theta) > a_0) = O(h^{-1})$. Hence $R_3(\theta) = O_p(N^{-1}h^{-1})$ from Lemma 5.6(ii). Since $\frac{1}{L_\theta} = \frac{1}{nb_0} + o_p(N^{-\frac{3}{2}+\xi})$, by Lemma 5.8,

$$R_3(\theta) = \frac{1}{nb_0}\sum_{i\in\mathcal{S}}\sum_{j\in\mathcal{G}} \kappa_{ij}^2(\theta)L_{q,\theta}(X_i)\epsilon_j^2 + \frac{1}{nb_0}M + o_p(n^{-\frac{3}{2}+\xi}), \tag{5.6}$$

where $M = \sum_{s\neq t} w_{st}\epsilon_s\epsilon_t$ with $w_{st} = \sum_{i\in\mathcal{S}} \kappa_{is}\kappa_{it}L_{q,\theta}(X_i)I(|K_i(\theta) > a_0)$. By Lemma 5.11, $EM^2 \le O(1)\sum_{s\neq t}\sum_{i\in\mathcal{S},j\in\mathcal{G}} E[\kappa_{is}(\theta)\kappa_{it}(\theta)\kappa_{js}(\theta)\kappa_{jt}(\theta)] = O(1)$. Hence $M = O_p(1)$. By Lemma 5.11, $E\left(\sum_{i\in\mathcal{S}} \kappa_{ij}^2(\theta)L_{q,\theta}(X_i)I(|K_i(\theta)| > a_0)\right)^2 = O(N^{-2}h^{-2})$. Thus, by Lemma 5.3, with $\mathcal{G}_k = \mathcal{G}\cap\mathcal{H}_k$, $k = 1, 2$,

$$R_3(\theta) = \frac{1}{nb_0}\sum_{k=1}^{2}\sum_{i\in\mathcal{S},j\in\mathcal{G}_k} \kappa_{ij}^2(\theta)L_{q,\theta}(X_i)\sigma_k^2 + O_p(N^{-1}). \tag{5.7}$$

Note that it suffices to examine the terms with $L_{q,\theta}(X_i) \neq 0$. Define index set $\mathcal{A} = \{i|L_{q,\theta}(X_i) \neq 0\}$. By Lemma 5.6(i) and the choice of h we have $K_i(\theta) = thf_\theta(\theta' X_i) + O_p(nh^{r+1})$ uniformly in i and hence, uniformly in $i \in \mathcal{A}, j$,

$$\kappa_{ij}(\theta) = \frac{K_h(\theta' X_j - \theta' X_i)}{thf_\theta(\theta' X_i)} + O_p(N^{-1}h^r). \tag{5.8}$$

Since $\sum_{j\in\mathcal{G}_k} K_h^2(\theta' X_j - \theta' X_i) = t_k h f_\theta(\theta' X_i) \int_{-1}^1 K^2(s)ds + O(Nh^2)$ uniformly in i, we have, for $k = 1, 2$, uniformly in i,

$$\sum_{j\in\mathcal{G}_k} \kappa_{ij}^2(\theta) = \sum_{j\in\mathcal{G}_k} \frac{K_h^2(\theta' X_j - \theta' X_i)}{\big(thf_\theta(\theta' X_i)\big)^2} + O_p(N^{-1}h^{r-1}) = \frac{t_k \int_{-1}^1 K^2(s)ds}{t^2 h f_\theta(\theta' X_i)} + O_p(N^{-1}).$$

The result follows by plugging the above into (5.7).

(iv) Easily we have $R_6(\theta) = O_p(N^{-1}h^{-1})$ since, by Lemma 5.11,

$$E\Big(\sum_{i\in\mathcal{S},j\in\mathcal{G}} L_{q,\theta}(X_i)\kappa_{ij}(\theta)I(|K_i(\theta) > a_0)\epsilon_i\epsilon_j\Big)^2 = O(h^{-2}). \tag{5.9}$$

Hence, with $\mathcal{F} = \mathcal{S}\cap\mathcal{G}$ and $\mathcal{A} = \{(i,j) \mid i\in\mathcal{S}, j\in\mathcal{G}, j\neq i\}$,

$$R_6(\theta) = -\frac{2}{nb_0}\Big(\sum_{i\in\mathcal{F}} \kappa_{ii}(\theta)L_{q,\theta}(X_i)\epsilon_i^2 + \sum_{(i,j)\in\mathcal{A}} \kappa_{ij}(\theta)L_{q,\theta}(X_i)\epsilon_i\epsilon_j\Big) + o_p(N^{-1}).$$

Since $E[\kappa_{ii}^2 L_{q,\theta}^2(X_i)I(|K_i(\theta) > a_0)] = O(N^{-2}h^{-2})$ uniformly in i, by Lemma 5.3 and (5.8), we have, with $\mathcal{F}_k = \mathcal{F}\cap\mathcal{H}_k$, $k = 1, 2$,

$$\sum_{i\in\mathcal{F}} \kappa_{ii}(\theta)L_{q,\theta}(X_i)\epsilon_i^2 = \frac{q_\theta K(0)}{th}(u_1\sigma_1^2 + u_2\sigma_2^2)\int_{-1}^1 L(s)ds + O_p(N^{-\frac{1}{2}}h^{-1}).$$

The result follows by observing that, with a method similar to that of the proof for Lemma 5.11(ii),

$$\sum_{(i,j)\in\mathcal{A}} \kappa_{ij}(\theta)L_{q,\theta}(X_i)\epsilon_i\epsilon_j = \sum_{(i,j)\in\mathcal{A}} \frac{K_h(\theta' X_j - \theta' X_i)}{thf_\theta(\theta' X_i)} L_{q,\theta}(X_i)\epsilon_i\epsilon_j + O_p(n^{-\frac{1}{2}}h^{-\frac{3}{2}}).$$

□

5.4 Proof of Theorem 2.3

Proof of Theorem 2.3.
Applying Theorem 2.2 repeatedly we get $T = \frac{2}{b_0 h}T_1 + \frac{1}{Nh}D_N + o_p(N^{-1+\xi})$ with $T_1 = \sum_{i=1}^N\sum_{j=1}^N c_{ij}w_{ij}\epsilon_i\epsilon_j$ and D_N given in (2.7). Here $w_{ij} = \frac{K_h(\theta' X_j - \theta' X_i)}{f_\theta(\theta' X_i)} L_{q,\theta}(X_i)$, $c_{ij} = 0$ for $i = j$ and

$$c_{ij} = \begin{cases} -\frac{1}{n_1 N} + \frac{1}{n_1^2} = \frac{n_2}{n_1^2 N}, & i,j = 1,\cdots,n_1; i\neq j; \\ -\frac{1}{n_1 N}, & i = 1,\cdots,n_1; j = n_1+1,\cdots,N; \\ -\frac{1}{n_2 N}, & i = n_1+1,\cdots,N; j = 1,\cdots,n_1; \\ -\frac{1}{n_2 N} + \frac{1}{n_2^2} = \frac{n_1}{n_2^2 N}, & i,j = n_1+1,\cdots,N; i\neq j. \end{cases}$$

Uniformly for $i \neq j$ (assume $i \in \mathcal{H}_k,\ j \in \mathcal{H}_s,\ k, s = 1, 2$), we have

$$E[w_{ij}^2 \epsilon_i^2 \epsilon_j^2] = E\Big(\frac{K_h(\theta' X_j - \theta' X_i)}{f_\theta(\theta' X_i)} L_{q,\theta}(X_i)\Big)^2 \sigma_k^2 \sigma_s^2$$
$$= \sigma_k^2 \sigma_s^2 \int \Big(\int K_h^2(u - v) f_\theta(v)\, dv\Big) L^2(\tfrac{u - c_\theta}{q_\theta}) \frac{1}{f_\theta(u)}\, du = a_{ks} h + O(h^2);$$

and, similarly we can show (again, assume $i \in \mathcal{H}_k,\ j \in \mathcal{H}_s,\ k, s = 1, 2$) $E[w_{ij} w_{ji} \epsilon_i^2 \epsilon_j^2] = a_{ks} h + O(h^2)$. Hence

$$\mathrm{Var}(T_1) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} (c_{ij}^2 + c_{ij} c_{ji}) a_{11} h + \sum_{i=1}^{n_1} \sum_{j=n_1+1}^{N} (c_{ij}^2 + c_{ij} c_{ji}) a_{12} h$$
$$+ \sum_{i=n_1+1}^{N} \sum_{j=1}^{n_1} (c_{ij}^2 + c_{ij} c_{ji}) a_{21} h + \sum_{i=n_1+1}^{N} \sum_{j=1}^{n_1} (c_{ij}^2 + c_{ij} c_{ji}) a_{22} h + O(N^{-2} h^2)$$
$$= \Big(\frac{2n_2^2}{n_1^2} a_{11} + (\frac{n_2}{n_1} + \frac{n_1}{n_2} + 1) a_{12} + \frac{2n_1^2}{n_2^2} a_{22}\Big) \frac{h}{N^2} + O(N^{-2} h^2).$$

By assumption (A4) and Lemma 5.11, the conditions of Lemma 5.1 are satisfied and thus the proof is completed. □

References

[1] Bowerman, B.L., and O'Connell, R.T. (1990), Linear Statical Models: An Applied Approach, Duxbury Press, California.

[2] Bowman, A., and Young, S. (1995), "Graphical Comparison of Nonparametric Curves," *Applied Statistics* 45, 83-98.

[3] Cubas, (2000)

[4] Dette, H., and Neumeyer, N. (2001), "Nonparametric Analysis of Covariance," *The Annals of Statistics,* 29, 1361-1400.

[5] Hall, P., and Hart, J.D. (1990), "Bootstrap Test for Difference Between Means in Nonparametric Regression," *Journal of the American Statistical Association,* 85, 1039-1049.

[6] Härdle, W., Hall, P., and Ichimura, H. (1993), "Optimal Smoothing in Single-Index Models," *The Annals of Statistics,* 21, 157-178.

[7] Härdle, W., and Marron, J.S. (1990), "Semiparametric Comparison of Regression Curves," *The Annals of Statistics,* 18, 63-89.

[8] Hart, J.D. (1997), Nonparametric Smoothing and Lack-of-Fit Tests, Springer, New York.

[9] Hristache, M., Juditsky, A., and Spokoiny, V. (2001), "Direct Estimation of the Index Coefficient in a Single-Index Model," *The Annals of Statistics,* 29, 595-623.

[10] Ichimura, H. (1993), "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models," *Journal of Econometrics,* 58, 71-120.

[11] King, E.C., Hart, J.D., and Wehrly, T.E. (1991), "Testing the Equality of Two Regression Curves Using Linear Smoothers," *Statistics & Probability Letters,* 12, 239-247.

[12] Klein, R.L., and Spady, R.H. (1993), "An Efficient Semiparametric Estimator for Binary Response Models," *Econometrica,* 61, 387-421.

[13] Kulasekera, K.B. (1995), "Testing the Equality of Regression Curves Using Quasi Residuals," *Journal of the American Statistical Association,* 90, 1085-1093.

[14] Kulasekera, K.B., and Lin, W. (2005), "On Variance Estimation of the Linear Single-Index Models," *Technical Report, Department of Mathematical Sciences, Clemson University.*

[15] Kulasekera, K.B., and Wang, J. (1997), "Smoothing Parameter Selection for Power Optimality in Testing of Regression Curves," *Journal of the American Statistical Association,* 92, 500-511.

[16] ——(2001), "A Test of Equality of Regression Curves using Gâteaux scores," *Australian & New Zealand Journal of Statistics,* 43, 89-100.

[17] Neumeyer, N., and Dette, H. (2003), "Nonparametric Comparison of Regression Curves: An Empirical Process Approach," *The Annals of Statistics,* 31, 880-920.

[18] Powell, J.L., Stock, J.H., and Stoker, T.M. (1989), "Semiparametric Estimation of Index Coefficients," *Econometrica,* 57, 1403-1430.

[19] Stoker, T.M. (1986), "Consistent Estimation of Scaled Coefficients," *Econometrica,* 54, 1461-1481.

[20] Stute, W., and Zhu, L.X. (2005), "Nonparametric Checks for Single-Index Models," *The Annals of Statistics,* 33, 1048-1083.

[21] Wu, C.F.Y. (1986), "Jacknife, Bootstrap and Other Resampling Methods in Regression Analysis (with discussion)", *The Annals of Statistics,* 14, 1261-1350.

[22] Xia, Y., Li, W.K., Tong, H., and Zhang, D. (2004), "A Goodness-of-Fit Test for Single-Index Models (with discussion)," *Statistica Sinica,* 14, 1-39.

[23] Yin, X., and Cook, R.D. (2005), "Direction Estimation in Single-Index Regression," *Biometrika,* 92, 371-384.

[24] Young, S.G., and Bowman, A.W. (1995), "Nonparametric Analysis of Covariance," *Biometrics,* 51, 920-931.

[25] Yu, Y., and Ruppert, D. (2002), "Penalized Spline Estimation for Partially Linear Single-Index Models," *Journal of the American Statistical Association,* 97, 1042-1054.