On Variance Estimation for the Single-Index Models

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

June 9, 2006

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 examine the estimation for the error variance. Two estimators are proposed, one is the traditional MSE-type estimator, and the other is similar to the difference-based error variance estimator in the literature of nonparametric regression. We allow both fixed design and random design for these two estimators. The asymptotic normality of both estimators is established and an empirical study is conducted 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

$$Y_i = m(X_i) + \epsilon_i, \quad i = 1, \cdots, n, \tag{1.1}$$

for some function $m(\cdot)$ where the covariate X' is a p-dimensional vector, Y is the response variable, and ϵ's are independent errors with zero mean and bounded variance. We allow for both fixed and random covariate designs in this article. We say that the model follows a linear single-index model (SIM) if $m(x) = g(\theta' x)$, where $g(\cdot)$ is an unknown univariate smooth function and $\theta \in \mathbb{R}^p$ is an unknown index vector. For identifiability purposes, we typically assume that $\|\theta\| = 1$ and the first nonzero element of θ is positive. A general condition of the uniqueness of the representation $m(x) = g(\theta' x)$ has been discussed by Lin and Kulasekera (2006). The SIM has been extensively investigated due to its appeal in its ability for dimension reduction (e.g. Powell, Stock and Stoker 1989, Duan and Li

1991, Ichimura 1993, Härdle, Hall and Ichimura 1993, Xia, Tong and Li 1999, Hristache, Juditsky and Spokoiny 2001, Yu and Ruppert 2002, Stute and Zhu 2005). In all the literature, researchers focus on estimating the index coefficients and on the asymptotic properties of the estimators of the mean function.

An equally important issue is the variance estimation in a SIM. This has not been discussed to a satisfactory length in the literature. In this paper we propose two variance estimators for the SIM. One is an MSE-type estimator, i.e. a properly weighted sum of squares of the residuals, and the second estimator is a weighted U-statistic type estimator. We establish the asymptotic normality for both estimators under a set of broad conditions. Note that although the covariate is multidimensional, the function $g(\cdot)$ is univariate. If the index vector θ is known, then one can easily use an optimal difference estimator by Hall, Kay and Titterington (1990) or Munk, Bissantz, Wagner and Freitag (2005) to estimate the error variance. However, since θ is unknown, we cannot directly use a classical difference estimator for a SIM because the ordering of $\theta' X_i$ is not readily available. It is also noteworthy that difference based estimators in a general nonparametric model have slow convergence rates when the dimension is high (Munk et al. 2005). We shall show that, in spite of the additional estimation of the index vector θ, the variance estimators proposed here have very similar asymptotic properties to the MSE-type estimators and difference-based estimators in a univariate nonparametric regression setting.

Our first estimator is an MSE-type statistic (or a kernel-based estimator; Hall and Marron 1990). It is the minimum $\hat{d}$ of $\hat{d}(\alpha)$ with respect to α, where $\hat{d}(\alpha)$ is a suitable sample version of $d(\alpha) = E\big(Y - E(Y|\alpha' X)\big)^2$, e.g.,

$$\hat{d}(\alpha) = \frac{1}{n}\sum_{i=1}^{n} w_i(\alpha)\big(Y_i - \hat{g}_\alpha(\alpha' X_i)\big)^2.$$

Here $\hat{g}_\alpha(t)$ can be any suitable estimator of $E(Y|\alpha' X = t)$ and $w_i(\alpha)$'s are suitable weights. This type of variance estimator is currently a popular choice in the SIM literature. Both a weighted and an unweighted version of $\hat{d}(\alpha)$ above were investigated by Ichimura (1993) who concentrated on the estimator $\hat{\theta}$ of the index vector θ. Härdle et al. (1993) analyzed a version of $\hat{d}(\alpha)$ through a decomposition method for bandwidth selection purposes. The asymptotic normality of $\inf_\alpha \hat{d}(\alpha)$ for some version of $\hat{d}(\alpha)$ in a partially linear SIM was obtained by Xia et al. (1999) under homogeneous errors assumption. Some very restrictive conditions were assumed by these authors. For example, they examined $\inf_\alpha \hat{d}(\alpha)$ only in a very small neighborhood of θ (within $O(n^{-t})$ of θ for some $t > 0$ and sample size n), which is quite impractical since θ is unknown and minimizing $\hat{d}(\alpha)$ over such a neighborhood is not possible. They also assumed the existence of very high (unspecified but as large as needed) moments of the errors. Our method does not require these restrictions and the normality of $\inf_\alpha \hat{d}(\alpha)$ is built for cases including heteroscedastic errors. Moreover, the aforementioned authors only include the data (X_i, Y_i) with $X_i \in A$ for the construction of $\hat{d}(\alpha)$ where A is a set over which $\alpha' X$ has density bounded away from zero for all α. Since A is made free of α, this can discard too many data points from the original sample. When the errors are homogeneous, we use a set $A(\alpha)$ that depends on α in constructing the proposed $\hat{d}(\alpha)$. This approach uses as many data points as possible.

The second estimator we propose resembles both a U-statistic and a difference estimator. It is the minimum (over α) of the sum of weighted squared differences

$$\sum_{i,j} w_{ij}(\alpha;h)(Y_i - Y_j)^2,$$

where the weights $w_{ij}(\alpha;h)$ are selected such that they are small for large values (controlled by the bandwidth parameter h) of $|\alpha' X_i - \alpha' X_j|$. Since the weight functions are nonzero only when the $\theta' X$ values associated with the two responses are close to each other, this estimator is quite similar to the difference-based estimator in the literature mainly for data with univariate covariate (Hall, Kay and Titterington 1990, 1991, Kulasekera and Gallagher 2002, Munk et al. 2005). Note that this type of estimators work only for homogeneous errors. Compared to the MSE-type estimator above, no smoothing is required for this estimator and thus smaller bandwidths are desired (see Section 2). One major advantage of this estimator is that it does not involve the problem of boundary effect as does the first MSE-type estimator.

We organize the paper as follows. In Section 2, we give our construction of the estimators and their asymptotic properties. A small simulation study examining the finite sample behavior of the statistics is given in Section S:variance-sml. All the proofs of the technical results are deferred to Section 4.

2 Variance Estimators

Suppose we have i.i.d. replicates (X_i, Y_i) coming from model (1.1) with $m(x) = g(\theta' x)$. We allow for both fixed and random covariate designs in this treatment. For simplicity of presentation, only random design case is presented in full detail. We assume X to be bounded with probability one and the index vector $\theta \in D$ where

$$D = \big\{\theta \in \mathbb{R}^p \mid \|\theta\| = 1,\, \theta_1 > 0\big\}. \tag{2.1}$$

Note that this definition is slightly different from what is given in the introduction where one only assumes the first nonzero element of θ positive. The reason for this modification is that we need θ to be an interior point of D. This is assumed in most of the existing literature. Since we can always move the most significant variable to be considered as the first variable, the assumption $\theta_1 > 0$ is reasonable. Let $S = S_X$ denote the support of X and, WLOG assume that S is contained in the unit ball $B(0,1)$.

2.1 MSE-type Estimator

2.1.1 Random Design Case

Let $d(\alpha) = E\big(Y - g_\alpha(\alpha' X)\big)^2$ where $g_\alpha(u) = E(Y|\alpha' X = u)$, which can be estimated, for example, by the kernel estimator

$$\hat{g}_\alpha(u) = \frac{\sum_{i=1}^n Y_i K_h(\alpha' X_i - u)}{\sum_{i=1}^n K_h(\alpha' X_i - u)}, \tag{2.2}$$

where $K_h(t) = K(t/h)$ for some proper kernel function K, and bandwidth parameter h. Now, we can construct a sample analog of $d(\alpha)$ like $\sum_{i=1}^n \big[Y_i - \hat{g}_\alpha(\alpha' X_i)\big]^2/n$ and minimize that to get an estimator of θ. However, the data can be sparse near the boundary of the domain of $\alpha' X$ for any given α. This can cause severe bias problems in the estimation of $g_\alpha(\cdot)$. Thus, we propose to use a weighted sum in constructing the sample version $\hat{d}(\alpha)$ of $d(\alpha)$ so that we avoid estimating g_α near the boundaries of the corresponding domain of $\alpha' X$.

Let the support of X be denoted by S_X and, for an $\alpha \in D$, let c_α and $2w_\alpha$ denote the center and width of the set $\{\alpha' x \mid x \in S_X\}$. We shall use an X_i in the construction of $\hat{d}(\alpha)$ only if $\alpha' X_i$ is not too far away from the center c_α. To do this, fix a constant $q < 1$ 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 α in the construction of the sample analog for $d(\alpha)$. In particular, let

$$\hat{d}(\alpha) = \frac{\sum_{i=1}^n \Big(Y_i - \hat{g}_\alpha(\alpha' X_i)\Big)^2 L_{q,\alpha}(X_i)}{\sum_{i=1}^n L_{q,\alpha}(X_i)}, \tag{2.3}$$

where

$$L_{q,\alpha}(x) = L\Big(\frac{\alpha' x - c_\alpha}{q_\alpha}\Big), \quad \forall x \in S_X, \tag{2.4}$$

for some nonnegative weight function L supported on $(-1, 1)$. Then we can estimate the index vector θ by $\hat{\theta} = \arg\inf_\alpha \hat{d}(\alpha)$. The properties of $\hat{\theta}$ for some special choice of the weight function $L_{q,\alpha}$ has been studied by Ichimura (1993). We will concentrate on analyzing the statistic $\hat{d}(\hat{\theta})$ itself which is a reasonable error variance estimator for homogeneous errors.

Remark 2.1. Hall (1989), Ichimura (1993) and Härdle, Hall and Ichimura (1993) used an indicator function $I(x \in \mathcal{X})$ in place of $L_{q,\alpha}(x)$ above, where $\mathcal{X}$ is a subset of our S_X, on which, $f_\alpha(\alpha' x)$ is bounded away from zero for all $\alpha \in D$. There the set $\mathcal{X}$ was not explicitly given. The set $\mathcal{X}$ can be constructed as the intersection of all the sets S_α where $S_\alpha = \{x | L_{q,\alpha}(x) > 0\}$. In some sense, more information is utilized to construct $\hat{d}(\alpha)$ since for every α the set S_α contains more covariate values than the set $\mathcal{X}$ in the above articles. A similar idea to ours was used by Xia, et al. (2004) in their construction of a Goodness-of-Fit test for a SIM.

2.1.2 Fixed-Design Case

For fixed-design models, instead of defining the projection function $g_\alpha(\cdot)$ using conditional expectation, let $g_\alpha(\cdot; m)$ minimize the L_2 distance

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

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

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

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

$$g_\alpha(y_1; m) = \frac{\int_{S(y_1)} m(A'y) dy_2 \cdots dy_p}{\int_{S(y_1)} dy_2 \cdots dy_p}. \tag{2.5}$$

In Lemma 4.10(iv) we show that (2.2) is also consistent for $g_\alpha(\cdot)$ and hence the same error variance estimator $\hat{d}(\hat{\theta})$ as in the random design case can be used for the fixed design case as well.

2.1.3 Model Assumptions and Main Results

The following technical assumptions are used in establishing the asymptotic theory of the proposed estimators of the error variance. These assumptions or stronger versions have been used by many authors in the past for establishing the asymptotic properties of $\hat{\theta}$.

(A1) The mean function $m(x) = E(Y|X = x)$ is r-times continuously differentiable and $m(\cdot)$, $m'(\cdot)$, $m''(\cdot)$ are bounded by a generic constant C. The function $m(x)$ is not constant on the support of $L_{q,\theta}(\cdot)$ where $L_{q,\theta}(\cdot)$ is defined in (2.4). The true index vector $\theta \in D$ where D is as given in (2.1).

(A2) The domain $S = S_X$ of X is closed, bounded and convex, which contains at least one interior ball B_0 with radius $w_0 > 0$. Without loss of generality, assume that $S_X \subset B(0, 1)$ where $B(0, 1)$ is the unit ball in $\mathbb{R}^p$. In 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. The density function $f(\cdot)$ is twice continuously differentiable. Also, there exist constants $0 < c_1 < c_2 < \infty$ such that

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

In fixed-design case, $(x_1, \cdots, x_n)$ is a point set of low discrepancy (Niederreiter, 1992) and the function $f_\alpha(\cdot)$ is defined as

$$f_\alpha(t) = \int_{\{y \in AS | y_1 = t\}} dy_2 \cdots dy_p,$$

where A is an orthogonal matrix with first row α';

(A3) The error ϵ has at least v (to be specified later) moments.

(A4) 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 $t \in (-1, 1)$.

(A5) The bandwidth h is such that $h = O(n^{-\beta})$ for some $\beta \in (0, \frac{1}{3})$.

Remark 2.2. The definition of D guarantees that θ is an interior point of D. This condition was used by Ichimura (1993) as well. Assmption (A2) guarantees that the data points are dense on the support of X. Härdle et al. (1993) and Xia et al. (1999) assume the existence of sufficiently high (their method could not produce a moderate number) of moments of ϵ. Here we only assume the existence of finite number of moments of ϵ for the MSE-type estimator and for the U-statistic estimator. See Theorem 2.3 and Theorem 2.6 for detail.

The following result gives a representation of the quantity $\hat{d}(\alpha)$ when minimized with respect to α.

Theorem 2.3. *Let $\hat{\theta}$ minimizes $\hat{d}(\alpha)$ above over D and let the bandwidth h be proportional to n^{-a} with $\frac{1}{8} < a \leq \frac{1}{5}$. Let assumptions (A1-A5) hold and $v \geq \max(2u, 8)$ where u is the smallest positive even integer greater than $\frac{3a}{1-4a}p$. When the error variances are homogeneous, i.e. $Var(\epsilon|x) = Var(\epsilon)$, we have, for all $\xi > 0$,*

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

If the errors are heteroscedastic, then the above holds provided that the weight function $L_{q,\alpha}(\cdot)$ is free of α.

Proof. This is a direct consequence of Lemma 4.13. □

Corollary 2.4. *Let the conditions of Theorem 2.3 hold. Let*

$$a_0 = E[\sigma^2(X) L_{q,\theta}(X)] \quad \text{and} \quad b = E[L_{q,\theta}(X)].$$

Define $\sigma^2(x) = Var(\epsilon|X = x)$ and $\lambda(x) = E(\epsilon^4|X = x)$. Then we have

$$\sqrt{n}\big(\hat{d}(\hat{\theta}) - \sigma_1^2\big) \xrightarrow{\mathcal{D}} N(0, \tau^2),$$

where, $\sigma_1^2 = a_0/b$ and

$$\tau^2 = \frac{1}{b^4} E\big[b^2 \lambda(X) L_{q,\theta}^2(X) - 2a_0 b\, \sigma^2(X) L_{q,\theta}^2(X) + a_0^2 L_{q,\theta}^2(X)\big].$$

For fixed design models, any quantity taking the form $E[A(X)]$ should be replaced by $\int_S A(x)dx$.

Proof. Suppose U_i's are independent with mean u, V_i's are independent with mean v, and they all have finite variances. Then we have

$$\frac{\sum_{i=1}^n U_i}{\sum_{i=1}^n V_i} - \frac{u}{v} = \frac{1}{nv^2} \sum_{i=1}^n \big(vU_i - uV_i\big) + o_p(n^{-1/2}).$$

Thus, by Theorem 2.3,

$$\begin{aligned}\hat{d}(\hat{\theta}) &= \frac{\sum_{i=1}^{n}\epsilon_i^2 L_{\theta,q}(X_i)}{\sum_{i=1}^{n} L_{\theta,q}(X_i)} + o_p(n^{-\frac{1}{2}}) \\ &= \frac{a_0}{b} + \frac{1}{nb^2}\sum_{i=1}^{n}\bigl(b\epsilon_i^2 L_{\theta,q}(X_i) - a_0 L_{\theta,q}(X_i)) + o_p(n^{-1/2}).\end{aligned}$$

The result now follows from the classical Central Limit Theorem. □

Remark 2.5. It is easily seen that for homogeneous error variances with $\mathrm{Var}(\epsilon|x) = \sigma^2$ we have $\sigma_1^2 = \sigma^2$, and the asymptotic variance τ^2 in Corollary 2.4 is reduced to

$$\tau^2 = \frac{E[\mathrm{Var}(\epsilon^2|X)L_{q,\theta}^2(X)]}{[E[L_{q,\theta}(X)]]^2}, \quad \text{or} \quad \tau^2 = \frac{\int_S \mathrm{Var}(\epsilon^2|x)L_{q,\theta}^2(x)dx}{\left[\int_S L_{q,\theta}(x)dx\right]^2},$$

for random design and fixed design models respectively. In the latter case, we can easily construct an asymptotic confidence interval for σ^2 and test hypothesis like $H_0 : \sigma^2 = \sigma_0^2$.

2.2 U-Statistic Type Estimator

For models with homogeneous error variance σ^2, other than the MSE-based variance estimator, especially in univariate design case (see Munk et al. 2005 for detailed discussion), a popular alternative is the difference-based estimator (Hall, Kay and Titterington 1990). Since SIM is essentially one-dimensional, we can also construct a difference-based estimator. For example, we may order Y_i's according to an ascending order of $\hat{\theta}'X_i$ where $\hat{\theta}$ is any index vector estimator. Denote such ordered Y's by $Y_{(1)}, Y_{(2)}, \cdots, Y_{(n)}$. Then, if we were to follow the classical difference method for univariate nonparametric regression models, a quantity like $\sum(Y_{(i)} - Y_{(i-1)})^2/(2n)$ can be used to estimate σ^2. Since the asymptotic properties of such a quantity are difficult to analyze unless an external estimator $\hat{\theta}$ of θ is used, we propose the following U-statistic type estimator. Let $K_h(\cdot)$ be defined as in the previous subsection and let

$$b(\alpha) = \frac{\sum_{i\neq j}(Y_i - Y_j)^2 K_h(\alpha'X_i - \alpha'X_j)}{2\sum_{i\neq j} K_h(\alpha'X_i - \alpha'X_j)}.$$

Note that the statistic $b(\alpha)$ includes both a bias component and a variance component. When α gets closer to the true index vector θ, the bias component becomes negligible and the variance component dominates. Hence we propose to use the infimum of $b(\alpha)$ over $\alpha \in D$ as the estimator for σ^2. It is noteworthy that, when α is close to θ, $b(\alpha)$ is close to a first-order difference-based estimator in the sense that we take the difference of the response variables only when the covariate values (actually $\theta'X$ values for SIM) are close to each other and hence the effect of mean function is canceled in each difference $Y_i - Y_j$. We can prove the following theorem that gives the asymptotic properties of $\inf_\alpha b(\alpha)$.

Theorem 2.6. *Let asumptions (A1-A5) hold. Take* h *proportional to* n^{-a} *with* $\frac{1}{4} < a < \frac{1}{3}$*. Take* $v \geq \max(2u_1, u_2, 8)$*, where* u_1, u_2 *are the smallest even integers such that*

$$u_1 \geq \frac{4a-1}{1-3a}p, \quad u_2 \geq \frac{5a-1}{2(1-3a)}p.$$

Suppose the error variances are homogeneous with $Var(\epsilon|X=x) = \sigma^2$*. Let* $\tilde{\theta}$ *minimize* $b(\alpha)$ *over* $\alpha \in D$*. Then,*

$$b(\tilde{\theta}) = \frac{\sum_{i=1}^{n} \epsilon_i^2 \tilde{K}_i(\theta)}{\sum_{i=1}^{n} \tilde{K}_i(\theta)} + o_p(n^{-\frac{1}{2}}),$$

where $\tilde{K}_i(\theta) = \sum_{j \neq i} K_h(\theta' X_j - \theta' X_i)$*,* $1 \leq i \leq n$*.*

Proof. This is a direct consequence of Lemma 4.16. ☐

Corollary 2.7. *Let the conditions of Theorem 2.6 hold and let* $\lambda(x) = E(\epsilon^4|X=x)$*. Then* $\sqrt{n}(b(\tilde{\theta}) - \sigma^2) \xrightarrow{\mathcal{D}} N(0, \tau_1^2)$*, where*

$$\tau_1^2 = \frac{E[(\lambda(X) - \sigma^4) f_\theta^2(\theta' X)]}{[E(f_\theta(\theta' X))]^2}.$$

For fixed design models, any quantity taking the form $E[A(X)]$ *should be replaced by* $\int_S A(x)dx$*.*

Proof. By Lemma 4.9(ii), for all $\xi > 0$,

$$\Big(\sum_{i=1}^{n} \tilde{K}_i(\theta)\Big)^{-1} = \Big(\sum_{i=1}^{n} nhf_\theta(\theta' X_i)\Big)^{-1} + o_p(n^{-\frac{5}{2}+\xi} h^{-1}).$$

Hence,

$$b(\tilde{\theta}) - \sigma^2 = \frac{\sum_{i=1}^{n} (\epsilon_i^2 - \sigma^2) f_\theta(\theta' X_i)}{\sum_{i=1}^{n} f_\theta(\theta' X_i)} + o_p(n^{-\frac{1}{2}}).$$

The result now follows from the classical Central Limit Theorem. ☐

Remark 2.8. When the covariate is one-dimensional with fixed design, the asymptotic variance of the MSE-type estimator (Hall and Marron, 1990) for the constant error variance $\text{Var}(\epsilon)$ is $\frac{1}{n}\text{Var}(\epsilon^2)$ and the asymptotic variance of the mth order difference-based estimator (Hall, Kay and Titterington 1990) is $\frac{1}{n}(\text{Var}(\epsilon^2) + m^{-1}\sigma^4)$. For the proposed MSE-type estimator, if $\sigma^2(x) = \sigma^2$ and $\lambda(x) = \lambda$, by taking $L_{q,\theta}(\cdot)$ converging to one as the sample size increases, the asymptotic variance can be reduced to $\frac{1}{n}\text{Var}(\epsilon^2)$. For the U-statistic estimator, when the variance of $f_\theta(\theta' X)$ is relatively small compare to its expectation, τ_1^2 is close to $\text{Var}(\epsilon^2)$ as well. Thus, even after estimating the index vector θ, the variance estimators we propose have very desirable asymptotic properties which are comparable to those for the univariate design case.

2.3 Applications

There are many immediate applications of the above results, e.g. constructing a confidence interval for σ^2 or conducting a hypothesis testing about σ^2. In this subsection we propose a test for the equality of the error variances for two single-index models, i.e.,

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{and} \quad H_a : \sigma_1^2 \neq \sigma_2^2. \tag{2.6}$$

Let $\hat{d}_k$ be the version of $\hat{d}(\hat{\theta})$ defined in Section 2.1.1 for the k-th model, which by Corollary 2.4, is consistent for σ_k^2 with

$$\sqrt{n_k}(\hat{d}_k - \sigma_k^2) \xrightarrow{\mathcal{D}} N\Big(0, \frac{E[\lambda_k(X) L_{q,\theta}^2(X)]}{\big(E[L_{q,\theta}(X)]\big)}\Big), \tag{2.7}$$

where n_k is the size of the k-th sample and $\lambda_k(x) = \mathrm{Var}(\epsilon^2|X)$ for the kth model, $k = 1, 2$. This immediately leads to the test statistic

$$T^* = \sqrt{N}\big(\hat{d}_1 - \hat{d}_2\big), \tag{2.8}$$

where $N = n_1 + n_2$. By (2.7) and the independence of $\hat{d}_1$ and $\hat{d}_2$ we can easily obtain the next theorem.

Theorem 2.9. *Let assumptions (A1–A6) hold and let T^* be given as in (2.8). Under H_0, $T^*/\tau_N \xrightarrow{\mathcal{D}} N(0, 1)$, where*

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

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

Hence we propose to reject H_0 when $|T^*|/\hat{\tau}_N$ is larger than the upper-$\frac{\alpha}{2}$ quantile of the standard normal distribution, where $\hat{\tau}_N$ is any consistent estimator of τ_N.

3 Empirical Study

3.1 Simulation Study

We conducted a simple simulation study for assessing the finite sample properties of the two estimators as well as the test for equal variances of two models. The errors were taken to be normal with mean zero and variance σ^2. We used the quadratic kernel function $K(u) = \frac{3}{4}(1 - u^2)I(|u| \leq 1)$ for both estimators and, we used the $L(\cdot)$ function

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

to create the weight function $L_{q,\alpha}$ for $\hat{d}(\alpha)$. The quantities c_α and w_α were estimated by

$$\hat{c}_\alpha = \tfrac{1}{2}\big(\max_i \alpha' X_i + \min_i \alpha' X_i\big) \quad \text{and} \quad \hat{w}_\alpha = \tfrac{1}{2}\big(\max_i \alpha' X_i - \min_i \alpha' X_i\big).$$

Several q values (from 0.85 to 0.95) were used and they all gave similar results. We only present the results with $q = 0.95$.

For both estimators, the selection of the bandwidth parameter is a critical step. In the simulation study, the bandwidth h for $\hat{d}(\hat{\theta})$ was chosen by a simple cross validation method. Namely, we selected h_0 by minimizing $\hat{d}_{cv}(\alpha; h)$ which is the version of $\hat{d}(\alpha)$ where in each term $Y_i - \hat{g}_\alpha(\alpha' X_i)$, the estimator $\hat{g}_\alpha(\alpha' X_i)$ is now computed by leaving out the ith observation. To keep such a completely data-driven h_0 value in a reasonable range (i.e. of order $n^{-1/5}$), we took h_1, the bandwidth for the first estimator $\hat{d}$, as the value in the interval $[\frac{1}{20}n^{-1/5}, \frac{1}{3}n^{-1/5}]$ that is closest to h_0. Since h_0 can be shown to be of order $O(n^{-\frac{1}{5}})$ and the order of the bandwidth parameter for $b(\tilde{\theta})$ is about $O(n^{-\frac{1}{3}})$ as given by Theorem 2.6, we can use bandwidth values like $h_1^{5/3}$ for evaluating $b(\alpha)$ Following a similar argument as in the case of $\hat{d}$, our choice of h_2, the bandwidth for the second estimator $b(\tilde{\theta})$, was the value in the interval $[\frac{1}{20}n^{-1/3}, \frac{1}{3}n^{-1/3}]$ that was closest to $h_1^{5/3}$. All simulations were conducted using R language. The results presented are based on 1000 simulations for Example 3.1 and 2000 simulations for Examples 3.2–3.4.

Example 3.1. For the two estimators, the p covariates were taken to be independent with X_i being uniform distribution between 0 and 1. The two functions $g_1(t) = 2\sin(\pi t)$ and $g_2(t) = 2\sin(2\pi t)$ were examined. For each $g(\cdot)$ function, we took dimensions $p = 2, 3$, sample sizes $n = 25, 50, 100$ and error variances $\sigma^2 = 0.09, 0.25$. For $p = 2$, we took $\theta = (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})'$ and for $p = 3$, we took $\theta = (\frac{\sqrt{2}}{2}, -\frac{1}{2}, \frac{1}{2})'$.

In Table 1 and Table 2 we present the sample averages and the sample standard deviations of $\hat{d}(\hat{\theta})$ and $b(\tilde{\theta})$ values. Results in these two tables show that both the estimators $\hat{d}(\hat{\theta})$ and $b(\tilde{\theta})$ perform well as estimators of the error variance in a SIM for all cases examined. Apart from numerical difficulties, the accuracy does not seem to depend on p and the signal g. The overall performance of the U-statistic type estimator is slightly better than the MSE-type estimator. It should be noted that, when computing time is of concern, a difference-based estimator is a lot faster to compute (Munk, et al. 2005). Usually the minimization takes a longer time when p gets large for both estimators. In such cases, we can plug in an estimator $\hat{\theta}$ obtained using another fast index estimation method proposed in the literature and the asymptotic behavior does not change.

It is noteworthy that our simulations indicate possible underestimation for the error variances for the sample sizes we have examined. This problem is more significant when the sample size is small. This is perhaps due to possible over-fitting in estimating the regression function which could be corrected by two ways. The first is to use a more adaptive (data-driven) method of bandwidth selection. The other is to use the "delete-one" estimate for the regression function(i.e. leave the ith observation out when constructing $\hat{Y}_i$), which is asymptotically equivalent to regular kernel estimate but could reduce bias for finite samples. Using other estimates for the regression function, e.g. the locally linear kernel estimator which is locally asymptotically minimax (Fan 1993), might help solve this problem as well. More simulations are needed to show the effect of these modifications.

Example 3.2 (Uniform Design; $p = 2$). In this example we investigate the finite sample performance of the test for equal variances. We consider parameter combinations

Table 1: Mean(standard deviation) of $\hat{d}(\hat{\theta})$ and $b(\tilde{\theta})$ with $\sigma^2 = 0.09$

		$g_1(t) = 2\sin(\pi t)$		$g_2(t) = 2\sin(2\pi t)$	
	n	$\hat{d}(\hat{\theta})$	$b(\tilde{\theta})$	$\hat{d}(\hat{\theta})$	$b(\tilde{\theta})$
$p = 2$	25	.076(.030)	.082(.033)	.097(.041)	.088(.040)
	50	.078(.019)	.084(.024)	.087(.025)	.087(.028)
	100	.082(.013)	.088(.016)	.085(.017)	.088(.018)
$p = 3$	25	.069(.031)	.065(.030)	.098(.045)	.076(.037)
	50	.076(.019)	.076(.022)	.096(.028)	.083(.026)
	100	.082(.013)	.083(.016)	.092(.018)	.087(.018)

Table 2: Mean(standard deviation) of $\hat{d}(\hat{\theta})$ and $b(\tilde{\theta})$ with $\sigma^2 = 0.25$

		$g_1(t) = 2\sin(\pi t)$		$g_2(t) = 2\sin(2\pi t)$	
	n	$\hat{d}(\hat{\theta})$	$b(\tilde{\theta})$	$\hat{d}(\hat{\theta})$	$b(\tilde{\theta})$
$p = 2$	25	.192(.070)	.204(.088)	.213(.077)	.217(.096)
	50	.214(.050)	.228(.059)	.218(.056)	.229(.068)
	100	.224(.037)	.236(.043)	.228(.040)	.237(.046)
$p = 3$	25	.172(.064)	.158(.072)	.200(.079)	.176(.085)
	50	.204(.050)	.199(.056)	.223(.057)	.213(.065)
	100	.223(.036)	.223(.040)	.232(.038)	.230(.044)

$(\sigma_1^2, \sigma_2^2) = (0.5, 0.25), (0.25, 0.25)$ in each of the following function/index vector settings:

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

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

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

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

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

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

The results are presented in Table 3. We observe that under H_0, although slowly, in most cases the rejection rates converge to the nominal levels as sample sizes increase; and under H_a the test can detect the difference in error variances satisfactorily and the power increases rather quickly with sample sizes.

Table 3: Simulated rejection rates for Example 3.2 with sample size $n_1 = n_2 = n$.

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

Table 4: Simulated rejection rates for Example 3.3 with sample size $n_1 = n_2 = n$.

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

Example 3.3 (Normal Design; $p = 3$). We also investigated the test for equal variances for dimension $p = 3$ using normal covariates design. All three covariates are generated independently from a normal distribution with mean 0 and standard deviation 0.5. We consider $\theta_1 = \theta_2 = (\sqrt{3}/3, \sqrt{3}/3, \sqrt{3}/3)$ and $(\sigma_1^2, \sigma_2^2) = (0.5, 0.25)$ and $(0.25, 0.25)$ in each of the following two settings:

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

The rejection rates for testing the equality of error variances are given in Table 4. We observe similar performances as in the previous three examples except that larger sample sizes are required to achieve the same nominal levels, which is not surprising since the dimension for the index vector has increased.

Example 3.4 (Unbalanced Sample Sizes). For the test of equal variances, when sample sizes are quite different, using a common bandwidth as we did in the previous examples leads to rejection rates under H_0 significantly higher than the nominal levels.

Table 5: Simulated rejection rates of the second test for Example 3.4.

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

To accommodate for unequal bandwidths we proceeded as follows. Since our asymptotic theory requires that the bandwidth parameter h should be proportional to $N^{-1/4}$, we first select an initial bandwidth parameter h_0 by the same cross-validation criteria as in the previous examples. Then let

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

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

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

The simulated rejection rates for testing (2.6) are given in Table 5.

3.2 Real Data Example

We applied our estimation methods to a real data example to show its usage. Here we use the FAT data from Bowerman and O'Connell (1990). The data consists of the response variable FAT (the percentage of fat in pork bellies) and 10 predictor variables for a sample of 45 pork carcasses. It was shown that six of these covariates gave a well fitting linear model to the responses. These covariates were X_1; an average of three measures of back fat thickness, X_2; a muscling score for the carcass, X_3; an average of three measures of fat depth opposite the tenth rib, X_4; live weight of the carcass, X_5; the average measure of leanness of three cross sections of the belly, and X_6; total weight of the belly. Since the linear model assumption is reasonable for this data, we applied the proposed variance estimation methods to obtain the error variance estimators. The CV method introduced in the simulation section was again used to select the bandwidths to get $h_1 = 0.156$ and $h_2 = 0.045$, respectively for the two estimators. The resulting estimators were $\hat{d}(\hat{\theta}) = 5.27$

and $b(\tilde{\theta}) = 3.45$. Note that the MSE reported by Bowerman and O'Connell (1990) with these six variables in a classical linear model is 6.03. We believe that both our estimators are smaller than the traditional MSE because a SIM can provide a better fit for the data than the classical linear model and hence a more precise estimate for the error variance. The test statistic for the equal variance test is $T^* = -0.01$ and hence we fail to reject their equality.

4 Proofs

The proofs of Theorem 2.3 and 2.6 require a few technical lemmas. We first give a few general results in Section 4.1 that are consequences of the Bernstein's inequality. In Section 4.2 we list a few technical results for our model. A decomposition of the statistic $\hat{d}(\alpha)$ and the asymptotic performance of each component are given in Section 4.3. The analysis of the second estimator $b(\tilde{\theta})$ is given in Section 4.4.

To simplify the presentation, we introduce some notation first. For fixed choice of kernel function K, bandwidth parameter h, index-vector α and $i = 1, \cdots, n$, let

$$K_i(\alpha) = \sum_{j=1}^{n} K_h\big(\alpha'(X_j - X_i)\big); \qquad L_\alpha = \sum_{i=1}^{n} L_{\alpha,q}(X_i);$$

$$\kappa_{ij}(\alpha) = K_h\big(\alpha'(X_j - X_i)\big)/K_i(\alpha); \qquad l_{i\alpha} = \frac{L_{\alpha,q}(X_i)}{L_\alpha};$$

$$K_{mi}(\alpha) = \sum_{j=1}^{n} \kappa_{ij}(\alpha) m(X_j); \qquad K_{ei}(\alpha) = \sum_{j=1}^{n} \kappa_{ij}(\alpha)\epsilon_j.$$

And, for any sequence $\{c_i\}$, we use the summation notation

$$\sum_{i\alpha} c_i = \sum_{i=1}^{n} c_i l_{i,\alpha}.$$

For a general r-th order kernel function $K(\cdot)$, to avoid diminishing denominator, the quantity $l_{i\alpha}$ is modified into

$$l_{i\alpha} = \frac{L_{q,\alpha}(X_i) I(|K_i(\alpha)| > a_0)}{L_\alpha},$$

where a_0 is an arbitrarily fixed positive constant.

4.1 A Few General Results

We start from the next definition.

Definition 4.1. The distinct-index number t for a sequence of pairs $\{(i_s, j_s)\}_{s=1}^{n}$ is the maximum number of pairs, say $\mathcal{A} = \{(u_1, v_1), \cdots, (u_t, v_t)\}$ from this sequence such that $u_k \neq v_k$, $k = 1, \cdots, t$ and each pair (u_k, v_k) has an index, either u_k or v_k, that is not contained in any other pair in $\mathcal{A}$.

Lemma 4.2 (Bernstein's Inequality). *Let* $Z_1, \cdots, Z_n$ *be independent with* $\sigma_1^2 + \cdots + \sigma_n^2 < \infty$. *If* $P(|Z_i - EZ_i| \le b) = 1, i = 1, \cdots, n$, *then*

$$P\Big(\Big|\frac{1}{n}\sum_{i=1}^{n}(Z_i - EZ_i)\Big| \ge \epsilon\Big) \le 2\exp\Big\{-\frac{n^2\epsilon^2}{2\sum_{i=1}^{n}\sigma_i^2 + \frac{2}{3}bn\epsilon}\Big\}.$$

Lemma 4.3. *Let* A *be any subset of the unit sphere in* $\mathbb{R}^p$. *Suppose* $Z_{i1}, \cdots, Z_{in}$ *are independent conditional on* X_i *for every fixed* i *and*

(i) $Var\big(Z_{ij}(\alpha)|X_i\big) \le \sigma_{nj}^2$;

(ii) $P\big(|Z_{ij}(\alpha)| \le b_n\big) = 1$;

(iii) $\big\|\frac{d}{d\alpha}Z_{ij}(\alpha)\big\| \le Cn^r$ *for some constants* C, r;

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

$$\sup_{\alpha\in A}\max_{1\le i\le n}\Big|\sum_{j=1}^{n}\big(Z_{ij}(\alpha) - E_iZ_{ij}(\alpha)\big)\Big| = o_p(\sigma_n n^{\xi} + b_n n^{\xi}),$$

where $E_iZ_{ij}(\alpha) = E[Z_{ij}(\alpha)|X_i]$ *and* $\sigma_n^2 = \sum_{j=1}^{n}\sigma_{nj}^2$. *The result holds for constant* X_i*'s.*

Proof. It suffices to show the result for random X_i's. 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 $|\alpha - \tilde{\alpha}| \le O(n^{-\frac{s}{p}})$. Now, for every fixed realization of the random variables we have that

$$\begin{aligned}
&\sup_{\alpha\in A}\max_{1\le i\le n}\Big|\sum_{j=1}^{n}\big(Z_{ij}(\alpha) - E_iZ_{ij}(\alpha)\big)\Big| - \sup_{\alpha\in\mathcal{A}_n}\max_{1\le i\le n}\Big|\sum_{j=1}^{n}\big(Z_{ij}(\alpha) - E_iZ_{ij}(\alpha)\big)\Big| \\
&\le \Big|\sum_{j=1}^{n}\big(Z_{tj}(\beta) - E_tZ_{tj}(\beta)\big)\Big| - \sup_{\alpha\in\mathcal{A}_n}\max_{1\le i\le n}\Big|\sum_{j=1}^{n}\big(Z_{ij}(\alpha) - E_iZ_{ij}(\alpha)\big)\Big| \\
&\le \Big|\sum_{j=1}^{n}\big(Z_{tj}(\beta) - E_tZ_{tj}(\beta)\big)\Big| - \Big|\sum_{j=1}^{n}\big(Z_{tj}(\tilde{\beta}) - E_tZ_{tj}(\tilde{\beta})\big)\Big| = O\big(n^{-\frac{s}{p}+r+1}\big), \qquad (4.1)
\end{aligned}$$

where $(\beta, t) = \arg\sup_{\alpha\in A}\max_{1\le i\le n}\big|\sum_{i=1}^{n}\big(Z_{ij}(\alpha) - EZ_{ij}(\alpha)\big)\big|$ and $\tilde{\beta} \in \mathcal{A}_n$ is chosen such that $|\beta - \tilde{\beta}| \le O(n^{-\frac{s}{p}})$. By Bernstein's inequality,

$$\begin{aligned}
&P\Big(\sup_{\alpha\in\mathcal{A}_n}\max_{1\le i\le n}\Big|\sum_{i=1}^{n}\big(Z_{ij}(\alpha) - E_iZ_{ij}(\alpha)\big)\Big| \ge a_n\Big) \\
&\quad\le \sup_{\alpha\in\mathcal{A}_n}\max_{1\le i\le n}P\Big(\Big|\sum_{i=1}^{n}\big(Z_{ij}(\alpha) - E_iZ_{ij}(\alpha)\big)\Big| \ge a_n\Big) \\
&\quad\le n^{1+s}\cdot\int 2\exp\Big\{-\frac{a_n^2}{2\sum_{j=1}^{n}\sigma_{nj}^2 + 2b_na_n}\Big\}f_{X_i}(t)dt \\
&\quad= 2\exp\Big\{-\frac{a_n^2}{2\sigma_n^2 + 2b_na_n} + (1+s)\ln n\Big\},
\end{aligned}$$

which converges to zero provided

$$\lim_{n\to\infty}\Big(\frac{a_n^2}{2\sigma_n^2+2b_na_n}-(1+s)\ln n\Big)=\infty. \tag{4.2}$$

A sufficient condition for (4.2) is $a_n = n^{\xi}(\sigma_n + b_n)$ for any $\xi > 0$. Hence, for all $\xi > 0$,

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

From (4.1), the proof is completed by taking s large enough. □

Lemma 4.4. *Suppose $\{(\epsilon_i, X_i), 1 \le i \le n)\}$ are independent pairs and $Var(\epsilon_i|X_i)$ are bounded by σ^2. Let $B_i(\alpha) = \sum_{j=1}^{n} w_{ij}(\alpha)\epsilon_j$, where $w_{ij}(\alpha) = w(X_i, X_j; \alpha)$ and, for every fixed i, $\{w_{ij}(\alpha), j \ne i\}$ are independent conditioned on X_i. Suppose for some $v > 0$, uniformly in i, j, α, we have*

$$E|\epsilon_i^v|\le M,\quad E\big(w_{ij}^2(\alpha)|X_i\big)\le v_{ni},\quad |w_{ij}(\alpha)|\le c_n,\quad \Big\|\frac{\partial}{\partial\alpha}w_{ij}(\alpha)\Big\|\le Cn^a,$$

for some constants $M, C, a > 0$. Suppose $c_n = O(n^c)$ and $v_n = \sum_{i=1}^{n} v_{ni} = O(n^c)$ for some constant c. Then we have

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

The result holds for constant X_i's.

Proof. It suffices to show the result for random X_i's. Let $\epsilon_{j1} = \epsilon_j I_{\{-C_1n^t\le\epsilon_j\le C_2n^t\}}$ and $\epsilon_{j2} = \epsilon_j - \epsilon_{j1}$ for some positive constants $C_1, C_2 \in (1, 2)$ and t such that $E(\epsilon_{j1}|X_j) = 0$. Let $B_{ik}(\alpha) = \sum_{j=1}^{n} w_{ij}(\alpha)\epsilon_{jk}$, $k = 1, 2$. By Hölder's inequality and Chebyshev's inequality we get

$$\begin{aligned}E|\epsilon_{j2}|^2 &= E\big|\epsilon_j^2 I_{\{\epsilon_j<-C_1n^t \text{ or } \epsilon_j\ge C_2n^t\}}\big| \le E\big|\epsilon_j^2 I_{\{|\epsilon_j|\ge n^t\}}\big|\\ &\le \Big(E|\epsilon_j^2|^{\frac{v}{2}}\Big)^{\frac{2}{v}}\Big(EI_{\{|\epsilon_j|\ge n^t\}}\Big)^{1-\frac{2}{v}} = \Big(E|\epsilon_j|^v\Big)^{\frac{2}{v}}\Big(P(|\epsilon_j|>n^t)\Big)^{1-\frac{2}{v}}\\ &\le \big(E|\epsilon_j|^v\big)^{\frac{2}{v}}\Big(\frac{E|\epsilon_j|^v}{n^{vt}}\Big)^{1-\frac{2}{v}} = \frac{E|\epsilon_j|^v}{n^{(v-2)t}}.\end{aligned} \tag{4.3}$$

Hence, by Cauchy-Schwartz inequality,

$$\begin{aligned}E\sum_{i=1}^{n}\sup_{\alpha}|B_{i2}(\alpha)|^2 &= \sum_{i=1}^{n}E\sup_{\alpha}\Big(\sum_{j=1}^{n}w_{ij}(\alpha)\epsilon_{j2}\Big)^2\\ &\le \sum_{i=1}^{n}E\Big(\sum_{j=1}^{n}\sup_{\alpha}|w_{ij}(\alpha)|\cdot|\epsilon_{j2}|\Big)^2\\ &\le \sum_{i=1}^{n}E\Big(\sum_{j=1}^{n}c_n|\epsilon_{j2}|\Big)^2 \le O(n^{3-(v-2)t}c_n^2).\end{aligned} \tag{4.4}$$

Since $Ew_{ij}(\alpha)\epsilon_{j1} = 0$, $|w_{ij}(\alpha)\epsilon_{j1}| \le 2c_n n^t$ and $\mathrm{Var}\big(w_{ij}(\alpha)\epsilon_{j1}|X_i\big) \le v_{ni}\sigma^2$, by Lemma 4.3 we have, for all $\xi > 0$,

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

Hence, $\sum_{i=1}^n \sup_{\alpha\in A} B_{i1}^2(\alpha) = o_p\big(n^{1+\xi}v_n + c_n^2 n^{1+2t+\xi}\big)$. From (4.4) the proof is completed by taking $t = \frac{2}{v}$. □

Lemma 4.5. *Suppose ϵ_i's are independent conditional on X (which can be contant). Suppose ϵ_i has at least v moments and $E(\epsilon_i|X) = 0$. Let $Z_i(\alpha) = Z_i(X;\alpha)$ be such that $\sup_{\alpha,i}\big\|\frac{d}{d\alpha}Z_i(\alpha)\big\| \le Cn^s$ for some constants C and s and $\sup_{\alpha,i}|Z_i(\alpha)| = O_p(b)$. If $v \ge 4$, then*

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

Proof. Let $A_k(\alpha) = \sum_{i=1}^n Z_i(\alpha)\epsilon_{ik}$, $k = 1, 2$, with ϵ_{jk} defined as in the proof of Lemma 4.4. By Cauchy-Schwartz inequality and (4.3),

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

For $A_1(\alpha)$, by the same discretization technique we used in the proof of Lemma 4.3 it suffices to take the sup over a discrete set $\mathcal{A}$ of size n^a for some sufficiently large constant $a > 0$ (note that ϵ_{i1}'s are bounded by n^t for some constant t). For arbitrary $\xi > 0$, let

$$E_n = \{x \mid \sup_{\alpha,i}|Z_i(x;\alpha)| < bn^{\xi}\}.$$

Easily $P(X \notin E_n) = o(1)$. For any $\gamma > 0$ by Bernstein's inequality we have

$$\begin{aligned}P\Big(\sup_{\alpha\in\mathcal{A}}\big|A_1(\alpha)\big| > \gamma\Big) &= E\Big[E\Big(I(\sup_{\alpha\in\mathcal{A}}\big|A_1(\alpha)\big| > \gamma) \mid X\Big)\Big]\\ &\le \int_{E_n} E\Big(I(\sup_{\alpha\in\mathcal{A}}\big|A_1(\alpha)\big| > \gamma) \mid X\Big)f_X(x)dx + o(1)\\ &\le \sum_{\alpha\in\mathcal{A}}\int_{E_n} P\Big(\Big|\sum_{i=1}^n Z_i(x;\alpha)\epsilon_{i1}\Big| > \gamma\Big)f_X(x)dx + o(1)\\ &\le 2\exp\Big\{-\frac{\gamma^2}{C_1nb^2 + bn^t\gamma} + a\ln n\Big\} + o(1),\end{aligned}$$

which goes to zero provided $\gamma = (n^{\frac{1}{2}} + n^t)bn^{\xi}$ for any $\xi > 0$. This shows that, for all $\xi > 0$, $\sup_{\alpha\in\mathcal{A}}\big|A_1(\alpha)\big| = o_p\big((n^{\frac{1}{2}} + n^t)bn^{\xi}\big)$; and the proof is completed by taking $t = \frac{2}{v}$. □

4.2 Technical Results

The next lemma lists some technical results we need to prove the main theorems. We will 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 α, matrix A and set E, let $\alpha' E$ denote the set $\{\alpha' x | x \in E\}$, and let AE denote the set $\{Ax | x \in E\}$. In all cases, $E[A(X)]$ should be understood as $\int_S A(x)dx$ for fixed designs.

Lemma 4.6. *Let assumptions (A1–A4) hold. Then, for both designs,*

(i) *$g_\alpha(\alpha' x)$ and $m_\alpha(\alpha' x)$ are twice continuously differentiable with respect to α;*

(ii) *the function $f_\alpha(\cdot)$, as given in (A2), is uniformly bounded away from zero, say larger than c_q, for all α and $t \in (c_\alpha - q_\alpha, c_\alpha + q_\alpha)$;*

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

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

And for random designs we have,

(iv) *the density function $\phi_\alpha(\cdot)$ of $\alpha' X_1 - \alpha' X_2$ satisfies*

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

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

Proof. **(i)** It's straight forward by writing the two terms in the form of integral. This has been used by Ichimura (1993) among others.

(ii) Consider the random design first. Let A be an orthogonal matrix whose first row is α'. Let $Z = AX$ and $S_Z = AS_X$. Then $\alpha' X = Z_1$ and

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

where $S(t) = \{(z_2, \cdots, z_p) | (t, z_2, \cdots, z_p) \in S_Z\}$ 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 α 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 $\alpha' x_0 \le t$. 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_α, which, by geometry, 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 $c_\alpha + w_\alpha - \alpha' x_0 \leq 2$. Finally it follows from the convexity of S_X that $B(x_{\alpha,t}, r_{\alpha,t}) \subset S_X$. Note that for fixed designs, $f_\alpha(\cdot)$ is exactly the function $\tilde{S}(\cdot)$ defined above and hence the result follows as well.

(iii) 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 $w_\alpha = (u(\alpha) - l(\alpha))/2$. Since

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

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

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

for some constants C, C_1 uniformly in $x \in S_X$ and in α_1, α_2.

(iv) Since $q_\alpha = q w_\alpha \geq q w_0 = q_0$, from (ii) we have

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

provided $|u| < q_0/2$.

□

Remark 4.7. Some of the results in this lemma have been used as assumptions by Ichimura (1993) and Härdle, Hall and Ichimura (1993) among others. It is also 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.

Lemma 4.8. *Denote the support of $\alpha' X$ by $\alpha' S_X = [c_\alpha - w_\alpha, c_\alpha + w_\alpha] = [l_\alpha, u_\alpha]$. Let Assumptions (A2) and (A4) hold. Then, for random design models, with $i \neq j$,*

(i) *if $K(\cdot)$ is an r-th order kernel function, uniformly in $\{(\alpha, i) \mid |\alpha' X_i - c_\alpha| \leq w_\alpha - h\}$ we have*

$$E_i K_h(\alpha' X_j - \alpha' X_i) = h f_\alpha(\alpha' X_i) \int_{-1}^{1} K(s) ds + O(h^{r+1});$$

(ii) *uniformly in $\{(\alpha, i) \mid L_{q,\alpha}(X_i) \neq 0, \|\alpha - \theta\| \leq \delta\}$ we have*

$$E_i[K_h(\alpha' X_j - \alpha' X_i) - K_h(\theta' X_j - \theta' X_i)] = O_p(h\delta);$$

the result also holds uniformly in $\{(\alpha, i) \mid |\alpha' X_i - c_\alpha| \leq w_\alpha - \gamma_n\}$ for all $\gamma_n > h$, provided $\delta = o(\gamma_n - h)$;

(iii) *uniformly in i and $\alpha \in B(\theta, \delta)$ we have*

$$E_i[K_h(\alpha' X_j - \alpha' X_i) - K_h(\theta' X_j - \theta' X_i)] = O_p(\delta).$$

The above results hold for fixed designs with $E_i[f(X_i, X_j)]$ replaced by $\int_S f(X_i, x)dx$ and $O_p(\cdot)$ by $O(\cdot)$.

Proof. The proofs for the two designs are identical and we present the proof for the random design models only.

(i) Note that

$$E_i K_h(\alpha' X_j - \alpha' X_i) = \int_{\max(\alpha' X_i - h, l_\alpha)}^{\min(\alpha' X_i + h, u_\alpha)} K\Big(\frac{t - \alpha' X_i}{h}\Big) f_\alpha(t) dt. \tag{4.5}$$

When $|\alpha' X_i - c_\alpha| \leq w_\alpha - h$, (4.5) is uniformly reduced to

$$\begin{aligned} E_i K_h(\alpha' X_j - \alpha' X_i) &= h \int_{-1}^{1} K(s) f_\alpha(\alpha' X_i + hs) ds \\ &= h f_\alpha(\alpha' X_i) \int_{-1}^{1} K(s) ds + O(h^{r+1}). \end{aligned} \tag{4.6}$$

(ii) Since $L_{q,\alpha}(X_i) \neq 0$, $|\alpha' X_i - c_\alpha| \leq q w_\alpha < w_\alpha - h$ and (4.6) follows. When $\delta = O(1)$, $E_i K_h(\alpha' X_j - \alpha' X_i) = O(h)$ and thus

$$E_i[K_h(\alpha' X_j - \alpha' X_i) - K_h(\theta' X_j - \theta' X_i)] = O(h) = O(h\delta).$$

When $\delta = o(1)$, since $\|\alpha - \theta\| < \delta$, we have

$$\begin{aligned} |\theta' X_i - c_\theta| &\leq |\alpha' X - c_\alpha| + O(\delta) \\ &\leq q w_\alpha + O(\delta) = q w_\theta + O(\delta) \leq w_\theta - h. \end{aligned}$$

Hence, by (4.6),

$$\begin{aligned} &E_i[K_h(\alpha' X_j - \alpha' X_i) - K_h(\theta' X_j - \theta' X_i)] \\ &\quad = h \int_{-1}^{1} K(s) \big[f_\alpha(\alpha' X_i + hs) - f_\theta(\theta' X_i + hs)\big] ds = O(h\delta). \end{aligned}$$

For the second result, it suffices to note that, uniformly in $\{(\alpha, i) \mid |\alpha' X_i - c_\alpha| \leq w_\alpha - \gamma_n\}$, provided $\delta = o(\gamma_n)$ and $\gamma_n > h$,

$$\begin{aligned} |\theta' X_i - c_\theta| &\leq |\alpha' X - c_\alpha| + O(\delta) \\ &\leq w_\alpha - \gamma_n + O(\delta) = w_\theta - \gamma_n + O(\delta) \leq w_\theta - h. \end{aligned}$$

(iii) It follows from the observation that

$$
\begin{aligned}
E_i K_h(\alpha' X_j - \alpha' X_i) &= \int_{c_\alpha - w_\alpha}^{c_\alpha + w_\alpha} K\Big(\frac{t_\alpha - \alpha' X_i}{h}\Big) f_\alpha(t)dt \\
&= h \int_{\frac{1}{h}[c_\alpha - w_\alpha - \alpha' X_i]}^{\frac{1}{h}[c_\alpha + w_\alpha - \alpha' X_i]} K(s) f_\alpha(\alpha' X_i + hs)ds \\
&= h \int_{-1}^{1} K(s) f_\alpha(\alpha' X_i + hs)ds + h \int_{\frac{c_\alpha - w_\alpha - \alpha' X_i}{h}}^{-1} K(s) f_\alpha(\alpha' X_i + hs)ds \\
&\qquad + h \int_{1}^{\frac{c_\alpha + w_\alpha - \alpha' X_i}{h}} K(s) f_\alpha(\alpha' X_i + hs)ds.
\end{aligned}
$$

$\square$

Lemma 4.9. *Let Assumptions (A2) and (A4) hold. Then, for random designs,*

(i) *if $K(\cdot)$ is an r-th order kernel function,*

$$\sup_{\{(\alpha,i) | |\alpha' X_i - c_\alpha| \le w_\alpha - h\}} \Big| K_i(\alpha) - nh f_\alpha(\alpha' X_i) \int_{-1}^{1} K(s)ds \Big| = o_p(n^{\frac{1}{2}+\xi}\sqrt{h}) + O_p(nh^{r+1});$$

(ii) $\inf_{\{(\alpha,i)|L_{q,\alpha}(X_i)\neq 0\}} |K_i(\alpha)| \ge c_K nh + o_p(nh)$;

(iii) $\inf_\alpha L_\alpha \ge c_L n + o_p(n)$;

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

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

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

(vii) $\inf_\alpha \big|\sum_{i=1}^n K_i(\alpha)\big| \ge c_M n^2 h + o_p(n^2 h)$;

(viii) *if $h \ge O(n^{-\frac{1}{3}})$, for any $\xi > 0$ we have*

$$\sup_{\{(\alpha,i) | \|\alpha - \theta\| \le \delta,\, L_{q,\alpha}(X_i)\neq 0\}} \big|K_i(\alpha) - K_i(\theta)\big| = o_p\big(n^{1+\xi} h\delta\big);$$

$$\sup_{\{(\alpha,i) | \|\alpha - \theta\| \le \delta,\, |\alpha' X_i - c_\alpha| < w_\alpha - n^\xi \delta - h\}} \big|K_i(\alpha) - K_i(\theta)\big| = o_p\big(n^{1+\xi} h\delta\big);$$

and

$$\sup_{i,j,\alpha\in B(\theta,\delta)} \big|K_i(\alpha) - K_i(\theta)\big| = o_p(n^{1+\xi}\delta);$$

(ix) *with $\delta_n = c_K nh - 2nh^{\frac{3}{2}}$ and $L_{q,\alpha}(X_i) \neq 0$, we have*

$$P(|K_i(\alpha)| < \delta_n) = o(n^{-t}), \quad \forall\, t > 0;$$

(x) *for any choice of fixed constant* a_0 *and a sequence* $c_n = o(1)$*, we have*

$$\sup_{\{(\alpha,i)||\alpha' X_i - c_\alpha|<q_\alpha + c_n\}} I(|K_i(\alpha)| < a_0) = o_p(n^{-t}),$$

for all $t > 0$*;*

where c_L, c_M *are some positive constants,* $I_K = |\int_{-1}^{1} K(s)ds|$*,* $c_q > 0$ *is the quantity specified in Lemma 4.6(ii) and* $c_K = c_q I_K$*. The results (i-viii) hold for fixed designs with* $O_p(\cdot)$ *replaced by* $O(\cdot)$ *and* $o_p(\cdot)$ *by* $o(\cdot)$*; results (ix-x) hold or fixed designs with the right hand sides replaced by 0 f when* n *is sufficiently large.*

Proof. The proofs for the two designs are very similar and we present the proof for the random designs only.

(i) Let $Z_{ij}(\alpha) = K_h\big(\alpha'(X_j - X_i)\big)$. Then $K_i(\alpha) = \sum_{j=1}^{n} Z_{ij}(\alpha)$. Given X_i, $Z_{ij}(\alpha)$ are i.i.d. and bounded. Now, uniformly in i, j, α,

$$\begin{aligned} E(Z_{ij}^2(\alpha)|X_i) &= \int K^2\Big(\frac{t - \alpha' X_i}{h}\Big) f_\alpha(t)dt \\ &\le h \int_{-1}^{1} K^2(s) f_\alpha(\alpha' X_i + hs)ds = O(h). \end{aligned}$$

Also, $\|\partial Z_{ij}(\alpha)/\partial\alpha\| = O(h^{-1})$ uniformly in i, j, α. Hence, by Lemma 4.3 with $\sigma_{nj}^2 = O(h)$ and $b_n = O(1)$, for all $\xi > 0$,

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

By Lemma 4.8(i), uniformly in $\{(\alpha, i) \mid |\alpha' X_i - c_\alpha| \le w_\alpha - h\}$,

$$\begin{aligned} E_i K_i(\alpha) &= K(0) + (n-1) E_i K_h(\alpha' X_j - \alpha' X_i) \\ &= nh f_\alpha(\alpha' X_i) \int_{-1}^{1} K(s)ds + O(nh^{r+1}), \end{aligned}$$

and the result follows.

(ii) Note that $L_{q,\alpha} \ne 0$ indicates that $|\alpha' X_i - c_\alpha| \le q w_\alpha < w_\alpha - h$. Hence, from (i) above and Lemma 4.6(ii),

$$\begin{aligned} &\inf_{\{(\alpha,i)|L_{q,\alpha}(X_i)\ne 0\}} |K_i(\alpha)| \\ &\ge \inf_{\{(\alpha,i)|L_{q,\alpha}(X_i)\ne 0\}} \Big|nh f_\alpha(\alpha' X_i) \int_{-1}^{1} K(s)ds\Big| \\ &\qquad - \sup_{\{(\alpha,i)|L_{q,\alpha}(X_i)\ne 0\}} \Big|K_i(\alpha) - nh f_\alpha(\alpha' X_i) \int_{-1}^{1} K(s)ds\Big| \\ &\ge c_q I_K nh + o_p(nh). \end{aligned}$$

When $\delta = o(1)$ and $L_{q,\alpha} \ne 0$,

$$\begin{aligned} |\theta' X_i - c_\theta| &\le |\alpha' X - c_\alpha| + O(\delta) \\ &\le q w_\alpha + O(\delta) = q w_\theta + O(\delta) \le w_\theta - h. \end{aligned}$$

The second result now follows from (i) above and an argument almost identical to the first result.

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

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

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

(iv) Note that

$$\begin{aligned} &|K_h(\alpha' X_j - \alpha' X_i)| \\ &\quad \leq \Big| |K_h(\alpha' X_j - \alpha' X_i)| - E_i|K_h(\alpha' X_j - \alpha' X_i)| \Big| + E_i|K_h(\alpha' X_j - \alpha' X_i)|. \end{aligned}$$

Hence a proof identical to (i) but with $Z_{ij}(\alpha) = |K_h(\alpha' X_j - \alpha' X_i)|$ gives that

$$\begin{aligned} &\sup_\alpha \max_{\{1\leq i\leq n\}} \sum_{j=1}^{n} |K_h(\alpha' X_j - \alpha' X_i)| \\ &\quad \leq o_p(\sqrt{nh}n^\xi) + \sup_\alpha \max_{\{1\leq i\leq n\}} \sum_{j=1}^{n} E_i|K_h(\alpha' X_j - \alpha' X_i)| \\ &\quad \leq o_p(\sqrt{nh}n^\xi) + \sup_\alpha \max_{\{1\leq i\leq n\}} nh \int_{-1}^{1} |K(s)| f_\alpha(\alpha' X_i + hs) ds = O_p(nh). \end{aligned}$$

The proof for the second result is almost identical.

(v) It follows from (ii) and (iv) above that

$$\begin{aligned} &\sup_{\{(\alpha,i)|L_{q,\alpha}(X_i)\neq 0\}} \sum_{j=1}^{n} |\kappa_{ij}(\alpha)| \\ &\quad \leq \frac{\sup_{\{(\alpha,i)|L_{q,\alpha}(X_i)\neq 0\}} \sum_{j=1}^{n} |K_h(\alpha' X_j - \alpha' X_i)|}{\inf_{\{(\alpha,i)|L_{q,\alpha}(X_i)\neq 0\}} |K_i(\alpha)|} = \frac{O_p(nh)}{c_K nh + o_p(nh)} = O_p(1). \end{aligned}$$

(vi) It follows from (ii) and (iv) above that

$$\begin{aligned} &\sup_\alpha \max_{\{1\leq j\leq n\}} \sum_{1\leq i\leq n; L_{q,\alpha}(X_i)\neq 0} |\kappa_{ij}(\alpha)| \\ &\quad \leq \frac{\sup_{(\alpha,j)} \sum_{i=1}^{n} |K_h(\alpha' X_j - \alpha' X_i)|}{\inf_{\{(\alpha,i)|L_{q,\alpha}(X_i)\neq 0\}} |K_i(\alpha)|} = \frac{O_p(nh)}{c_K nh + o_p(nh)} = O_p(1). \end{aligned}$$

(vii) Note that

$$\inf_\alpha \Big| \sum_{i=1}^{n} K_i(\alpha) \Big| \geq \inf_\alpha \Big| \sum_{|\alpha' X_i - c_\alpha| \leq w_\alpha - h} K_i(\alpha) \Big| - \sup_\alpha \Big| \sum_{|\alpha' X_i - c_\alpha| > w_\alpha - h} K_i(\alpha) \Big|.$$

By (i) above we have

$$\begin{aligned}
&\inf_\alpha \Big| \sum_{|\alpha' X_i - c_\alpha| \le w_\alpha - h} K_i(\alpha) \Big| \\
&\quad \ge \inf_\alpha \Big| \sum_{|\alpha' X_i - c_\alpha| \le w_\alpha - h} nh f_\alpha(\alpha' X_i) \Big| - \sup_\alpha \sum_{|\alpha' X_i - c_\alpha| \le w_\alpha - h} |K_i(\alpha) - nh f_\alpha(\alpha' X_i)| \\
&\quad \ge nh \cdot \inf_\alpha \Big| \sum_{i=1}^n f_\alpha(\alpha' X_i) I(|\alpha' X_i - c_\alpha| \le w_\alpha - h) \Big| + o_p(n^2 h).
\end{aligned}$$

A method identical to (ii) above yields that, for some constant $c_M > 0$,

$$\inf_\alpha \Big| \sum_{i=1}^n f_\alpha(\alpha' X_i) I(|\alpha' X_i - c_\alpha| \le w_\alpha - h) \Big| \ge c_M n + o_p(n).$$

Thus,

$$\inf_\alpha \Big| \sum_{|\alpha' X_i - c_\alpha| \le w_\alpha - h} K_i(\alpha) \Big| \ge c_M n^2 h + o_p(n^2 h).$$

Now, a method identical to (iv) above gives

$$\sup_\alpha \sum_{i=1}^n I(|\alpha' X_i - c_\alpha| > w_\alpha - h) = O_p(nh).$$

Hence, by (iv) above,

$$\begin{aligned}
\sup_\alpha \Big| \sum_{|\alpha' X_i - c_\alpha| > w_\alpha - h} K_i(\alpha) \Big| &\le \sup_{\alpha, i} |K_i(\alpha)| \cdot \sup_\alpha \sum_{i=1}^n I(|\alpha' X_i - c_\alpha| > w_\alpha - h) \\
&\le O_p(nh) \cdot O_p(nh) = O_p(n^2 h^2).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\inf_\alpha \Big| \sum_{i=1}^n K_i(\alpha) \Big| &\ge nh \cdot \inf_\alpha \Big| \sum_{i=1}^n f_\alpha(\alpha' X_i) I(|\alpha' X_i - c_\alpha| \le w_\alpha - h) \Big| + o_p(n^2 h) \\
&\ge c_M n^2 h + o_p(n^2 h).
\end{aligned}$$

(viii) Let

$$\begin{aligned}
&\mathcal{A}_\delta = \{(\alpha, i) \mid \|\alpha - \theta\| \le \delta,\ L_{q,\alpha}(X_i) \ne 0\}; \\
&\mathcal{B}_\delta = \{(\alpha, i) \mid \|\alpha - \theta\| \le \delta,\ |\alpha' X_i - c_\alpha| < w_\alpha - n^\xi \delta - h\};
\end{aligned}$$

and let E_i denote the conditional expectation given X_i. Define

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

and $Z_{ij}(\alpha) = A_{ij}(\alpha) - E_i A_{ij}(\alpha)$. Then

$$\begin{aligned}\sup_{(\alpha,i)\in\mathcal{A}_\delta}\big|K_i(\alpha)-K_i(\theta)\big| &= \sup_{(\alpha,i)\in\mathcal{A}_\delta}\Big|\sum_{j=1}^n A_{ij}(\alpha)\Big| \\ &\le \sup_{(\alpha,i)\in\mathcal{A}_\delta}\Big|\sum_{j=1}^n Z_{ij}(\alpha)\Big| + \sup_{(\alpha,i)\in\mathcal{A}_\delta}\Big|\sum_{j=1}^n E_i A_{ij}(\alpha)\Big|.\end{aligned}$$

Note that uniformly in i, j and $\alpha \in B(\theta, \delta)$ we have $A_{ij}(\alpha) \le O(h^{-1}\delta)$ and

$$\mathrm{Var}(A_{ij}(\alpha)|X_i) \le E_i\Big(K_h(\alpha' X_j - \alpha' X_i) - K_h(\theta' X_j - \theta' X_i)\Big)^2 \le O(h^{-1}\delta^2).$$

By Lemma 4.3 with $\sigma_n^2 = O(nh^{-1}\delta^2)$ and $b_n = O(h^{-1}\delta)$ we have

$$\sup_{(\alpha,i)\in B(\theta,\delta)}\Big|\sum_{j=1}^n Z_{ij}(\alpha)\Big| = o_p(\sqrt{nh^{-1}}\delta n^{\xi} + h^{-1}\delta n^{\xi}).$$

By Lemma 4.8(ii), $E_i A_{ij}(\alpha) = O(h\delta)$ uniformly in $(\alpha, i) \in \mathcal{A}_\delta$ and j. Hence, if $h \ge O(n^{-\frac{1}{3}})$, we have

$$\sup_{(\alpha,i)\in\mathcal{A}_\delta}\big|K_i(\alpha)-K_i(\theta)\big| \le o_p(\sqrt{nh^{-1}}\delta n^{\xi} + h^{-1}\delta n^{\xi} + nh\delta) = o_p(n^{1+\xi}h\delta).$$

Similarly we have

$$\sup_{(\alpha,i)\in\mathcal{B}_\delta}\big|K_i(\alpha)-K_i(\theta)\big| \le o_p(n^{1+\xi}h\delta).$$

Finally, by Lemma 4.8(iii), $E_i A_{ij}(\alpha) = O(\delta)$ uniformly in i, j and $\alpha \in B(\theta, \delta)$. Thus,

$$\sup_{i,j,\alpha\in B(\theta,\delta)}\big|K_i(\alpha)-K_i(\theta)\big| \le o_p(\sqrt{nh^{-1}}\delta n^{\xi} + h^{-1}\delta n^{\xi} + n\delta) = o_p(n^{1+\xi}\delta).$$

(ix) Since $L_{q,\alpha}(X_i) \ne 0$, $|\alpha' X_i - c_\alpha| \le q_\alpha$ and hence, by Lemma 4.6(ii), $f_\alpha(\alpha' X_i) \ge c_q$. Note that

$$|K_i(\alpha)| \ge \big|E_i K_i(\alpha)\big| - \big|K_i(\alpha) - E_i K_i(\alpha)\big|.$$

Hence, by Lemma 4.8(i) and Bernstein's inequality

$$\begin{aligned}P(|K_i(\alpha)| < \delta_n) &\le P\Big(\big|E_i K_i(\alpha)\big| - \big|K_i(\alpha) - E_i K_i(\alpha)\big| < \delta_n\Big) \\ &= P\Big(nhf_\alpha(\alpha' X_i)I_K + O(nh^{r+1}) - \big|K_i(\alpha) - E_i K_i(\alpha)\big| < c_q I_K nh - 2nh^{\frac{3}{2}}\Big) \\ &\le P\Big(\big|K_i(\alpha) - E_i K_i(\alpha)\big| > [f_\alpha(\alpha' X_i) - c_q]I_K nh + nh^{\frac{3}{2}}\Big) \\ &\le P\Big(\big|K_i(\alpha) - E_i K_i(\alpha)\big| > nh^{\frac{3}{2}}\Big) \\ &\le 2\exp\Big\{-\frac{(nh^{\frac{3}{2}})^2}{C_1 nh + nh^2}\Big\} = o(n^{-t}),\end{aligned}$$

for all $t > 0$.

(x) Let $\mathcal{B} = \{(\alpha, i) \mid |\alpha' X_i - c_\alpha| < q_\alpha + c_n\}$ and partition it into n^{s+1} (s is decided below) small cells evenly. Take one point from each cell and let $\mathcal{B}_n$ denote the collection of them. Then for every $(\alpha, i) \in \mathcal{A}$ we can find one $(\tilde{\alpha}, i) \in \mathcal{A}_n$ such that $\|\alpha - \tilde{\alpha}\| = O(n^{-\frac{s}{p}})$ uniformly in α. By taking s large enough, we have, uniformly in i and α,

$$\big||K_i(\alpha)| - |K_i(\tilde{\alpha})|\big| \le O(N \cdot n^{-\frac{s}{p}} \cdot h^{-1}) < a_0;$$

and hence,

$$\Big|\inf_{(\alpha,i)\in\mathcal{B}} |K_i(\alpha)| - \inf_{(\alpha,i)\in\mathcal{B}_n} |K_i(\alpha)|\Big| < a_0.$$

Therefore, by Lemma 4.9(ix),

$$\begin{aligned}
\sup_{(\alpha,i)\in\mathcal{B}} I(|K_i(\alpha)| < a_0) &\le I\big(\inf_{(\alpha,i)\in\mathcal{B}} |K_i(\alpha)| < a_0\big) \\
&\le I\Big(\inf_{(\alpha,i)\in\mathcal{B}_n} |K_i(\alpha)| - \big|\inf_{(\alpha,i)\in\mathcal{B}} |K_i(\alpha)| - \inf_{(\alpha,i)\in\mathcal{B}_n} |K_i(\alpha)|\big| < a_0\Big) \\
&\le I\big(\inf_{(\alpha,i)\in\mathcal{B}_n} |K_i(\alpha)| < 2a_0\big) = o_p(n^{-t}),
\end{aligned}$$

for all $t > 0$ and the proof is completed.

$\square$

Lemma 4.10. *Suppose assumptions (A1–A5) hold and $v \ge 6$. Let $Z_{ij}(\alpha) = K_h(\alpha' X_j - \alpha' X_i)$. Let $U_i(\alpha)$ be independent of ϵ_i's. Suppose $\sup_{i,\alpha} |U_i(\alpha)| = O_p(u_n)$. Then, for both designs,*

(i) $\sum_{i=1}^n \sup_{\|\alpha-\theta\|<\delta} \Big|\sum_{j=1}^n \big(Z_{ij}(\alpha) - Z_{ij}(\theta)\big)\epsilon_j\Big|^2 = o_p\big(n^{2+\xi} h^{-1} \delta^2\big)$*;*

(ii) $\sum_{i=1}^n \sup_\alpha \Big|\sum_{j=1}^n Z_{ij}(\alpha)\epsilon_j\Big|^2 = o_p(n^{2+\xi} h)$*; and*

$$\sum_{i=1}^n \sup_{\{\alpha \mid L_{q,\alpha}(X_i)\neq 0\}} \Big|\sum_{j=1}^n \kappa_{ij}(\alpha)\epsilon_j\Big|^2 = o_p(n^\xi h^{-1});$$

(iii) $\sup_{\|\alpha-\theta\|\le\delta} \sum_{\{i \mid L_{q,\alpha}(X_i)\neq 0\}} \Big(\sum_{j=1}^n \big[\kappa_{ij}(\alpha) - \kappa_{ij}(\theta)\big]\epsilon_j\Big)^2 = o_p(h^{-3}\delta^2 n^\xi)$*, provided* $\delta = o(1)$*;*

(iv) $\sup_\alpha \max_{\{i \mid L_{q,\alpha}(X_i)\neq 0\}} |K_{mi}(\alpha) - g_\alpha(\alpha' X_i)| = o_p\big(h^2 n^\xi\big)$*; moreover, if* $K(\cdot)$ *is an* r*-th order kernel function, this upper bound can be improved to* $o_p\big((h^r + n^{-\frac{1}{2}} h^{\frac{1}{2}}) n^\xi\big)$ *(here* $o_p(\cdot)$ *should be changed into* $o(\cdot)$ *for fixed designs);*

(v) $\sum_{i\alpha} K_{ei}(\alpha) U_i(\alpha) = o_p(n^{-\frac{1}{2}+\xi} u_n)$*.*

Proof. The proofs for the two designs are very similar for (i), (ii), (iii) and (v). We only present the proof for the random design models in these cases.

(i) Let $w_{ij}(\alpha) = Z_{ij}(\alpha) - Z_{ij}(\theta)$. Note that uniformly in i, j and $\alpha \in B(\theta, \delta)$ we have $\sum_{i=1}^n E_i w_{ij}^2(\alpha) \le v_n = O(nh^{-1}\delta^2)$ and $|w_{ij}(\alpha)| \le c_n = O(\delta h^{-1})$. By Lemma 4.4,

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

Clearly the upper bound can be reduced to $o_p\big(n^{2+\xi}h^{-1}\delta^2\big)$ provided $h \ge O(n^{-\frac{1}{3}})$ and $v \ge 6$.

(ii) Since uniformly in i, j, α we have $\sum_{i=1}^n E_i Z_{ij}^2(\alpha) \le v_n = O(nh)$ and $|Z_{ij}(\alpha)| \le c_n = O(1)$, by Lemma 4.4,

$$\sum_{i=1}^n \sup_{\alpha}\Big|\sum_{j=1}^n Z_{ij}(\alpha)\epsilon_j\Big|^2 \le o_p\big(n^{1+\xi} nh + n^{1+\frac{4}{v}+\xi}\big) = o_p\big(n^{2+\xi}h\big);$$

The second result follows from Lemma 4.9(ii).

(iii) Note that

$$\kappa_{ij}(\alpha) - \kappa_{ij}(\theta) = \frac{K_i(\theta) - K_i(\alpha)}{K_i(\alpha)K_i(\theta)} Z_{ij}(\alpha) + \frac{Z_{ij}(\alpha) - Z_{ij}(\theta)}{K_i(\theta)}.$$

From (i,ii) above and Lemma 4.9(ii,viii),

$$\begin{aligned}
&\sup_{\|\alpha-\theta\|\le\delta} \sum_{\{i|L_{q,\alpha}(X_i)\neq 0\}} \Big(\sum_{j=1}^n \big[\kappa_{ij}(\alpha) - \kappa_{ij}(\theta)\big]\epsilon_j\Big)^2 \\
&\le 2 \sup_{\|\alpha-\theta\|\le\delta} \sum_{\{i|L_{q,\alpha}(X_i)\neq 0\}} \Big(\frac{K_i(\theta) - K_i(\alpha)}{K_i(\alpha)K_i(\theta)}\Big)^2 \Big(\sum_{j=1}^n Z_{ij}(\alpha)\epsilon_j\Big)^2 \\
&\qquad + 2 \sup_{\|\alpha-\theta\|\le\delta} \sum_{\{i|L_{q,\alpha}(X_i)\neq 0\}} \Big(\frac{1}{K_i(\theta)}\Big)^2 \Big(\sum_{j=1}^n \big[Z_{ij}(\alpha) - Z_{ij}(\theta)\big]\epsilon_j\Big)^2 \\
&\le o_p\Big(\frac{n^{1+\xi}h\delta}{(nh)^2}\Big)^2 o_p\big(n^{2+\xi}h\big) + O_p\Big(\frac{1}{(nh)^2}\Big) o_p\big(n^{2+\xi}h^{-1}\delta^2 + n^{1+\frac{4}{v}+\xi}h^{-2}\delta^2\big) \\
&= o_p\big((h^{-3} + n^{\frac{4}{v}-1}h^{-4})\delta^2 n^{\xi}\big).
\end{aligned}$$

Clearly the upper bound is reduced to $o_p(h^{-3}\delta^2 n^{\xi})$ if $h \ge O(n^{-\frac{1}{3}})$ and $v \ge 6$.

(iv) First consider random designs. Note that $g_\alpha(\alpha' X_i) = E\big(m(X_i)|\alpha' X_i\big)$ and

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

Let $Z_{ij}(\alpha) = \big[m(X_j) - E\big(m(X_i)|\alpha' X_i\big)\big] K_h(\alpha' X_j - \alpha' X_i)$. Given $\alpha' X_i$, $Z_{ij}(\alpha)$ $(j \neq i)$ are i.i.d., uniformly bounded and, uniformly in i and α,

$$\begin{aligned}
E(Z_{ij}^2(\alpha)|X_i) &= E\Big[E\Big(Z_{ij}^2(\alpha) \mid (\alpha' X_i, \alpha' X_j)\Big) \mid X_i\Big] \\
&= E\Big(\big(g_\alpha(\alpha' X_j) - g_\alpha(\alpha' X_i)\big)^2 K_h(\alpha' X_j - \alpha' X_i)\Big) = O(h^3).
\end{aligned}$$

By Lemma 4.3 with $\sigma_n^2 = O(nh^3)$ and $b_n = O(1)$, we have

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

With $L_{q,\alpha}(X_i) \neq 0$, i.e. $s_i = \alpha' X_i \in c_\alpha \pm q_\alpha$, by Taylor's expansion,

$$\begin{aligned} E_i Z_{ij}(\alpha) &= E\big[E\big(Z_{ij}(\alpha)|\alpha' X_j, X_i\big)|X_i\big] \\ &= E\Big[\big(g_\alpha(\alpha' X_j) - g_\alpha(s_i)\big)K_h(\alpha' X_j - s_i)|X_i\Big] \\ &= h\int_{-1}^1 \big(g_\alpha(s_i + uh) - g_\alpha(s_i)\big)K(u) f_\alpha(s_i + uh)du \\ &= h\int_{-1}^1 \big(uh \cdot g_\alpha^{(1)}(s_i) + O(h^2)\big)K(u)\big(f_\alpha(s_i) + O(h)\big)du. \qquad (4.7) \end{aligned}$$

Thus, for $j \neq i$, $\sup_\alpha \max_{\{i|L_{q,\alpha}(X_i)\neq 0\}} |E_i Z_{ij}(\alpha)| = O(h^3)$. This bound can be improved to $O(h^{r+1})$ if $K(\cdot)$ is an r-th order kernel function and a higher order Taylor's expansion is used. By Lemma 4.9(ii),

$$\begin{aligned} \sup_\alpha \max_{\{i|L_{q,\alpha}(X_i)\neq 0\}} |K_{mi} - g_\alpha(\alpha' X_i)| &= o_p\big(\frac{1}{nh}(\sqrt{nh^3} + 1 + nh^{r+1})n^\xi\big) \\ &= o_p\big((h^r + n^{-\frac{1}{2}}h^{\frac{1}{2}})n^\xi\big). \end{aligned}$$

Note that this upper bound order can be reduced to $o_p(h^2 n^\xi)$ if $r = 2$.

For fixed design models, note that, for all $\xi > 0$,

$$\begin{aligned} K_{mi}(\alpha) - g_\alpha(\alpha' X_i) &= \frac{1}{K_i(\alpha)}\sum_{j=1}^n \big[m(X_j) - g_\alpha(\alpha' X_i)\big]K_h(\alpha' X_j - \alpha' X_i) \\ &= \frac{1}{K_i(\alpha)} n\int_S \big[m(x) - g_\alpha(\alpha' X_i)\big]K_h(\alpha' x - \alpha' X_i)dx + o(n^{-1+\xi}). \end{aligned}$$

Again let $s_i = \alpha' X_i$. Since

$$\int_S g_\alpha(s_i)K_h(\alpha' x - s_i)dx = \int_{-1}^1 g_\alpha(s_i) f_\alpha(u) K_h(u - s_i)du,$$

and, noticing (2.5),

$$\int_S m(x)K_h(\alpha' x - s_i)dx = \int_{-1}^1 f_\alpha(u) g_\alpha(u) K_h(u - s_i)du,$$

we have

$$\int_S \big[m(x) - g_\alpha(s_i)\big]K_h(\alpha' x - s_i)dx = \int_{-1}^1 \big[g_\alpha(u) - g_\alpha(s_i)\big] f_\alpha(u) K_h(u - s_i)du.$$

The result now follows from an argument identical to the random design case.

(v) Write $\sum_{i\alpha} K_{ei}(\alpha)U_i(\alpha) = \sum_{j=1}^n Z_j(\alpha)\epsilon_j$, where

$$Z_j(\alpha) = \frac{1}{L_\alpha}\sum_{i=1}^n \kappa_{ij}(\alpha)U_i(\alpha)L_{q,\alpha}(X_i)I(|K_i(\alpha)| > a_0).$$

By Lemma 4.9(iii,vi), $\sup_{j,\alpha}|Z_j(\alpha)| = o_p(n^{-1}u_n)$. The result now follows from Lemma 4.5.

□

Lemma 4.11. *Let u be any fixed integer. For pairs $\{(i_s, j_s)_{s=1}^u\}$, define*

$$b_{i\alpha} = L_{q,\alpha}(X_i)I\big(|K_j(\alpha)| > a_0\big),$$

and let t be the distinct-index number for $\{(i_s, j_s)_{s=1}^u\}$ as defined in Definition 4.1. Then, for random designs, uniformly in $\{(i_s, j_s)\}$ and $\alpha \in B(\theta, \delta)$,

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

(ii) $E\Big|\kappa_{i_1j_1}(\alpha)\kappa_{i_2j_2}(\alpha)\cdots\kappa_{i_uj_u}(\alpha)\cdot b_{i_1\alpha}\cdots b_{i_u\alpha}\Big| = O\big(h^t(Nh)^{-u}\big)$;

(iii) $E\big|[\kappa_{i_1j_1}(\alpha) - \kappa_{i_1j_1}(\theta)]\cdots[\kappa_{i_sj_s}(\alpha) - \kappa_{i_sj_s}(\theta)]\cdot b_{i_1\alpha}\cdots b_{i_u\alpha}\big| = O\big(\delta^u(Nh^2)^{-u}\big)$, *provided* $\delta = o(1)$.

The result (iii) holds for fixed designs with ordinary understanding of expectation, i.e. E(a)=a for any constant a.

Proof. For simplicity of presentation, we give the proof for $u = 2$. The proof for general u is almost identical. The first result follows immediately from direct verification. Now we show (ii). Let $Z_{ij}(\alpha) = K_h(\alpha'X_j - \alpha'X_i)$. Then $\kappa_{ij}(\alpha) = Z_{ij}(\alpha)/K_i(\alpha)$. Note that

$$\kappa_{ij}(\alpha)b_{i\alpha} = \frac{K_h(\alpha'X_j - \alpha'X_i)}{K_i(\alpha)}L_{q,\alpha}(X_i)I\big(|K_j(\alpha)| > a_0\big)$$

are uniformly bounded by $O(1)$. Hence, by Lemma 4.9(ix), uniformly we have, with $\delta_n = O(Nh)$ and for all $t > 0$,

$$\begin{aligned}
&E\Big|\kappa_{i_1j_1}(\alpha)\kappa_{i_2j_2}(\alpha)b_{i_1\alpha}b_{i_2\alpha}\Big| \\
&\quad\le E\Big|\kappa_{i_1j_1}(\alpha)\kappa_{i_2j_2}(\alpha)b_{i_1\alpha}b_{i_2\alpha}I\big(|K_{i_1}(\alpha)| \ge \delta_n\big)I\big(|K_{i_2}(\alpha)| \ge \delta_n\big)\Big| \\
&\qquad+ E\Big|\kappa_{i_1j_1}(\alpha)\kappa_{i_2j_2}(\alpha)b_{i_1\alpha}b_{i_2\alpha}I\big(|K_{i_1}(\alpha)| < \delta_n\big)\Big| \\
&\qquad+ E\Big|\kappa_{i_1j_1}(\alpha)\kappa_{i_2j_2}(\alpha)b_{i_1\alpha}b_{i_2\alpha}I\big(|K_{i_2}(\alpha)| < \delta_n\big)\Big| \\
&\quad\le \frac{1}{\delta_n^2}E|Z_{i_1j_1}(\alpha)Z_{i_2j_2}(\alpha)| + o(n^{-2}) = O\big(h^t(Nh)^{-2}\big).
\end{aligned} \tag{4.8}$$

For the last part, we have

$$\begin{aligned}
&E\big|[\kappa_{i_1j_1}(\alpha) - \kappa_{i_1j_1}(\theta)][\kappa_{i_2j_2}(\alpha) - \kappa_{i_2j_2}(\theta)]b_{i_1\alpha}b_{i_2\alpha}\big| \\
&\quad= E\Big|\Big(\frac{Z_{i_1j_1}(\alpha) - Z_{i_1j_1}(\theta)}{K_{i_1}(\alpha)} + Z_{i_1j_1}(\theta)\frac{K_{i_1}(\theta) - K_{i_1}(\alpha)}{K_{i_1}(\alpha)K_{i_1}(\theta)}\Big) \\
&\qquad\cdot\Big(\frac{Z_{i_2j_2}(\alpha) - Z_{i_2j_2}(\theta)}{K_{i_2}(\alpha)} + Z_{i_2j_2}(\theta)\frac{K_{i_2}(\theta) - K_{i_2}(\alpha)}{K_{i_2}(\alpha)K_{i_2}(\theta)}\Big)b_{i_1\alpha}b_{i_2\alpha}\Big|.
\end{aligned}$$

The result for fixed designs now follow by Lemma 4.9(ii,viii) and the observation that $Z_{ij}(\alpha) - Z_{ij}(\theta)$ is uniformly bounded by $O(h^{-1}\delta)$. For random designs, by Lemma 4.8(iii), uniformly in i and $\alpha \in B(\theta, \delta)$, uniformly for all integers $s \le u$,

$$E\big([K_i(\alpha) - K_i(\theta)]^s \mid X_i\big) = E_i\Big(\sum_{j=1}^{n}[Z_{ij}(\alpha) - Z_{ij}(\theta)]\Big)^s$$
$$= \sum_{j_1 j_2 \cdots j_s} E_i[Z_{ij_1}(\alpha) - Z_{ij_1}(\theta)] \cdots E_i[Z_{ij_s}(\alpha) - Z_{ij_s}(\theta)] + o(n\delta)^s = O(n^s\delta^s),$$

which gives

$$E\big([K_i(\alpha) - K_i(\theta)]^s\big) = O(n^s\delta^s). \tag{4.9}$$

Also note that uniformly in i, j and $\alpha \in B(\theta, \delta)$,

$$Z_{ij}(\alpha) - Z_{ij}(\theta) = O\big(\frac{\delta}{h}\big).$$

Therefore, when $\delta = o(1)$, by Lemma 4.9(ii), a method similar to showing (4.8) gives,

$$E\Big|\frac{Z_{i_1j_1}(\alpha) - Z_{i_1j_1}(\theta)}{K_{i_1}(\alpha)} Z_{i_2j_2}(\theta)\frac{K_{i_2}(\theta) - K_{i_2}(\alpha)}{K_{i_2}(\alpha)K_{i_2}(\theta)} b_{i_1\alpha}b_{i_2\alpha}\Big|$$
$$\le O\big(\frac{\delta}{h}\big) \cdot E\Big|\frac{K_{i_2}(\theta) - K_{i_2}(\alpha)}{K_{i_1}(\alpha)K_{i_2}(\alpha)K_{i_2}(\theta)} b_{i_1\alpha}b_{i_2\alpha}\Big|$$
$$= O\big(\frac{\delta}{h}\big) \cdot O\Big(\frac{n\delta}{(nh)^3}\Big) + o(n^{-3}) = O\Big(\frac{\delta^2}{(nh^2)^2}\Big).$$

We can similarly show that the other three terms are of the same order. The result for $u > 2$ follows by similar argument coupled with the observation that

$$E\big|[K_{i_1}(\alpha) - K_{i_1}(\theta)] \cdots [K_{i_u}(\alpha) - K_{i_u}(\theta)]\big| = O(n^u\delta^u), \tag{4.10}$$

which follows by Cauchy-Schwartz inequality and (4.9). □

4.3 Decomposition of $\hat{d}(\alpha)$

Write $m(X_i) = m_\alpha(X_i) + g_\alpha(\alpha' X_i)$ and $\hat{g}_\alpha(\alpha' X_i) = \sum_{j=1}^{n} \kappa_{ij}(\alpha) Y_j = K_{mi}(\alpha) + K_{ei}(\alpha)$. Then we have

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

where

$$\begin{aligned}
R_0(\alpha) &= \sum_{i\alpha} \epsilon_i^2; & R_1(\alpha) &= \sum_{i\alpha} m_\alpha^2(X_i);\\
R_2(\alpha) &= \sum_{i\alpha} \Big(g_\alpha(\alpha' X_i) - K_{mi}(\alpha)\Big)^2; & R_3(\alpha) &= \sum_{i\alpha} K_{ei}^2(\alpha);\\
R_4(\alpha) &= 2\sum_{i\alpha} m_\alpha(X_i)\epsilon_i; & R_5(\alpha) &= 2\sum_{i\alpha}\Big(g_\alpha(\alpha' X_i) - K_{mi}(\alpha)\Big)\epsilon_i;\\
R_6(\alpha) &= -2\sum_{i\alpha}\sum_{j=1}^{n} \kappa_{ij}(\alpha)\epsilon_i\epsilon_j; & R_7(\alpha) &= 2\sum_{i\alpha} m_\alpha(X_i)\Big(g_\alpha(\alpha' X_i) - K_{mi}(\alpha)\Big);\\
R_8(\alpha) &= -2\sum_{i\alpha} m_\alpha(X_i)K_{ei}(\alpha); & R_9(\alpha) &= -2\sum_{i\alpha}\Big(g_\alpha(\alpha' X_i) - K_{mi}(\alpha)\Big)K_{ei}(\alpha).
\end{aligned}$$

Lemma 4.12. *Suppose assumptions (A1–A5) hold and $v \geq 8$. Then, for any $\xi > 0$, for random designs,*

R0) *when $Var(\epsilon|x) = Var(\epsilon)$, $\sup_{\beta\in D}\sup_{\|\alpha-\beta\|\leq\delta} |R_0(\alpha) - R_0(\beta)| = o_p\big(n^{-\frac{1}{2}+\xi}\delta\big)$; for heteroscedastic errors, the result holds when $L_{q,\alpha}$ is free of α;*

R1) *for any fixed constant $t \in (0, \frac{1}{2})$, there exists a constant $c_t > 0$ such that*

$$\inf_{\|\alpha-\theta\|>t} R_1(\alpha) > c_t + o_p(1);$$

moreover, there exists a constant $c_1 > 0$ such that, uniformly in α,

$$c_1\|\alpha-\theta\|^2 + O_p(\|\alpha-\theta\|^3) \leq R_1(\alpha) \leq O_p(\|\alpha-\theta\|^2);$$

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

R3) $\sup_{\alpha\in D} |R_3(\alpha)| = o_p\big(n^{-1+\xi}h^{-1}\big)$*; moreover, when $\delta = o(1)$, $\sup_{\|\alpha-\theta\|\leq\delta} |R_3(\alpha) - R_3(\theta)| = o_p(N^{-1+\xi}h^{-3}\delta^2 + N^{-1+\xi}h^{-2}\delta)$;*

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

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

R6) $\sup_{\alpha} |R_6(\alpha)| = o_p(n^{-\frac{1}{2}+\xi}h^{-\frac{1}{2}})$*; and* $\sup_{\|\alpha-\theta\|\leq\delta} |R_6(\alpha) - R_6(\theta)| = o_p\big(N^{-\frac{1}{2}}\delta\big)$*, provided $\delta = o(1)$, $h = O(N^{-a})$ and $v \geq 2u$ where u is the smallest positive even integer greater than $\frac{3a}{1-4a}p$;*

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

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

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

The above results hold for fixed designs with $O_p(\cdot)$ replaced by $O(\cdot)$ and $o_p(\cdot)$ replaced by $o(\cdot)$ for terms R_1, R_2 and R_7.

Proof. The proof for two designs are quite similar and we only present the proof for the random design except for (R1). In all cases, Lemma 4.9(iii) is applied to take care of each of the denominators.

(R0) When $L_{q,\alpha}(x)$ is free of α, clearly $R_0(\alpha) - R_0(\beta) = 0$. Now suppose $\mathrm{Var}(\epsilon|x) = \sigma^2$. Note that

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

Since

$$L_{q,\alpha}(X_i) L_\beta - L_{q,\beta}(X_i) L_\alpha = \big(L_{q,\alpha}(X_i) - L_{q,\beta}(X_i) \big) L_\beta + L_{q,\beta}(X_i)(L_\beta - L_\alpha)$$

which is uniformly bounded by $C_1 n\delta$ for some constant C_1, and $\epsilon^2 - \sigma^2$ has mean zero with at least four moments, the result follows from Lemma 4.5 and Lemma 4.9.

(R1) Since $m(x) = g(\theta' x)$ we have $g(\theta' X) = g(\alpha' X) + (\theta - \alpha)' X \cdot g^{(1)}(\alpha' X) + o(\|\alpha - \theta\|)$, and hence

$$m_\alpha(X) = g(\theta' X) - E(g(\theta' X)|\alpha' X) = (\theta - \alpha)' \mu(X) + O(\|\alpha - \theta\|^2), \qquad (4.11)$$

where $\mu(x) = \big(x - E(X|\theta' X = \theta' x) \big) g^{(1)}(\theta' x)$. Note that $\inf_{x \in B_0} L_{q,\alpha}(x) \geq c_{01}$ for some constant $c_{01} > 0$, where B_0 is the interior ball specified in Assumption (A2). Hence,

$$\begin{aligned} R_1(\alpha) = \sum_{i\alpha} m_\alpha^2(X_i) &\geq \frac{c_{01}}{L_\alpha} \sum_{i=1}^{n} m_\alpha^2(X_i) I(X_i \in B_0) \\ &= \frac{c_{01} n}{L_\alpha} \sum_{i=1}^{n} E[m_\alpha^2(X_i) I(X_i \in B_0)] \\ &\qquad - \frac{c_{01} n}{L_\alpha} \Big| \sum_{i=1}^{n} \Big(m_\alpha^2(X_i) I(X_i \in B_0) - E[m_\alpha^2(X_i) I(X_i \in B_0)] \Big) \Big|. \end{aligned}$$

Since $m(\cdot)$ is not constant on S_X, WLOG we may assume that $m(\cdot)$ is not constant on B_0. By the identifiability of single index models (Lin and Kulasekera 2006), $m_\alpha(\cdot)$ is not constant on B_0 for any $\alpha \neq \theta$ ($\alpha \in D$). Hence, by the continuity of $E[m_\alpha^2(X) I(X \in B_0)]$,

$$\inf_{\|\alpha - \theta\| > t} E[m_\alpha^2(X) I(X \in B_0)] = \tilde{c}_t,$$

for some constant $\tilde{c}_t > 0$. By (4.11) and Lemma 4.3,

$$\sup_{\|\alpha - \theta\| \leq \delta} \Big| \sum_{i=1}^{n} \Big(m_\alpha^2(X_i) I(X_i \in B_0) - E[m_\alpha^2(X_i) I(X_i \in B_0)] \Big) \Big| = o_p(n^{\frac{1}{2} + \xi} \delta^2).$$

Thus we have

$$\inf_{\|\alpha - \theta\| > t} R_1(\alpha) > c_t + o_p(1),$$

for some constant $c_t > 0$.

The upper bound of the second result follows directly from (4.11). For the lower bound,

$$
\begin{aligned}
R_1(\alpha) &= \sum_{i\alpha} m_\alpha^2(X_i) \geq \frac{c_{01}}{L_\alpha} \sum_{i=1}^n m_\alpha^2(X_i) I(X_i \in B_0) \\
&= \frac{c_{01}}{L_\alpha} \sum_{i=1}^n [(\alpha-\theta)'\mu(X_i) I(X_i \in B_0)]^2 + O_p(\|\alpha-\theta\|^3) \\
&= \frac{c_{01} n}{L_\alpha} (\alpha-\theta)' W (\alpha-\theta) + O_p(\|\alpha-\theta\|^3),
\end{aligned}
$$

where $W = E\big(\mu(X)\mu(X)' \cdot I(X \in B_0)\big)$. 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 $\Lambda = \operatorname{diag}(\lambda_1, \cdots, \lambda_p)$ with $\lambda_1 \geq \cdots \lambda_{p-1} > \lambda_p = 0$. Since $\theta'\mu(X) = 0$ a.s., we have $W\theta = 0$, i.e. θ is an eigenvector corresponding to $\lambda_p = 0$. Hence the last row of Q can be taken to be θ. Let $b = Q(\alpha - \theta)$. Then $b_p = \theta'(\alpha-\theta) = -\|\alpha-\theta\|^2/2$. Observing that

$$
\|\alpha+\theta\|^2 + \|\alpha-\theta\|^2 = 4,
$$

and

$$
\inf_{\alpha \in D} \|\alpha+\theta\|^2 \geq \inf_{\alpha \in D} (\alpha_1+\theta_1)^2 \geq \theta_1^2,
$$

we have

$$
\|b\|^2 - b_p^2 = \|\alpha-\theta\|^2 - \frac{1}{4}\|\alpha-\theta\|^4 = \frac{1}{4}\|\alpha+\theta\|^2\|\alpha-\theta\|^2 \geq \frac{\theta_1^2}{4}\|\alpha-\theta\|^2.
$$

Thus

$$
\begin{aligned}
R_1(\alpha) &= \frac{n}{L_\alpha}(\alpha-\theta)'Q'\Lambda Q(\alpha-\theta) + o(\|\alpha-\theta\|^2) = \frac{n}{L_\alpha} b'\Lambda b + o(\|\alpha-\theta\|^2) \\
&\geq \frac{n}{L_\alpha}\lambda_{p-1}(\|b\|^2 - b_p^2) + o(\|\alpha-\theta\|^2) \\
&\geq c_1\|\alpha-\theta\|^2 + o_p(\|\alpha-\theta\|^2),
\end{aligned}
$$

where $c_1 = \theta_1^2\lambda_{p-1}/(4U)$ with U being an upper bound for $L(\cdot)$. It is left to show $\operatorname{rank}(W) = p-1$. Take $\beta \perp \theta$. It suffices to show $\beta' W \beta > 0$. Note that

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

By assumption (A1), 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 constant $C_1 > 0$. Hence $\beta' W\beta \geq C_2 t$ for some constant $C_2 > 0$ where

$$
t = \int_B \beta'\big(x - E(X|\theta'X = \theta'x)\big)\big(x - E(X|\theta'X = \theta'x)\big)'\beta \, dx.
$$

Note that $t = 0$ if and only if, when $X \in B$, $\beta' X = \xi(\theta' X)$ for some function $\xi(\cdot)$, which is impossible since $(\theta' X, \beta' X)$ has a positive joint density on B. This completes the proof.

Next, we consider fixed-design models. Recall that $g_\alpha(\cdot)$ was defined in (2.5). Let $S(t) = \{y \in AS \mid y_1 = t\}$. Since $A\alpha = (1, 0, \cdots, 0)'$, we have that, under H_0,

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

where

$$r(\alpha, t) = \frac{1}{f_\alpha(t)} \int_{S(t)} A' y \, dy_2 \cdots dy_p.$$

Hence,

$$\begin{aligned}
\sum_{i=1}^n m_\alpha^2(x_i) L_{q,\alpha}(x_i) &= n \int_S \big(g(\theta' x) - g_\alpha(\alpha' x)\big)^2 L_{q,\alpha}(x) dx + o(n^\xi) \\
&= n \int \big(g^{(1)}(\theta' x) \cdot (\theta - \alpha)'[x - r(\theta, \theta' x)]\big)^2 L_{q,\theta}(x) dx + o(n\|\theta - \alpha\|^2),
\end{aligned}$$

which gives the first part immediately. For the second part, 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,

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

where

$$W(\theta) = \int_B [x - r(\theta, \theta' x)][x - r(\theta, \theta' x)]' dx.$$

Note that $\alpha' r(\alpha, t) = t$. Hence $\theta' W(\theta)\theta = 0$. Let A_0 be an orthogonal matrix with first row θ' and other rows β_k', $k = 2, \cdots, p$. Then we have

$$\begin{aligned}
\beta_k' W(\theta)\beta_k &= \int_B [\beta_k' x - \beta_k' r(\theta, \theta' x)]^2 dx \\
&= \int_{A_0 B} [y_k - \beta_k' r(\theta, y_1)]^2 dy > 0.
\end{aligned}$$

Hence W has rank $p - 1$ and the result follows as in random design case.

(R2) It's a direct consequence of Lemma 4.10(iv).

(R3) The first result follows from Lemma 4.10(ii) and Lemma 4.9(ii). For the second result, write $R_3(\alpha) = \sum_{i=1}^{3} U_i(\alpha)$ where

$$U_1(\alpha) = \frac{1}{L_\alpha} \sum_{i=1}^{n} \Big(\sum_{j=1}^{n} \big(\kappa_{ij}(\alpha) - \kappa_{ij}(\theta)\big)\epsilon_j \Big)^2 b_{i\alpha}$$

$$U_2(\alpha) = \frac{2}{L_\alpha} \sum_{i=1}^{n} \Big(\sum_{j=1}^{n} \big(\kappa_{ij}(\alpha) - \kappa_{ij}(\theta)\big)\epsilon_j \Big) \Big(\sum_{j=1}^{n} \kappa_{ij}(\alpha)\epsilon_j \Big) b_{i\alpha}$$

$$U_3(\alpha) = \sum_{i\alpha} \Big(\sum_{j=1}^{n} \kappa_{ij}(\theta)\epsilon_j \Big)^2 = U_{31}(\alpha) + U_{32}(\alpha) + U_{33},$$

with $b_{i\alpha} = L_{q,\alpha}(X_i) I\big(|K_j(\alpha)| > a_0\big)$ and

$$U_{31}(\alpha) = \frac{1}{L_\alpha} \sum_{i=1}^{n} \Big(\sum_{j=1}^{n} \kappa_{ij}(\theta)\epsilon_j \Big)^2 (b_{i\alpha} - b_{i\theta}),$$

$$U_{32}(\alpha) = \Big(\frac{1}{L_\alpha} - \frac{1}{L_\theta} \Big) \sum_{i=1}^{n} \Big(\sum_{j=1}^{n} \kappa_{ij}(\theta)\epsilon_j \Big)^2 b_{i\theta},$$

$$U_{33} = \frac{1}{L_\theta} \sum_{i=1}^{n} \Big(\sum_{j=1}^{n} \kappa_{ij}(\theta)\epsilon_j \Big)^2 b_{i\theta}.$$

Since $U_{33} = R_3(\theta)$,

$$|R_3(\alpha) - R_3(\theta)| \le |U_1(\alpha)| + |U_2(\alpha)| + |U_{31}(\alpha)| + |U_{32}(\alpha)|. \tag{4.12}$$

Note that by Lemma 4.6(iii) and Lemma 4.9(x),

$$\begin{aligned} \big|b_{i\alpha} - b_{i\theta}\big| &\le \big|L_{q,\alpha}(X_i) - L_{q,\theta}(X_i)\big| + \big|I(|K_i(\alpha)| < a_0) - I(|K_i(\alpha)| < a_0)\big| L_{q,\theta}(X_i) \\ &\le O(\delta) + o_p(n^{-1}) = O_p(\delta). \end{aligned} \tag{4.13}$$

Hence, by Lemma 4.9(iii) and Lemma 4.10(iv),

$$\sup_{\|\alpha-\theta\|<\delta} U_1(\alpha) \le O_p(n^{-1}) \cdot o_p(h^{-3}\delta^2 n^{\xi}) = o_p(n^{-1+\xi}h^{-3}\delta^2);$$

$$\sup_{\|\alpha-\theta\|<\delta} U_2(\alpha) \le O_p(n^{-1}) \cdot \sqrt{o_p(h^{-3}\delta^2 n^{\xi}) \cdot o_p(n^{\xi}h^{-1})} = o_p(n^{-1+\xi}h^{-2}\delta)$$

$$\sup_{\|\alpha-\theta\|<\delta} U_{31}(\alpha) \le O_p(n^{-1}) \cdot o_p(n^{\xi}h^{-1}) \cdot O(\delta) = o_p(n^{-1+\xi}h^{-1}\delta);$$

$$\sup_{\|\alpha-\theta\|<\delta} U_{32}(\alpha) \le O_p(n^{-1}\delta) \cdot o_p(n^{\xi}h^{-1}) = o_p(n^{-1+\xi}h^{-1}\delta).$$

The result now follows from (4.12).

(R4) The result follows from (4.11) and Lemma 4.5.

(R5) The result follows from Lemma 4.10(iv) and Lemma 4.5.

(R6) Note that by Lemma 4.9(iii), $R_6(\alpha) = O_p(n^{-1})\tilde{R}_6(\alpha)$, where

$$\tilde{R}_6(\alpha) = \sum_{i=1}^{n}\sum_{j=1}^{n} \kappa_{ij}(\alpha) L_{q,\alpha}(X_i)\epsilon_i\epsilon_j.$$

By Cauchy-Schwartz inequality, Lemma 4.9(ii) and Lemma 4.10(ii) we have

$$\begin{aligned}
\sup_{\alpha} |\tilde{R}_6(\alpha)|^2 &\le \sup_{\alpha} \sum_{i=1}^{n} \epsilon_i^2 \cdot \sup_{\alpha} \sum_{i=1}^{n} \Big(\sum_{j=1}^{n} \kappa_{ij}\epsilon_j\Big)^2 L_{q,\alpha}^2(X_i) \\
&\le O(n) \cdot \sup_{\alpha} \frac{L_{q,\alpha}^2(X_i)}{K_i(\alpha)^2} \cdot \sup_{\alpha} \sum_{i=1}^{n} \Big(\sum_{j=1}^{n} K_h(\alpha' X_j - \alpha' X_i)\epsilon_j\Big)^2 \\
&= O(n) \cdot O_p\Big(\frac{1}{(nh)^2}\Big) \cdot o_p(n^{2+\xi}h) = o_p(n^{1+\xi}h^{-1}).
\end{aligned}$$

Hence

$$\sup_{\alpha} |R_6(\alpha)| = O_p(n^{-1}) \cdot o_p(n^{\frac{1}{2}+\xi}h^{-1}) = o_p(n^{-\frac{1}{2}+\xi}h^{-1}).$$

For the second result, note that

$$R_6(\alpha) - R_6(\theta) = \frac{1}{L_\alpha}[V_1(\alpha) + V_2(\alpha)] + \frac{L_\theta - L_\alpha}{L_\theta L_\alpha} V_3(\theta), \tag{4.14}$$

where $V_3(\theta) = \sum_{i,j} \kappa_{ij}(\theta) b_{i\theta}\epsilon_i\epsilon_j$ and

$$\begin{aligned}
V_1(\alpha) &= \sum_{i,j} [\kappa_{ij}(\alpha) - \kappa_{ij}(\theta)] b_{i\alpha}\epsilon_i\epsilon_j, \\
V_2(\alpha) &= \sum_{i,j} \kappa_{ij}(\theta)[b_{i\alpha} - b_{i\theta}]\epsilon_i\epsilon_j.
\end{aligned}$$

By Lemma 4.11,

$$E\Big(\sum_{i,j} \kappa_{ij}(\theta) b_{i\theta}\epsilon_i\epsilon_j\Big)^2 = O(h^{-1}). \tag{4.15}$$

Hence

$$V_3(\theta) = O_p(h^{-\frac{1}{2}}). \tag{4.16}$$

For $V_1(\alpha)$ and $V_2(\alpha)$ we apply the discretization technique again. Let the discrete set of $\alpha \in B(\theta, \delta)$ be denoted by $\mathcal{A}$ with size n^c (hence each cell has diameter of order $n^{-c/p}\delta$). Note that

$$\begin{aligned}
&V_1(\alpha) + V_2(\alpha) - V_1(\beta) - V_2(\beta) \\
&= \sum_{i=1}^{n} \epsilon_i b_{i\alpha} \sum_{j=1}^{n} [\kappa_{ij}(\alpha) - \kappa_{ij}(\beta)]\epsilon_j + \sum_{i=1}^{n} \epsilon_i [b_{i\alpha} - b_{i\beta}] \sum_{j=1}^{n} \kappa_{ij}(\beta)\epsilon_j.
\end{aligned}$$

By Cauchy-Schwartz inequality and Lemma 4.10(ii,iii),

$$\begin{aligned}
&\sup_{\|\alpha-\beta\|\le N^{-\frac{c}{p}}\delta}\Big|\sum_{i=1}^{n}\epsilon_i b_{i\alpha}\sum_{j=1}^{n}[\kappa_{ij}(\alpha)-\kappa_{ij}(\beta)]\epsilon_j\Big| \\
&\qquad\le \sup_{\|\alpha-\beta\|\le N^{-\frac{c}{p}}\delta}\Big(\sum_{i=1}^{n}\epsilon_i^2\Big)^{\frac{1}{2}}\cdot\Big(\sum_{i=1}^{n}\Big[\sum_{j=1}^{n}[\kappa_{ij}(\alpha)-\kappa_{ij}(\beta)]\epsilon_j\Big]^2 b_{i\alpha}^2\Big)^{\frac{1}{2}} \\
&\qquad= O_p(n^{\frac{1}{2}})\cdot o_p\big(h^{-\frac{3}{2}}n^{-\frac{c}{p}+\xi}\delta\big);
\end{aligned}$$

and, from (4.13),

$$\begin{aligned}
&\sup_{\|\alpha-\beta\|\le n^{-\frac{c}{p}}\delta}\Big|\sum_{i=1}^{n}\epsilon_i[b_{i\alpha}-b_{i\beta}]\sum_{j=1}^{n}\kappa_{ij}(\beta)\epsilon_j\Big| \\
&\qquad\le \sup_{\|\alpha-\beta\|\le n^{-\frac{c}{p}}\delta}\Big(\sum_{i=1}^{n}\epsilon_i^2\Big)^{\frac{1}{2}}\cdot\Big(\sum_{i=1}^{n}\Big[\sum_{j=1}^{n}\kappa_{ij}(\beta)\epsilon_j\Big]^2(b_{i\alpha}-b_{i\beta})^2\Big)^{\frac{1}{2}} \\
&\qquad= O_p(n^{\frac{1}{2}})\cdot o_p\big(h^{-\frac{1}{2}}n^{\xi}\big)\cdot O_p\big(n^{-\frac{c}{p}}\delta\big);
\end{aligned}$$

Hence,

$$\begin{aligned}
&\sup_{\|\alpha-\theta\|\le\delta}\big|V_1(\alpha)+V_2(\alpha)\big|-\sup_{\alpha\in\mathcal{A}}\big|V_1(\alpha)+V_2(\alpha)\big| \\
&\qquad\le \sup_{\|\alpha-\beta\|\le n^{-\frac{c}{p}}\delta}\big|V_1(\alpha)+V_2(\alpha)-V_1(\beta)-V_2(\beta)\big| = o_p\big(h^{-\frac{3}{2}}n^{\frac{1}{2}-\frac{c}{p}+\xi}\delta\big). \qquad (4.17)
\end{aligned}$$

By Lemma 4.11 we have, when $\delta=o(1)$,

$$\begin{aligned}
E[V_1(\alpha)]^u &= E\Big(\sum_{i,j}[\kappa_{ij}(\alpha)-\kappa_{ij}(\theta)]b_{i\alpha}\epsilon_i\epsilon_j\Big)^u \\
&= \sum_{i_1j_1\cdots i_uj_u}E\big([\kappa_{i_1j_1}(\alpha)-\kappa_{i_1j_1}(\theta)]\cdots[\kappa_{i_uj_u}(\alpha)-\kappa_{i_uj_u}(\theta)]b_{i_1\alpha}\cdots b_{i_u\alpha}\big) \\
&\qquad\qquad\cdot E(\epsilon_{i_1}\epsilon_{j_1}\cdots\epsilon_{i_u}\epsilon_{j_u}) \\
&= O\Big(n^u\frac{\delta^u}{(nh^2)^u}\Big).
\end{aligned}$$

Hence, for all $u\in\mathbb{N}$,

$$\sup_{\alpha\in\mathcal{A}}E[V_1(\alpha)]^u = O(\delta^u h^{-2u}).$$

Similarly we can show $\sup_{\alpha\in\mathcal{A}}E[V_2(\alpha)]^u = O(\delta^u h^{-u})$ for all $u\in\mathbb{N}$. Thus, for $k=1,2$ and all even integer u,

$$P\big(\sup_{\alpha\in\mathcal{A}}|V_k(\alpha)|>a_n\big)\le n^c\frac{\sup_{\alpha\in\mathcal{A}}E[V_k(\alpha)]^u}{a_n^u}\le O\Big(\frac{1}{a_n}n^{\frac{c}{u}}h^{-2}\delta\Big)^u,$$

which converges to zero provided $a_n=n^{\frac{c}{u}+\xi}h^{-2}\delta$ for any $\xi>0$. Therefore we have

$$\sup_{\alpha\in\mathcal{A}}|V_k(\alpha)| = o_p(n^{\frac{c}{u}+\xi}h^{-1}\delta), \quad k=1,2.$$

Thus, from (4.14), (4.16) and (4.17) above,

$$\begin{aligned}\sup_{\|\alpha-\theta\|\le\delta}|R_6(\alpha)-R_6(\theta)| &= \sup_{\|\alpha-\theta\|\le\delta}\Big|\frac{1}{L_\alpha}[V_1(\alpha)+V_2(\alpha)]+\frac{L_\theta-L_\alpha}{L_\theta L_\alpha}V_3(\theta)\Big| \\ &\le O(n^{-1})\sup_{\|\alpha-\theta\|\le\delta}|V_1(\alpha)+V_2(\alpha)|+O_p(n^{-1}h^{-\frac{1}{2}}\delta) \\ &\le o_p\big(h^{-\frac{3}{2}}n^{-\frac{1}{2}-\frac{c}{p}+\xi}\delta\big)+o_p(n^{-1+\frac{c}{u}+\xi}h^{-2}\delta)+O_p(n^{-1}h^{-\frac{1}{2}}\delta) \\ &= o_p\big(\delta n^{-\frac{1}{2}}\big),\end{aligned}$$

provided

$$\frac{3}{2}a-\frac{1}{2}-\frac{c}{p}<-\frac{1}{2};\quad -1+\frac{c}{u}+2a<-\frac{1}{2};\quad -1+\frac{1}{2}a<-\frac{1}{2}.$$

This can be achieved by taking $c>\frac{3}{2}ap$ and $u>\frac{2c}{1-4a}>\frac{3ap}{1-4a}$. Note that the moment method for $V_k(\alpha)$ terms requires $v\ge 2u$.

(R7) This is a direct consequence of Lemma 4.10(iv) and (4.11).

(R8) It follows from (4.11) and Lemma 4.10(v).

(R9) The result follows from Lemma 4.10(iv) and Lemma 4.5.

□

Lemma 4.13. *Take h to be proportional to n^{-a}. Under the conditions of Theorem 2.3 we have, for all $\xi>0$, $\|\hat\theta-\theta\|=o_p\big(n^{-2a+\xi}\big)$ and $|R_0(\hat\theta)-R_0(\theta)|+\sum_{i=1}^{9}R_i(\hat\theta)=o_p(n^{-4a+\xi})$.*

Proof. Since $\hat d(\hat\theta)$ minimizes $\hat d(\alpha)$, we have $\hat d(\hat\theta)\le\hat d(\theta)$, i.e. $\sum_{i=0}^{9}R_i(\hat\theta)\le\sum_{i=0}^{9}R_i(\theta)$. Since $R_1(\theta)=0$ and $h=O(n^{-a})$ with $\frac{1}{8}<a\le\frac{1}{5}$, by Lemma 4.12,

$$R_1(\hat\theta)\le\sum_{i\ne1}\big|R_i(\hat\theta)-R_i(\theta)\big|\le\sum_{i\ne1}\sup_{\|\alpha-\theta\|\le\delta}\big|R_i(\alpha)-R_i(\theta)\big|=o_p\big(n^{-2a+\xi}\big),\tag{4.18}$$

where we currently take $\delta=O(1)$. The upper bound of (4.18) can be improved to

$$R_1(\hat\theta)\le o_p\big(n^{-4a+\xi}+\delta n^{-2a+\xi}\big),\tag{4.19}$$

if we take $\delta=o(1)$. We now show $\hat\theta\xrightarrow{P}\theta$. Otherwise there exists a constant $t>0$ such that $P(A)>0$ where

$$A=\{\|\hat\theta-\theta\|>t\}.$$

By the first result of Lemma 4.12(R1),

$$\begin{aligned}P\big(R_1(\hat\theta_1)>c_t\big)&\ge P\big(\{R_1(\hat\theta_1)>c_t\}\cap A\big)\\ &\ge P\big(\{\inf_{\|\alpha-\theta\|>t}R_1(\alpha)>c_t\}\cap A\big)\ge P(A)+o(1),\end{aligned}$$

where $c_t>0$ is as specified in Lemma 4.12(R1). This contradicts (4.18). Hence $\hat\theta\xrightarrow{P}\theta$. By (4.18) and the second result of Lemma 4.12(R1) we have $\|\hat\theta-\theta\|=o_p\big(n^{-a+\xi}\big)$. Now, fix any $0 < \xi_0 < a$. Let $\delta_1 = n^{-a+\xi_0}$ and let $B_n = \{\|\hat{\theta} - \theta\| \leq \delta_1\}$. Then $P(B_n) \to 1$ and, by (4.19) with $\delta = \delta_1$, for all $\gamma > 0$ and $\xi > \xi_0$,

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

Hence $\|\hat{\theta} - \theta\| = o_p\big(n^{-\frac{3a}{2}+\xi}\big)$ for all $\xi > 0$. Repeating this procedure we have $\|\hat{\theta} - \theta\| = o_p\big(n^{-2a+\xi}\big)$, for all $\xi > 0$. Similarly we can show $\tilde{R}_i(\hat{\theta}) = o_p(n^{-4a+\xi})$, where $\tilde{R}_i(\hat{\theta})$ denotes $|R_0(\hat{\theta}) - R_0(\theta)|$ for $i = 0$ and denotes $R_i(\hat{\theta})$ for $i = 1, 2, \cdots, 9$. This completes the proof. □

4.4 Decomposition of $b(\alpha)$

Now we assume $\mathrm{Var}(\epsilon|X) = \sigma^2$. Let $K_\alpha = \sum_{i=1}^n \tilde{K}_i(\alpha)$, where

$$\tilde{K}_i(\alpha) = \sum_{j\neq i} K_h(\alpha' X_j - \alpha' X_i), \quad 1 \leq i \leq n.$$

Using the notation $\sum_{ij\alpha} c_{ij} = \sum_{i=1}^n \sum_{j\neq i} c_{ij} K_h(\alpha' X_i - \alpha' X_j)$ for any sequence $\{c_i\}$, the numerator of $b(\alpha)$ can be written as

$$\begin{aligned} b(\alpha) &= \frac{1}{2K_\alpha} \sum_{ij\alpha} (Y_i - Y_j)^2 \\ &= \frac{1}{2K_\alpha} \sum_{ij\alpha} \Big([\epsilon_i - \epsilon_j] + [m_\alpha(x_i) - m_\alpha(x_j)] + [g_\alpha(\alpha' X_i) - g_\alpha(\alpha' X_j)]\Big)^2 \\ &= \frac{1}{2K_\alpha} \sum_{i=0}^{6} Q_i(\alpha), \end{aligned}$$

where $Q_0(\alpha) = 2\sum_{i=1}^n \epsilon_i^2 \tilde{K}_i(\alpha)$ and

$$\begin{aligned} Q_1(\alpha) &= \sum_{ij\alpha} [m_\alpha(X_i) - m_\alpha(X_j)]^2; \\ Q_2(\alpha) &= \sum_{ij\alpha} [g_\alpha(\alpha' X_i) - g_\alpha(\alpha' X_j)]^2; \\ Q_3(\alpha) &= 2\sum_{ij\alpha} [m_\alpha(X_i) - m_\alpha(X_j)](\epsilon_i - \epsilon_j); \\ Q_4(\alpha) &= 2\sum_{ij\alpha} [g_\alpha(\alpha' X_i) - g_\alpha(\alpha' X_j)](\epsilon_i - \epsilon_j); \\ Q_5(\alpha) &= 2\sum_{ij\alpha} [g_\alpha(\alpha' X_i) - g_\alpha(\alpha' X_j)][m_\alpha(X_i) - m_\alpha(X_j)]; \\ Q_6(\alpha) &= -2\sum_{ij\alpha} \epsilon_i \epsilon_j. \end{aligned}$$

Lemma 4.14. *Let Assumptions (A2) and (A4) hold. Then, for random designs,*

(i) $|K_\alpha - K_\theta| = o_p(n^{2+\xi} h\delta) + o_p(n^{2+\xi}\delta^2)$*;*

(ii) $\sup_{\|\alpha-\theta\|<\delta}\big|\sum_{i=1}^n [K_i(\alpha) - K_i(\theta)]\epsilon_i\big| = o_p\big(n^{2+\xi}h^2\delta\big) + o_p\big(n^{2+\xi}h^{\frac{3}{2}}\delta^{\frac{3}{2}}\big)$*, provided that* ϵ *has up to* u *moments where* $u \geq 4$ *is an even integer and* $u \geq \frac{4a-1}{1-3a}p$*;*

(iii) $\sup_{\alpha\in B(\theta,\delta)} |Q_6(\alpha) - Q_6(\theta)| \leq o_p(n^{2+\xi}h^2\delta)$*, provided that* ϵ *has up to* u *moments where* $u \geq 6$ *is an even integer and* $u \geq \frac{5a-1}{2(1-3a)}p$*.*

The results hold for fixed designs with $o_p(\cdot)$ *replaced by* $o(\cdot)$ *in (i).*

Proof. The proofs for two designs are almost identical.

(i) Let $\gamma = n^\xi\delta + h$. By Lemma 4.9(viii) we have

$$\begin{aligned}|K_\alpha - K_\theta| &\leq \sum_{|\alpha' X_i - c_\alpha|\leq w_\alpha-\gamma} |\tilde{K}_i(\alpha) - \tilde{K}_i(\theta)| + \sum_{|\alpha' X_i - c_\alpha|> w_\alpha-\gamma} |\tilde{K}_i(\alpha) - \tilde{K}_i(\theta)| \\ &\leq o_p(n^{2+\xi}h\delta) + o_p(n^{1+\xi}\delta)\sum_{i=1}^n I\big(|\alpha' X_i - c_\alpha| > w_\alpha - \gamma\big).\end{aligned}$$

Similar to the proof for Lemma 4.9(viii) we can show

$$\sum_{i=1}^n I\big(|\alpha' X_i - c_\alpha| > w_\alpha - \gamma\big) = o_p(n^{1+\xi}\gamma).$$

Hence

$$|K_\alpha - K_\theta| \leq o_p(n^{2+\xi}h\delta) + o_p(n^{2+\xi}\delta^2).$$

(ii) Apply the discretization technique. Let the discrete set of $\alpha \in B(\theta,\delta)$ be denoted by $\mathcal{A}$ with size n^c (hence each cell has diameter of order $n^{-c/p}\delta$). Let

$$A(\alpha) = \Big|\sum_{i=1}^n [K_i(\alpha) - K_i(\theta)]\epsilon_i\Big|.$$

By Lemma 4.9(viii) and Lemma 4.5,

$$\begin{aligned}\sup_{\alpha\in B(\theta,\delta)} A(\alpha) - \sup_{\alpha\in\mathcal{A}} A(\alpha) &\leq \sup_{\|\alpha-\beta\|<n^{-\frac{c}{p}}\delta} \big|A(\alpha) - A(\beta)\big| \\ &\leq \sup_{\|\alpha-\beta\|<n^{-\frac{c}{p}}\delta} \Big|\sum_{i=1}^n [K_i(\alpha) - K_i(\beta)]\epsilon_i\Big| \\ &\leq o_p\big(n^{\frac{3}{2}-\frac{c}{p}+\xi}\delta\big).\end{aligned}$$

Note that

$$\begin{aligned}\sup_{\alpha\in\mathcal{A}} A(\alpha) \leq &\sup_{\alpha\in\mathcal{A}}\Big|\sum_{i=1}^n [K_i(\alpha) - K_i(\theta)] I(|\alpha' X_i - c_\alpha| \leq w_\alpha - \lambda)\epsilon_i\Big| \\ &+ \sup_{\alpha\in\mathcal{A}}\Big|\sum_{i=1}^n [K_i(\alpha) - K_i(\theta)] I(|\alpha' X_i - c_\alpha| > w_\alpha - \lambda)\epsilon_i\Big|.\end{aligned}$$

Take $\lambda = n^{\xi}\delta + h$ for any $\xi > 0$. By Lemma 4.9(viii)

$$\sup_{i,\alpha}\Big|[K_i(\alpha) - K_i(\theta)]I(|\alpha' X_i - c_\alpha| \le w_\alpha - \lambda)\Big| = o_p(n^{1+\xi}h\delta).$$

Hence, by Lemma 4.5, provided $u \ge 4$,

$$\sup_{\alpha\in\mathcal{A}}\Big|\sum_{i=1}^{n}[K_i(\alpha) - K_i(\theta)]I(|\alpha' X_i - c_\alpha| \le w_\alpha - \lambda)\epsilon_i\Big| = o_p\big(n^{\frac{3}{2}+\xi}h\delta\big).$$

Let

$$B(\alpha) = \sum_{i=1}^{n}[K_i(\alpha) - K_i(\theta)]I(|\alpha' X_i - c_\alpha| > w_\alpha - \lambda)\epsilon_i.$$

From (4.10), for any even integer u, uniformly in $\alpha \in B(\theta, \delta)$,

$$E\big[B(\alpha)\big]^u = O\big(n^{\frac{u}{2}} \cdot (n\delta)^u \cdot \lambda^{\frac{u}{2}}\big) = O\big(n^{\frac{3}{2}}\delta\lambda^{\frac{1}{2}}\big)^u.$$

Thus,

$$P\big(\sup_{\alpha\in\mathcal{A}}|B(\alpha)| > a_n\big) \le n^c\frac{E\big[B(\alpha)\big]^u}{a_n^u} = \Big(\frac{n^{\frac{3}{2}+\frac{c}{u}}\delta\lambda^{\frac{1}{2}}}{a_n}\Big)^u,$$

which goes to zero if $a_n = n^{\frac{3}{2}+\frac{c}{u}+\xi}\delta\lambda^{\frac{1}{2}}$. This shows,

$$\sup_{\alpha\in\mathcal{A}}|B(\alpha)| = o_p\big(n^{\frac{3}{2}+\frac{c}{u}+\xi}\delta\lambda^{\frac{1}{2}}\big).$$

Therefore,

$$\begin{aligned}\sup_{\alpha\in B(\theta,\delta)} A(\alpha) &\le o_p\big(n^{\frac{3}{2}-\frac{c}{p}+\xi}\delta\big) + o_p\big(n^{\frac{3}{2}+\xi}h\delta\big) + o_p\big(n^{\frac{3}{2}+\frac{c}{u}+\xi}\delta\lambda^{\frac{1}{2}}\big)\\ &\le o_p\big(n^{\frac{3}{2}-\frac{c}{p}+\xi}\delta\big) + o_p\big(n^{\frac{3}{2}+\xi}h\delta\big) + o_p\big(n^{\frac{3}{2}+\frac{c}{u}+\xi}\delta^{\frac{3}{2}}\big) + o_p\big(n^{\frac{3}{2}+\frac{c}{u}+\xi}\delta h^{\frac{1}{2}}\big).\end{aligned}$$

Take c and u such that

$$\frac{3}{2} - \frac{c}{p} \le 2 - 2a, \qquad \frac{3}{2} + \frac{c}{u} \le 2 - \frac{3}{2}a \quad \text{and} \quad \frac{3}{2} + \frac{c}{u} - \frac{1}{2}a \le 2 - 2a;$$

i.e., $c \ge (2a - \frac{1}{2})p$ and $u \ge \frac{2c}{1-3a} \ge \frac{4a-1}{1-3a}p$. Then

$$\sup_{\alpha\in B(\theta,\delta)} A(\alpha) \le o_p\big(n^{2+\xi}h^2\delta\big) + o_p\big(n^{2+\xi}h^{\frac{3}{2}}\delta^{\frac{3}{2}}\big).$$

(iii) Apply the discretization technique. Let the discrete set of $\alpha \in B(\theta, \delta)$ be denoted by $\mathcal{A}$ with size n^c (hence each cell has diameter of order $n^{-c/p}\delta$). Let $A_{ij}(\alpha) = K_h(\alpha' X_i - \alpha' X_j)$. By Lemma 4.10(i) we have

$$\begin{aligned}&\sup_{\|\alpha-\beta\|<n^{-\frac{c}{p}}\delta}|Q_6(\alpha) - Q_6(\beta)| = 2\sup_{\|\alpha-\beta\|<n^{-\frac{c}{p}}\delta}\Big|\sum_{i=1}^{n}\Big(\sum_{j\ne i}[A_{ij}(\alpha) - A_{ij}(\beta)]\epsilon_j\Big)\epsilon_i\Big|\\ &\le 2\Big(\sum_{i=1}^{n}\epsilon_i^2\Big)^{\frac{1}{2}}\Big(\sup_{\|\alpha-\beta\|<n^{-\frac{c}{p}}\delta}\sum_{i=1}^{n}\Big(\sum_{j=1}^{n}[A_{ij}(\alpha) - A_{ij}(\beta)]\epsilon_j\Big)^2\Big)^{\frac{1}{2}}\\ &\le O_p(n^{\frac{1}{2}}) \cdot o_p\big(n^{1-\frac{c}{p}+\xi}h^{-\frac{1}{2}}\delta\big) = o_p\big(n^{\frac{3}{2}-\frac{c}{p}+\xi}h^{-\frac{1}{2}}\delta\big).\end{aligned}$$

Now, for any even integer u,

$$
\begin{aligned}
P\big(\sup_{\alpha\in\mathcal{A}}|Q_6(\alpha)-Q_6(\theta)|>a_n\big) &\le n^c\cdot\frac{E[Q_6(\alpha)-Q_6(\theta)]^u}{a_n^u}\\
&\le n^c\cdot O_p\Big(\frac{n^u(h^{-1}\delta)^u}{a_n^u}\Big)\\
&= O_p\Big(\frac{n^{1+\frac{c}{u}}h^{-1}\delta}{a_n}\Big)^u,
\end{aligned}
$$

which goes to zero if $a_n=n^{1+\frac{c}{u}+\xi}h^{-1}\delta$ for any $\xi>0$. Hence,

$$
\sup_{\alpha\in B(\theta,\delta)}|Q_6(\alpha)-Q_6(\theta)|\le o_p\big(n^{\frac{3}{2}-\frac{c}{p}+\xi}h^{-\frac{1}{2}}\delta\big)+o_p\big(n^{1+\frac{c}{u}+\xi}h^{-1}\delta\big).
$$

Take c and u such that

$$
\frac{3}{2}-\frac{c}{p}+\frac{1}{2}a\le 2-2a \quad\text{and}\quad 1+\frac{c}{u}+a\le 2-2a;
$$

i.e., $c\ge(\frac{5}{2}a-\frac{1}{2})p$ and $u\ge\frac{c}{1-3a}\ge\frac{5a-1}{2(1-3a)}p$. Then,

$$
\sup_{\alpha\in B(\theta,\delta)}|Q_6(\alpha)-Q_6(\theta)|\le o_p(n^{2+\xi}h^2\delta).
$$

□

Lemma 4.15. *Let all the sup below be taken over the set $B(\theta,\delta)=\{\alpha\mid\|\alpha-\theta\|\le\delta\}$ for arbitrarily fixed $\delta>0$. Suppose $h\ge O(n^{-\frac{1}{3}})$ and $v\ge 6$. Under Assumptions (A1)-(A5) we have, for random designs and for all $\xi>0$,*

(i) *provided that $v\ge\max(2u,8)$ where u is an even integer and $u\ge\frac{4a-1}{1-3a}p$,*

$$
\sup|K_\alpha^{-1}Q_0(\alpha)-K_\theta^{-1}Q_0(\theta)|=o_p(n^{-\frac{1}{2}+\xi}h^{-1}\delta^2)+o_p(n^\xi h\delta)+o_p(n^{-\frac{1}{2}+\xi}h^{-\frac{1}{2}}\delta^{\frac{3}{2}})
$$

(ii) *for any fixed constant $t>0$, there exists a constant $e_t>0$ such that*

$$
\inf_{\|\alpha-\theta\|>t}Q_1(\alpha)>e_tn^2h+o_p(n^2h);
$$

moreover, there exists a constant $c_2>0$ such that

$$
c_2n^2h\|\alpha-\theta\|^2+O_p(n^2h\|\alpha-\theta\|^3)\le Q_1(\alpha)\le O_p(n^2h\|\alpha-\theta\|^3);
$$

(iii) $\sup|Q_2(\alpha)|=O_p(n^2h^3)$*;* $\sup|Q_3(\alpha)|=o_p(n^{\frac{3}{2}+\xi}h\delta)$*;* $\sup|Q_4(\alpha)|=o_p(n^{\frac{3}{2}+\xi}h^2)$*; and* $\sup|Q_5(\alpha)|=O_p(n^2h^2\delta)$*;*

(iv) *provided that $v\ge u$ where $u\ge 6$ is the smallest even integer greater than $\frac{5a-1}{2(1-3a)}p$,*

$$
\sup|K_\alpha^{-1}Q_6(\alpha)-K_\theta^{-1}Q_6(\theta)|=o_p(n^\xi h\delta)+o_p\big(n^{-1+\xi}h^{-\frac{3}{2}}\delta^2\big).
$$

The results hold for fixed designs with $O_p(\cdot)$ replaced by $O(\cdot)$ and $o_p(\cdot)$ replaced by $o(\cdot)$ for terms Q_1, Q_2 and Q_5.

Proof. The proofs for two designs are almost identical. We use notation $Z_i(\alpha)$ for different random variables in different parts of the proof. And we apply Lemma 4.9(ii) in all cases to handle the denominators.

(i) Let $\lambda_i = \epsilon_i^2 - \sigma^2$. By Lemma 4.9(vii-viii) and Lemma 4.14 we have

$$\begin{aligned}
&\sup_\alpha \Big| \frac{Q_0(\alpha)}{K_\alpha} - \frac{Q_0(\theta)}{K_\theta} \Big| \\
&\quad \le \sup_\alpha \Big| \frac{K_\theta - K_\alpha}{K_\theta K_\alpha} \sum_{i=1}^n \tilde{K}_i(\alpha)\lambda_i \Big| + \sup_\alpha \Big| \frac{1}{K_\theta} \sum_{i=1}^n [\tilde{K}_i(\alpha) - \tilde{K}_i(\theta)]\lambda_i \Big| \\
&\quad \le o_p\Big(\frac{n^{2+\xi}\delta(h+\delta)}{(n^2h)^2} \Big) \sup_\alpha \Big| \sum_{i=1}^n \tilde{K}_i(\alpha)\lambda_i \Big| + o_p\big(\frac{1}{n^2h}\big) \sup_\alpha \Big| \sum_{i=1}^n [\tilde{K}_i(\alpha) - \tilde{K}_i(\theta)]\lambda_i \Big|.
\end{aligned}$$

By Lemma 4.9(iv) and Lemma 4.5 we have, provided $v \ge 8$,

$$\sup_\alpha \Big| \sum_{i=1}^n \tilde{K}_i(\alpha)\lambda_i \Big| \le o_p(n^{\frac{3}{2}+\xi}h);$$

and by Lemma 4.14(ii) we have

$$\sup_\alpha \Big| \sum_{i=1}^n [\tilde{K}_i(\alpha) - \tilde{K}_i(\theta)]\lambda_i \Big| \le o_p\big(n^{2+\xi}h^2\delta\big) + o_p\big(n^{\frac{3}{2}+\xi}h^{\frac{1}{2}}\delta^{\frac{3}{2}}\big).$$

Thus,

$$\sup_\alpha \Big| \frac{Q_0(\alpha)}{K_\alpha} - \frac{Q_0(\theta)}{K_\theta} \Big| \le o_p(n^{-\frac{1}{2}+\xi}h^{-1}\delta^2) + o_p(n^\xi h\delta) + o_p(n^{-\frac{1}{2}+\xi}h^{-\frac{1}{2}}\delta^{\frac{3}{2}}).$$

(ii) A method similar to Lemma 4.12(R1) gives the first result. For the second result, write $Q_1(\alpha) = 2Q_{11}(\alpha) - 2Q_{12}(\alpha)$ where,

$$\begin{aligned}
Q_{11}(\alpha) &= \sum_{i\ne j} m_\alpha^2(X_i)K_h(\alpha' X_i - \alpha' X_j) = \sum_{i=1}^n m_\alpha^2(X_i)K_i(\alpha); \\
Q_{12}(\alpha) &= \sum_{i\ne j} m_\alpha(X_i)m_\alpha(X_j)K_h(\alpha' X_i - \alpha' X_j).
\end{aligned}$$

By Lemma 4.9(i,viii) we have

$$\begin{aligned}
Q_{11}(\alpha) &= \sum_{i=1}^n m_\alpha^2(X_i)K_i(\alpha) = \sum_{i=1}^n m_\alpha^2(X_i)K_i(\theta) + o_p(n^{2+\xi}h\|\alpha-\theta\|^3) \\
&= nh \sum_{i=1}^n m_\alpha^2(X_i)f_\theta(\theta' X_i) \int_{-1}^1 K(s)ds + o_p(n^{2+\xi}h\|\alpha-\theta\|^3) \\
&= nh(\alpha-\theta)' \Big[\sum_{i=1}^n \mu(X_i)\mu(X_i)' f_\theta(\theta' X_i) \int_{-1}^1 K(s)ds \Big] (\alpha-\theta) \\
&\qquad + o_p(n^{2+\xi}h\|\alpha-\theta\|^3),
\end{aligned}$$

where $\mu(x) = \big(x - E(X|\theta'X = \theta'x)\big)g^{(1)}(\theta'x)$ as defined in (4.11). Similar to the proof for Lemma 4.12(R1), we can find a positive constant c_2 such that

$$c_2 n^2 h\|\alpha - \theta\|^2 + o_p(n^{2+\xi}h\|\alpha - \theta\|^3) \le Q_{11}(\alpha) \le O_p(n^2 h\|\alpha - \theta\|^2).$$

By (4.11), uniformly in α,

$$Q_{12}(\alpha) = (\alpha - \theta)'\Big(\sum_{i=1} \mu(X_i)A_i(\theta)'\Big)(\alpha - \theta) + O_p(n^2 h\|\alpha - \theta\|^3),$$

where $A_i(\theta) = \sum_{j\neq i} Z_{ij}(\theta)$ with $Z_{ij}(\theta) = \mu(X_j)K_h(\theta'X_i - \theta'X_j)$. Observing that $E[Z_{ij}(\theta)|X_i] = 0$ and $E[Z_{ij}^2(\theta)|X_i] = O(h)$ uniformly in i, j, by Lemma 4.3 we have $\max_{1\le i\le n}|A_i(\theta)| = o_p(\sqrt{nh}n^{\xi})$. Hence $\sup|Q_{12}(\alpha)| \le o_p(n^{\frac{3}{2}+\xi}h^{\frac{1}{2}}\delta^2)$ and the result follows. Note that the analogous result of $A_i(\theta)$ in fixed designs follows by observing that

$$\begin{aligned}
&\int_S [x - r(\theta, \theta'x)]g^{(1)}(\theta'x)K_h(\theta'x - t)dx \\
&\quad = \int_{AS} [A'y - r(\theta, y_1)]g^{(1)}(y_1)K_h(y_1 - t)dy \\
&\quad = \int_{-1}^{1}\Big(\int_{S(y_1)} A'y\, dy_2\cdots dy_p - r(\theta, y_1)f_\alpha(y_1)\Big)g^{(1)}(y_1)K_h(y_1 - t)dy_1 \\
&\quad = 0.
\end{aligned}$$

(iii) The results for Q_2 and Q_5 follow from Taylor's expansion directly. Write $Q_3(\alpha) = 2\sum_{i=1}^n Z_i(\alpha)\epsilon_i$ where

$$Z_i(\alpha) = \sum_{j\neq i}[m_\alpha(X_i) - m_\alpha(X_j)]K_h(\alpha'X_i - \alpha'X_j).$$

Since $\sup|Z_i(\alpha)| = O_p(nh\delta)$, by Lemma 4.5 we have $\sup|Q_3(\alpha)| = o_p(n^{\frac{3}{2}+\xi}h\delta)$. Similarly we can write $Q_4(\alpha) = 2\sum_{i=1}^n \tilde{Z}_i(\alpha)\epsilon_i$ where

$$\tilde{Z}_i(\alpha) = \sum_{j\neq i}[g_\alpha(\alpha'X_i) - g_\alpha(\alpha'X_j)]K_h(\alpha'X_i - \alpha'X_j).$$

Since $\sup|\tilde{Z}_i(\alpha)| = O_p(nh^2)$, by Lemma 4.5 we have $\sup|Q_3(\alpha)| = o_p(n^{\frac{3}{2}+\xi}h^2)$.

(iv) Since $EQ_6^2(\theta) = O(n^2 h)$, we have $Q_6(\theta) = O_p(nh^{\frac{1}{2}})$. By Lemma 4.9(vii,viii) and Lemma 4.14(i,iii)

$$\begin{aligned}
\sup\Big|\frac{Q_6(\alpha)}{K_\alpha} - \frac{Q_6(\theta)}{K_\theta}\Big| &\le \sup\Big|\frac{Q_6(\alpha) - Q_6(\theta)}{K_\alpha}\Big| + \sup\Big|\frac{K_\theta - K_\alpha}{K_\theta K_\alpha}Q_6(\theta)\Big| \\
&\le o_p\Big(\frac{n^{2+\xi}h^2\delta}{n^2 h}\Big) + o_p\Big(\frac{n^{2+\xi}h\delta + n^{2+\xi}\delta^2}{(n^2h)^2}nh^{\frac{1}{2}}\Big) \\
&= o_p(n^{\xi}h\delta) + o_p\big(n^{-1+\xi}h^{-\frac{1}{2}}\delta\big) + o_p\big(n^{-1+\xi}h^{-\frac{3}{2}}\delta^2\big) \\
&= o_p(n^{\xi}h\delta) + o_p\big(n^{-1+\xi}h^{-\frac{3}{2}}\delta^2\big).
\end{aligned}$$

□

Lemma 4.16. *Take h proportional to n^{-a} with $\frac{1}{4} < a < \frac{1}{3}$. Take $v \geq \max(2u_1, u_2, 8)$, where u_1, u_2 are the smallest even integers such that*

$$u_1 \geq \frac{4a-1}{1-3a}p, \quad u_2 \geq \frac{5a-1}{2(1-3a)}.$$

Let $\tilde{\theta}$ minimize $b(\alpha)$ over D. Then $\|\tilde{\theta}-\theta\| = o_p(hn^{\xi})$ and $b(\tilde{\theta}) = (2K_\theta)^{-1}Q_0(\theta) + o_p(h^2 n^{\xi})$.

Proof. First note that by Lemma 4.9(iv,vii) we have $\sup_\alpha |K_\alpha^{-1}| \leq O_p(n^{-2}h^{-1})$ and $\inf_\alpha |K_\alpha^{-1}| \geq c_M^{-1} n^{-2}h^{-1} + o_p(n^{-2}h^{-1})$ for some constant $c_M > 0$. Since $b(\tilde{\theta})$ minimizes $b(\alpha)$, we have $b(\tilde{\theta}) \leq b(\theta)$, i.e.

$$\sum_{i=0}^{6} K_{\tilde{\theta}}^{-1} Q_i(\tilde{\theta}) \leq \sum_{i=0}^{6} K_\theta^{-1} Q_i(\theta).$$

Observing $Q_1(\theta) = 0$, by Lemma 4.15 we have

$$\begin{aligned} Q_1(\tilde{\theta}) &\leq K_{\tilde{\theta}} \sum_{i \neq 1} \sup_{\|\alpha-\theta\| \leq \delta} \left| K_\alpha^{-1} Q_i(\alpha) - K_\theta^{-1} Q_i(\theta) \right| \\ &\leq o_p\big([n^{-\frac{1}{2}} h^{-1} \delta^2 + n^{-\frac{1}{2}} h^{-\frac{1}{2}} \delta^{\frac{3}{2}} + h\delta + h^2] n^{\xi} \big), \end{aligned}$$

where we initially take $\delta = \delta_0 = O(1)$. By Lemma 4.15(ii), using a method similar to that of the proof of Lemma 4.13 we get $\|\tilde{\theta} - \theta\| = o_p(hn^{\xi})$ and $|b(\tilde{\theta}) - b(\theta)| \leq o_p(h^2 n^{\xi})$. By Lemma 4.15 and the fact that $EQ_6^2(\theta) = O(n^2 h)$, we have $b(\tilde{\theta}) = (2K_\theta)^{-1} Q_0(\theta) + O_p(h^2)$ and the result follows. □

References

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

[2] Duan, N., and Li, K.C. (1991), "Slicing Regression: A Link Free Regression Method," *The Annals of Statistics,* 19, 505-530.

[3] Fan, J. (1993), "Local Linear Regression Smoothers and their Minimax Efficiencies," *The Annals of Statistics,* 21, 196-216.

[4] Hall, P. (1989), "On Projection Pursuit Regression," *The Annals of Statistics,* 17, 573-588.

[5] Hall, P., Kay, J.W., and Titterington, D.M. (1990), "Asymptotically Optimal Difference-based Estimation of Variance in Nonparametric Regression," *Biometrika,* 77, 521-528.

[6] —— (1991), "On Estimation of Noise Variance in Two-Dimensional Signal Processing," *Advances in Applied Probability,* 23, 476–495.

[7] Hall, P., and Marron, J.S. (1990), "On Variance Estimation in Nonparametric Regression," *Biometrika,* 77, 415-419.

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

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

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

[11] Kulasekera, K.B., and Gallagher, C. (2002), "Variance Estimation in Nonparametric Regression", *Communications in Statistics, Theory and Methods*, 31, 1373-1383.

[12] Lin, W., and Kulasekera, K.B. (2006), "Identifiability of Single-Index Models and Additive-Index Models," *Biometrika,* (to appear).

[13] Munk, A., Bissantz, N., Wagner, T., and Freitag, G. (2005), "On Difference-Based Variance Estimation in Nonparametric Regression when the Covariate is High Dimensional", *Jounal of Royal Statistical Society,* B 67, 19-41.

[14] Niederreiter, H. (1992), Random Number Generation and Quasi-Monte Carlo Methods, Vermont: Capital City Press.

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

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

[17] Xia, Y., Li, W.K., Tong, H., and Zhang, D. (2004), "A Goodness-of-Fit Test for Single-Index Models (with discussion)," ß14, 1-39.

[18] Xia, Y.C., Tong, H., and Li, W.K. (1999), "On Extended Partially Linear Single-Index Models," *Biometrika,* 86, 831-842.

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