

# Error Variance Estimation and Testing for the Single-Index Model

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

## Abstract

Single-index models (SIMs) provide one way of reducing the dimension in regression analysis. The statistical literature has mainly focused on estimating the index coefficients, the mean function, and their asymptotic properties. In this paper we propose two tests for the hypothesis that a sample comes from a linear SIM. In the process we also develop an asymptotically normal estimator for the error variance in a SIM. The test statistics are based on the estimator of the error variance under the null hypothesis and they can be large in magnitude under the alternative hypothesis. We examine the asymptotic properties of the proposed methods and conduct an empirical study to evaluate their finite sample performances.

## 1 Introduction

In many situations experimenters collect multivariate observations from individuals. Regression analysis is a common tool that can be used to model such data. Suppose our data  $(Y_i, X_i)$  follow the classical regression model

$$Y_i = m(X_i) + \epsilon_i, \quad i = 1, \dots, n, \quad (1.1)$$

for some function  $m(\cdot)$  where the covariate  $X$ 's are  $p$ -dimensional vectors,  $Y$  is the response variable, and  $\epsilon$ 's are independent errors with zero mean and bounded common variance  $\sigma^2$ . We allow for both fixed and random covariate designs in this article. We say that the model follows a general single-index model (SIM) if  $m(x) = g(h(x, \theta))$ , where  $g(\cdot)$  is an unknown univariate smooth function and  $h(\cdot)$  is known up to a parameter  $\theta \in \mathbb{R}^p$ . For identifiability purposes, we typically assume  $\|\theta\| = 1$  with its first nonzero element being positive. The model is said to be linear if  $h(x, \theta) = \theta'x$  and the model is said to be partially linear if the mean function takes the form  $m(x) = \beta'x + g(\theta'x)$  with  $\beta \perp \theta$ . The SIM has been extensively investigated due to its appeal in its ability for dimension reduction. In most applications the user assumes

that the data follow a SIM and mainly focuses on estimating the index coefficients  $\theta$ . However, when not much is known in advance, it may be desirable to judge the appropriateness of the use of a SIM for a given experimental setting. In this paper we address this issue. Namely, we propose a test that can be used to check whether the data follow a SIM. In the process, if the SIM is the correct model, the proposed procedure gives an estimator of the error variance  $\sigma^2$  and the index vector  $\theta$  automatically.

Among the researchers who have studied the estimation of the index parameter, Ichimura (1993) gives a least squares estimator  $\hat{\theta}$  of  $\theta$  in SIM and proves the asymptotic normality of  $\hat{\theta}$  under conditions that are very similar to those assumed here. Härdle, Hall and Ichimura (1993) consider decomposing a squared error type quantity that is very similar to the one used by Ichimura (1993) to obtain an estimator of the index vector and the optimal bandwidth. However, these authors impose a set of stringent conditions on the index vector and the bandwidth. A LS estimator of  $(\beta, \theta)$  in a partially linear SIM is discussed by Xia, Tong and Li (1999) while Yu and Ruppert (2002) give a penalized spline estimator with a fixed number of knots. The spline method essentially assumes that the index function has a fixed representation for the given finite number of knots and estimates the coefficients of this representative function, thus reducing the problem to a finite dimensional setting. Their argument is that such representations incur smaller approximation errors compared with the estimation errors. Other methods of estimating the index vector in SIM include, for example, the semiparametric inverse regression method by Duan and Li (1991), the average derivative (AD) method suggested by Powell, Stock and Stoker (1989) and iterative AD method suggested by Hristache, Juditsky and Spokoiny (2001). In all these articles the main attempt was to establish the large sample properties of the proposed estimator of the index vector. Some methods are superior in computational aspects than others (see Yu and Ruppert, 2002).

In this article we propose tests to determine whether a given sample follows a linear SIM. That is we develop a test for

$$H_0 : m(x) = g(\theta'x) \text{ for some } \theta \text{ and } g,$$

against

$$H_a : \text{not } H_0,$$

where  $m$  is the regression function in (1.1) above. This is a generalization of the widely discussed hypotheses testing problem where one tests the linearity of the model; see Fan and Huang (2001) and references therein. In the process of developing the tests we obtain a natural, asymptotically normal estimator for the error variance in a SIM. The variance estimation in a SIM has not been formally addressed in the literature to our knowledge. The classical difference type estimators for error variance in nonparametric regression with several covariates use the closeness of the design points and the higher order smoothness of the regression function (Kulasekera and Gallagher 2002, Spokoiny, 2002) to obtain consistent estimators of the error variance. The asymptotic normality for such estimators is very difficult to establish (see Spokoiny, 2002). In a SIM, although the design points become univariate after the transformation  $\theta'X$ , the estimation of  $\theta$  is needed to implement a difference method. Then, the closeness of design points

cannot be guaranteed in order to have a diminishing effect of the signal in a difference estimator; see conditions for the proof of asymptotic normality of difference estimators in Hall et al., (1991). Hence one resorts to using estimators based on sum of squared errors. Asymptotic properties of such estimators of error variance in the context of SIMs have not been fully investigated.

Our test is motivated from the fact that if the null hypothesis is true, then  $d(\alpha) = E(Y - E(Y|\alpha'X))^2$  should be just the error variance for the true vector  $\theta$  and it would be larger than the error variance for all other  $\alpha$ . Moreover, if we minimize the above expectation with respect to  $\alpha$ , the minimized value will be larger than  $\sigma^2$  under the alternative hypothesis. Thus, by minimizing an estimator of  $d(\alpha)$ , we should be able to get a statistic that can indicate possible departures from the null hypothesis. In the following we first give an estimator  $\hat{d}(\alpha)$  of  $E(Y - E(Y|\alpha'X))^2$ . Then we show that  $\inf_{\alpha} \hat{d}(\alpha)$  is a  $\sqrt{n}$ -consistent estimator of the error variance  $\sigma^2$  under  $H_0$ , and is consistent for  $\sigma^2 + d_{m,L}$  for some constant  $d_{m,L} > 0$  under  $H_a$ . In fact, we show that  $\sqrt{n}[\inf_{\alpha} \hat{d}(\alpha) - \sigma^2]$  converges in distribution to a mean zero normal random variable under the null hypothesis. This type of results for variance estimation in a SIM has not been given prior to this to our knowledge.

If  $\sigma^2$  was known, then an obvious test statistic for the above hypotheses would be a normalized difference between the minimized value  $\inf_{\alpha} \hat{d}(\alpha)$  and  $\sigma^2$ . One would then reject the null hypothesis for large values of such a test statistic which is shown to be asymptotically normal under  $H_0$ . Since  $\sigma^2$  is typically unknown we use a slightly different approach to construct test statistics for the above hypotheses. In particular, we propose to construct two test statistics; one using two versions of  $\inf_{\alpha} \hat{d}(\alpha)$  based on two weight functions and the other exploiting the idea that under the null hypothesis, a re-projection of residuals resulting from the first projection of the responses does not significantly reduce the sum of squares of the errors. In the sequel we give a detailed description of the proposed ideas and a few practical guidelines for the user.

We organize the paper as follows. In Section 2.1, we give a mathematical motivation of the statistics for both random and fixed-design models. The asymptotic normality of the estimator of the error variance and each of the proposed test statistics under the null hypothesis is shown in Section 2.2. The results of a simulation study examining the finite sample behavior of the proposed techniques along with an application to a real data set are given in Section 3. The proofs of the results stated in the main body of the article use standard arguments. However, they involve tedious calculations. For the sake of completeness we provide detailed proofs in several Appendices.

## 2 Hypothesis Testing Problem

Suppose we have data  $(X_i, Y_i)$  coming from model (1.1), where the mean function  $m(\cdot)$  is a  $p$ -variate smooth function and the errors ( $\epsilon$ 's) are independent with zero mean and finite common variance  $\sigma^2$ . For simplicity we assume  $X$  to be bounded with probability one, and, without loss of generality, we may assume the support of  $m(\cdot)$  is contained in  $S$  where

$$S = \{x \in \mathbb{R}^p \mid \|x\| \leq 1\}.$$

We use the following precise definition of a single index model.

**Definition 2.1.** A  $p$ -variate function  $m(\cdot) \in L_2(S)$  is said to be a single-index function with index vector  $\theta$  if  $m(x) = g(\theta'x)$  a.e. for some  $g \in L_2([-1, 1])$  and  $\theta \in D$  where

$$D = \{\theta \in \mathbb{R}^p \mid \|\theta\| = 1 \text{ with first nonzero component positive}\}.$$

The restriction that  $\theta \in D$  is for identifiability purposes. Under the assumption that  $m$  is continuous and non-constant, we can show that if  $m$  is a single-index function, then the representation  $m(x) = g(\theta'x)$  is unique. Formally, we have

**Theorem 2.2.** *Suppose  $m(\cdot)$  is a non-constant continuous function on  $S$ . If*

$$m(x) = g(\alpha'x) = h(\beta'x), \quad \forall x \in S,$$

for some  $g, h \in C([-1, 1])$  and some  $\alpha, \beta \in D$ , then  $g = h$  and  $\alpha = \beta$ .

Xia, Tong and Li (1999) provide a proof of the uniqueness for partially linear SIM assuming that  $g$  is twice differentiable. We prove the above theorem only assuming a much weaker condition of continuity of  $g$ . This is given in the appendix. Now, restricting  $m(\cdot)$  to be continuous and non-constant, our hypothesis testing problem can be formally stated as

$$H_0 : m(x) \text{ is a single-index function} \quad \text{v.s.} \quad H_a : m(x) \text{ is not a single-index function.} \quad (2.1)$$

## 2.1 Motivation

In this section we try to provide motivation(s) and the mathematical reasoning for developing the variance estimator and test statistics for testing the hypotheses (2.1).

### 2.1.1 Random Design Case

For random design model, we may define the projection of  $m(\cdot)$  in the direction of  $\alpha$  by

$$g_\alpha(\alpha'x; m) = E(m(X) \mid \alpha'X = \alpha'x), \quad (2.2)$$

and let the remainder function be

$$m_\alpha(x) = m(x) - g_\alpha(\alpha'x; m).$$

We can quantify the difference between  $m(\cdot)$  and  $g_\alpha(\cdot; m)$  by

$$d(\alpha; m) = E\left(m_\alpha(X)\right)^2 = \int \left(m(x) - g_\alpha(\alpha'x; m)\right)^2 P_X(dx).$$

Later in the article we shall use notation  $g_\alpha(\cdot)$  and  $d(\alpha)$  instead of  $g_\alpha(\cdot, m)$  and  $d(\alpha, m)$  respectively, when there is no confusion. The properties of conditional expectation give us

$$d(\alpha; m) = \inf_{\psi \in L_2([-1, 1])} E\left(\psi(\alpha'X) - m(X)\right)^2,$$

where  $L_2([-1, 1])$  is the class of square integrable functions on  $[-1, 1]$ . A natural measure of the deviation from a single index model is given by

$$d_m = \inf_{\alpha \in D} d(\alpha; m).$$

In particular, suppose  $m$  is a non-constant continuous function, in which case we have shown that, if  $m$  is a single index function,  $m$  has a unique representation, say,  $m(x) = g(\theta'x)$ . Then it is seen that  $g_{\theta}(\theta'x) = g(\theta'x)$  and  $d_m = d(\theta; g(\theta'x)) = 0$ . In general, we have the following result.

**Lemma 2.3.** *Under the above notation, for any continuous function  $m(\cdot)$  on  $S$ ,  $d(\alpha; m)$  is continuous in  $\alpha$ . Under  $H_0$  and  $H_a$  above,  $d_m = 0$  and  $d_m > 0$  respectively.*

*Proof.* Notice that  $g_{\alpha}(\alpha'x) = E(m(X)|\alpha'X = \alpha'x)$  is continuous in  $\alpha$  (this has been assumed or used by Hall (1989) and Ichimura (1993) among others) and therefore  $d(\alpha; m)$  is continuous in  $\alpha$ . By the definition of  $D$ ,  $d_m$  can be attained at some point

$$\theta_m = \arg\{\inf_{\alpha} d(\alpha; m)\}. \quad (2.3)$$

Then  $d_m = d(\theta_m; m) = 0$  if and only if  $Em_{\theta_m}^2(X) = 0$ , which is true if and only if  $m(x) = g(\theta'_m x)$  for some  $g$ . This completes the proof.  $\square$

**Remark 2.4.** Notice that  $g_{\theta_m}(\theta'_m \cdot)$  is the first projective approximation of  $m(\cdot)$  as discussed by Hall (1989).

Hence, if we can estimate  $d_m$ , then a large value (compared to some standard, of course) of  $\hat{d}_m$  indicates the departure from a SIM and thus the rejection of  $H_0$ . We can estimate  $g_{\alpha}(\cdot)$  by a kernel estimator (see Ichimura, 1993)

$$\hat{g}_{\alpha}(u) = \frac{\sum_{i=1}^n Y_i K_h(\alpha'X_i - u)}{\sum_{i=1}^n K_h(\alpha'X_i - u)},$$

where  $K_h(t) = K(t/h)$  for some proper kernel function  $K$ , and  $h$  is a smoothing parameter that typically tends to 0 as the sample size increases. Now, we can construct a sample version of  $d(\alpha)$  such as  $\sum_{i=1}^n (Y_i - \hat{g}_{\alpha}(\alpha'X_i))^2 / n$  and minimize that to get estimators of  $\theta_m$  and the minimized distance. However, in these situations, the data can be sparse near the boundary of the domain of  $\alpha'X$  for a given  $\alpha$ . That is, if  $a_{\alpha} \leq \alpha'X \leq b_{\alpha}$ , we may have very few data points  $(Y_i, X_i)$  with corresponding  $\alpha'X_i$  values around a point  $u$  if  $u$  is close to either end of the interval  $[a_{\alpha}, b_{\alpha}]$ . Thus, we propose to use a weighted sum in constructing the sample version of  $d(\alpha)$  so that we avoid estimating  $g_{\alpha}$  near the boundaries of the corresponding domain of  $\alpha'X$ . In particular, we use a version of the form

$$\hat{d}(\alpha) = \frac{\sum_{i=1}^n (Y_i - \hat{g}_{\alpha}(\alpha'X_i))^2 L_{q,\alpha}(X_i)}{\sum_{i=1}^n L_{q,\alpha}(X_i)},$$

where  $L_{q,\alpha}(x)$  controls the  $x$ 's that are considered for the weighted sum. A formal description of  $\hat{d}(\alpha)$  using specific conditions for  $L$  is given in Section 2.2.

Now, let  $\hat{\theta} = \arg \inf_{\alpha} \hat{d}(\alpha)$ . Then a sample analog of  $d_m$  is given by  $\hat{d}(\hat{\theta})$ . In the next section we show that  $\hat{d}(\hat{\theta})$  has a decomposition  $\hat{d}(\hat{\theta}) = \sigma^2 + d_{m,L} + o_p(1)$  where  $d_{m,L}$  is a weighted version of  $d_m$  given by

$$d_{m,L} = \inf_{\alpha \in D} \frac{E(m_{\alpha}^2(X)L_{q,\alpha}(X))}{EL_{q,\alpha}(X)}.$$

Therefore, one can construct a useful estimator of the error variance of a SIM and check departures from the null hypothesis above by using a standardized statistic involving  $\hat{d}(\hat{\theta})$ .

### 2.1.2 Fixed Design Case

In the following we propose to modify the ideas used in the random design case for a fixed-design case. Instead of defining the projection function  $g_{\alpha}(\cdot)$  using conditional expectation, let  $g_{\alpha}(\cdot; m)$  minimize the  $L_2$  distance

$$\int_S (m(x) - \psi(\alpha'x))^2 dx.$$

To obtain an explicit expression of  $g_{\alpha}(\cdot)$ , let  $A$  be an orthogonal matrix with first row  $\alpha'$ . The transformation  $y = Ax$  yields

$$\begin{aligned} \int_S (m(x) - \psi(\alpha'x))^2 dx &= \int_S (m(A'y) - \psi(y_1))^2 dy \\ &= \int_{-1}^1 \left( \int_{\tilde{S}(y_1)} (m(A'y) - \psi(y_1))^2 dy_2 \cdots dy_p \right) dy_1, \end{aligned}$$

where  $\tilde{S}(t) = \{y \in S \mid y_1 = t\}$ . Hence it suffices to minimize the inner integral for every  $y_1$ , which gives

$$g_{\alpha}(y_1; m) = \frac{\int_{\tilde{S}(y_1)} m(A'y) dy_2 \cdots dy_p}{\int_{\tilde{S}(y_1)} dy_2 \cdots dy_p}. \quad (2.4)$$

Let  $m_{\alpha}(x) = m(x) - g_{\alpha}(\alpha'x; m)$  and

$$d(\alpha) = \int_S (m(x) - g_{\alpha}(\alpha'x; m))^2 dx,$$

and  $\theta_m = \arg \inf_{\alpha} d(\alpha)$ . Clearly  $\theta_m = \theta$  and  $\inf_{\alpha} d(\alpha) = 0$  under  $H_0$ . From (2.4),  $d(\alpha)$  is continuous in  $\alpha$  and thus, following similar arguments that lead to Lemma 2.3,  $d_m > 0$  under  $H_a$ .

Similar to the random design case, we seek a sample analog of  $d_m$ . Although  $g_{\alpha}$  is no longer the conditional expectation  $E(m(X)|\alpha'X)$  as in random-design models, the statistic

$$\hat{g}_{\alpha}(u) = \frac{\sum_{i=1}^n Y_i K_h(\alpha'x_i - u)}{\sum_{i=1}^n K_h(\alpha'x_i - u)}$$

is still a good estimator of  $g_\alpha(u)$ . This fact is justified by the next lemma, which is proved in the Appendix.

**Lemma 2.5.** *Under the above notation and a few regularity conditions (see Assumptions A1-A5),*

$$\sup_{\alpha \in D} \sup_{L_{q,\alpha}(x) \neq 0} \left| \frac{\sum_{i=1}^n m(x_i) K_h(\alpha' x_i - \alpha' x)}{\sum_{i=1}^n K_h(\alpha' x_i - \alpha' x)} - g_\alpha(\alpha' x) \right| = O(h^2),$$

where  $L_{q,\alpha}(\cdot)$  is the control kernel function given in Section 2.2.

Hence a sample analog of  $d_m$  can still be  $\hat{d}(\hat{\theta})$  as in random design models, where  $\hat{\theta} = \arg \inf_{\alpha} \hat{d}(\alpha)$  and we can develop the inference following similar steps as in the random design case.

## 2.2 Variance Estimator and Test Statistics

In this section we formally develop the variance estimator and the test statistics and provide their asymptotic distributions. The proofs of stated results are given in the Appendix.

Let the support of  $X$  be denoted by  $S_X$  and, for an  $\alpha \in D$ , let  $c_\alpha$  and  $2w_\alpha$  denote the center and width of the set  $\{\alpha' x \mid x \in S_X\}$ . We shall use  $X_i$  in the construction of  $\hat{d}(\alpha)$  only if  $\alpha' X_i$  is not too far away from the center  $c_\alpha$ . We use the following approach to eliminate boundary effects instead of using boundary kernels. First, fix a constant  $0 < q < 1$  as a width control parameter and let  $q_\alpha = q \cdot w_\alpha$ . Now, we consider all  $X$ 's such that  $\alpha' X \in [c_\alpha - q_\alpha, c_\alpha + q_\alpha]$  for a given  $\alpha$  in the construction of  $\hat{d}(\alpha)$  proposed in the previous section. In particular, let

$$\hat{d}(\alpha) = \frac{\sum_{i=1}^n \left( Y_i - \hat{g}_\alpha(\alpha' X_i) \right)^2 L_{q,\alpha}(X_i)}{\sum_{i=1}^n L_{q,\alpha}(X_i)}, \quad (2.5)$$

where

$$L_{q,\alpha}(x) = L\left(\frac{\alpha' x - c_\alpha}{q_\alpha}\right), \quad \forall x \in S_X,$$

for some kernel  $L$  supported on  $(-1, 1)$ . Notice that this choice of the kernel  $L$  enables us to consider the values of  $\alpha' X$  that are not too close to the boundary of the set  $\{\alpha' x \mid x \in S_X\}$ .

We shall make the following technical assumptions that are subsequently used in developing various properties of the proposed test statistic.

**(A1)** The mean function  $m(\cdot)$  is twice continuously differentiable and  $m(\cdot)$ ,  $m'(\cdot)$ ,  $m''(\cdot)$  are bounded by a generic constant  $C$ . Under  $H_0$ , the true index vector  $\theta \in D$  where, for the remainder of the paper,

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

**(A2)** In the random-design case, 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.  $f(\cdot)$  is twice continuously differentiable. The domain  $S_X \subset S$  of  $X$  is closed and convex, which contains at least one interior ball with radius  $w_0 > 0$ . Also, there exist constants  $0 < c_1 < c_2 < \infty$  such that

$$c_1 \leq f(x) \leq c_2, \quad \forall x \in S_X.$$

**(A3)** In the fixed-design case,  $S_x = S$  and  $(x_1, \dots, x_n)$  is a point set of low discrepancy (Niederreiter, 1992).

**(A4)** The error  $\epsilon$  is symmetric about zero with at least  $v$  ( $\geq 10$ ) moments and is independent of  $X$  in random-design models.

**(A5)** The kernel functions  $K(\cdot)$  and  $L(\cdot)$  are bounded, symmetric non-negative functions supported on  $(-1, 1)$ .  $K(\cdot)$  is continuously differentiable and  $L(\cdot)$  is Lipschitz continuous of order 1.  $L(t)$  is non-increasing in  $|t|$  and  $L(t) > 0$  for all  $t \in (-1, 1)$ . The bandwidth  $h$  is such that  $h = O(n^{-\beta})$  for some  $\beta \in (0, \frac{1}{3})$ .

The modification of  $D$  in the first assumption ensures that  $\theta$  is an interior point of  $D$ . This condition was implicitly used by Ichimura (1993), who assumed that  $\theta$  is an interior point of the parameter space  $\Theta_0$  which is a subset of the original set  $D$ .

A few immediate consequences of the above assumptions are summarized in the next lemma which is proved in the Appendix. Some of the results in this lemma have been used as assumptions in Ichimura (1993) and Härdle, Hall and Ichimura (1993) among others. However, it should be noted that the set of  $X$  values used in Ichimura (1993) and other LS estimation methods is assumed to be in a set that depends on the sample size, although the explicit construction of such a set has not been given. In our approach, by choosing the function  $L$ , we explicitly rule out points in the domain of the covariates that are close to the boundary and may cause the underlying densities of  $\alpha'X$  to be near zero. This is a technical condition that needs to be satisfied for asymptotic results. Ichimura (1993) and Hall (1989) used an indicator function  $I(x \in \mathcal{X})$  in place of  $L_{q,\alpha}(x)$ , where  $\mathcal{X}$  is some subset of our  $S_X$ , on which,  $f_\theta(\theta'x)$  is bounded away from zero. There the set  $\mathcal{X}$  was not explicitly given. It can be constructed as the intersection of all the sets  $S_\alpha$  where  $S_\alpha = \{x | L_{q,\alpha}(x) > 0\}$ . In some sense, we are utilizing more information to construct  $\hat{d}$  since for every  $\alpha$  we use a set  $S_\alpha$  and each  $S_\alpha$  contains more covariate values than  $\mathcal{X}$  above. Most literature on SIMs use moment conditions on the errors that are similar to ours.

The next lemma lists some technical results we need to prove the main theorems. The proof of the lemma is given in Appendix A.

**Lemma 2.6.** *Under Assumptions (A1)-(A5), we have*

- (i)  $g_\alpha(\alpha'x)$  and  $m_\alpha(\alpha'x)$  are twice continuously differentiable with respect to  $\alpha$ ;
- (ii)  $\sup_\alpha q_\alpha < 1$  and  $\sup_\alpha |c_\alpha \pm q_\alpha| < 1$ ;
- (iii) for  $q \geq 1 - w_0$ , there exists a ball  $B(x_0, v_0)$  such that  $\alpha'B(x_0, v_0) \subset (c_\alpha \pm q_\alpha)$  for all  $\alpha$ ;
- (iv) the density  $f_\alpha(t)$  is uniformly bounded away from zero, say larger than  $c_q$ , for all  $\alpha$  and  $t \in (c_\alpha - q_\alpha, c_\alpha + q_\alpha)$ ;

(v) there exists a positive constant  $C$  such that, uniformly for  $x \in S_X$ ,

$$|L_{q,\alpha_1}(x) - L_{q,\alpha_2}(x)| \leq C\|\alpha_1 - \alpha_2\|, \quad \forall \alpha_1, \alpha_2 \in D;$$

(vi) the density function  $\phi_\alpha(\cdot)$ , say, of  $\alpha'X_1 - \alpha'X_2$  satisfies

$$\phi_\alpha(u) \geq c_q^2 q_0, \quad \forall \alpha, \forall |u| < q_0/2,$$

where  $c_q > 0$  is the quantity specified in (iv) above and  $q_0 = qw_0$ .

It is noteworthy that Ichimura (1993) assumes that  $g_\alpha(\alpha'x)$  is three times continuously differentiable and Härdle et al. (1993) use the property that it is at least continuously differentiable. The following result gives the properties of the quantity  $\hat{d}(\alpha)$  when minimized with respect to  $\alpha$ . Note that this minimization is essentially the same as that in LS estimation of the single-index parameter (Ichimura, 1993). However, the results in the literature address only the properties of the resulting  $\hat{\theta}$  rather than the minimized squared error. It is seen that the minimization yields an asymptotically normal estimator for the error variance in a SIM.

**Theorem 2.7.** Let  $\hat{\theta}$  minimize  $\hat{d}(\alpha)$  over  $D$  and let  $h = O(n^{-\frac{1}{5}})$ . Then, under Assumptions (A1)-(A5) and for  $\xi > 0$  we have the following:

(i) under  $H_0$ ,

$$\hat{d}(\hat{\theta}) = \frac{\sum_{i=1}^n \epsilon_i^2 L_{\theta,q}(X_i)}{\sum_{i=1}^n L_{\theta,q}(X_i)} + o_p(n^{-\frac{4}{5}+\xi});$$

and  $\sqrt{n}(\hat{d}(\hat{\theta}) - \sigma^2) \xrightarrow{\mathcal{D}} N(0, \tau)$ , where, for random design models

$$\tau = \frac{\text{Var}(\epsilon^2) E[L_{q,\theta}^2(X)]}{[E[L_{q,\theta}(X)]]^2};$$

and for fixed design models

$$\tau = \frac{\text{Var}(\epsilon^2) \int_S L_{q,\theta}^2(x) dx}{[\int_S L_{q,\theta}(x) dx]^2}.$$

(ii) under  $H_a$ ,

$$\hat{d}(\hat{\theta}) = \frac{\sum_{i=1}^n \epsilon_i^2 L_{q,\theta_m}(X_i)}{\sum_{i=1}^n L_{q,\theta_m}(X_i)} + d_{m,L} + o_p(n^{-\frac{2}{5}+\xi}),$$

where, for random-design models

$$d_{m,L} = \inf_{\alpha \in D} \frac{E(m_\alpha^2(X) L_{q,\alpha}(X))}{E L_{q,\alpha}(X)};$$

and for fixed-design models

$$d_{m,L} = \inf_{\alpha \in D} \frac{\int m_\alpha^2(x) L_{q,\alpha}(x) dx}{\int L_{q,\alpha}(x) dx},$$

and  $\theta_m$  is the vector  $\alpha$  such that  $d_{m,L}$  is achieved.

**Remark 2.8.** Note that under  $H_a$ ,  $\hat{d}(\hat{\theta})$  above can be written as  $\hat{d}(\hat{\theta}) = \sigma^2 + d_{m,L} + o_p(n^{-\frac{2}{5}+\xi})$ .  $\hat{d}(\hat{\theta})$  is an estimator of  $\sigma^2$  that is similar to a minimized averaged squared error. The proof of this result involves several steps in which one has to systematically remove terms that can be ignored under the null hypothesis of a SIM. We give this proof in detail in the Appendix. The finite sample performance is evaluated using some simulations in the sequel.

Now, we turn to developing a test statistic for the above hypotheses based on  $\hat{d}(\hat{\theta})$ . Since  $\sigma^2$  is unknown,  $\hat{d}(\hat{\theta})$  cannot be used directly as a test statistic to test the hypothesis of a single index model. We can take one of the following two approaches to create a test statistic. Both approaches lead to *Von-Neumann* type statistics which are differences between two consistent variance estimators under the null hypothesis.

In the first we note that under the null hypothesis, minimization of  $\hat{d}(\alpha)$  with respect to  $\alpha$  for any weight function  $L$  produces a consistent estimator of  $\sigma^2$ . Hence we propose the following construction. Let  $L_{q,\alpha,1}$  and  $L_{q,\alpha,2}$  be any two suitably selected (different) weight functions in the a construction of  $\hat{d}(\alpha)$  in (2.5) above. Now, let

$$\hat{d}_r(\alpha) = \frac{\sum_{i=1}^n \left( Y_i - \hat{g}_\alpha(\alpha' X_i) \right)^2 L_{q,\alpha,r}(X_i)}{\sum_{i=1}^n L_{q,\alpha,r}(X_i)}, \quad r = 1, 2$$

and obtain  $\hat{d}_r(\hat{\theta}_r), r = 1, 2$  minimizing  $\hat{d}_r(\alpha), r = 1, 2$  respectively. Then, we propose to use a standardized form of

$$T = \sqrt{n}(\hat{d}_1(\hat{\theta}_1) - \hat{d}_2(\hat{\theta}_2))$$

as our test statistic. Although the two estimators  $\hat{d}_1(\hat{\theta}_1)$  and  $\hat{d}_2(\hat{\theta}_2)$  are not necessarily independent, it can be shown that, under the null hypothesis,  $T$  converges in distribution to a mean zero normal variable as  $n \rightarrow \infty$ . The statistic tends to be large in absolute value for departures from the null hypothesis as shown in the proof of the following theorem. Based on these observations, we propose to reject the hypothesis of a single index model for large absolute values of  $T$ . The asymptotic critical points for such a test can be obtained using a normal distribution according to the following theorem.

**Theorem 2.9.** *Let  $T$  be constructed as above with  $h = O(n^{-\frac{1}{5}})$ . Let  $d_{m,L,r}$  be versions of  $d_{m,L}$  defined in Theorem 2.7, with  $L_{q,\alpha}$  replaced by  $L_{q,\alpha,r}, r = 1, 2$ . Then under  $H_0$  and Assumptions (A1)-(A5) we have*

$$T \xrightarrow{\mathcal{D}} N(0, \tau'),$$

where

$$\tau' = \text{Var}(\epsilon^2) \cdot E \left( \frac{L_{q,\theta,1}(X)}{a_1} - \frac{L_{q,\theta,2}(X)}{a_2} \right)^2,$$

with  $a_r = E(L_{q,\theta,r}[X]), r = 1, 2$ ; and, under  $H_a$   $|T|$  diverges to infinity at the rate of  $\sqrt{n}$  provided  $d_{m,L,1} - d_{m,L,2} \neq 0$ .

Thus, our test would be to reject the null hypothesis if  $\frac{|T|}{\sqrt{\hat{\tau}'}} > z_{\alpha/2}$  where  $z_{\alpha/2}$  is the upper  $\alpha/2$  cutoff point in a standard normal distribution and  $\hat{\tau}'$  is a suitable consistent estimator of the asymptotic null variance  $\tau'$ . The choices of the two functions  $L_{q,\alpha,1}$  and  $L_{q,\alpha,2}$  are discussed in the sequel. It should be noted that the power of a test using  $T$  depends on the choice of these two functions, giving them a role similar to that of the scores in a linear rank statistic (Randles and Wolfe, 1979).

The second test is motivated by the fact that if the model is actually a SIM, then the residuals  $\hat{\epsilon}_i = Y_i - \hat{g}(\hat{\theta}'X)$  would be ‘close’ to the  $\epsilon_i$ ’s for large sample sizes and  $\hat{d}(\hat{\theta})$  given in Theorem 2.7 is a weighted mean squared error (MSE) after a first projection of the responses. Therefore, the difference between  $\hat{d}(\hat{\theta})$  and the MSE for projecting the  $\hat{\epsilon}_i$ ’s should be small under the null hypothesis. However, a projection of these  $\hat{\epsilon}_i$ ’s is almost the same as projecting a set of iid random variables. Thus, a reasonable MSE for the second projection (i.e., projection of  $\hat{\epsilon}_i$ ’s) is

$$W = \frac{\sum_{i=1}^n (\hat{\epsilon}_i - \bar{\epsilon})^2 I(X_i)}{\sum_{i=1}^n I(X_i)}, \quad (2.6)$$

where

$$\bar{\epsilon} = \frac{\sum_{i=1}^n \hat{\epsilon}_i I(X_i)}{\sum_{i=1}^n I(X_i)}, \quad (2.7)$$

and we take into account that  $\hat{\epsilon}_i$ ’s may have a nonzero mean in finite samples. Here  $I(x) = I(x \in E)$  for some set  $E \subset S$  chosen to avoid boundary problems. Under  $H_a$ ,  $\hat{\epsilon}_i$ ’s are not close to the errors  $\epsilon_i$ ’s and  $W$  is consistent for  $\sigma^2 + c$  for some constant  $c > 0$  different from  $d_{m,L}$  above. This leads to using a rescaled version of the statistic

$$T^* = \sqrt{n}(\hat{d}(\hat{\theta}) - W). \quad (2.8)$$

as a test statistic and rejecting the null hypothesis for large absolute values of it. The quantity  $\hat{d}(\hat{\theta}) - W$  under the null hypothesis can be thought of as the difference in the mean squares of the errors after the first and the second projection in a two step projection pursuit regression (PPR). Under the alternative hypothesis, this difference is estimating  $d_{m,L} - d_0$  where  $d_0 \approx E[m_{\theta_m}(X) - E(m_{\theta_m}(X))]^2$ . The critical points for a test based on  $T^*$  are obtained using a normal distribution and this result is summarized in the next theorem which is proved in the appendix.

**Theorem 2.10.** *Let  $T^*$  be constructed as above. Under  $H_0$ ,  $T^* \xrightarrow{D} N(0, \tau'')$  where*

$$\tau'' = \text{Var}(\epsilon^2) \cdot E\left(\frac{L_{q,\theta}(X)}{b_1} - \frac{I(X)}{b_2}\right)^2,$$

with  $b_1 = E[L_{q,\theta}(X)]$  and  $b_2 = E[I(X)]$  and under  $H_a$ ,  $T^*$  diverges to infinity with rate  $\sqrt{n}$ .

**Remark 2.11.** A major requirement of the proposed two tests above is their consistency. For  $T$ , this requires the existence of two suitable functions  $L_{q,\alpha,1}$  and  $L_{q,\alpha,2}$  such that the power of the proposed test converges to 1 under an alternative and that the

power is non-trivial for suitable local alternatives. Note that for a given  $L_{q,\alpha}$ ,  $d_{m,L}$  can be thought of as  $\inf_{\alpha} E_{\psi_{\alpha}}[m_{\alpha}^2(\tilde{X})]$  where  $\tilde{X}$  follows a density

$$\psi_{\alpha}(t) = \frac{L_{q,\alpha}(t)f(t)}{\int L_{q,\alpha}(t)f(t)dt}.$$

Now, if the two functions  $L_{q,\alpha,1}$  and  $L_{q,\alpha,2}$  produce very different expectations, the test based on such  $L$ 's would be consistent. The consistency of  $T^*$  follows from a similar argument.

**Remark 2.12.** The testing procedure automatically produces consistent estimators of the single index vector  $\theta$  if the null hypothesis is not rejected. The minimization method we use for calculating  $\hat{d}$  is very similar to the LS methods proposed in the literature. One may use the estimator  $\hat{\theta}$  in  $T^*$  directly or use  $\hat{\theta}_1$  and  $\hat{\theta}_2$  as good starting values for finding a final estimator of  $\theta$  when using  $T$ .

In practice a user has to decide or estimate several things in constructing the test statistics  $T$  and  $T^*$ . For  $T$ , these are  $h$  (the bandwidth for estimating  $g_{\alpha}$ ),  $q$  (the window width for weight functions  $L_{q,\alpha,r}$ ,  $r = 1, 2$ ),  $\tau'$  and the two weight functions. For  $T^*$ , the user has to select a  $L$ -function, the values  $h$  and  $q$  and estimate  $\tau''$ . Apart from choosing the weight function(s), it is not uncommon for a user to have to either select or estimate the smoothing parameters and estimate the null asymptotic variance in a nonparametric decision making problem. Almost all nonparametric regression methods inherit the smoothing parameter selection problem.

The value of  $h$  for the estimation of  $g_{\alpha}$  in constructing either test statistic can be chosen using one of the popular bandwidth selectors like GCV. Since we are estimating a univariate function for a given  $\alpha$  it would be effective to have a stochastic bandwidth selected using one of the popular yet computationally efficient methods. The value  $q$  itself is a bandwidth. The selection of  $q$  again has to be done to achieve some type of optimality in selecting the sample points to be included in the evaluation of  $\hat{d}$ . If each of the covariates  $X_i, i = 1, \dots, p$  has roughly the same domain, then the intersection of all  $p$  domains would provide a reasonable choice of  $q$ . If not, one may estimate  $\hat{\theta}$  from the whole sample and estimate  $w_{\theta}$  by  $w_{\hat{\theta}}$  and take  $q = (w_{\hat{\theta}} - h)/w_{\hat{\theta}}$  for the bandwidth  $h$  that is being used above. Since the estimator  $\hat{\theta}$  is consistent for  $\theta$ , the best projective vector (see proof of Theorem 2.7), the large sample properties of the test statistic would not change if we were to use a  $q$  based on an argument of this type. Data based smoothing parameter selection for power optimality is a difficult problem (Kulasekera and Wang, 1997 and 1998). A complete discussion of such issues is out of the scope of this paper.

The choice of the two functions  $L_{q,\alpha,1}$  and  $L_{q,\alpha,2}$  is a major part of developing  $T$ . The choice of the two functions  $L_{q,\alpha,r}$ ,  $r = 1, 2$  has no impact on the distribution of the test statistic under the null hypothesis. However, the power of the test is affected by selection of these functions. In a simple minded selection, one may choose the most important covariate and split the domain of the covariates into two parts based on the values of that covariate; such as low and high values and choose  $L$ 's to correspond to the indicators of these two sets. For example, for any  $L_{q,\alpha}$  we may define

$$L_{q,\alpha,r}(x) = L_{q,\alpha}(x) \cdot I(x \in S_r), \quad r = 1, 2,$$

where  $I(\cdot)$  is the indicator function. Alternatively, we can estimate the best projection vector by  $\hat{\theta}$  using the whole sample and choose two regions based on the values of  $\hat{\theta}'X$ . The null asymptotic properties would hold under such a partition due to the consistency of the estimator of the index vector  $\theta$ .

Finally, the estimation of the null asymptotic variance  $\tau'$  of  $T$  requires the estimation of  $\text{Var}[\epsilon^2]$  under  $H_0$ . One can fit a SIM for the whole data set and use the residuals of the best fit to estimate the variance of the square of the error. If any information on the error structure is known, one may use the residuals to enhance this primitive estimator.

## 3 Empirical Study

### 3.1 Simulation Results

We conducted a broad simulation study for assessing the finite sample properties of the proposed methods in a random-design model. Although we performed many calculations examining various aspects of the tests, we present only a selected few results showing the efficiency of the proposed procedures. In the sequel we comment on other aspects examined in this study without providing numerical details. The general model is written as  $Y = m(X) + \epsilon$  where the mean function  $m(x)$  takes the form

$$m(x) = g(\theta'x) + f(x).$$

The  $p$  covariates were taken to be independent with  $X_i$  being normal with mean  $i$  and standard deviation  $\frac{1}{2}$ ,  $i = 1, \dots, p$ , where we used  $p = 2, 3$  and  $4$ . The errors were taken to have  $\epsilon \sim N(0, \sigma^2)$  distributions. In the first part of the study, we took  $f(x) \equiv 0$  and examined the performance of  $\hat{d}(\hat{\theta})$  as the estimator of the error variance and the size of test statistics under  $H_0$ . In the second part we chose different types of functions for  $f$  and evaluate the empirical power of the tests. In all cases, we selected the true index vector  $\theta$  to be such that  $\theta_2 = \dots = \theta_p = \frac{1}{\sqrt{3p}}$  and  $\|\theta\| = 1$ . In this study we considered two  $g(\cdot)$  functions,

$$g_1(t) = 5 \sin(2\pi t), \quad \text{and} \quad g_2(t) = \frac{2}{p} t^2.$$

We used the quadratic kernel function  $K(u) = \frac{3}{4}(1-u^2)I(|u| \leq 1)$  and the  $L(\cdot)$  function

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

to create  $L_{q,\alpha}$  functions given in (2.5) for our simulations.

In evaluating the performance of  $\hat{d}(\hat{\theta})$  as an estimator of the error variance, we used the full sample with several  $q$  and  $\sigma^2$  values. The bandwidth  $h$  was chosen using a simple cross validation as described in the sequel. We used two  $g$  functions for several values of  $p$ . The results given in Table 1 are the averages and the sample standard deviations of  $\hat{d}(\hat{\theta})$  values for 500 simulations with  $q = 0.95$  and  $\sigma^2 = 0.09$  and  $0.25$ . All other results lead to similar conclusions and hence are omitted here.

For testing we constructed two versions of  $T$ , one we refer to as  $T_1$ , using the above  $L$  to create

$$L_{q,\alpha,1}(x) = L_{q,\alpha}(x) \cdot I(x_1 < 1) \quad \text{and} \quad L_{q,\alpha,2}(x) = L_{q,\alpha}(x) \cdot I(x_1 > 1).$$

where the quantities  $c_\alpha$  and  $w_\alpha$ , specified in Section 2.2, were estimated by

$$c_\alpha = \frac{1}{2}(\max_i \alpha' X_i + \min_i \alpha' X_i) \quad \text{and} \quad w_\alpha = \frac{1}{2}(\max_i \alpha' X_i - \min_i \alpha' X_i).$$

The second version of  $T$  (we call it  $T_2$ ) is constructed similarly but with

$$L_{q,\alpha,1}(x) = L_{q,\alpha}(x) \quad \text{and} \quad L_{q,\alpha,2}(x) = K_{q,\alpha}(x),$$

where  $K$  is the quadratic kernel given above. Finally,  $T^*$  was constructed using the  $L(\cdot)$  function given above and taking

$$I(x) = I(|x_1 - 1| < 0.9) \cdot I(|x_2 - 2| < 0.9) \cdot I(|x_3 - 3| < 1),$$

for calculating  $W$ . We used several values of  $q$ ; the results for  $q = 0.95$  will be presented here. Clearly the bandwidth  $h$  will play a major role in the size and the power of each test. In this simulation study, the bandwidth  $h$  for each of the three test statistics was chosen by a simple cross validation. For  $T^*$ , we selected  $h$  by minimizing  $\hat{d}_{cv}(h)$  which is the version of  $\hat{d}(\hat{\theta})$  where, in each term  $Y_i - \hat{g}_\alpha(\alpha' X_i)$ , the estimator  $\hat{g}(\cdot)$  is computed by leaving out the  $i$ th observation. For  $T_1$  and  $T_2$ , since there are two different  $\hat{d}(\hat{\theta})$ 's involved,  $h$  was selected by minimizing the sum of the two corresponding  $\hat{d}_{cv}(h)$ 's. For an asymptotic size 0.1 test, the null hypothesis was rejected when  $|\frac{T_i}{\sqrt{\hat{\tau}'_i}}| > 1.645$ ,  $i = 1, 2$  and  $|\frac{T^*}{\sqrt{\hat{\tau}''}}| > 1.645$  where  $\hat{\tau}'_i$  is the estimated asymptotic variance  $\tau'_i$  for  $T_i$ ,  $i = 1, 2$  and  $\hat{\tau}''$  is the estimated  $\tau''$  for  $T^*$ .

To examine the power of our tests under  $H_a$ , three  $f(\cdot)$  functions were considered corresponding to  $g_1$ , namely,

$$f_1(x) = x_1^2, \quad f_2(x) = 2 \cos(x_1) \quad \text{and} \quad f_3(x) = \frac{1}{3}e^{x_1};$$

and three other  $f(\cdot)$  functions were considered corresponding to  $g_2$

$$f_4(x) = -x_1^2, \quad f_5(x) = 2 \cos(x_1) \quad \text{and} \quad f_6(x) = -\frac{1}{2}e^{x_1}.$$

We ran 500 simulations for each combination of  $p = 2, 3, 4$ ,  $n = 50, 100, 200$  and  $g_i$ ,  $i = 1, 2$  for size calculations and for  $(g_i, f_j)$  given above for power calculations. We examined several values of  $q$  in each case and the results for  $q = 0.95$  are reported here. The results for other  $q$  values were very similar. All simulations were done using  $R$  language.

The empirical sizes are reported in Tables 2 and 3 and the empirical powers are reported in Tables 4, 5 and 6.

Results in Table 1 show that the estimator  $\hat{d}(\hat{\theta})$  performs well as an estimator of the error variance in a SIM for all dimensions we have examined. Apart from numerical difficulties, the accuracy does not seem to depend on  $p$  and the signal  $g$ .

	$n$	$g_1$		$g_2$	
		$\sigma^2 = 0.09$	$\sigma^2 = 0.25$	$\sigma^2 = 0.09$	$\sigma^2 = 0.25$
$p = 2$	50	.188(.097)	.300(.108)	.076(.020)	.206(.053)
	100	.138(.049)	.272(.071)	.083(.014)	.227(.036)
	200	.109(.025)	.254(.042)	.086(.010)	.237(.026)
$p = 3$	50	.171(.085)	.278(.106)	.074(.020)	.198(.051)
	100	.126(.044)	.258(.060)	.081(.014)	.220(.038)
	200	.108(.022)	.250(.038)	.085(.010)	.235(.027)
$p = 4$	50	.163(.079)	.257(.097)	.071(.021)	.186(.050)
	100	.127(.043)	.254(.060)	.079(.015)	.220(.038)
	200	.109(.028)	.248(.038)	.084(.010)	.231(.026)

Table 1: Mean(standard deviation) of  $\hat{d}(\hat{\theta})$  for  $m(x) = g_i(\theta'x)$ ,  $i = 1, 2$

Test Statistic	$n$	$p = 2$	$p = 3$	$p = 4$
$T_1$	50	.158	.166	.198
	100	.116	.126	.116
	200	.130	.116	.130
$T_2$	50	.130	.156	.122
	100	.146	.168	.122
	200	.186	.148	.184
$T^*$	50	.150	.106	.110
	100	.104	.090	.122
	200	.134	.106	.088

Table 2: Empirical Size for 0.1 Level Tests with  $m(x) = g_1(\theta'x)$

Test Statistic	$n$	$p = 2$	$p = 3$	$p = 4$
$T_1$	50	.152	.138	.154
	100	.138	.132	.118
	200	.112	.110	.130
$T_2$	50	.108	.102	.094
	100	.118	.136	.104
	200	.120	.116	.150
$T^*$	50	.086	.092	.082
	100	.084	.098	.084
	200	.098	.110	.104

Table 3: Empirical Size for 0.1 Level Tests with  $m(x) = g_2(\theta'x)$

		$f_1$			$f_2$			$f_3$		
	$n$	$p = 2$	$p = 3$	$p = 4$	$p = 2$	$p = 3$	$p = 4$	$p = 2$	$p = 3$	$p = 4$
$T_1$	50	.154	.190	.236	.238	.162	.142	.160	.162	.192
	100	.274	.434	.542	.310	.252	.252	.156	.204	.282
	200	.664	.870	.870	.562	.502	.466	.240	.384	.528
$T_2$	50	.184	.126	.160	.110	.124	.132	.134	.120	.120
	100	.238	.240	.214	.184	.168	.182	.178	.196	.182
	200	.312	.348	.356	.224	.230	.208	.254	.236	.312
$T^{*}$	50	.268	.294	.210	.218	.144	.154	.146	.170	.178
	100	.602	.496	.384	.434	.332	.186	.268	.260	.288
	200	.916	.852	.616	.752	.658	.404	.608	.570	.492

Table 4: Empirical Power for 0.1 Level Tests with  $m(x) = g_1(\theta'x) + f_i(x)$ ,  $i = 1, 2, 3$

		$f_4$			$f_5$			$f_6$		
	$n$	$p = 2$	$p = 3$	$p = 4$	$p = 2$	$p = 3$	$p = 4$	$p = 2$	$p = 3$	$p = 4$
$T_1$	50	.262	.244	.220	.442	.384	.232	.180	.222	.204
	100	.412	.350	.250	.674	.648	.378	.238	.250	.276
	200	.520	.570	.286	.830	.906	.662	.328	.328	.320
$T_2$	50	.098	.130	.096	.152	.174	.176	.114	.136	.128
	100	.126	.116	.124	.176	.150	.168	.143	.128	.148
	200	.116	.122	.142	.256	.190	.126	.172	.176	.190
$T^{*}$	50	.262	.336	.396	.526	.382	.186	.160	.192	.284
	100	.404	.606	.756	.798	.684	.250	.204	.350	.470
	200	.676	.908	.954	.970	.898	.424	.386	.562	.772

Table 5: Empirical Power for 0.1 Level Tests with  $m(x) = g_2(\theta'x) + f_i(x)$ ,  $i = 4, 5, 6$

	$n$	$p = 2$	$p = 3$	$p = 4$
$T_1$	50	.136	.202	.332
	100	.142	.186	.486
	200	.144	.174	.526
$T_2$	50	.116	.190	.158
	100	.156	.228	.256
	200	.150	.314	.446
$T^{*}$	50	.132	.126	.168
	100	.098	.216	.250
	200	.144	.346	.434

Table 6: Empirical Power for 0.1 Level Tests with  $m(x) = g_2(\theta'x) + f_1(x)$

In examining the size results, it appears that both tests seem to have reasonable size properties in general. The empirical size becomes close to the nominal size in most cases as the sample size increases. In  $T_2$ , the size seems to increase slightly when  $n$  gets large. This is most likely due to selecting a sub optimal bandwidth leading to poor estimation of the underlying function  $g$ , thus the test is mistakingly finding a departure from the null model although the departure is due to the bias of estimation of  $g$ . Specifically, the finite sample performance of all tests, either size or power, depends on the choice of  $L_{q,\alpha,r}$ ,  $r = 1, 2$ , and the selection of the bandwidth parameter  $h$ . In our simulations it was seen that there always exists a set of  $h$  values (unless the noise-signal ratio is very small) such that all three finite sample null distributions of  $T_1$ ,  $T_2$  and  $T^*$  are close to a standard normal distribution. However, these  $h$  values were not all in the same range for the three tests and therefore it was not easy to find the most suitable  $h$  for all the tests. That lead to using the ad hoc CV criterion that seem to work for most but not all cases we examined. As the sample size increases, the set of  $h$  values that produced test statistic values close to standard normal samples differed somewhat from those selected by our CV criterion in some cases, especially for  $T_2$ . The CV criterion used here produces a stochastic  $h$  that has its own asymptotic properties, which may not be optimal for testing a hypothesis. Thus it is conceivable that as the sample size increases the size of the test departs further from the nominal level due to a systematic bias of estimation of  $g$  caused by the wrong  $h$ .

Examining the tables for power, all the tests seem to detect departures from the null reasonably well, the power increasing as the sample size does.  $T^*$  seems to have the largest power in most cases examined. It is noteworthy that, the closer  $L_{q,\alpha,1}$  is to  $L_{q,\alpha,2}$ , the smaller is  $|d_{m,L,1} - d_{m,L,2}|$  and therefore the power. In some sense the two  $L_{q,\alpha,r}$  functions in  $T_1$  are as different as they could be, due to the indicator functions over two disjoint sets. But the functions  $L$  and  $K$  differ little. As can be seen from Tables 3 and 4, the power of  $T_1$  is higher than that of  $T_2$ .

However, in constructing  $T_1$  in the simulation study we actually partitioned the known domain of  $X$ . In practice this partition (and thus the choice of indicator functions) has to be data driven unless the domain of  $X$  is known beforehand. It is clear that this partition should be made as optimally as possible so that the two versions of  $\hat{d}(\hat{\theta})$  would be able to detect departures from the null effectively. At the same time, the partition should include sufficient data in each group so that artifacts of the partition will not lead to the rejection of  $H_0$  falsely.

One more noteworthy aspect of these power simulations is the following. Although we generated the data from a model of type  $m(x) = g(\theta'x) + f(x)$ , it should be noted that the departure of the model  $m$  from the class of SIMs is not specified by the function  $f(\cdot)$  itself. Instead, it is the remainder function from the first projection of  $m$  onto the class of all SIM's and this remainder may not necessarily be  $f$ . It is this remainder function that determines the power (for a specific test). We plotted the quantities  $m(x) = g_2(\theta'x) + f_i(x)$  for  $i = 1$  and  $i = 5$  functions against  $u = \theta'X_i$  for a sample of 200 values of  $X_i$ 's. It can be seen in Figure 1 that, although  $f_1$  and  $f_5$  are comparable in their magnitudes when plotted against  $\theta'X$ , the behavior is different when added to  $g_2$ . It is a lot easier to detect the departure of the function  $g_2(\theta'x) + f_5(x)$  from the class of SIM's than the departure of the function  $g_2(\theta'x) + f_1(x)$  from the SIMs.

For the same reason the power does not have to decrease as the dimension  $p$  increases because the remainder function may be more significant for large  $p$ . This can also be seen from Tables 3 and 4.

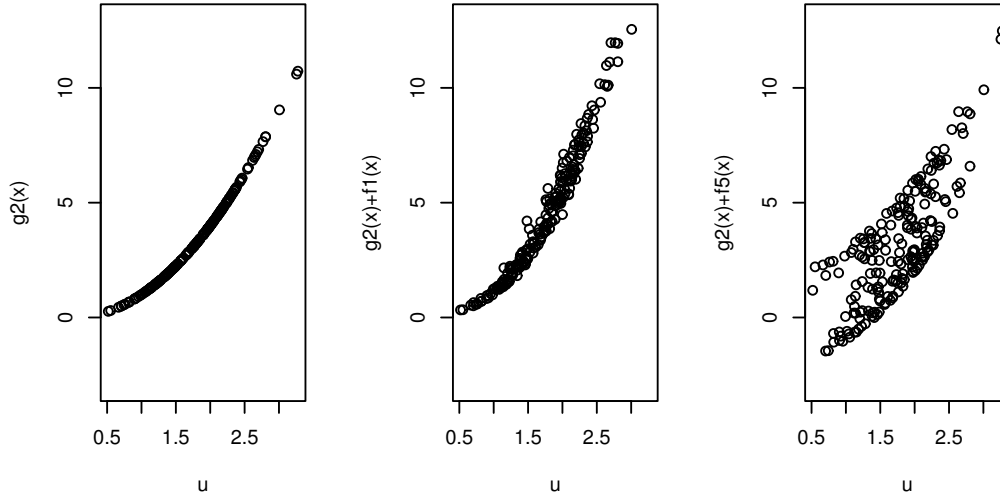


Figure 1: Plots of  $g_2 + f_1$  (panel 2) and  $g_2 + f_5$  (panel 3) against  $\theta'X$

### 3.2 Data Analysis

We analyzed a real data set using the proposed method. These were collected at a semi-urban hospital in Greenville, SC, for a study on the growth patterns of babies of drug addicted mothers. In the whole data collection experimenters were interested in comparing the growth patterns of children of crack cocaine addicted mothers against those of normal mothers. In this part of our analysis we use the data for the babies of crack addicted mothers. The data set includes 111 cases where the birth height  $H$  of each baby is recorded against several covariates including  $x_1$  : mother's weight gain during pregnancy and  $x_2$  : mother's age at delivery. Our goal in this analysis is to see if a model of type  $H = g(\theta_1 x_1 + \theta_2 x_2) + \epsilon$  will fit for the data  $(H, x_1, x_2)$ .

In order to test the hypothesis that the model function is a SIM, we took  $q = 0.95$  with the  $L$  function and the kernel function  $K$  as given in the simulation section above. To pick a suitable bandwidth, we used cross validation as described in the simulations for each test. The calculated test statistics (after dividing by the estimated asymptotic standard deviation) and the corresponding bandwidth pairs for the three tests  $T_1, T_2$  and  $T^*$  described above were  $(0.806, 2.53), (-0.379, 3.00)$  and  $(0.263, 2.036)$ , none leading to a rejection of the null hypothesis that the data follow a SIM. The

estimated index vector was  $\hat{\theta} = (.85, .53)'$  obtained from  $T^*$  with the error variance estimated as  $\hat{d}(\hat{\theta}) = 8.06$ . Other choices of  $q$  ranging from 0.85 – 0.95 produced similar results. A scatter plot of  $u = \hat{\theta}'x$  versus the heights as well as the lowest smoothed curve  $\hat{g}$  is given in Figure 2.

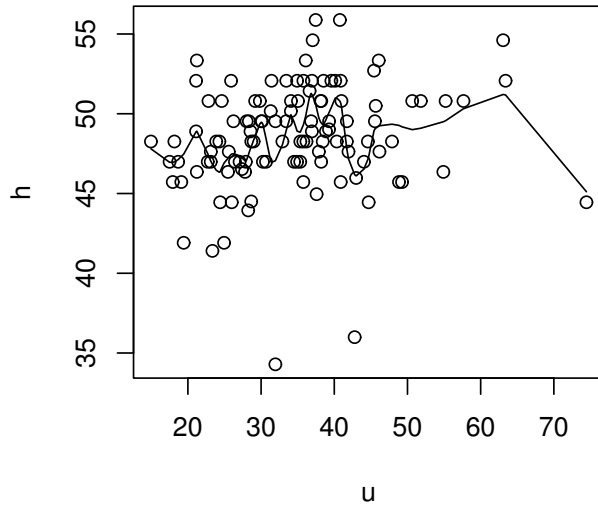


Figure 2: Smoothing of Heights against  $u = \hat{\theta}'x$  in Growth Data

## 4 Conclusion

In this article we developed an estimator for the error variance  $\sigma^2$  in a SIM and two tests to check whether a given set of responses follow a SIM. The estimator for  $\sigma^2$  has the classical *root* –  $n$  properties including the asymptotic normality. The tests for checking whether the model is a SIM were shown to be consistent against local alternatives of order  $n^{-1/4}$  where the test statistics are asymptotically normal under the null hypothesis. The tests require user specified quantities  $q$  and  $h$  where our simulations show the role of  $q$  in both power and size is almost the same as long as  $q$  is taken to be close to 1. The simulations indicated that these tests have reasonable power and good size in a variety of situations.

A major requirement for these methods is the choice of the  $L$  functions. The role of these functions is similar to the scores in many linear rank statistics analyses. These scores essentially decide whether the tests have non trivial power against certain alternatives. Of course for functions  $m$  for which  $d_{m,L,1} - d_{m,L,2} = 0$  for a given  $(L_1, L_2)$ ,

the power of the first test is trivial. This is also the same for the second test for a similar condition. A way to avoid trivial power in the first test would be to decide on  $L$  functions based on data. Alternatively, one may examine several statistics of type  $T_1$  as in the simulations where one decomposes the domain of each covariate and get a test statistic  $T_i, i = 1, \dots, p$  and uses  $\tilde{T} = \min_i T_i$ . However, it would be difficult to obtain the sampling properties of such a combination.

## Appendix

### A Proofs of Theorem 2.2 and Lemma 2.6

We shall denote the ball centered at  $x$  with radius  $r$  by  $B(x, r)$ , the dimension of which is the same as  $x$ . For dimension matching vector  $\alpha$ , matrix  $A$  and set  $E$ , let  $\alpha'E$  denote the set  $\{\alpha'x | x \in E\}$ , and let  $AE$  denote the set  $\{Ax | x \in E\}$ .

#### *Proof of Theorem 2.2.*

First we show  $g = h$ . If  $\alpha = \beta$ , since  $\{\alpha'x | x \in S\} = [-1, 1]$ , it is clear that  $g = h$  on  $[-1, 1]$ . Now suppose  $\alpha \neq \beta$ . Let  $\theta = \frac{\alpha + \beta}{\|\alpha + \beta\|}$ . Then  $\alpha'\theta = \beta'\theta$ . Put

$$u = \alpha'\theta = \beta'\theta = \sqrt{\frac{1 + \alpha'\beta}{2}}.$$

Since  $\alpha \neq \pm\beta$ , we have  $u \in (0, 1)$ . It is seen that  $\alpha'\beta = 2u^2 - 1$ . Take any  $c \in [-1, 1]$ . Then

$$g(cu) = g(\alpha'(c\theta)) = h(\beta'(c\theta)) = h(cu),$$

giving  $g(t) = h(t), \forall t \in [-u, u]$ . Now consider  $t \in (u, 1]$  and put

$$a(t) = \frac{t - u}{1 - u} \in (0, 1].$$

Let

$$x_t = a(t)\alpha + (1 - a(t))\theta, \quad \text{and} \quad y_t = a(t)\beta + (1 - a(t))\theta.$$

Clearly  $\alpha'x_t = \beta'y_t = t$  and  $\beta'x_t = \alpha'y_t$ . Now, let

$$b(t) = \beta'x_t = \alpha'y_t = a(t)\alpha'\beta + (1 - a(t))u = -(1 + 2u)t + 2u(1 + 2u).$$

By definition,

$$g(t) = g(\alpha'x_t) = h(\beta'x_t) = h(b(t)),$$

and

$$h(t) = h(\beta'y_t) = g(\alpha'y_t) = g(b(t)).$$

Hence  $g(t) = h(t)$  if  $b(t) \in [-u, u]$ . However, note that

$$b(t) \leq b(u) = u, \quad \forall t \in (u, 1],$$

and  $b(t) \geq -u$  if and only if  $t \leq u + \frac{2u}{2u+1}$ . This proves  $g(t) = h(t)$  for  $t \in (u, u + v]$  where  $v = \frac{2u}{2u+1}$ . By symmetry we obtain

$$g(t) = h(t), \quad \forall t \in [-u_1, u_1],$$

where  $u_1 = u + \frac{2u}{2u+1}$ . Provided  $u_{i-1} < 1$ , repeating the above procedure we obtain

$$g(t) = h(t), \quad \forall t \in [-u_i, u_i],$$

where

$$u_i = u_{i-1} + \frac{2u_{i-1}}{2u_{i-1} + 1} \geq u_{i-1} + v \geq u + iv.$$

Since  $v > 0$ , eventually  $u_i$  will exceed 1 and thus  $g = h$  on  $[-1, 1]$ .

Now we show  $\alpha = \beta$ . By the requirement that the first nonzero-components of  $\alpha$  and  $\beta$  are positive, it's not possible that  $\alpha = -\beta$ . Hence if  $\alpha \neq \beta$ , then  $|\alpha'\beta| < 1$  and we have, noting that  $h$  is continuous, for all  $t \in [-1, 1]$ ,

$$g(t) = g(\alpha'(t\alpha)) = h(\beta'(t\alpha)) = h(t(\alpha'\beta)) = \dots = h(t(\alpha'\beta)^n) = h(0),$$

which is a constant function. This contradicts the assumption that  $m(\cdot)$  is not constant on  $S$ .  $\square$

**Proof of Lemma 2.6.**

- (i) It's straight forward by writing the two terms in the form of integral. This has been used by Ischimura (1993) among others.
- (ii) Easily  $q_\alpha = q \cdot w_\alpha \leq q < 1$ . For the second result, since  $B(x_0, w_0) \subset S_X$ , we get  $\alpha'B(x_0, w_0) \subset (c_\alpha \pm w_\alpha)$ . Hence  $\inf_\alpha w_\alpha \geq w_0$ . Now it suffices to note that  $(c_\alpha \pm q_\alpha) \subset (c_\alpha \pm w_\alpha) \subset (-1, 1)$  and  $w_\alpha - q_\alpha = (1 - q)w_\alpha \geq (1 - q)w_0$ .
- (iii) Take  $v_0 < q - (1 - w_0)$ . If  $\alpha'B(x_0, v_0) \not\subset (c_\alpha \pm q_\alpha)$ , then

$$w_0 - v_0 \leq w_\alpha - q_\alpha = (1 - q)w_\alpha \leq 1 - q,$$

which is a contradiction.

- (iv) Let  $A$  be an orthogonal matrix whose first row is  $\alpha'$ . Let  $Z = AX$  and  $S_Z = AS_X$ . Then  $\alpha'X = Z_1$  and

$$f_\alpha(z_1) = \int_{S(z_1)} f(A'z) dz_2 \cdots dz_p \geq c_1 \int_{S(z_1)} dz_2 \cdots dz_p = c_1 \tilde{S}(z_1),$$

where  $S(t) = \{(z_2, \dots, z_p) | (t, z_2, \dots, z_p) \in S_Z\}$  and  $\tilde{S}(t)$  is the area of the hyperplane  $\{z \in S_Z | z_1 = t\}$  which equals the area of the hyperplane  $\{x \in S_x | \alpha'x = t\}$ . We now show this area  $\tilde{S}(t)$  is bounded away from zero uniformly in  $\alpha$  and  $t \in (c_\alpha \pm q_\alpha)$ . It suffices to show there exists  $B(x_{\alpha,t}, r_{\alpha,t}) \subset S_X$  with

$$\alpha'x_{\alpha,t} = t \quad \text{and} \quad \inf_\alpha \inf_{t \in c_\alpha \pm q_\alpha} r_{\alpha,t} > 0.$$

Take the ball  $B(x_0, w_0)$  and, without loss of generality, suppose  $\alpha'x_0 \leq t$ . Let  $x_\alpha \in S_X$  be such that  $\alpha'x_\alpha = c_\alpha + w_\alpha$ . Then the desired ball is tangent interiorly to the cone constructed from  $B(x_0, w_0)$  and  $x_\alpha$ , which is specified by

$$x_{\alpha,t} = \frac{c_\alpha + w_\alpha - t}{c_\alpha + w_\alpha - \alpha'x_0}x_0 + \frac{t - \alpha'x_0}{c_\alpha + w_\alpha - \alpha'x_0}x_\alpha,$$

and

$$r_{\alpha,t} = \frac{c_\alpha + w_\alpha - t}{c_\alpha + w_\alpha - \alpha'x_0}w_0.$$

The uniform lower bound follows since  $c_\alpha + w_\alpha - t \geq w_\alpha - q_\alpha \geq (1 - q)w_0$  and  $c_\alpha + w_\alpha - \alpha'x_0 \leq 2$ . The fact that  $B(x_{\alpha,t}, r_{\alpha,t}) \subset S_X$  follows from the convexity of  $S_X$ .

(v) Let

$$l(\alpha) = \inf_{x \in S_X} \alpha'x \quad \text{and} \quad u(\alpha) = \sup_{x \in S_X} \alpha'x.$$

Then  $c_\alpha = (l(\alpha) + u(\alpha))/2$  and  $w_\alpha = (u(\alpha) - l(\alpha))/2$ . Note that

$$|l(\alpha_1) - l(\alpha_2)| = |\inf \alpha_1'x - \inf \alpha_2'x| \leq \sup |\alpha_1'x - \alpha_2'x| \leq \|\alpha_1 - \alpha_2\|,$$

gives that  $l(\alpha)$  is Lipschitz continuous with order 1 and so is  $u(\alpha)$ . Hence, noticing  $\inf_\alpha w_\alpha \geq w_0$ ,

$$|L_{q,\alpha_1}(x) - L_{q,\alpha_2}(x)| \leq C \left\| \frac{\alpha_1'x - c_{\alpha_1}}{qw_{\alpha_1}} - \frac{\alpha_2'x - c_{\alpha_2}}{qw_{\alpha_2}} \right\| \leq C_1 \|\alpha_1 - \alpha_2\|,$$

for some constants  $C, C_1$  uniformly in  $x \in S_X$  and in  $\alpha_1, \alpha_2$ .

(vi) By (ii) we have

$$\phi_\alpha(u) = \int f_\alpha(u+v)f_\alpha(v)dv \geq \int_{c_\alpha - \frac{q_0}{2}}^{c_\alpha + \frac{q_0}{2}} f_\alpha(u+v)c_q dv \geq \int_{c_\alpha - \frac{q_0}{2}}^{c_\alpha + \frac{q_0}{2}} c_q^2 dv = c_q^2 q_0,$$

provided  $|u| < q_0/2$ .

□

## B Decomposition of $\hat{d}(\alpha)$

Proofs of Theorems 2.7, 2.9 and 2.10 require a few results that examine the convergence properties of  $\hat{d}(\alpha)$ . We state these as technical lemmas. To simplify the presentation, we introduce some notation first. For fixed choice of kernel function  $K$ , bandwidth parameter  $h$ , index-vector  $\alpha$  and  $i = 1, \dots, n$ , let

$$\begin{aligned} K_i(\alpha) &= \sum_{j=1}^n K_h(\alpha'(X_j - X_i)); & L_\alpha &= \sum_{i=1}^n L_{\alpha,q}(X_i); \\ \kappa_{ij}(\alpha) &= K_h(\alpha'(X_j - X_i))/K_i(\alpha); & l_{i\alpha} &= \frac{L_{\alpha,q}(X_i)}{L_\alpha}; \\ K_{mi}(\alpha) &= \sum_{j=1}^n \kappa_{ij}(\alpha)m(X_j); & \sum_{i\alpha} c_i &= \sum_{i=1}^n c_i l_{i,\alpha}, \end{aligned}$$

where  $\{c_i\}$  above is any sequence. Noticing  $m(X_i) = m_\alpha(X_i) + g_\alpha(\alpha' X_i)$  and

$$\hat{g}_\alpha(\alpha' X_i) = \sum_{j=1}^n \kappa_{ij}(\alpha) Y_j = K_{mi}(\alpha) + \sum_{j=1}^n \kappa_{ij}(\alpha) \epsilon_j,$$

we write

$$\begin{aligned} \hat{d}(\alpha) &= \sum_{i\alpha} \left( Y_i - \hat{g}_\alpha(\alpha' X_i) \right)^2 = \sum_{i\alpha} \left( m_\alpha(X_i) + (g_\alpha(\alpha' X_i) - K_{mi}(\alpha)) + \epsilon_i - \sum_{j=1}^n \kappa_{ij}(\alpha) \epsilon_j \right)^2 \\ &= \sum_{i\alpha} \epsilon_i^2 + \sum_{i\alpha} m_\alpha^2(X_i) + \sum_{i\alpha} \left( g_\alpha(\alpha' X_i) - K_{mi}(\alpha) \right)^2 + \sum_{i\alpha} \left( \sum_{j=1}^n \kappa_{ij}(\alpha) \epsilon_j \right)^2 \\ &\quad + 2 \sum_{i\alpha} m_\alpha(X_i) \epsilon_i + 2 \sum_{i\alpha} \left( g_\alpha(\alpha' X_i) - K_{mi}(\alpha) \right) \epsilon_i - 2 \sum_{i\alpha} \sum_{j=1}^n \kappa_{ij}(\alpha) \epsilon_i \epsilon_j \\ &\quad + 2 \sum_{i\alpha} m_\alpha(X_i) \left( g_\alpha(\alpha' X_i) - K_{mi}(\alpha) \right) - 2 \sum_{i\alpha} m_\alpha(X_i) \left( \sum_{j=1}^n \kappa_{ij}(\alpha) \epsilon_j \right) \\ &\quad - 2 \sum_{i\alpha} \left( g_\alpha(\alpha' X_i) - K_{mi}(\alpha) \right) \left( \sum_{j=1}^n \kappa_{ij}(\alpha) \epsilon_j \right) \\ &= \sum_{i=0}^9 R_i(\alpha); \end{aligned}$$

where

$$\begin{aligned} R_0(\alpha) &= \sum_{i\alpha} \epsilon_i^2; & R_1(\alpha) &= \sum_{i\alpha} m_\alpha^2(X_i); \\ R_2(\alpha) &= \sum_{i\alpha} \left( g_\alpha(\alpha' X_i) - K_{mi}(\alpha) \right)^2; & R_3(\alpha) &= \sum_{i\alpha} \left( \sum_{j=1}^n \kappa_{ij}(\alpha) \epsilon_j \right)^2; \\ R_4(\alpha) &= 2 \sum_{i\alpha} m_\alpha(X_i) \epsilon_i; & R_5(\alpha) &= 2 \sum_{i\alpha} \left( g_\alpha(\alpha' X_i) - K_{mi}(\alpha) \right) \epsilon_i; \\ R_6(\alpha) &= -2 \sum_{i\alpha} \sum_{j=1}^n \kappa_{ij}(\alpha) \epsilon_i \epsilon_j; & R_7(\alpha) &= 2 \sum_{i\alpha} m_\alpha(X_i) \left( g_\alpha(\alpha' X_i) - K_{mi}(\alpha) \right); \\ R_8(\alpha) &= -2 \sum_{i\alpha} m_\alpha(X_i) \left( \sum_{j=1}^n \kappa_{ij}(\alpha) \epsilon_j \right); \end{aligned}$$

and

$$R_9(\alpha) = -2 \sum_{i\alpha} \left( g_\alpha(\alpha' X_i) - K_{mi}(\alpha) \right) \left( \sum_{j=1}^n \kappa_{ij}(\alpha) \epsilon_j \right).$$

Then we have the following two lemmas, Lemma 1 and Lemma 2, establishing the large sample properties of  $R_i(\alpha)$ ,  $i = 0, \dots, 9$  and  $R_i(\hat{\theta})$ ,  $i = 0, \dots, 9$ . Once these results are established, we can examine the large sample behavior of  $\hat{d}(\hat{\theta})$ .

**Lemma 1.** Suppose Assumption (A1)-(A5) holds. Then under  $H_0$  with  $\theta$  being the true index vector, we have that, for any  $\xi > 0$ ,

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

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

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

**R2)**  $\sup_{\alpha \in D} |R_2(\alpha)| = o_p(h^4 n^\xi);$

**R3)**  $\sup_{\alpha \in D} |R_3(\alpha)| = o_p(n^{-1 + \xi} h^{-1});$

**R4+R5)**  $\sup_{\|\alpha - \theta\| \leq \delta} |R_4(\alpha) + R_5(\alpha)| = o_p(n^{-\frac{1}{2} + \xi} (\delta + h^2));$

**R6)**  $\sup_{\|\alpha - \theta\| \leq \delta} |R_6(\alpha)| = o_p(n^{-\frac{1}{2} + \xi} h^{-\frac{1}{2}} \delta + n^{-1 + \frac{2}{v} + \xi} h^{-2} \delta);$

**R7)**  $\sup_{\|\alpha - \theta\| \leq \delta} |R_7(\alpha)| = o_p(h^2 \delta n^\xi);$

**R8)**  $\sup_{\|\alpha - \theta\| \leq \delta} |R_8(\alpha)| = o_p(n^{-\frac{1}{2} + \xi} h^{-\frac{1}{2}} \delta);$

**R9)**  $\sup_{\alpha \in D} |R_9(\alpha)| = o_p(n^{-\frac{1}{2} + \xi} h^{\frac{3}{2}}).$

Under  $H_a$  (in which case  $\theta$  is arbitrary), the above results hold for  $R_0, R_2, R_3, R_6$  and  $R_9$ ; they hold for  $R_4 + R_5, R_7, R_8$  with  $\delta$  replaced by  $O(1)$ ; and  $R_1(\alpha) \geq c_0 + o_p(1)$  uniformly in  $\alpha$  for some constant  $c_0 > 0$ . The results hold for both random-design and fixed-design models.

**Lemma 2.** Suppose  $v \geq 10$  and we take  $h = O(n^{-\frac{1}{5}})$ . Let  $\hat{\theta}$  minimize  $\hat{d}(\alpha)$  over  $D$ . Then under Assumption (A1)-(A5), we have, for all  $\xi > 0$ ,

(i) under  $H_0$ ,  $\|\hat{\theta} - \theta\| = o_p(n^{-\frac{2}{5} + \xi});$

(ii) under  $H_0$ ,  $\sum_{i=1}^9 R_i(\hat{\theta}) = o_p(n^{-\frac{4}{5} + \xi});$

(iii) under  $H_a$ ,  $\sum_{i=2}^9 R_i(\hat{\theta}) = o_p(n^{-\frac{2}{5} + \xi});$

(iv)  $R_0(\hat{\theta}) = R_0(\theta) + o_p(n^{-\frac{9}{10} + \xi}).$

We shall develop a series of results (Lemma B.1-B.9) that are used in the proofs of Lemmas 1 and 2. In the next subsection each of these Lemmas are stated and proved. In some proofs we use generic constants  $C_1, C_2$  etc. for notational convenience. The proofs of Lemmas 1 and 2 are then given in Section B.2.

## B.1 Preparatory Lemmas

We start with the following version of Bernstein's inequality.

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

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

Next we give a few results that we shall often use, the first of which utilizes a discretization technique, which has been used by, for example, Ichimura (1993), Klein and Spady (1993), and Härdle et. al. (1993).

**Lemma B.2 (Discretization Technique).** *Let  $A$  be any subset of the unit sphere in  $\mathbb{R}^p$ . Suppose  $Z_{i1}, \dots, Z_{in}$  are independent conditioned on  $X_i$  for every fixed  $i$  and*

- (i)  $\text{Var}(Z_{ij}(\alpha)|X_i) \leq \sigma_{nj}^2$ ;
- (ii)  $P(|Z_{ij}(\alpha)| \leq b_n) = 1$ ;
- (iii)  $\left\| \frac{d}{d\alpha} Z_{ij}(\alpha) \right\| \leq Cn^r$  for some constants  $C, r$ ;

for all  $i, j = 1, \dots, n$ , and for all  $\alpha \in A$ . Then, for all  $\xi > 0$ ,

$$\sup_{\alpha \in A} \max_{1 \leq i \leq n} \left| \sum_{j=1}^n (Z_{ij}(\alpha) - E_i Z_{ij}(\alpha)) \right| = o_p(\sigma_n n^\xi + b_n n^\xi),$$

where  $E_i Z_{ij}(\alpha) = E[Z_{ij}(\alpha)|X_i]$  and  $\sigma_n^2 = \sum_{j=1}^n \sigma_{nj}^2$ . The result holds for fixed design by treating  $X_i$  as constants.

*Proof.* It suffices to show the random-design case. Partition  $A$  into  $n^s$  small parts evenly and take one point from each part. Let  $\mathcal{A}_n$  denote the collection of these representative points. Then for every  $\alpha \in A$  we can find a  $\tilde{\alpha} \in \mathcal{A}_n$  such that  $|\alpha - \tilde{\alpha}| \leq 2\sqrt{pn}^{-\frac{s}{p}}$ . Now, for every fixed realization of the random variables we have that

$$\begin{aligned} & \sup_{\alpha \in A} \max_{1 \leq i \leq n} \left| \sum_{j=1}^n (Z_{ij}(\alpha) - E_i Z_{ij}(\alpha)) \right| - \sup_{\alpha \in \mathcal{A}_n} \max_{1 \leq i \leq n} \left| \sum_{j=1}^n (Z_{ij}(\alpha) - E_i Z_{ij}(\alpha)) \right| \\ & \leq \left| \sum_{j=1}^n (Z_{i_0j}(\alpha_0) - E_{i_0} Z_{i_0j}(\alpha_0)) \right| - \left| \sum_{j=1}^n (Z_{i_0j}(\tilde{\alpha}_0) - E_{i_0} Z_{i_0j}(\tilde{\alpha}_0)) \right| \\ & \leq 2n \cdot Cn^r \cdot 2\sqrt{pn}^{-\frac{s}{p}} = 4Cn^{-\frac{s}{p}+r+1}, \end{aligned} \tag{B.1}$$

where

$$(\alpha_0, i_0) = \arg \sup_{\alpha \in A} \max_{1 \leq i \leq n} \left| \sum_{j=1}^n (Z_{ij}(\alpha) - E Z_{ij}(\alpha)) \right|,$$

and  $\tilde{\alpha}_0 \in \mathcal{A}_n$  is chosen such that  $|\alpha_0 - \tilde{\alpha}_0| \leq 2\sqrt{pn}^{-\frac{s}{p}}$ . By Bernstein's inequality,

$$\begin{aligned} & P\left( \sup_{\alpha \in \mathcal{A}_n} \max_{1 \leq i \leq n} \left| \sum_{j=1}^n (Z_{ij}(\alpha) - E_i Z_{ij}(\alpha)) \right| \geq a_n \right) \\ & \leq \sum_{\alpha \in \mathcal{A}_n} \sum_{i=1}^n P\left( \left| \sum_{j=1}^n (Z_{ij}(\alpha) - E_i Z_{ij}(\alpha)) \right| \geq a_n \right) \\ & \leq n^{1+s} \cdot \int 2 \exp\left\{ -\frac{a_n^2}{2 \sum_{j=1}^n \sigma_{nj}^2 + 2b_n a_n} \right\} f_{X_i}(t) dt \\ & = 2 \exp\left\{ -\frac{a_n^2}{2\sigma_n^2 + 2b_n a_n} + (1+s) \ln n \right\}. \end{aligned}$$

Hence, for all  $\xi > 0$ ,

$$\sup_{\alpha \in \mathcal{A}_n} \max_{1 \leq i \leq n} \left| \sum_{i=1}^n (Z_i(\alpha) - EZ_i(\alpha)) \right| = o_p(\sigma_n n^\xi + b_n n^\xi),$$

Noticing (B.1), the proof is completed by taking  $s$  large enough such that

$$n^{-\frac{s}{p}+r+1} = o(\sigma_n n^\xi + b_n n^\xi).$$

□

**Lemma B.3.** *Let  $\epsilon_i$ 's be independent and*

$$B_i(\alpha) = \sum_{j=1}^n w_{ij}(\alpha) \epsilon_j \quad \text{and} \quad B_{ir}(\alpha) = \sum_{j=1}^n w_{ij}(\alpha) \epsilon_{jr}, \quad r = 1, 2; \quad (\text{B.2})$$

where

$$\epsilon_{j1} = \epsilon_j I_{\{|\epsilon_j| \leq n^t\}}, \quad \epsilon_{j2} = \epsilon_j I_{\{|\epsilon_j| > n^t\}}. \quad (\text{B.3})$$

Suppose  $w_{ij}(\alpha) = w(X_i, X_j, \alpha)$ ,  $j \neq i$ , are independent given  $X_i$ ,  $X_i$ 's and  $\epsilon_j$ 's are independent, and

$$E_i(w_{ij}^2(\alpha)) \leq v_{ni}, \quad |w_{ij}(\alpha)| \leq c_n, \quad \text{and} \quad \left\| \frac{\partial}{\partial \alpha} w_{ij}(\alpha) \right\| \leq C n^r,$$

for some constants  $C, r$  uniformly in  $i, j, \alpha$ . Let  $v_n = \sum_{i=1}^n v_{ni}$ . Then, if  $E|\epsilon_i^v| \leq M$  for some common finite constant  $M$  and  $c_n = O(n^s)$ ,  $v_n = O(n^t)$  for some constants  $s, t$ , we have,

(i)

$$E|\epsilon_{j2}| \leq \frac{E|\epsilon_j|^v}{n^{(v-1)t}} \quad \text{and} \quad E\epsilon_{j2}^2 \leq \frac{E|\epsilon_j|^v}{n^{(v-2)t}};$$

(ii) for all  $\xi > 0$ ,

$$\sum_{i=1}^n \sup_{\alpha \in A} |B_{i1}(\alpha)| = o_p(n^{1+\xi} \sqrt{v_n} + n^{1+t+\xi} c_n),$$

and

$$\sum_{i=1}^n \sup_{\alpha \in A} |B_{i1}(\alpha)|^2 = o_p(n^{1+\xi} v_n + n^{1+2t+\xi} c_n^2);$$

(iii)

$$E \sum_{i=1}^n \sup_{\alpha \in A} |B_{i2}(\alpha)| = O(n^{2-(v-1)t} c_n); \quad \text{and} \quad E \sum_{i=1}^n \sup_{\alpha \in A} |B_{i2}(\alpha)|^2 = O(n^{3-(v-2)t} c_n^2).$$

(iv)

$$\sum_{i=1}^n \sup_{\alpha \in A} |B_i(\alpha)| = o_p(n^{1+\xi} \sqrt{v_n} + n^{1+\frac{1}{v}+\xi} c_n),$$

and

$$\sum_{i=1}^n \sup_{\alpha \in A} |B_i(\alpha)|^2 = o_p(n^{1+\xi} v_n + n^{1+\frac{4}{v}+\xi} c_n^2);$$

The above results hold for fixed-design as well with  $w_{ij}(\alpha)$  considered as constants.

*Proof.* From the proof below we can see that it suffices to show the random-design case. Using Hölder's inequality and Chebyshev's inequality we get

$$E|\epsilon_{j2}| \leq (E|\epsilon_j|^v)^{\frac{1}{v}} (P(|\epsilon_j| > n^t))^{1-\frac{1}{v}} \leq (E|\epsilon_j|^v)^{\frac{1}{v}} \left( \frac{E|\epsilon_j|^v}{n^{vt}} \right)^{1-\frac{1}{v}} = \frac{E|\epsilon_j|^v}{n^{(v-1)t}};$$

and

$$E|\epsilon_{j2}|^2 \leq (E|\epsilon_j|^v)^{\frac{2}{v}} (P(|\epsilon_j| > n^t))^{1-\frac{2}{v}} \leq (E|\epsilon_j|^v)^{\frac{2}{v}} \left( \frac{E|\epsilon_j|^v}{n^{vt}} \right)^{1-\frac{2}{v}} = \frac{E|\epsilon_j|^v}{n^{(v-2)t}}.$$

Now, for all  $i$  and all  $\alpha \in A$ , we have

$$\left\| \frac{\partial}{\partial \alpha} B_{i1}(\alpha) \right\| \leq n^t \sum_{j=1}^n \left\| \frac{\partial}{\partial \alpha} w_{ij}(\alpha) \right\| \leq C n^{1+t+r};$$

and

$$\left\| \frac{\partial}{\partial \alpha} B_{i1}^2(\alpha) \right\| \leq |2B_{i1}(\alpha)| \cdot \left\| \frac{\partial}{\partial \alpha} B_{i1}(\alpha) \right\| \leq 2n^{1+t} c_n \cdot C n^{1+t+r} = 2c_n C n^{2+2t+r}.$$

Again we construct the discrete set  $A_n$  of size  $n^s$  as in the proof of Lemma B.2 and then we have

$$\sum_{i=1}^n \sup_{\alpha \in A} |B_{i1}(\alpha)| - \sum_{i=1}^n \sup_{\alpha \in A_n} |B_{i1}(\alpha)| \leq n \cdot C n^{1+t+r} \cdot 2\sqrt{pn}^{-\frac{s}{p}}; \quad (\text{B.4})$$

and

$$\sum_{i=1}^n \sup_{\alpha \in A} |B_{i1}(\alpha)|^2 - \sum_{i=1}^n \sup_{\alpha \in A_n} |B_{i1}(\alpha)|^2 \leq n \cdot 2c_n C n^{2+2t+r} \cdot 2\sqrt{pn}^{-\frac{s}{p}}. \quad (\text{B.5})$$

By Bernstein's inequality, noticing  $E w_{ij}(\alpha) \epsilon_{jr} = 0$  and

$$|w_{ij}(\alpha) \epsilon_{jr}| \leq c_n n^t, \quad \text{Var}(w_{ij}(\alpha) \epsilon_{jr} | X_i) \leq v_{ni} \sigma^2,$$

we have

$$\begin{aligned}
& P\left(\sum_{i=1}^n \sup_{\alpha \in A_n} |B_{i1}(\alpha)| > a_n \epsilon\right) \\
& \leq \sum_{i=1}^n \sum_{\alpha \in A_n} P\left(\left|\sum_{j \neq i} w_{ij}(\alpha) \epsilon_{j1}\right| > \frac{1}{n} a_n \epsilon\right) \\
& \leq n^{1+s} \int 2 \exp\left\{-\frac{(\frac{1}{n} a_n \epsilon)^2}{2v_n \sigma^2 + 2c_n n^t (\frac{1}{n} a_n \epsilon)}\right\} f(x) dx \\
& = 2 \exp\left\{-\frac{a_n^2 \epsilon^2}{2\sigma^2 n^2 v_n + 2c_n n^{1+t} a_n \epsilon} + (1+s) \ln n\right\};
\end{aligned}$$

and

$$\begin{aligned}
& P\left(\sum_{i=1}^n \sup_{\alpha \in A_n} B_{i1}^2(\alpha) > a_n \epsilon\right) \\
& \leq \sum_{i=1}^n \sum_{\alpha \in A_n} P\left(B_{i1}^2(\alpha) > \frac{1}{n} a_n \epsilon\right) \\
& = \sum_{i=1}^n \sum_{\alpha \in A_n} P\left(\left|\sum_{j \neq i} w_{ij}(\alpha) \epsilon_{j1}\right| > \sqrt{\frac{1}{n} a_n \epsilon}\right) \\
& \leq n^{1+s} \int 2 \exp\left\{-\frac{\frac{1}{n} a_n \epsilon}{2v_n \sigma^2 + 2c_n n^t \sqrt{\frac{1}{n} a_n \epsilon}}\right\} f(x) dx \\
& = 2 \exp\left\{-\frac{a_n \epsilon}{2\sigma^2 n v_n + 2c_n n^t \sqrt{n a_n \epsilon}} + (1+s) \ln n\right\}.
\end{aligned}$$

Thus, noticing right hand sides of (B.4) and (B.5) are of smaller rate by taking  $s$  large enough, we have proved (ii). For (iii),

$$E \sum_{i=1}^n \sup_{\alpha} |B_{i2}(\alpha)| \leq \sum_{i=1}^n E \sum_{j=1}^n \left(\sup_{\alpha} |w_{ij}(\alpha) \epsilon_{j2}|\right) \leq n^2 c_n E|\epsilon_{j2}| = O(n^{2-(v-1)t} c_n);$$

and

$$E \sum_{i=1}^n \sup_{\alpha} |B_{i2}(\alpha)|^2 \leq \sum_{i=1}^n E \left(\sum_{j=1}^n \sup_{\alpha} |w_{ij}(\alpha) \cdot \epsilon_{j2}|\right)^2 \leq n c_n^2 \cdot n \sum_{j=1}^n E|\epsilon_{j2}^2| = O(n^{3-(v-2)t} c_n^2).$$

Finally, taking  $t = \frac{1}{v}$  for the first order case and taking  $t = \frac{2}{v}$  for the second order case, (ii) and (iii) yield (iv).  $\square$

Now we introduce a few results that will simplify the proofs of Lemmas 1-2. All the inf or sup, if not specified, will be taken over any subset  $A$  of the unit sphere in  $\mathbb{R}^p$ .

**Lemma B.4.** Let  $Z_i(\alpha)$ 's and  $\epsilon_i$ 's be independent. Suppose  $Z_i(\alpha)$ 's are bounded by  $b$  uniformly in  $\alpha$  and  $i$ , and  $\epsilon_i$ 's are independent with mean zero and at least  $v$  moments. Then, if  $\|\frac{d}{d\alpha}Z_{ij}(\alpha)\| \leq Cn^s$  for some constants  $C, s$ , and  $v \geq 4$ , we have

$$\sup_{\alpha} \left| \sum_{i=1}^n Z_i(\alpha)\epsilon_i \right| = o_p(n^{\frac{1}{2}+\xi}b).$$

Moreover, if  $Z_i$ 's are independent when conditioned on some random variable  $X$  (which is independent with  $\epsilon_j$ 's as well), and uniformly we have  $\sum_{i=1}^n E[Z_i^2(\alpha)|X] \leq \sigma_n^2$ , then

$$\sup_{\alpha} \left| \sum_{i=1}^n Z_i(\alpha)\epsilon_i \right| = o_p(n^{\frac{2}{v}+\xi}b + \sigma_n n^{\xi}).$$

*Proof.* Let  $A_r = \sum_{i=1}^n Z_i(\alpha)\epsilon_{ir}$ ,  $r = 1, 2$ , with  $\epsilon_{ir}$  defined as in (B.3). Then Lemma B.2 with  $\sigma_n = \sqrt{nb}$  and  $b_n = n^t b$  gives

$$\sup_{\alpha} |A_1| = o_p(\sqrt{nb}n^{\xi} + n^t b n^{\xi}). \quad (\text{B.6})$$

And, by Lemma B.3,

$$\sup_{\alpha} |A_2| \leq \sup_{\alpha} \sqrt{\sum_{i=1}^n Z_i^2(\alpha) \sum_{i=1}^n \epsilon_{i2}^2} \leq O_p(\sqrt{nb^2 \cdot n^{1-(v-2)t}}) = O_p(n^{1-\frac{v-2}{2}t}b).$$

Hence

$$\sup_{\alpha} |A_1 + A_2| = o_p((n^{1-\frac{v-2}{2}t} + \sqrt{n} + n^t)bn^{\xi}).$$

Taking  $t = \frac{2}{v}$  gives the first result. For the second result, it suffices to notice that (B.6) now becomes

$$\sup_{\alpha} |A_1| = o_p(\sigma_n n^{\xi} + n^t b n^{\xi}).$$

□

**Lemma B.5.** Under Assumption (A1)-(A5), we have

$$\inf_{\alpha} \min_{\{i|L_{q,\alpha}(X_i) \neq 0\}} K_i(\alpha) \geq c_K n h + o_p(nh) \quad \text{and} \quad \inf_{\alpha} L_{\alpha} \geq c_L n + o_p(n),$$

where  $c_K, c_L$  are some positive constants. The result also holds for fixed-design models with  $o_p(\cdot)$  replaced by  $o(\cdot)$ .

*Proof.* We first look at random-design models. Let  $Z_{ij}(\alpha) = K_h(\alpha'(X_j - X_i))$ . Then  $K_i(\alpha) = \sum_{j=1}^n Z_{ij}(\alpha)$ . Given  $X_i, Z_{ij}(\alpha)$  are i.i.d. and bounded. Now,

$$E(Z_{ij}^2(\alpha)|X_i) = \int K^2\left(\frac{t - \alpha'X_i}{h}\right) f_{\alpha}(t) dt \leq h \int_{-1}^1 K^2(s) f_{\alpha}(\alpha'X_i + hs) ds \leq Ch,$$

for some constant  $C$  uniformly in  $i, j, \alpha$ . Also,  $\|\partial Z_{ij}(\alpha)/\partial \alpha\| = O(h^{-1})$  uniformly. Hence, by Lemma B.2 with  $\sigma_{nj}^2 = Ch$  and  $b_n = O(1)$ , for all  $\xi > 0$ ,

$$\sup_{\alpha} \max_{1 \leq i \leq n} |K_i(\alpha) - E_i K_i(\alpha)| = o_p(\sqrt{nhn^\xi}).$$

When  $L_{q,\alpha}(X_i) \neq 0$ , we have  $\alpha' X_i \in (c_\alpha \pm q_\alpha)$  and hence, by Lemma 2.6 (iv),

$$\begin{aligned} E_i K_i(\alpha) &= K(0) + (n-1) \int_{-1}^1 K\left(\frac{t - \alpha' X_i}{h}\right) f_\alpha(t) dt \\ &= K(0) + (n-1)h \int_{-1}^1 K(u) f_\alpha(\alpha' X_i + hu) du \\ &= nh f_\alpha(\alpha' X_i) + O(nh^2) \geq c_K nh + O(nh^2), \end{aligned}$$

where  $c_K = \inf_{\alpha} \inf_{t \in c_\alpha \pm q_\alpha} f_\alpha(t) > 0$ . Finally,

$$\begin{aligned} \inf_{\alpha \in A} \min_{\{i | L_{q,\alpha}(X_i) \neq 0\}} K_i(\alpha) &= \inf_{\alpha \in A} \min_{\{i | L_{q,\alpha}(X_i) \neq 0\}} |K_i(\alpha) - E_i K_i(\alpha) + E_i K_i(\alpha)| \\ &\geq \inf_{\alpha \in A} \min_{\{i | L_{q,\alpha}(X_i) \neq 0\}} (|E_i K_i(\alpha)| - |K_i(\alpha) - E_i K_i(\alpha)|) \\ &\geq \inf_{\alpha \in A} \min_{\{i | L_{q,\alpha}(X_i) \neq 0\}} E_i K_i(\alpha) - \sup_{\alpha \in A} \max_{1 \leq i \leq n} |K_i(\alpha) - E_i K_i(\alpha)| \\ &\geq c_K nh + O(nh^2) - o_p(\sqrt{nhn^\xi}), \end{aligned}$$

which completes the proof.

For fixed-design models we have

$$\begin{aligned} K_i(\alpha) &= \sum_{j=1}^n K_h(\alpha' x_j - \alpha' x_i) = n \int_S K_h(\alpha' x - \alpha' x_i) dx + O((\log n)^p) \\ &= n \int_{-1}^1 K_h(y_1 - \alpha' x_i) \phi(y_1) dy_1 + o(nh) = nh \int_{-1}^1 K(u) \phi(\alpha' x_i + uh) du + o(nh) \\ &= nh \phi(\alpha' x_i) + O(nh^3) + o(nh), \end{aligned}$$

where

$$\phi(t) = \int_{t^2 + y_2^2 + \dots + y_p^2 \leq 1} dy_2 dy_3 \dots dy_p = w_p (1 - t^2)^{\frac{n}{2}},$$

with  $w_p$  being some constant only depends on  $p$ . Since  $L_{q,\alpha}(X_i) \neq 0$  indicates that  $\alpha' X_i \in (c_\alpha \pm q_\alpha)$ , by Lemma 2.6 (ii),  $|\alpha' x_i| \leq Q$  for some constant  $Q < 1$ . Thus

$$\inf_{\alpha} \min_{\{i | L_{q,\alpha}(X_i) \neq 0\}} \phi(\alpha' x_i) \geq \phi(Q) = c_L > 0,$$

which completes the proof.

Noticing that  $L(t)$  is non-increasing in  $|t|$  and  $q_\alpha = qw_\alpha \geq qw_0$ , we have

$$L_\alpha = \sum_{i=1}^n L\left(\frac{\alpha' X_i - c_\alpha}{q_\alpha}\right) \geq \sum_{i=1}^n L\left(\frac{\alpha' X_i - c_\alpha}{qw_0}\right).$$

Hence a proof similar to above (with  $h$  replaced by  $qw_0$ ,  $E_i$  replaced by  $E$ , and so on) gives the desired result.  $\square$

**Lemma B.6.** Under Assumption (A1)-(A5), we have, for all  $\xi > 0$ ,

$$\sup_{\alpha} \max_{\{i|L_{q,\alpha}(X_i) \neq 0\}} |K_{mi}(\alpha) - g_{\alpha}(\alpha'X_i)| = o_p(h^2n^{\xi}).$$

For fixed design models, the right hand side is  $O(h^2)$ .

*Proof.* First consider random design case. Notice that  $g_{\alpha}(\alpha'X_i) = E(m(X_i)|\alpha'X_i)$  and

$$K_{mi}(\alpha) - E(m(X_i)|\alpha'X_i) = \frac{1}{K_i(\alpha)} \sum_{j=1}^n [m(X_j) - E(m(X_i)|\alpha'X_i)] K_h(\alpha'X_j - \alpha'X_i).$$

Let

$$Z_{ij}(\alpha) = [m(X_j) - E(m(X_i)|\alpha'X_i)] K_h(\alpha'X_j - \alpha'X_i).$$

Then

$$\begin{aligned} \sup_{\alpha} \max_{\{i|L_{q,\alpha}(X_i) \neq 0\}} |K_{mi} - g_{\alpha}(\alpha'X_i)| &= \sup_{\alpha} \max_{1 \leq i \leq n} \frac{1}{K_i(\alpha)} \left| \sum_{j=1}^n Z_{ij}(\alpha) \right| \\ &\leq \frac{\sup_{\alpha} \max_{1 \leq i \leq n} \left| \sum_{j=1}^n (Z_{ij}(\alpha) - E_i Z_{ij}(\alpha)) \right| + \sup_{\alpha} \max_{\{i|L_{q,\alpha}(X_i) \neq 0\}} |nE_i Z_{ij}(\alpha)|}{\inf_{\alpha} \min_{\{i|L_{q,\alpha}(X_i) \neq 0\}} K_i(\alpha)}. \end{aligned}$$

Now, with  $L_{q,\alpha}(X_i) \neq 0$ , or  $s_i = \alpha'X_i \in c_{\alpha} \pm q_{\alpha}$ ,

$$\begin{aligned} E_i Z_{ij}(\alpha) &= E(Z_{ij}(\alpha)|X_i) = E[E(Z_{ij}(\alpha)|\alpha'X_j, X_i)|X_i] \\ &= E\left[(g_{\alpha}(\alpha'X_j) - g_{\alpha}(\alpha'X_i)) K_h(\alpha'X_j - \alpha'X_i)|X_i\right] \\ &= \int_{c_{\alpha}-w_{\alpha}}^{c_{\alpha}+w_{\alpha}} (g_{\alpha}(t) - g_{\alpha}(s_i)) K\left(\frac{t-s_i}{h}\right) f_{\alpha}(t) dt \\ &= h \int_{-1}^1 (g_{\alpha}(s_i + uh) - g_{\alpha}(s_i)) K(u) f_{\alpha}(s_i + uh) du \end{aligned}$$

Hence, noticing  $K(\cdot)$  is an even function, a simple Taylor's expansion yields that, for  $j \neq i$ ,

$$\sup_{\alpha} \max_{\{i|L_{q,\alpha}(X_i) \neq 0\}} |E_i Z_{ij}(\alpha)| \leq C_1 h^3,$$

for some constant  $C_1$ . Given  $\alpha'X_i$ ,  $Z_{ij}(\alpha)$  ( $j \neq i$ ) are i.i.d., uniformly bounded and  $E(Z_{ij}^2(\alpha)|\alpha'X_i) \leq C_2 h^3$  for some constant  $C_2$  uniformly in  $i$  and  $\alpha$ . By Lemma B.2 with  $\sigma_n^2 = C_2 n h^3$  and  $b_n = O(1)$ , we have

$$\sup_{\alpha} \max_{1 \leq i \leq n} \left| \sum_{j=1}^n (Z_{ij}(\alpha) - E_i Z_{ij}(\alpha)) \right| = o_p(\sqrt{nh^3}n^{\xi} + n^{\xi}), \quad \forall \xi > 0.$$

Thus,

$$\sup_{\alpha} \max_{1 \leq i \leq n} |K_{mi}(\alpha) - g_{\alpha}(\alpha'X_i)| \leq o_p(n^{-1}h^{-1}(\sqrt{nh^3} + 1 + nh^3)n^{\xi}) = o_p(h^2n^{\xi}).$$

Now we consider fixed-design models. Let  $u = \alpha'x$ . Notice that

$$\begin{aligned}
\sum_{i=1}^n m(x_i)K\left(\frac{\alpha'x_i - u}{h}\right) &= n \int m(x)K\left(\frac{\alpha'x - u}{h}\right)dx + O((\log n)^p) \\
&= n \int m(A'y)K\left(\frac{y_1 - u}{h}\right)dy + o(nh^3) \\
&= n \int_{u-h}^{u+h} K\left(\frac{y_1 - u}{h}\right) \left( \int_{S(y_1)} m(A'y)dy_2 \cdots dy_p \right) dy_1 + o(nh^3) \\
&= n \int_{u-h}^{u+h} K\left(\frac{y_1 - u}{h}\right) \phi(y_1) dy_1 + o(nh^3) \\
&= nh \int_{-1}^1 K(z)\phi(u + hz)dz + o(nh^3) = nh\phi(u) + O(nh^3),
\end{aligned}$$

uniformly in  $u$  and  $\alpha$ , where  $\phi(u) = \int_{S(u)} m(A'y)dy_2 \cdots dy_p$  with  $S(u) = \{y \in S | y_1 = u\}$ . Taking  $m(\cdot) \equiv 1$  the above yields

$$\sum_{i=1}^n K\left(\frac{\alpha'x_i - u}{h}\right) = c_p(1 - u)^{\frac{p-1}{2}} \cdot nh + O(nh^3),$$

uniformly in  $u$  and  $\alpha$ , where  $c_p$  is a positive constant which only depends on  $p$ . Hence, uniformly for  $u = \alpha'x$  with  $L_{q,\alpha}(x) \neq 0$ ,

$$\frac{\sum_{i=1}^n m(x_i)K_h(\alpha'x_i - u)}{\sum_{i=1}^n K_h(\alpha'x_i - u)} = \frac{nh \int_{S(u)} m(A'y)dy_2 \cdots dy_p + O(nh^3)}{nh \int_{S(u)} dy_2 \cdots dy_p + O(nh^3)} = g_\alpha(u) + O(h^2).$$

□

**Lemma B.7.** *For any  $\xi > 0$  we have*

$$\sup_{\|\alpha - \theta\| \leq \delta} \max_{1 \leq i \leq n} |K_i(\alpha) - K_i(\theta)| = o_p(nh\delta n^\xi).$$

More generally,

$$\sup_{\|\alpha - \theta\| \leq \delta} \max_{1 \leq i \leq n} \left| \sum_{j=1}^n (K_h(\alpha'X_j - \alpha'X_i) - K_h(\theta'X_j - \theta'X_i))m(X_j) \right| = o_p(nh\delta n^\xi).$$

For fixed-design models, the right hand side is  $O(nh\delta)$ .

*Proof.* We shall only give the proof for the special case, i.e.  $m(\cdot) \equiv 1$ . The proof for the general  $m(\cdot)$  is almost identical. Let  $E_i$  denote the conditional expectation given  $X_i$ . Then

$$\begin{aligned}
&\sup_{\|\alpha - \theta\| \leq \delta} \max_{1 \leq i \leq n} |K_i(\alpha) - K_i(\theta)| \\
&\leq \sup_{\|\alpha - \theta\| \leq \delta} \max_{1 \leq i \leq n} \left| [K_i(\alpha) - K_i(\theta)] - E_i(K_i(\alpha) - K_i(\theta)) \right| \\
&\quad + \sup_{\|\alpha - \theta\| \leq \delta} \max_{1 \leq i \leq n} |E_i(K_i(\alpha) - K_i(\theta))|.
\end{aligned}$$

Now,

$$\left| E_i(K_i(\alpha) - K_i(\theta)) \right| \leq \sum_{j=1}^n \left| E_i [K_h(\alpha' X_j - \alpha' X_i) - K_h(\theta' X_j - \theta' X_i)] \right| \leq C_1 n h \delta,$$

for some constant  $C_1$  uniformly in  $X_i$  and  $\alpha \in B(\theta, \delta)$ . Let

$$Z_{ij}(\alpha) = K_h(\alpha' X_j - \alpha' X_i) - K_h(\theta' X_j - \theta' X_i) - E_i(K_h(\alpha' X_j - \alpha' X_i) - K_h(\theta' X_j - \theta' X_i)).$$

Then given  $X_i$ ,  $Z_{ij}(\alpha)$ ,  $j \neq i$ , are i.i.d. with mean zero. Also  $Z_{ij}(\alpha)$  are uniformly bounded by  $C_2 h^{-1} \delta$  everywhere for some constant  $C_2$  and

$$\text{Var}(Z_{ij}(\alpha) | X_i) \leq E_i \left( K_h(\alpha' X_j - \alpha' X_i) - K_h(\theta' X_j - \theta' X_i) \right)^2 \leq C_3 h \delta^2,$$

uniformly in  $i, j$  and  $\alpha \in B(\theta, \delta)$  for some constant  $C_3$ . Applying Lemma B.2 with  $\sigma_n^2 = C_3 n h \delta^2$  and  $b_n = C_2 h^{-1} \delta$  we have

$$\begin{aligned} & \sup_{\|\alpha - \theta\| \leq \delta} \max_{1 \leq i \leq n} \left| [K_i(\alpha) - K_i(\theta)] - E_i(K_i(\alpha) - K_i(\theta)) \right| \\ &= \sup_{\|\alpha - \theta\| \leq \delta} \max_{1 \leq i \leq n} \left| \sum_{j=1}^n (Z_{ij}(\alpha) - E Z_{ij}(\alpha)) \right| = o_p(\sqrt{n h \delta n^\xi} + h^{-1} \delta n^\xi), \end{aligned}$$

which completes the proof since  $h^{-1} \leq O(nh)$ .

For fixed-design models,

$$\begin{aligned} K_i(\alpha) - K_i(\theta) &= \sum_{j=1}^n \left( K_h(\alpha' x_j - \alpha' x_i) - K_h(\theta' x_j - \theta' x_i) \right) \\ &= n \int \left( K_h(\alpha' x - \alpha' x_i) - K_h(\theta' x - \theta' x_i) \right) dx + O((\log n)^p) \\ &= n h \int \left( K(\alpha' y) - K(\theta' y) \right) dy + o(n^\xi) = O(n h \delta). \end{aligned}$$

□

**Lemma B.8.** *In both random-design and fixed-design models, we have*

(i)

$$\begin{aligned} & \sum_{i=1}^n \sup_{\|\alpha - \theta\| < \delta} \left| \sum_{j=1}^n (K_h(\alpha' X_j - \alpha' X_i) - K_h(\theta' X_j - \theta' X_i)) \epsilon_j \right| \\ &= o_p((n^{\frac{3}{2}} \sqrt{h} + n^{1+\frac{1}{v}} h^{-1}) \delta n^\xi); \\ & \sum_{i=1}^n \sup_{\|\alpha - \theta\| < \delta} \left| \sum_{j=1}^n (K_h(\alpha' X_j - \alpha' X_i) - K_h(\theta' X_j - \theta' X_i)) \epsilon_j \right|^2 \\ &= o_p((n^2 h + n^{1+\frac{4}{v}} h^{-2}) \delta^2 n^\xi); \end{aligned}$$

(ii)

$$\begin{aligned} \sum_{i=1}^n \sup_{\alpha} \left| \sum_{j=1}^n K_h(\alpha' X_j - \alpha' X_i) \epsilon_j \right| &= o_p(n^{1.5+\xi} \sqrt{h}); \\ \sum_{i=1}^n \sup_{\alpha} \left| \sum_{j=1}^n K_h(\alpha' X_j - \alpha' X_i) \epsilon_j \right|^2 &= o_p(n^{2+\xi} h); \end{aligned}$$

(iii)

$$\begin{aligned} \sup_{\|\alpha-\theta\| \leq \delta} \sum_{i=1}^n \left| \sum_{j=1}^n (\kappa_{ij}(\alpha) - \kappa_{ij}(\theta)) \epsilon_j \right| &= o_p((\sqrt{nh^{-1}} + n^{\frac{1}{v}} h^{-2}) \delta n^{\xi}); \\ \sup_{\|\alpha-\theta\| \leq \delta} \sum_{i=1}^n \left( \sum_{j=1}^n (\kappa_{ij}(\alpha) - \kappa_{ij}(\theta)) \epsilon_j \right)^2 &= o_p((h^{-1} + n^{\frac{4}{v}-1} h^{-4}) \delta^2 n^{\xi}). \end{aligned}$$

*Proof.* We shall provide the proof for random-design models only. The proof for fixed-design models is almost identical.

(i) Let

$$w_{ij}(\alpha) = K_h(\alpha' X_j - \alpha' X_i) - K_h(\theta' X_j - \theta' X_i).$$

Notice that  $\sum_{i=1}^n E_i w_{ij}^2(\alpha) \leq v_n = C_1 n h \delta^2$  and  $|w_{ij}(\alpha)| \leq c_n = C_2 \delta h^{-1}$  for some constants  $C_1, C_2$  uniformly in  $i, j, \alpha$ . Then Lemma B.3 yields that

$$\sum_{i=1}^n \sup_{\|\alpha-\theta\| < \delta} \left| \sum_{j=1}^n w_{ij}(\alpha) \epsilon_j \right| = o_p(n^{1+\xi} \sqrt{n \delta^2 h} + n^{1+\frac{1}{v}+\xi} \delta h^{-1});$$

and

$$\sum_{i=1}^n \sup_{\|\alpha-\theta\| < \delta} \left| \sum_{j=1}^n w_{ij}(\alpha) \epsilon_j \right|^2 = o_p(n^{1+\xi} n \delta^2 h + n^{1+\frac{4}{v}+\xi} (\delta h^{-1})^2).$$

(ii) Now let  $w_{ij}(\alpha) = K_h(\alpha' X_j - \alpha' X_i)$  with  $\sum_{i=1}^n E_i w_{ij}^2(\alpha) \leq v_n = C_3 n h$  and  $|w_{ij}(\alpha)| \leq c_n = M$  for some constants  $C_3$  uniformly in  $i, j, \alpha$ . Then Lemma B.3 yields that

$$\sum_{i=1}^n \sup_{\|\alpha-\theta\| < \delta} \left| \sum_{j=1}^n w_{ij}(\alpha) \epsilon_j \right| = o_p(n^{1+\xi} \sqrt{nh} + n^{1+\frac{1}{v}+\xi});$$

and

$$\sum_{i=1}^n \sup_{\|\alpha-\theta\| < \delta} \left| \sum_{j=1}^n w_{ij}(\alpha) \epsilon_j \right|^2 = o_p(n^{1+\xi} n h + n^{1+\frac{4}{v}+\xi}).$$

(iii) Notice that we have

$$\begin{aligned} &\kappa_{ij}(\alpha) - \kappa_{ij}(\theta) \\ &= \frac{K_i(\theta) - K_i(\alpha)}{K_i(\alpha) K_i(\theta)} K_h(\alpha' X_j - \alpha' X_i) - \frac{1}{K_i(\theta)} (K_h(\alpha' X_j - \alpha' X_i) - K_h(\theta' X_j - \theta' X_i)). \end{aligned}$$

Hence, by Lemma B.5, Lemma B.7 and (i), (ii) above, we have

$$\begin{aligned}
& \sup_{\|\alpha-\theta\|} \sum_{i=1}^n \left| \sum_{j=1}^n (\kappa_{ij}(\alpha) - \kappa(\theta)) \epsilon_j \right| \\
& \leq O_p\left(\frac{1}{(nh)^2}\right) \sup_{\|\alpha-\theta\| \leq \delta} \max_{1 \leq i \leq n} |K_i(\alpha) - K_i(\theta)| \sum_{i=1}^n \left| \sum_{j=1}^n K_h(\alpha' X_j - \alpha' X_i) \epsilon_j \right| \\
& \quad + O_p\left(\frac{1}{nh}\right) \sup_{\|\alpha-\theta\| \leq \delta} \sum_{i=1}^n \left| \sum_{j=1}^n (K_h(\alpha' X_j - \alpha' X_i) - K_h(\theta' X_j - \theta' X_i)) \epsilon_j \right| \\
& \leq o_p\left(\frac{1}{n^2 h^2} \cdot nh\delta \cdot n^{1.5} \sqrt{hn}^\xi\right) + o_p\left(\frac{1}{nh} \cdot [n^{1.5} \sqrt{h} + n^{1+\frac{1}{v}} h^{-1}] \delta n^\xi\right) \\
& = o_p([\sqrt{nh^{-1}} + n^{\frac{1}{v}} h^{-2}] \delta n^\xi);
\end{aligned}$$

and

$$\begin{aligned}
& \sup_{\|\alpha-\theta\| \leq \delta} \sum_{i=1}^n \left( \sum_{j=1}^n (\kappa_{ij}(\alpha) - \kappa_{ij}(\theta)) \epsilon_j \right)^2 \\
& \leq \sup_{\|\alpha-\theta\| \leq \delta} \sum_{i=1}^n \left[ 2 \left( \frac{K_i(\theta) - K_i(\alpha)}{K_i(\alpha) K_i(\theta)} \right)^2 \left( \sum_{j=1}^n K_h(\alpha' X_j - \alpha' X_i) \epsilon_j \right)^2 \right. \\
& \quad \left. + 2 \left( \frac{1}{K_i(\theta)} \right)^2 \left( \sum_{j=1}^n [K_h(\alpha' X_j - \alpha' X_i) - K_h(\theta' X_j - \theta' X_i)] \epsilon_j \right)^2 \right] \\
& \leq o_p\left(\frac{n^{1+\xi} h \delta}{(nh)^2}\right)^2 o_p(n^{2+\xi} h) + O_p\left(\frac{1}{(nh)^2}\right) o_p(n^{2+\xi} \delta^2 h + n^{1+\frac{4}{v}+\xi} (\delta h^{-1})^2) \\
& = o_p((h^{-1} + n^{\frac{4}{v}-1} h^{-4}) \delta^2 n^\xi).
\end{aligned}$$

□

**Lemma B.9.** *For random-design models only, uniformly in  $i, j$  ( $i \neq j$ ) and  $\alpha$ , we have  $E\kappa_{ij}^2(\alpha) = O(\frac{1}{n^2 h})$ .*

*Proof.* Let  $Z_j = K_h(\alpha' X_j - \alpha' X_i)$  and  $S = \sum_{k \neq i, j} Z_k$  with joint density  $f_{zs}$ . Then,

$$\begin{aligned}
E\kappa_{ij}^2(\alpha) &= E \frac{Z_j^2}{(\sum_{k=1}^n Z_k)^2} \leq E \frac{Z_j^2}{(Z_j + S)^2} \\
&= \int_{|v| \geq \delta_n} \frac{u^2}{(u+v)^2} f_{zs}(u, v) dudv + \int_{|v| < \delta_n} \frac{u^2}{(u+v)^2} f_{zs}(u, v) dudv \\
&\leq \int_{|v| \geq \delta_n} \frac{u^2}{v^2} f_{zs}(u, v) dudv + \int_{|v| < \delta_n} f_{zs}(u, v) dudv \\
&\leq EZ_j^2 \frac{1}{\delta_n^2} + P(S < \delta_n).
\end{aligned}$$

Notice that  $ES = c_\alpha(n-2)h$  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  $\delta_n = c_\alpha(n-2)h - (nh)^{\frac{2}{3}}$ . Then, conditioned on  $\alpha'X_i$  and using Bernstein's inequality, we have

$$\begin{aligned} P(S < \delta_n) &= P(S - ES < \delta_n - ES) = P(S - ES < -(nh)^{\frac{2}{3}}) \\ &\leq P(|S - ES| > (nh)^{\frac{2}{3}}) \leq 2 \exp\left\{-\frac{(nh)^{\frac{4}{3}}}{C_1nh + (nh)^{\frac{2}{3}}}\right\}. \end{aligned}$$

Hence,

$$E\kappa_{ij}^2(\alpha) \leq \frac{C_2h}{(c_0(n-2)h - (nh)^{\frac{2}{3}})^2} + o(n^{-2}) = O\left(\frac{1}{n^2h}\right),$$

for some constant  $C_2$ , where, by Lemma 2.6,

$$c_0 = \inf_{\alpha} c_\alpha = \inf_{\alpha} \int_{-1}^1 K(s)\phi_\alpha(s)ds \geq \int_{-q_0/2}^{q_0/2} K(s)c_q^2q_0 ds > 0.$$

□

## B.2 Proofs of Lemma 1 and Lemma 2

We give the proofs of Lemma 1 and Lemma 2 here. All the results will be first stated for the random-design models. We shall comment on the conditions under which the results hold for fixed-design models and give the proof if it is very different from the random-design version.

We begin with the proof of Lemma 1. Since the proof of part (i) is different from the rest, we separate this part out.

### **Proof of Lemma 1 (Term R1).**

Under  $H_0$  with  $m(x) = g(\theta'x)$  we have

$$g(\theta'X) = g(\alpha'X) + (\theta - \alpha)'X \cdot g^{(1)}(\alpha'X) + T(X, \alpha),$$

where  $T(X) = o(\|\alpha - \theta\|)$  uniformly everywhere due to the boundedness of  $g''(\cdot)$ . Hence

$$m_\alpha(X) = g(\theta'X) - E(g(\theta'X)|\alpha'X) = (\theta - \alpha)'(X - E(X|\alpha'X))g^{(1)}(\alpha'X) + T_1(X, \alpha),$$

and Lemma B.5 gives

$$R_1(\alpha) = \sum_{i\alpha} m_\alpha^2(X_i) = \sum_{i\alpha} ((\theta - \alpha)' \mu(X_i, \alpha))^2 + T_2(X, \alpha) = \frac{1}{L_\alpha} \sum_{i=1}^n r_{X_i}(\alpha) + T_2(X, \alpha),$$

where  $T_1(X, \alpha) = o(\|\alpha - \theta\|)$ ,  $T_2(X, \alpha) = o(\|\alpha - \theta\|^2)$  uniformly everywhere,  $r_x(\alpha) = [(\alpha - \theta)' \mu(x, \alpha)]^2 L_{q,\alpha}(x)$  and

$$\mu(x, \alpha) = (x - E(X|\alpha'X = \alpha'x))g^{(1)}(\alpha'x).$$

The upper bound is obvious. To show the lower bound, since  $E(X|\alpha'X = \alpha'x)$  is differentiable with respect to  $\alpha$ , we have

$$\frac{\partial r_x(\alpha)}{\partial \alpha} = 2(\alpha - \theta)' \mu(x, \alpha) \left( \mu(x, \alpha) + \frac{\partial \mu(x, \alpha)}{\partial \alpha} (\alpha - \theta) \right) L_{q, \alpha}(x) + [(\alpha - \theta)' \mu(x, \alpha)]^2 \frac{\partial L_{q, \alpha}(x)}{\partial \alpha}.$$

Hence,  $\frac{\partial r_x(\theta)}{\partial \alpha} = 0$  and  $\frac{\partial^2 r_x(\theta)}{\partial \alpha \partial \alpha'} = 2\mu(x, \theta) \mu(x, \theta)' L_{q, \theta}(x)$ . By Taylor's expansion, noticing  $r_x(\theta) = 0$ , we have

$$r_x(\alpha) = (\alpha - \theta)' \mu(x, \theta) \mu(x, \theta)' (\alpha - \theta) L_{q, \theta}(x) + T_3(x, \alpha),$$

where  $T_3(x, \alpha) = o(\|\alpha - \theta\|^2)$  uniformly in  $x$  and in  $\alpha$ . Hence,

$$\begin{aligned} \sum_{i=1}^n r_{X_i}(\alpha) &= (\alpha - \theta)' \sum_{i=1}^n \mu(X_i, \theta) \mu(X_i, \theta)' (\alpha - \theta) L_{q, \theta}(X_i) \\ &= (\alpha - \theta)' (nW) (\alpha - \theta) + T_4(\alpha), \end{aligned}$$

where  $T_4(\alpha) = o_p(n\|\alpha - \theta\|^2)$  uniformly in  $\alpha$ , and

$$W = E(\mu(X, \theta) \mu(X, \theta)' \cdot L_{q, \theta}(X)).$$

Clearly  $W$  is positive semi-definite. Suppose  $W$  has rank  $p - 1$ . Then  $W = Q' \Lambda Q$  for some orthogonal matrix  $Q$  and diagonal matrix  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$  with  $\lambda_p = 0$ . Observe that  $\theta' W \theta = 0$ , i.e.,  $\theta$  is an eigenvector corresponding to  $\lambda_p = 0$ . Hence the last row of  $Q$  can be taken to be  $\theta$ . Let  $\lambda_m$  be the smallest nonzero eigenvalue of  $W$  and  $b = Q(\alpha - \theta)$ . Then  $b_p = \theta'(\alpha - \theta) = -\|\alpha - \theta\|^2/2$ . Noticing  $\|b\|^2 = \|\alpha - \theta\|^2$  and  $\|\alpha + \theta\|^2 + \|\alpha - \theta\|^2 = \|2\theta\|^2 = 4$ , we have  $\|b\|^2 - b_p^2 = \|\alpha + \theta\|^2 \|\alpha - \theta\|^2/4$ . Let  $r$  be such that  $B(-\theta, r) \subset D^c$ . Then

$$\inf_{\alpha \in D} \|\alpha + \theta\| \geq \inf_{\alpha \in B(-\theta, r)^c} \|\alpha - (-\theta)\| \geq r.$$

Thus,

$$\begin{aligned} R_1(\alpha) &= \frac{1}{L_\alpha} \left( (\alpha - \theta)' (nW) (\alpha - \theta) + T_4(\alpha) \right) = \frac{n}{L_\alpha} b' \Lambda b + \frac{T_4(\alpha)}{L_\alpha} \\ &\geq \frac{n}{L_\alpha} \lambda_m \sum_{i=1}^{p-1} b_i^2 + \frac{T_4(\alpha)}{L_\alpha} = \frac{n}{L_\alpha} \lambda_m (\|b\|^2 - b_p^2) + \frac{T_4(\alpha)}{L_\alpha} \\ &\geq c_1 \|\alpha - \theta\|^2 + o_p(\|\alpha - \theta\|^2), \end{aligned}$$

where  $c_1 = r\lambda_m/U$  with  $U$  being an upper bound for  $L(\cdot)$ .

It is left to show  $\text{rank}(W) = p - 1$ . Since  $\theta' \mu(x, \theta) \equiv 0$ , we have  $\theta' W \theta = 0$ . Now take  $\beta \perp \theta$ . Notice

$$\beta' W \beta = \int \beta'(x - E(X|\theta'X = \theta'x)) (x - E(X|\theta'X = \theta'x))' \beta (g^{(1)}(\theta'x))^2 L_{q, \theta}(x) f(x) dx.$$

By our assumptions,  $g(\cdot)$  is not constant in some interior part. Hence there exists a  $t \in (-1, 1)$  such that  $g^{(1)}(t) \neq 0$ . Take  $q > |t|$ . By the continuity of  $g^{(1)}$  and the

assumption on the function  $L(\cdot)$ , we can find a ball  $B$  on which  $|g^{(1)}(\theta'x)| \geq C_1$  and  $L_{q,\theta}(x) \geq C_1$  for some  $C_1 > 0$ . Hence

$$\beta'W\beta \geq C_2 \int_B \beta'(x - E(X|\theta'X = \theta'x))(x - E(X|\theta'X = \theta'x))' \beta dx, \quad (\text{B.7})$$

for some constant  $C_2 > 0$ . The right-hand side of B.7 becomes zero if and only if  $\beta'X = E(\beta'X|\theta'X) = \xi(\theta'X)$  when  $X \in B$ , for some function  $\xi(\cdot)$ . This cannot happen if  $\beta$  and  $\theta$  are orthogonal since  $(\theta'X, \beta'X)$  have a positive joint density on  $B$ . Hence  $\text{rank}(W) = p - 1$  and the proof is completed.

Next, we consider fixed-design models. Recall that  $g_\alpha(\cdot)$  was defined in Section 2.1.2. Noticing  $A\alpha = (1, 0, \dots, 0)'$  and  $\int y_i dy_i = 0$ , we have that, under  $H_0$ ,

$$\begin{aligned} g_\alpha(y_1) &= \frac{\int_{S(y_1)} g(\theta'A'y) dy_2 \cdots dy_p}{\int_{S(y_1)} dy_2 \cdots dy_p} = \frac{\int_{S(y_1)} g(y_1 + (\theta - \alpha)'A'y) dy_2 \cdots dy_p}{\int_{S(y_1)} dy_2 \cdots dy_p} \\ &= \frac{\int_{S(y_1)} (g(y_1) + g^{(1)}(y_1)(\theta - \alpha)'A'y + o(\|\theta - \alpha\|)) dy_2 \cdots dy_p}{\int_{S(y_1)} dy_2 \cdots dy_p} \\ &= g(y_1) + g^{(1)}(y_1) \cdot (\theta - \alpha)'(y_1\alpha) + o(\|\theta - \alpha\|). \end{aligned}$$

Hence,

$$\begin{aligned} \sum_{i=1}^n m_\alpha(x_i)^2 L_{q,\alpha}(x_i) &= n \int_S (g(\theta'x) - g_\alpha(\alpha'x))^2 L_{q,\alpha}(x) dx + O(nh^3) \\ &\geq n \int (g(\theta'x) - g(\alpha'x) - g^{(1)}(\alpha'x) \cdot (\theta - \alpha)'(\alpha\alpha'x) + o(\|\theta - \alpha\|))^2 L_{q,\alpha}(x) dx + O(nh^3) \\ &= n \int (g^{(1)}(\alpha'x) \cdot (\theta - \alpha)'(x - \alpha\alpha'x))^2 L_{q,\theta}(x) dx + o(n\|\theta - \alpha\|^2). \end{aligned}$$

Using the same argument as in random design case, there exists a  $C_3 > 0$  and an open ball  $B$  on which  $|g^{(1)}(\theta'x)| \geq C_3$  and  $L_{q,\theta}(x) \geq C_3$ . Thus, there exists a  $C_4 > 0$  such that  $|g^{(1)}(\alpha'x)| \geq C_3 - C_4\|\alpha - \theta\|$  and

$$\begin{aligned} \sum_{i=1}^n m_\alpha(x_i)^2 L_{q,\alpha}(x_i) &\geq (C_3 - C_4\|\alpha - \theta\|)^2 C_3 \cdot n \int_B ((\theta - \alpha)'(x - \alpha\alpha'x))^2 dx + o(n\|\theta - \alpha\|^2) \\ &= C_3^3 n (\theta - \alpha)'(I - \alpha\alpha') \int_B xx'dx (I - \alpha\alpha')(\theta - \alpha) + o(n\|\theta - \alpha\|^2). \end{aligned}$$

Let  $V = \int_B xx'dx$  and  $W(\alpha) = (I - \alpha\alpha')V(I - \alpha\alpha')$ . Then  $W(\alpha) = W(\theta) + O(\|\alpha - \theta\|)$  and hence

$$\sum_{i=1}^n m_\alpha(x_i)^2 L_{q,\alpha}(x_i) \geq C_3^3 n (\theta - \alpha)'W(\theta)(\theta - \alpha) + o(n\|\theta - \alpha\|^2).$$

Clearly  $\theta'W(\theta)\theta = 0$ . Take  $\beta \perp \theta$  and  $\beta \neq 0$ . Then  $\beta'W\beta = \beta'V\beta > 0$  since  $V$  is positive definite and  $\beta \neq 0$ . Hence  $W$  has rank  $p - 1$  and the proof is completed as in random design case.

The performance of  $R_1$  under  $H_a$  is rather straightforward. Since the result for the fixed design models follows directly by integral approximation (similar to the proof of Theorem 2.7 in Appendix C), we shall only show the result for the more difficult case, the random design models. Let  $U$  be an upper bound of the function  $L(\cdot)$ . Then

$$\begin{aligned} \inf_{\alpha} R_1(\alpha) &\geq \frac{\inf_{\alpha} \sum_{i=1}^n m_{\alpha}(X_i)^2 L_{q,\alpha}(X_i)}{\sup_{\alpha} \sum_{i=1}^n L_{q,\alpha}(X_i)} \\ &\geq \frac{1}{nU} \inf_{\alpha} \sum_{i=1}^n m_{\alpha}(X_i)^2 L_{q,\alpha}(X_i) \geq \frac{1}{nU} (A_1 - A_2), \end{aligned}$$

where

$$A_1 = \inf_{\alpha} \sum_{i=1}^n E(m_{\alpha}(X_i)^2 L_{q,\alpha}(X_i)),$$

and

$$A_2 = \frac{1}{nU} \sup_{\alpha} \left| \sum_{i=1}^n \left( m_{\alpha}(X_i)^2 L_{q,\alpha}(X_i) - E m_{\alpha}(X_i)^2 L_{q,\alpha}(X_i) \right) \right|.$$

Since  $A_1 = nc_1$  with  $c_1 = \inf_{\alpha} E m_{\alpha}(X)^2 L_{q,\alpha}(X) > 0$ , and, by Lemma B.2,  $A_2 = o_p(n^{\frac{1}{2}+\xi})$  for any  $\xi > 0$ , we get  $\inf_{\alpha} R_1(\alpha) \geq c_0 + o_p(1)$  with  $c_0 = c_1/U > 0$ .  $\square$

**Proof of Lemma 1 ( Terms R0 and R2-R9).**

Notice that the results of the previous lemmas for random-design and fixed-design cases take the same form except some difference like  $o_p(\cdot n^{\xi})$  replaced by  $O(\cdot)$ . Hence we shall state the proofs mainly using the random-design form. In each part, we shall first give the proof of the result under  $H_0$  followed by the proof under the alternative  $H_a$ .

**(R0)** Notice that

$$R_0(\alpha) - R_0(\beta) = \frac{1}{L_{\alpha}L_{\beta}} \sum_{i=1}^n (\epsilon_i^2 - \sigma^2) (L_{q,\alpha}(X_i)L_{\beta} - L_{q,\beta}(X_i)L_{\alpha}),$$

and  $L_{q,\alpha}(X_i)L_{\beta} - L_{q,\beta}(X_i)L_{\alpha} = (L_{q,\alpha}(X_i) - L_{q,\beta}(X_i))L_{\beta} + L_{q,\beta}(X_i)(L_{\beta} - L_{\alpha})$  is uniformly bounded by  $C_1 n \delta$  for some constant  $C_1$ . Since  $\epsilon^2 - \sigma^2$  has mean zero and at least four moments, the result follows by Lemma B.4 and Lemma B.5. Notice that the result holds under both hypotheses..

**(R2)** It's a direct consequence of Lemma B.5 and Lemma B.6 for both models and both hypotheses.

**(R3)** By Lemma B.5 and Lemma B.8 we have

$$\begin{aligned} \sup_{\alpha} |R_3(\alpha)| &\leq \frac{1}{\inf_{\alpha} L_{\alpha} \cdot \inf_{\alpha} \min_{\{i|L_{q,\alpha}(X_i) \neq 0\}} K_i(\alpha)} \sum_{i=1}^n \sup_{\alpha} \left( \sum_{j=1}^n K_h(\alpha'X_j - \alpha'X_i) \epsilon_j \right)^2 \\ &\leq O_p\left(\frac{1}{n \cdot (nh)^2}\right) \cdot o_p(n^{2+\xi}h) = o_p(h^{-1}n^{-1+\xi}). \end{aligned}$$

Notice that the result also holds under both hypotheses.

**(R4+R5)** Under  $H_a$ ,

$$R_4(\alpha) + R_5(\alpha) = \sum_{i\alpha} (m(X_i) - K_{mi}(\alpha))\epsilon_i.$$

Since  $m(X_i) - K_{mi}(\alpha)$  is uniformly bounded, by Lemma B.4 we have,

$$\sup_{\alpha} \left| \sum_{i=1}^n (m(X_i) - K_{mi}(\alpha))\epsilon_i \right| = o_p(n^{\frac{1}{2}+\xi}),$$

for all  $\xi > 0$ . Hence Lemma B.5 gives that

$$\sup_{\alpha} |R_4(\alpha) + R_5(\alpha)| = o_p(n^{-\frac{1}{2}+\xi}).$$

Now assume  $H_0$  hold. Clearly  $R_4(\theta) = 0$ . We shall show  $R_5(\theta) = O_p(n^{-\frac{1}{2}}h^2)$ . Note that under  $H_0$ ,  $g_{\theta}(\theta'x) = g(\theta'x)$  and, with  $E_{it}(\cdot)$  denoting  $E(\cdot|\theta'X_i = t)$ ,

$$\begin{aligned} & E\left(\sum_{i=1}^n (g(\theta'X_i) - K_{mi}(\theta))L_{q,\theta}(X_i)\epsilon_i\right)^2 \\ &= \sigma^2 \sum_{i=1}^n E(m(X_i) - K_{mi}(\theta))^2 L_{q,\theta}^2(X_i) \\ &= \sigma^2 \sum_{i=1}^n E\left(\sum_{j=1}^n \kappa_{ij}(\theta)(g(\theta'X_i) - g(\theta'X_j))L_{q,\theta}(X_i)\right)^2 \\ &\leq C_1 \sum_{i=1}^n \int_{c_{\theta}-q_{\theta}}^{c_{\theta}+q_{\theta}} E_{it}\left(\sum_{j=1}^n \kappa_{ij}(\theta)(g(t) - g(\theta'X_j))\right)^2 f_{\theta}(t)dt \\ &= C_1 \sum_{i=1}^n \int_{c_{\theta}-q_{\theta}}^{c_{\theta}+q_{\theta}} \left(\sum_{j=1}^n A_{ijt} + \sum_{j \neq k} B_{ijkt}\right) f_{\theta}(t)dt, \end{aligned}$$

where  $C_1$  is a positive constant,

$$A_{ijt} = E_{it}\left(\kappa_{ij}^2(\theta)(g(\theta'X_i) - g(\theta'X_j))^2\right),$$

and

$$B_{ijkt} = E_{it}\left(\kappa_{ij}(\theta)\kappa_{ik}(\theta)(g(\theta'X_i) - g(\theta'X_j))(g(\theta'X_i) - g(\theta'X_k))\right).$$

By Lemma B.9, uniformly in  $i, j$  and  $t$ ,

$$\begin{aligned} \int_{c_{\theta}-q_{\theta}}^{c_{\theta}+q_{\theta}} A_{ijt} f_{\theta}(t)dt &\leq E\left(\kappa_{ij}^2(\theta)(g(\theta'X_i) - g(\theta'X_j))^2\right) \\ &= O(h^2)E\kappa_{ij}^2(\theta) = O(n^{-2}h). \end{aligned}$$

Let  $W = \sum_{s \neq i, j, k} K_h(\theta' X_s - t)$  with density  $f_W(\cdot)$ . Since  $t \in c_\theta \pm (w_\theta - h)$  and  $K(\cdot)$  is an even function, using Taylor's expansion we have, uniformly in  $i$  and  $t$ ,

$$\begin{aligned}
B_{ijkt} &= E \frac{K_h(\theta' X_j - t) K_h(\theta' X_k - t) (g(\theta' X_j) - g(t)) (g(\theta' X_k) - g(t))}{(W + K_h(\theta' X_j - t) + K_h(\theta' X_k - t) + K(0))^2} \\
&= h^2 \iiint \frac{K(u) K(v) (g(t + uh) - g(t)) (g(t + vh) - g(t))}{(w + K(u) + K(v))^2} \\
&\quad \cdot f_\theta(t + uh) f_\theta(t + vh) f_W(w) du dv dw \\
&= h^2 \cdot O(h^4) \iiint_{\{W \geq \delta_n\} \cup \{W < \delta_n\}} \frac{K(u) K(v)}{(w + K(u) + K(v))^2} f_W(w) du dv dw \\
&= O(h^6) \left( \frac{C_2}{\delta_n^2} + P(W < \delta_n) \right),
\end{aligned}$$

where  $C_2$  is a constant and  $\delta_n$  is any sequence. A method similar to the proof of Lemma B.9 yields that  $P(W < \delta_n) = o(n^{-2})$  by taking, say,  $\delta_n = EW - (nh)^{\frac{2}{3}}$  which is of order  $O(nh)$  uniformly in  $t \in c_\theta \pm q_\theta$ . Hence, noticing that all the rates hold uniformly,

$$\begin{aligned}
&E \left( \sum_{i=1}^n (g(\theta' X_i) - K_{mi}(\theta)) L_{q, \theta}(X_i) \epsilon_i \right)^2 \\
&\leq C_1 \sum_{i=1}^n \left( \sum_{j=1}^n O(n^{-2}h) + \sum_{j \neq k} O(n^{-2}h^4) \right) f_\theta(t) dt = O(nh^4),
\end{aligned}$$

and, by Lemma B.5,

$$R_5(\theta) = O_p(n^{-1} \sqrt{nh^4}) = O_p(n^{-\frac{1}{2}} h^2).$$

Next we examine  $R_4(\alpha) + R_5(\alpha) - R_4(\theta) - R_5(\theta)$ . Let  $Z_i(\alpha) = m(X_i) - K_{mi}(\alpha)$ . Since  $m(\cdot)$  is bounded,  $K_{mi}(\alpha)$  and  $Z_i(\alpha)$  are both uniformly bounded. We can write

$$\begin{aligned}
R_4(\alpha) + R_5(\alpha) - R_4(\theta) - R_5(\theta) &= \sum_{i=1}^n \left( \frac{Z_i(\alpha) L_{q, \alpha}(X_i)}{L_\alpha} - \frac{Z_i(\theta) L_{q, \theta}(X_i)}{L_\theta} \right) \epsilon_i \\
&= A_1 + A_2 + A_3,
\end{aligned}$$

where

$$\begin{aligned}
A_1 &= \frac{L_\theta - L_\alpha}{L_\alpha L_\theta} \sum_{i=1}^n Z_i(\alpha) L_{q, \alpha}(X_i) \epsilon_i; \\
A_2 &= \frac{1}{L_\theta} \sum_{i=1}^n Z_i(\alpha) (L_{q, \alpha}(X_i) - L_{q, \theta}(X_i)) \epsilon_i; \\
A_3 &= \frac{1}{L_\theta} \sum_{i=1}^n (Z_i(\alpha) - Z_i(\theta)) L_{q, \theta}(X_i) \epsilon_i.
\end{aligned}$$

By Lemma B.4, Lemma B.5 and Lemma 2.6(v),  $A_1$  and  $A_2$  are both  $o_p(n^{-\frac{1}{2}+\xi}\delta)$ . For  $A_3$ , since  $Z_i(\alpha) - Z_i(\theta) = K_{mi}(\theta) - K_{mi}(\alpha)$ , we have

$$\begin{aligned} A_3 &= \frac{1}{L_\theta} \sum_{i=1}^n \left( \sum_{j=1}^n (\kappa_{ij}(\theta) - \kappa_{ij}(\alpha)) m(X_j) \right) L_{q,\theta}(X_i) \epsilon_i \\ &= \frac{1}{L_\theta} (A_{41} + A_{42} + A_{51} + A_{52}), \end{aligned}$$

where, for  $r = 1, 2$ ,

$$\begin{aligned} A_{4r} &= \sum_{i=1}^n \frac{K_i(\theta) - K_i(\alpha)}{K_i(\theta)} \left( \sum_{j=1}^n \kappa_{ij}(\alpha) m(X_j) \right) L_{q,\theta}(X_i) \epsilon_{ir}, \\ A_{5r} &= \sum_{i=1}^n \frac{K_{mi}(\alpha) K_i(\alpha) - K_{mi}(\theta) K_i(\theta)}{K_i(\theta)} L_{q,\theta}(X_i) \epsilon_{ir}, \end{aligned}$$

with  $\epsilon_{ir}$  defined as in (B.3). Let  $X = (X_1, \dots, X_n)$ . Define

$$B_q(X) = \min_{\{i | L_{q,\theta}(X_i) \neq 0\}} K_i(\theta),$$

and

$$D_\delta(X) = \sup_{\|\alpha - \theta\| \leq \delta} \max_{1 \leq i \leq n} |K_{mi}(\alpha) K_i(\alpha) - K_{mi}(\theta) K_i(\theta)|.$$

By Lemma B.5,  $B_q(X) \geq c_K n h + A_q(X)$  with  $A_q(X) = o_p(nh)$ ; and by Lemma B.7,  $D_\delta(X) = o_p(nh\delta n^\xi)$ . Hence,

$$\sup_{\|\alpha - \theta\| \leq \delta} |A_{52}| \leq \sqrt{\sum_{i=1}^n \frac{D_\delta^2(X)}{B_q^2(X)} \sum_{i=1}^n \epsilon_{i2}^2} \leq o_p(n^{1 - \frac{v-2}{2}t + \xi}\delta).$$

Similarly, we can show that the same bound works for  $A_{42}$ . For  $A_{51}$ , let  $A_\delta$  be a discrete set of size  $n^s$  constructed as in Lemma B.2 such that

$$\sup_{\|\alpha - \theta\| \leq \delta} |A_{51}| - \sup_{\alpha \in A_\delta} |A_{51}| \leq c_1 n^{-c_2 s},$$

for some positive constants  $c_1, c_2$ . Let

$$E_n = \{x \in \mathbb{R}^{np} \mid |A_q(x)| < \frac{1}{2} c_K n h\}$$

and, for any  $c_3 > 0$  let

$$F_n = \left\{ x \in \mathbb{R}^{np} \mid \sup_{\|\alpha - \theta\| \leq \delta} |D_\delta(x)| \leq c_3 n h \delta n^\xi \right\}.$$

Clearly  $P(X \in E_n \cap F_n) \rightarrow 1$ . By Bernstein's inequality, for all  $\gamma_n > 0$  and some constant  $c_4 > 0$  that is free of  $x$ ,

$$\begin{aligned}
& P\left(\sup_{\alpha \in A_\delta} |A_{51}| \geq \gamma_n\right) \\
&= \int_{E_n \cap F_n} P\left(\sup_{\alpha \in A_\delta} \left|\sum_{i=1}^n \frac{K_{mi}(\alpha)K_i(\alpha) - K_{mi}(\theta)K_i(\theta)}{K_i(\theta)} L_{q,\theta}(x_i)\epsilon_{ir}\right| \geq \gamma_n\right) dF_X(x) + o(1) \\
&\leq \int_{E_n \cap F_n} n^s 2 \exp\left\{-\frac{\gamma_n^2}{2c_4 n^{1+\xi}\delta^2 + c_4 \delta n^{t+\xi}\gamma_n}\right\} dF_X(x) + o(1) \\
&= 2 \exp\left\{-\frac{\gamma_n^2}{2c_4 n^{1+\xi}\delta^2 + c_4 \delta n^{t+\xi}\gamma_n} + s \ln n\right\},
\end{aligned}$$

which shows (by taking  $s$  large enough),

$$\sup_{\alpha \in A_\delta} |A_{51}| = o_p((\sqrt{n} + n^t)\delta n^\xi).$$

A slight change of the function  $D_\delta(x)$  and a similar proof gives the same rate for  $A_{41}$ . Taking  $t = 2/v$  above, we get the bound  $o_p(n^{-\frac{1}{2}+\xi}\delta)$  for  $A_3$  and the proof is completed.

**(R6)** The proofs for random-design and fixed-design models are identical in this case. First observe that, by Lemma B.9,

$$E\left(\sum_{i,j=1}^n L_{q,\theta}(X_i)\kappa_{ij}(\alpha)\epsilon_i\epsilon_j\right)^2 = O(h^{-1}).$$

Hence, by Lemma B.5,  $R_6(\theta) = O_p(n^{-1}h^{-\frac{1}{2}})$ . Next we examine  $R_6(\alpha) - R_6(\theta)$ . Notice that

$$\begin{aligned}
R_6(\alpha) - R_6(\theta) &= \frac{1}{L_\alpha L_\theta} \left( L_\theta \sum_{i,j=1}^n \kappa_{ij}(\alpha)\epsilon_i\epsilon_j L_{q,\alpha}(X_i) - L_\alpha \sum_{i,j=1}^n \kappa_{ij}(\theta)\epsilon_i\epsilon_j L_{q,\theta}(X_i) \right) \\
&= \sum_{i,j=1}^n \frac{\epsilon_i\epsilon_j}{L_\alpha L_\theta} \left( L_\theta [\kappa_{ij}(\alpha)L_{q,\alpha}(X_i) - \kappa_{ij}(\theta)L_{q,\theta}(X_i)] + (L_\theta - L_\alpha)\kappa_{ij}(\theta)L_{q,\theta}(X_i) \right) \\
&= \frac{1}{L_\alpha L_\theta} (A_1 + A_2 + A_3 + A_4),
\end{aligned}$$

where

$$\begin{aligned}
A_1 &= L_\theta \sum_{i=1}^n \epsilon_{i1} \sum_{j=1}^n [\kappa_{ij}(\alpha) L_{q,\alpha}(X_i) - \kappa_{ij}(\theta) L_{q,\theta}(X_i)] \epsilon_j, \\
A_2 &= L_\theta \sum_{i=1}^n \epsilon_{i2} \sum_{j=1}^n [\kappa_{ij}(\alpha) L_{q,\alpha}(X_i) - \kappa_{ij}(\theta) L_{q,\theta}(X_i)] \epsilon_j, \\
A_3 &= (L_\theta - L_\alpha) \sum_{i=1}^n L_{q,\theta}(X_i) \epsilon_{i1} \sum_{j=1}^n \kappa_{ij}(\theta) \epsilon_j, \\
A_4 &= (L_\theta - L_\alpha) \sum_{i=1}^n L_{q,\theta}(X_i) \epsilon_{i2} \sum_{j=1}^n \kappa_{ij}(\theta) \epsilon_j,
\end{aligned}$$

with  $\epsilon_{i1}$  and  $\epsilon_{i2}$  defined as in (B.3). Notice that we can write

$$\begin{aligned}
A_1 &= L_\theta \sum_{i=1}^n L_{q,\alpha}(X_i) \epsilon_{i1} \sum_{j=1}^n (\kappa_{ij}(\alpha) - \kappa_{ij}(\theta)) \epsilon_j \\
&\quad - L_\theta \sum_{i=1}^n (L_{q,\alpha}(X_i) - L_{q,\theta}(X_i)) \epsilon_{i1} \sum_{j=1}^n \kappa_{ij}(\theta) \epsilon_j;
\end{aligned}$$

and

$$\begin{aligned}
A_2 &= L_\theta \sum_{i=1}^n L_{q,\alpha}(X_i) \epsilon_{i2} \sum_{j=1}^n (\kappa_{ij}(\alpha) - \kappa_{ij}(\theta)) \epsilon_j \\
&\quad - L_\theta \sum_{i=1}^n (L_{q,\alpha}(X_i) - L_{q,\theta}(X_i)) \epsilon_{i2} \sum_{j=1}^n \kappa_{ij}(\theta) \epsilon_j.
\end{aligned}$$

By Lemma 2.6, Lemma B.3 and Lemma B.8 we have

$$\begin{aligned}
\sup_{\|\alpha-\theta\|\leq\delta} A_1 &\leq o_p(n^{1+t} \cdot (\sqrt{nh^{-1}} + n^{\frac{1}{v}} h^{-2}) \delta n^\xi) + o_p(n^{1+t} \delta \cdot \frac{1}{nh} \cdot n^{\frac{3}{2}+\xi} \sqrt{h}); \\
\sup_{\|\alpha-\theta\|\leq\delta} A_2 &\leq O(n) \left( \sum_{i=1}^n \epsilon_{i2}^2 \right)^{\frac{1}{2}} \cdot \left( \sup_{\|\alpha-\theta\|\leq\delta} \sum_{i=1}^n \left( \sum_{j=1}^n (\kappa_{ij}(\alpha) - \kappa_{ij}(\theta)) \epsilon_j \right)^2 \right)^{\frac{1}{2}} \\
&\quad + O_p(n\delta) \left( \sum_{i=1}^n \epsilon_{i2}^2 \right)^{\frac{1}{2}} \cdot \left( \sup_{\|\alpha-\theta\|\leq\delta} \sum_{i=1}^n \left( \sum_{j=1}^n \kappa_{ij}(\theta) \epsilon_j \right)^2 \right)^{\frac{1}{2}} \\
&= o_p\left( n \cdot n^{\frac{1-(v-2)t}{2}} \cdot (h^{-1} + n^{\frac{4}{v}-1} h^{-4})^{\frac{1}{2}} \delta n^\xi + n\delta \cdot n^{\frac{1-(v-2)t}{2}} \cdot \frac{n^{1+\xi} \sqrt{h}}{nh} \right); \\
\sup_{\|\alpha-\theta\|\leq\delta} A_3 &\leq o_p\left( n^{1+t} \delta \cdot \frac{1}{nh} n^{\frac{3}{2}+\xi} \sqrt{h} \right);
\end{aligned}$$

and

$$\begin{aligned} \sup_{\|\alpha-\theta\|\leq\delta} A_4 &\leq O_p(n\delta) \left( \sum_{i=1}^n \epsilon_{i2}^2 \right)^{\frac{1}{2}} \cdot \left( \sup_{\|\alpha-\theta\|\leq\delta} \sum_{i=1}^n \left( \sum_{j=1}^n \kappa_{ij}(\theta) \epsilon_j \right)^2 \right)^{\frac{1}{2}} \\ &= o_p(n\delta \cdot n^{\frac{1-(v-2)t}{2}} \cdot \frac{1}{nh} n^{1+\xi} \sqrt{h}). \end{aligned}$$

Combining the above terms and taking  $t = 0$  we get

$$\sup_{\|\alpha-\theta\|\leq\delta} |R_6(\alpha) - R_6(\theta)| \leq o_p(n^{-\frac{1}{2}+\xi} \delta h^{-\frac{1}{2}} + n^{-1+\frac{2}{v}+\xi} h^{-2} \delta).$$

Again, the result holds under both hypotheses..

**(R7)** In both cases, it is a direct consequence of Lemma B.6 and the fact that  $m_\alpha(x)$  is bounded by  $O(\|\alpha - \theta\|)$  everywhere under  $H_0$  and by  $O(1)$  under  $H_a$ .

**(R8)** Since  $m_\alpha(x)$  is bounded by  $O(\|\alpha - \theta\|)$  everywhere, from Lemma B.5 and Lemma B.8 we get

$$\sup_{\|\alpha-\theta\|\leq\delta} |R_8(\alpha)| \leq o_p\left(\frac{1}{n} \cdot \delta \cdot \frac{1}{nh} \sum_{i=1}^n \sup_{\alpha} \left| \sum_{j=1}^n K_h(\alpha' X_j - \alpha' X_i) \epsilon_j \right| \right) = o_p(\delta n^\xi \sqrt{nh^{-1}}).$$

Under  $H_a$ ,  $\delta$  is replaced by  $O(1)$ .

**(R9)** By Lemma B.5, B.6 and B.8 we have under both hypotheses

$$\sup_{\alpha \in D} |R_9(\alpha)| \leq o_p\left(\frac{1}{n} \cdot h^2 \cdot \frac{1}{nh} \cdot n^{1.5+\xi} \sqrt{h}\right) = o_p(n^{-\frac{1}{2}+\xi} h^{3/2}).$$

□

### ***Proof of Lemma 2.***

(i) Since  $\hat{d}(\hat{\theta})$  minimizes  $\hat{d}(\alpha)$ , we have  $\hat{d}(\hat{\theta}) \leq \hat{d}(\theta)$ , i.e.,

$$\sum_{i=0}^9 R_i(\hat{\theta}) \leq \sum_{i=0}^9 R_i(\theta).$$

Since  $R_1(\theta) = 0$  under  $H_0$ , by Lemma 1 and noticing  $n^{-\frac{1}{2}} \delta \leq n^{-\frac{1}{2}} h^{-\frac{1}{2}} \delta$ , we have

$$\begin{aligned} R_1(\hat{\theta}) &\leq \sum_{i \neq 1} |R_i(\hat{\theta}) - R_i(\theta)| \leq \sum_{i \neq 1} \sup_{\|\alpha-\theta\|\leq\delta} |R_i(\alpha) - R_i(\theta)| \\ &\leq o_p\left((h^4 + n^{-1} h^{-1} + n^{-\frac{1}{2}} h^{-\frac{1}{2}} \delta + n^{-1+\frac{2}{v}} h^{-2} \delta + h^2 \delta + n^{-\frac{1}{2}} h^{\frac{3}{2}}) n^\xi\right) \\ &= o_p\left((h^4 + n^{-1} h^{-1} + n^{-\frac{1}{2}} h^{\frac{3}{2}}) n^\xi + \delta (n^{-\frac{1}{2}} h^{-\frac{1}{2}} + n^{-1+\frac{2}{v}} h^{-2} + h^2) n^\xi\right), \end{aligned}$$

where we currently take  $\delta = \delta_0 = \sqrt{2}$ . The right hand side is balanced for  $h = O(n^{-\frac{1}{5}})$ , for which  $R_1(\hat{\theta}) \leq o_p(N(\delta) n^\xi)$  where

$$N(\delta) = n^{-\frac{4}{5}} + \delta n^{-\frac{2}{5}}.$$

From Lemma 1, there exists  $c_1 > 0$  such that

$$c_1 \|\hat{\theta} - \theta\|^2 + o(\|\hat{\theta} - \theta\|^2) \leq R_1(\hat{\theta}) \leq o_p(n^{-\frac{4}{5}+\xi} + \delta n^{-\frac{2}{5}+\xi}). \quad (\text{B.8})$$

Then (B.8) with  $\delta = \delta_0$  implies  $\|\hat{\theta} - \theta\| = o_p(n^{-\frac{1}{5}+\xi})$ . We claim that  $R_1(\hat{\theta}) \leq o_p(N(\delta_1)n^\xi) = o_p(n^{-\frac{3}{5}+\xi})$  where  $\delta_1 = O(n^{-\frac{1}{5}})$ . To see this, fix any  $\xi_0 > 0$  and let

$$B_n = \{\|\hat{\theta} - \theta\| \leq \delta_1\}.$$

Then  $P(B_n) \rightarrow 1$  and, for all  $\gamma > 0$ ,

$$\begin{aligned} & P\left(R_1(\hat{\theta}) \leq \gamma n^{-\frac{3}{5}+\xi}\right) \\ & \geq P\left(\sum_{i \neq 1} |R_i(\hat{\theta}) - R_i(\theta)| \leq \gamma n^{-\frac{3}{5}+\xi}\right) \\ & = P\left(\left\{\sum_{i \neq 1} |R_i(\hat{\theta}) - R_i(\theta)| \leq \gamma n^{-\frac{3}{5}+\xi}\right\} \cap B_n\right) + o(1) \\ & \geq P\left(\left\{\sum_{i \neq 1} \sup_{\|\alpha - \theta\| \leq \delta_1} |R_i(\alpha) - R_i(\theta)| \leq \gamma n^{-\frac{3}{5}+\xi}\right\} \cap B_n\right) + o(1) \\ & = P\left(\sum_{i \neq 1} \sup_{\|\alpha - \theta\| \leq \delta_1} |R_i(\alpha) - R_i(\theta)| \leq \gamma n^{-\frac{3}{5}+\xi}\right) + o(1) \rightarrow 1. \end{aligned}$$

Hence (B.8) with  $\delta = \delta_1$  implies  $\|\hat{\theta} - \theta\| = o_p(n^{-\frac{3}{10}+\xi})$ . Repeating this procedure we have  $\|\hat{\theta} - \theta\| = o_p(n^{-\frac{2}{5}+\xi})$ .

- (ii) The terms  $R_1, R_2, R_3$  and  $R_9$  are of order  $o_p(n^{-\frac{1}{2}})$  directly by Lemma 1. Now we consider terms  $R_4(\hat{\theta}) + R_5(\hat{\theta}), R_6(\hat{\theta}), R_7(\hat{\theta})$  and  $R_8(\hat{\theta})$ , which we shall denote by  $\tilde{R}_i(\hat{\theta})$  in the following. Let  $\gamma$  be any positive constant, let  $\xi > 0$  be sufficiently small and let

$$E_n = \{\|\hat{\theta} - \theta\| \leq n^{-\frac{2}{5}+\xi}\}.$$

Then  $P(E_n) \rightarrow 1$  and

$$\begin{aligned} P\left(\tilde{R}_i(\hat{\theta}) \leq \gamma n^{-\frac{1}{2}}\right) & = P\left(\{\tilde{R}_i(\hat{\theta}) \leq \gamma n^{-\frac{1}{2}}\} \cap E_n\right) + o(1) \\ & \geq P\left(\left\{\sup_{\|\alpha - \theta\| \leq n^{-\frac{2}{5}+\xi}} |\tilde{R}_i(\alpha)| \leq \gamma n^{-\frac{1}{2}}\right\} \cap E_n\right) + o(1) \\ & = P\left(\sup_{\|\alpha - \theta\| \leq n^{-\frac{2}{5}+\xi}} |\tilde{R}_i(\alpha)| \leq \gamma n^{-\frac{1}{2}}\right) + o(1) \rightarrow 1. \end{aligned}$$

which shows  $\tilde{R}_i(\hat{\theta}) = o_p(n^{-\frac{1}{2}})$  for these four terms.

- (iii) This is a direct consequence of Lemma 1.

(iv) Let  $\gamma$  be any positive constant and let  $E_n$  be defined as in (ii). Then

$$\begin{aligned}
& P\left(|R_0(\hat{\theta}) - R_0(\theta)| \leq \gamma n^{-\frac{9}{10}}\right) \\
&= P\left(\{|R_0(\hat{\theta}) - R_0(\theta)| \leq \gamma n^{-\frac{9}{10}}\} \cap E_n\right) + o(1) \\
&\geq P\left(\left\{\sup_{\|\alpha - \theta\| \leq n^{-\frac{2}{5} + \epsilon}} |R_0(\alpha) - R_0(\theta)| \leq \gamma n^{-\frac{9}{10}}\right\} \cap E_n\right) + o(1) \\
&= P\left(\sup_{\|\alpha - \theta\| \leq n^{-\frac{2}{5} + \epsilon}} |R_0(\alpha) - R_0(\theta)| \leq \gamma n^{-\frac{9}{10}}\right) + o(1) \rightarrow 1,
\end{aligned}$$

which completes the proof.  $\square$

## C Proofs of Theorem 2.7, Theorem 2.9 and Theorem 2.10

### *Proof of Theorem 2.7.*

Under  $H_0$ , by Lemma 1(i) and Lemma 2(ii), we have

$$\hat{d}(\hat{\theta}) = R_0(\theta) + o_p(n^{-\frac{4}{5} + \epsilon}).$$

Hence

$$\sqrt{n}(\hat{d}(\hat{\theta}) - \sigma^2) = \sqrt{n} \frac{\sum_{i=1}^n (\epsilon_i^2 - \sigma^2) L_{\theta, q}(X_i)}{\sum_{i=1}^n L_{\theta, q}(X_i)} + o_p(1),$$

and the asymptotic normality is immediate from the classical Central Limit Theorem.

Under  $H_a$ , by Lemma 2(iii) we have

$$\hat{d}(\hat{\theta}) = R_0(\hat{\theta}) + R_1(\hat{\theta}) + o_p(n^{-\frac{2}{5} + \epsilon}).$$

Since  $R_0(\theta_m) = \sigma^2 + o_p(n^{-\frac{1}{2} + \epsilon})$  and, by Lemma 1(i),  $R_0(\hat{\theta}) = R_0(\theta_m) + o_p(n^{-\frac{1}{2} + \epsilon})$ , we get

$$\hat{d}(\hat{\theta}) = R_0(\theta_m) + R_1(\hat{\theta}) + o_p(n^{-\frac{2}{5} + \epsilon}) = \sigma^2 + R_1(\hat{\theta}) + o_p(n^{-\frac{2}{5} + \epsilon}). \quad (\text{C.1})$$

Also, since  $R_1(\theta_m) = d_{m,L} + O(n^{-\frac{2}{5}})$ , we have  $\hat{d}(\theta_m) = \sigma^2 + d_{m,L} + o_p(n^{-\frac{2}{5} + \epsilon})$ . Then  $\hat{d}(\hat{\theta}) \leq \hat{d}(\theta_m)$  gives

$$R_1(\hat{\theta}) \leq d_{m,L} + o_p(n^{-\frac{2}{5} + \epsilon}). \quad (\text{C.2})$$

By (C.1), it suffices to show the other direction of (C.2). For fixed-design models,

$$\begin{aligned}
R_1(\hat{\theta}) &\geq \inf_{\alpha \in D} R_1(\alpha) = \inf_{\alpha \in D} \frac{\sum_{i=1}^n m_\alpha^2(x_i) L_{q, \alpha}(x_i)}{L_\alpha} \\
&= \inf_{\alpha \in D} \frac{n \int m_\alpha^2(x) L_{q, \alpha}(x) dx + O(nh^3)}{n \int L_{q, \alpha}(x) dx + O(nh^3)} = d_{m,L} + O(n^{-\frac{2}{5}}).
\end{aligned}$$

For random-design models,

$$\begin{aligned}
R_1(\hat{\theta}) &\geq \inf_{\alpha \in D} R_1(\alpha) \\
&= \inf_{\alpha \in D} \left( \frac{\sum_{i=1}^n [m_\alpha^2(X_i)L_{q,\alpha}(X_i) - \mu_\alpha]}{L_\alpha} + \frac{\mu_\alpha}{b_\alpha} + \frac{\mu_\alpha(nb_\alpha - L_\alpha)}{b_\alpha L_\alpha} \right) \\
&\geq \inf_{\alpha \in D} \frac{\mu_\alpha}{b_\alpha} - \sup_{\alpha \in D} \left| \frac{\mu_\alpha(nb_\alpha - L_\alpha)}{b_\alpha L_\alpha} \right| - \sup_{\alpha \in D} \left| \frac{\sum_{i=1}^n [m_\alpha^2(X_i)L_{q,\alpha}(X_i) - \mu_\alpha]}{L_\alpha} \right|, \quad (\text{C.3})
\end{aligned}$$

where

$$b_\alpha = EL_{q,\alpha}(X) \quad \text{and} \quad \mu_\alpha = E(m_\alpha^2(X)L_{q,\alpha}(X)).$$

Notice that  $\inf_{\alpha \in D} \frac{\mu_\alpha}{b_\alpha} = \frac{\mu_{\theta_m}}{b_{\theta_m}} = d_{m,L}$ . Now,

$$b_\alpha = EL_{q,\alpha}(X) = \int L\left(\frac{t-c_\alpha}{q_\alpha}\right) f_\alpha(t) dt = q_\alpha \int_{-1}^1 L(u) f_\alpha(c_\alpha + q_\alpha u) du \geq c_0 q_0 > 0,$$

which shows  $\inf_{\alpha} b_\alpha \geq c_0 q_0 > 0$ . By Lemma B.2 we have

$$\sup_{\alpha \in D} |L_\alpha - nb_\alpha| = o_p(n^{\frac{1}{2}+\xi}),$$

and

$$\sup_{\alpha \in D} \left| \sum_{i=1}^n (m_\alpha^2(X_i)L_{q,\alpha}(X_i) - \mu_\alpha) \right| = o_p(n^{\frac{1}{2}+\xi}).$$

Hence, noticing  $\sup_{\alpha} \mu_\alpha < \infty$  and applying Lemma B.5, (C.3) gives

$$R_1(\hat{\theta}) \geq \inf_{\alpha \in D} d_{m,L} + o_p(n^{-\frac{1}{2}+\xi}).$$

□

**Proof of Theorem 2.9.**

Denote the decompositions by  $\hat{d}_r(\hat{\theta}_r) = \sum_{i=0}^9 R_{ri}(\hat{\theta}_r)$ ,  $r = 1, 2$ . Then  $\hat{d}_1(\hat{\theta}_1) - \hat{d}_2(\hat{\theta}_2) = R_{10}(\theta) - R_{20}(\theta) + o_p(n^{-\frac{1}{2}})$ . Let  $a_r = EL_{q,\theta,r}(X)$ ,  $r = 1, 2$ . By Theorem 2.7(i),

$$\begin{aligned}
T &= \sqrt{n}(R_{10} - R_{20} + o_p(n^{-\frac{1}{2}})) \\
&= \sqrt{n} \left( \frac{1}{na_1} \sum_{i=1}^n (\epsilon_i^2 - \sigma^2) L_{q,\theta,1}(X_i) - \frac{1}{na_2} \sum_{i=1}^n (\epsilon_i^2 - \sigma^2) L_{q,\theta,2}(X_i) + R_1 + R_2 + o_p(n^{-\frac{1}{2}}) \right) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\epsilon_i^2 - \sigma^2) \left( \frac{1}{a_1} L_{q,\theta,1}(X_i) - \frac{1}{a_2} L_{q,\theta,2}(X_i) \right) + \sqrt{n}(R_1 + R_2) + o_p(1),
\end{aligned}$$

where, by the Law of Iterated Logarithm,

$$\begin{aligned}
R_r &= \frac{1}{\sum_{i=1}^n L_{q,\theta,r}(X_i)} \sum_{i=1}^n (\epsilon_i^2 - \sigma^2) L_{q,\theta,r}(X_i) - \frac{1}{na_r} \sum_{i=1}^n (\epsilon_i^2 - \sigma^2) L_{q,\theta,r}(X_i) \\
&= \frac{a_r - \frac{1}{n} \sum_{i=1}^n L_{q,\theta,r}(X_i)}{a_r \sum_{i=1}^n L_{q,\theta,r}(X_i)} \sum_{i=1}^n (\epsilon_i^2 - \sigma^2) L_{q,\theta,r}(X_i) \\
&= O_p(n^{-1}(\ln \ln n)^2), \quad r = 1, 2.
\end{aligned}$$

Hence

$$T = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\epsilon_i^2 - \sigma^2) \left( \frac{1}{a_1} L_{q,\theta,1}(X_i) - \frac{1}{a_2} L_{q,\theta,2}(X_i) \right) + o_p(1) \xrightarrow{\mathcal{D}} N(0, \tau').$$

Under  $H_a$ , by Theorem 2.7(ii) and Remark 2.8,  $T = \sqrt{n}(d_{m,L,1} - d_{m,L,2}) + o_p(n^{\frac{1}{10} + \xi})$  which diverges to infinity in absolute value at a rate of  $\sqrt{n}$ .  $\square$

**Proof of Theorem 2.10.**

We first assume  $H_0$  is true. Let  $W$  and  $\bar{\epsilon}$  be as defined in (2.6) and (2.7). The goal is to show

$$W = \frac{\sum_{i=1}^n \epsilon_i^2 I(X_i)}{\sum_{i=1}^n I(X_i)} + o_p(n^{-\frac{1}{2}}). \quad (\text{C.4})$$

Then the result follows by the traditional CLT. Let  $\xi$  be any positive number and

$$\bar{\epsilon} = \frac{\sum_{i=1}^n \epsilon_i I(X_i)}{\sum_{i=1}^n I(X_i)}.$$

Rewrite  $W$  as

$$W = \frac{\sum_{i=1}^n (\hat{\epsilon}_i - \bar{\epsilon})^2 I(X_i)}{\sum_{i=1}^n I(X_i)} = \frac{\sum_{i=1}^n \hat{\epsilon}_i^2 I(X_i)}{\sum_{i=1}^n I(X_i)} - \bar{\epsilon}^2.$$

By Lemma 2 we have, uniformly in  $i$  with  $I(X_i) \neq 0$ ,

$$\begin{aligned} |\hat{\epsilon}_i - \epsilon_i| &= |g(\theta' X_i) - \hat{g}_{\hat{\theta}}(\hat{\theta}' X_i)| \\ &= \frac{1}{K_i(\hat{\theta})} \left| \sum_{j=1}^n (g(\theta' X_i) - g(\theta' X_j) - \epsilon_j) K_h(\hat{\theta}' X_j - \hat{\theta}' X_i) \right| \\ &\leq \frac{1}{\inf_{\alpha} \min_{\{i | I(X_i) \neq 0\}} K_i(\alpha)} (M_{i1} + M_{i2} + M_{i3}), \end{aligned}$$

where

$$\begin{aligned} M_{i1} &= \left| \sum_{j=1}^n (g(\theta' X_i) - g(\theta' X_j)) K_h(\theta' X_j - \theta' X_i) \right|; \\ M_{i2} &= \left| \sum_{j=1}^n (g(\theta' X_i) - g(\theta' X_j)) (K_h(\hat{\theta}' X_j - \hat{\theta}' X_i) - K_h(\theta' X_j - \theta' X_i)) \right|; \\ M_{i3} &= \left| \sum_{j=1}^n \epsilon_j K_h(\hat{\theta}' X_j - \hat{\theta}' X_i) \right|. \end{aligned}$$

Note that  $M_{i1}$  is exactly the term  $|\sum_{j=1}^n Z_{ij}(\theta)|$  is the proof of Lemma B.6 and hence is uniformly of order  $O_p(nh^3)$ . By Lemma B.7 (used twice),  $M_{i2}$  is uniformly bounded by  $o_p(nh\delta n^\xi)$  with  $\delta = n^{-\frac{1}{3} + \xi}$  (since  $\|\hat{\theta} - \theta\| = o_p(n^{-\frac{1}{3} + \xi})$ ). And by Lemma B.4,

$$M_{i3} \leq \sup_{\alpha, i} \left| \sum_{j=1}^n K_h(\alpha' X_j - \alpha' X_i) \epsilon_j \right| = o_p(n^{\frac{1}{2} + \xi}).$$

Thus, by Lemma B.5

$$\max_{\{i|I(X_i)\neq 0\}} |\hat{\epsilon}_i - \epsilon_i| \leq o_p(n^{-1+\xi}h^{-1}) \cdot (O_p(nh^3) + o_p(n^{\frac{2}{3}+\xi}h) + o_p(n^{\frac{1}{2}+\xi})) = o_p(n^{-\frac{1}{2}+\xi}h^{-1}).$$

This immediately gives that

$$\bar{\epsilon}^2 = \bar{\epsilon}^2 + |\bar{\epsilon}| \cdot o_p(n^{-\frac{1}{2}+\xi}h^{-1}) + o_p(n^{-\frac{1}{2}+\xi}h^{-1})^2 = o_p(n^{-\frac{1}{2}}).$$

Noticing that

$$\frac{\sum_{i=1}^n \hat{\epsilon}_i^2 I(X_i)}{\sum_{i=1}^n I(X_i)} = \frac{\sum_{i=1}^n \epsilon_i^2 I(X_i)}{\sum_{i=1}^n I(X_i)} + \frac{\sum_{i=1}^n (\hat{\epsilon}_i - \epsilon_i) \epsilon_i}{\sum_{i=1}^n I(X_i)} + o_p(n^{-\frac{1}{2}}),$$

it suffices to show  $\sum_{i=1}^n (\hat{\epsilon}_i - \epsilon_i) \epsilon_i I(X_i) = o_p(\sqrt{n})$ . Now,

$$\sum_{i=1}^n (\hat{\epsilon}_i - \epsilon_i) \epsilon_i I(X_i) = \sum_{i=1}^n \sum_{j=1}^n (g(\theta' X_i) - g(\theta' X_j) - \epsilon_j) \kappa_{ij}(\hat{\theta}) \epsilon_i I(X_i) = A_1 + A_2,$$

where

$$A_1 = \sum_{i=1}^n \sum_{j=1}^n (g(\theta' X_i) - g(\theta' X_j)) \kappa_{ij}(\hat{\theta}) \epsilon_i,$$

and

$$A_2 = \sum_{i=1}^n \sum_{j=1}^n \kappa_{ij}(\hat{\theta}) \epsilon_i \epsilon_j I(X_i).$$

Notice that  $A_2 / \sum_{i=1}^n I(X_i)$  is a simplified version of  $R_6(\cdot)$  defined in the decomposition of  $\hat{d}(\cdot)$  (with  $L_{q,\alpha}(x)$  replaced by  $I(x)$ ). The result for  $R_6(\hat{\theta})$  in Lemma 2 also holds for  $A_2$ , i.e., under  $H_0$ ,  $A_2 / \sum_{i=1}^n I(X_i) = o_p(n^{-\frac{4}{5}+\xi})$ . For  $A_1$ , let

$$Z_i(\alpha) = \sum_{j=1}^n (g(\theta' X_i) - g(\theta' X_j)) \kappa_{ij}(\alpha), \quad i, j = 1 \cdots n.$$

When  $|\alpha - \theta| \leq \delta$ , in order for  $\kappa_{ij}(\alpha) \neq 0$ , uniformly we have  $|g(\theta' X_i) - g(\theta' X_j)| < C(\delta + h)$  for some constant  $C > 0$  and hence

$$|Z_i(\alpha)| \leq C(\delta + h) \sum_{j=1}^n \kappa_{ij}(\alpha) = C(\delta + h).$$

By Lemma B.4, noticing  $\|\hat{\theta} - \theta\| = o_p(n^{-\frac{1}{3}+\xi})$  and taking  $\delta = n^{-\frac{1}{3}+\xi}$ , we have

$$A_1 \leq \sup_{\|\alpha - \theta\| \leq \delta} \left| \sum_{i=1}^n Z_i(\alpha) \epsilon_i \right| = o_p(n^{\frac{1}{2}+\xi}h).$$

This finishes the proof of (C.4) under  $H_0$ .

Now we turn to  $H_a$ . Note that, uniformly in  $i$  with  $I(X_i) \neq 0$ ,  $\hat{\epsilon}_i = y_i - K_{mi}(\hat{\theta}) + o_p(1)$ . By Lemma B.6, uniformly in  $i$  with  $I(X_i) \neq 0$ ,  $K_{mi}(\hat{\theta}) = g_{\hat{\theta}}(\hat{\theta}' X_i) + o_p(h^2 n^\xi)$

for all  $\xi > 0$ . By Theorem 2.7,  $\hat{\theta} \xrightarrow{P} \theta_m$  when  $\theta_m$  is the unique minimizer of  $d_{m,L}$ . Hence, uniformly in  $i$  with  $I(X_i) \neq 0$ ,

$$\hat{\epsilon}_i = y_i - g_{\theta_m}(\theta'_m X_i) + o_p(1). \quad (\text{C.5})$$

By a method analogous to  $H_0$  case, we have

$$\sum_{i=1}^n (\hat{\epsilon}_i - \epsilon_i) \epsilon_i I(X_i) = o_p(n^{\frac{1}{2} + \xi}).$$

Hence,

$$\begin{aligned} \sum_{i=1}^n \hat{\epsilon}_i^2 I(X_i) &= \sum_{i=1}^n (\epsilon_i - \hat{\epsilon}_i)^2 I(X_i) + \sum_{i=1}^n \epsilon_i^2 I(X_i) + o_p(n) \\ &= \sum_{i=1}^n \epsilon_i^2 I(X_i) + \sum_{i=1}^n (m(X_i) - g_{\theta_m}(\theta'_m X_i))^2 I(X_i) + o_p(n) \end{aligned}$$

Finally, by (C.5),

$$\begin{aligned} \bar{\epsilon}^2 &= \frac{\left(\sum_{i=1}^n \hat{\epsilon}_i I(X_i)\right)^2}{\left(\sum_{i=1}^n I(X_i)\right)^2} \\ &= \frac{\left(\sum_{i=1}^n (\hat{\epsilon}_i - \epsilon_i) I(X_i)\right)^2}{\left(\sum_{i=1}^n I(X_i)\right)^2} + \frac{\left(\sum_{i=1}^n \epsilon_i I(X_i)\right)^2}{\left(\sum_{i=1}^n I(X_i)\right)^2} + \frac{2\left(\sum_{i=1}^n (\hat{\epsilon}_i - \epsilon_i) I(X_i)\right)\left(\sum_{i=1}^n \epsilon_i I(X_i)\right)}{\left(\sum_{i=1}^n I(X_i)\right)^2} \\ &= \frac{[E(m(X_i) - g_{\theta_m}(\theta'_m X_i))I(X_i)]^2}{[E I(X_i)]^2} + o_p(1). \end{aligned}$$

Thus,

$$\begin{aligned} W &= \frac{\sum_{i=1}^n \hat{\epsilon}_i^2 I(X_i)}{\sum_{i=1}^n I(X_i)} - \bar{\epsilon}^2 \\ &= \sigma^2 + \frac{E[(m(X) - g_{\theta_m}(\theta'_m X))^2 I(X)]}{E I(X)} - \frac{[E(m(X) - g_{\theta_m}(\theta'_m X))I(X)]^2}{[E I(X)]^2} + o_p(1), \end{aligned}$$

which, together with Theorem 2.7, completes the proof.  $\square$

## References

- [1] Duan, N., and Li, K.C. (1991), "Slicing Regression: A Link Free Regression Method," *The Annals of Statistics*, 19, 505-530.
- [2] Fan, J. (1996), "Test of Significance Based on Wavelet Thresholding and Neyman's Truncation," *Journal of the American Statistical Association*, 91, 674-688.

- [3] Fan, J., and Huang, L.S. (2001), "Goodness-of-fit tests for parametric regression models," *Journal of the American Statistical Association*, 96, 640–652.
- [4] Hall, P. (1989), "On Projection Pursuit Regression," *The Annals of Statistics*, 17, 573–588.
- [5] Hall, P., Kay, J.W., and Titterton, D.M. (1991), "On Estimation of Noise Variance in Two-Dimensional Signal Processing," *Advances in Applied Probability*, 23, 476–495.
- [6] Härdle, W., Hall, P., and Ichimura, H. (1993), "Optimal Smoothing in Single-Index Models," *The Annals of Statistics*, 21, 157–178.
- [7] 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.
- [8] Ichimura, H. (1993), "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models," *Journal of Econometrics*, 58, 71–120.
- [9] Kulasekera, K.B., and Gallagher, C. (2002), "Variance Estimation in Nonparametric Regression", *Communications in Statistics, Theory and Methods*, 31, 1373–1383.
- [10] 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.
- [11] Kulasekera, K.B., and Wang, J. (1998), "Bandwidth Selection in a Test of Equality of regression Curves", *Statistics & Probability Letters*, 37, 287–293.
- [12] Niederreiter, H. (1992), *Random Number Generation and Quasi-Monte Carlo Methods*, Vermont: Capital City Press.
- [13] Powell, J.L., Stock, J.H., and Stoker, T.M. (1989), "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57, 1403–1430.
- [14] Randles, R., and Wolfe, D. (1979), *Introduction to the Theory of Nonparametric Statistics*, Wiley, New York.
- [15] Spokoiny, V. (2002), "Variance Estimation for High-Dimensional Regression Models," *Journal of Multivariate Analysis*, 82, 111–133.
- [16] Xia, Y.C., Tong, H., and Li, W.K. (1999), "On Extended Partially Linear Single-Index Models," *Biometrika*, 86, 831–842.
- [17] Yu, Y., and Ruppert, D. (2002), "Penalized Spline Estimation for Partially Linear Single-Index Models," *Journal of the American Statistical Association*, 97, 1042–1054.
- [18] Delecroix, M., Härdle, W., and Hristache, M. (2003), "Efficient Estimation in Conditional Single-Index Regression," *Journal of Multivariate Analysis*, 86, 213–226.
- [19] Härdle, W., and Stoker, T. M. (1989), "Investigating Smooth Multiple Regression by the Method of Average Derivatives," *Journal of the American Statistical Association*, 84, 986–995.

- [20] Klein, R.L., and Spady, R.H. (1993), "An Efficient Semiparametric Estimator for Binary Response Models," *Econometrica*, 61, 387-421.
- [21] Powell, J.L., Stock, J.H., and Stoker, T.M. (1989), "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57, 1403-1430.