INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2006-24

Breaking the Curse of Dimensionality in Nonparametric Testing

P. LAVERGNE[1]
V. PATILEA[2]

The first author gratefully acknowledge support from the NSERC. We thank participants of ESRC 2005 (Bristol), CESG 2005 (Vancouver), ESF exploratory workshop on specification testing 2005 (Santander), and of seminars at the University of Toulouse and UBC for various comments. Special thanks to Jean-Marie Rolin for a very helpful suggestion.

[1] Simon Fraser University, Canada. *Address correspondence to*: Pascal Lavergne, Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby BC, V5A 1S6 Canada. Email : pascal lavergne@sfu.ca

[2] CREST-ENSAI. Email : patilea@ensai.fr

Abstract

For tests based on nonparametric methods, power crucially depends on the dimension of the conditioning variables, and specifically decreases with this dimension. This is known as the “curse of dimensionality." We propose a new general approach to nonparametric testing in high dimensional settings and we show how to implement it when testing for a parametric regression. The resulting test behaves against directional local alternatives almost as if the dimension of the regressors was one. It is also almost optimal against classes of one-dimensional alternatives for a suitable choice of the smoothing parameter. A simulation study shows that it outperforms the standard test by Zheng (1996).

Keywords : Curse of dimensionality, Testing, Nonparametric methods.

JEL classification : Primary C52; Secondary C14.

Résumé

Dans les tests d'adéquation basées sur un lissage non paramétrique, la puissance diminue avec la dimension du vecteur des variables explicatives. Ce phénomène, connu sous le nom du "fléau de la dimension", est dû à l'utilisation d'une estimation non paramétrique d'une fonction multivariée. Nous proposons une nouvelle méthode de test d'adéquation non paramétrique en grande dimension et nous montrons comment cette méthode s'applique dans le cas de la régression paramétrique. Le test proposé se comporte envers les alternatives locales à la Pitman comme si la dimension du vecteur de variables explicatives était égale à 1. Sous certaines conditions, le test est optimal envers les alternatives définies par des classes de fonctions dépendantes uniquement d'une combinaison linéaire de variables explicatives. Des simulations montrent que notre test se comporte mieux que le test classique de Zheng (1996).

Mots clefs : °Fléau de la dimension, test d'adéquation, méthodes non paramétriques.

1 Introduction

The expression "curse of dimensionality" refers to the poor performances of local smoothing methods for multivariate data. Because of the sparsity of data in multidimensional spaces, the behavior of nonparametric smooth estimators quickly deteriorates as the number of dimension increases, see Stone (1980). This issue is a prominent reason for the study of dimension-reduction models in nonparametric estimation. For instance, when the regression function depends only on a single linear index of the variables, the nonparametric estimator performs as in the one-dimensional case. The single-index regression model has been widely studied in econometrics, see e.g. Stoker (1986), Härdle and Stoker (1989), Powell, Stock and Stocker (1989), Ichimura (1993), Sherman (1994b), Delecroix, Hristache and Patilea (2006).

Many consistent specification tests of a (semi)parametric model contrast the latter model with a completely nonparametric one. Since nonparametric estimators suffer from the curse of dimensionality, so too do the power of the related tests. Specifically, most specification tests of a parametric regression are consistent against directional local alternatives that go further away from the null hypothesis when the dimension of the regressors increases, see for instance Härdle and Mammen (1993) and Zheng (1996). Another approach looks at the uniform consistency of the test against a class of regular alternatives, see Spokoiny (1996), Horowitz and Spokoiny (2001), Guerre and Lavergne (2002), and essentially reaches the same conclusion. The adverse effect of dimension on the tests' power is also found to be significant in practice, as illustrated by our simulations in Section 4.

Very little research has been aimed at alleviating the curse of dimensionality in testing. Zhu (2003) proposed a dimension-reduction type test for a parametric regression, but his null hypothesis is actually the independence of residuals and regressors. This is too strong a hypothesis for econometric application, in particular as data often exhibit conditional volatility. After we wrote a first version of this paper, we discovered a former work on testing a linear parametric regression by Zhu and Li (1998), who put forward a similar idea that we develop here, but did not study the related test.

The purpose of this paper is to propose a new approach for nonparametric testing in high dimensional settings. In Section 2, we propose a general approach that could apply to many nonparametric testing problems. Before entering into details, let us give a flavor of our general approach and of our results. Many testing problems consider a null hypothesis of the form

$$H_0 : \ \mathbb{E}\left[U(\theta_0)|X\right] = 0 \quad \text{almost surely (a.s.)}, \tag{1.1}$$

where θ_0 is an unknown parameter to be estimated and $X \in \mathbb{R}^q$. That is, we want to check whether a zero conditional moment restriction holds for almost any value of the variables X. Our proposal to reduce the dimension of the problem is to use single linear indices $X'\beta$ as a conditioning variable instead of X, and then to look for the direction β that makes $\mathbb{E}\left[U(\theta_0)|X'\beta\right]$ furthest away from zero. Our dimension-reduction approach is thus the testing counterpart of single-index model in estimation, with the fundamental difference that the function under test does not need to depend on a single-index only. Indeed, our first main finding, as stated in Lemma 2.1 and its corollary, is that our approach yields a test which is consistent against *any* nonparametric alternative. To show the potential benefits of our approach, we apply it to testing for the parametric form of a regression in Section 3. Our second main finding is that the resulting test of a parametric regression behaves against general nonparametric directional alternatives *almost* as if the dimension of the regressors was one. Hence there is a cost to dimensionality, but this cost is low and is paid only once, as it does not increase with the dimension of the regressors. Our third main finding is that for a suitable choice of the smoothing parameter, the test is almost optimal against classes of single-index alternatives. This is at odds with standard multidimensional tests, which have the same detection properties irrespective of the dimension of the alternatives. Finally, our simulation study in Section 4 show that our test outperforms the standard "multidimensional" test of Zheng (1996).

2 Dimension reduction in nonparametric testing

2.1 Testing against nonparametric alternatives

A common feature of many nonparametric tests is to consider a zero conditional moment restriction of the form (1.1). The unknown parameter θ_0 can be of finite or infinite dimension and should be estimated either before constructing the test or at the same time. Many testing problems can be recast into this framework. We detail here some important ones, and first the one that we will look at in Section 3.

EXAMPLE 1: TESTING FOR A PARAMETRIC REGRESSION. In that case, $U(\theta) = Y - \mu(X;\theta)$, where $\mu(\cdot;\cdot)$ belongs to a parametric family and θ belongs to a subset of $\mathbb{R}^d$. Tests using smoothing methods have been proposed by Härdle and Mammen (1993), Hong and White (1995), and Zheng (1996), among others, see Hart (1997) for a review. Against nonparametric directional alternatives of the form

$$\mathbb{E}\left[Y|X\right] = \mu(X;\theta_0) + r_n\delta(X)\,,$$

r_n should be of higher order than $n^{-1/2}h^{-q/4}$ to obtain consistency. When looking at the uniform consistency of the test against a class of alternatives of known smoothness s, consistency requires that the alternatives lie at distance $n^{-2s/4s+q}$ from the null hypothesis, see Guerre and Lavergne (2002). When the smoothness index s is unknown, the so-called adaptive rate is less by a small factor, see Spokoiny (1996), Horowitz and Spokoiny (2001), and Guerre and Lavergne (2005).

Another class of consistent tests for a parametric regression is based on various transforms of the cumulative process of residuals obtained from estimation of the parametric model, see in particular Bierens (1982, 1990) and Stute (1997). Theoretical results are mixed: while such tests do not theoretically suffer from the curse of dimensionality under directional alternatives, they exhibit poor performances against sets of regular alternatives, see Guerre and Lavergne (2002).

EXAMPLE 2: TESTING CONDITIONAL MOMENT RESTRICTIONS. In econometrics, we are frequently interested in conditional moment restrictions beyond the regression case. In our framework, $U(\theta) = \rho(Y, X, \theta)$, where $\rho(\cdot, \cdot, \cdot)$ is a multivariate function known up to a finite-dimensional parameter θ. A simple instance is testing for homoscedasticity in a parametric regression model, where $\rho(Y, X, \theta, \sigma^2) = \{Y - \mu(X, \theta)\}^2 - \sigma^2$. Delgado, Dominguez and Lavergne (2006) provide several more instances. Stinchcombe and White (1998) and Whang (2001) study single conditional moment restrictions, Donald, Imbens and Newey (2003) and Delgado, Dominguez and Lavergne (2006) study multiple ones.

Example 3: Testing nonparametric restrictions. In this context, θ is infinite-dimensional. When testing for additivity, $U(\theta) = Y - \sum_{l=1}^{q} m_l(X_l)$, where the unknown univariate functions $m_l(\cdot)$ are properly normalized, see Gozalo and Linton (2001). When testing for a single-index model, $U(\theta) = Y - m(X'\beta)$, for an unknown β and an unknown univariate function $m(\cdot)$, see Fan and Li (1996), Stute and Zhu (2005), and Xia and al. (2005). When testing for the significance of some regressors Z in a nonparametric regression on $X = (X_1, Z)$, $U(\theta) = Y - \mathbb{E}(Y|X_1)$, see Fan and Li (1996), Lavergne and Vuong (2000), Aït-Sahalia, Bickel and Stoker (2001), Delgado and Gonzalez-Manteiga (2001) and Lavergne (2001). Chen and Fan (1999) consider other types of nonparametric restrictions.

Another class of testing problems is closely related to our framework. Consider for instance testing for a parametric conditional distribution function. The null hypothesis is then

$$H_0 \ : \ \mathbb{E}\left[\mathbb{I}(Y \le y) - F(y|X, \theta_0)|X\right] = 0 \quad \text{a.s.} \qquad \text{for all } y \in \mathcal{Y} \quad \text{for some } \theta_0 \ ,$$

where $F(\cdot|X, \theta)$ is a parametric conditional cumulative distribution function and $\mathbb{I}(\cdot)$ denotes the indicator function, see Andrews (1997). Here, one faces a set of conditional moment restrictions indexed not only by (the random) X, but also by (the non-random) y. Such a pattern also appears in other instances such as testing for conditional independence, see Delgado and Gonzalez-Manteiga (2001). Though we do not pursue this issue, our approach could be extended to these more general hypotheses, for instance by rewriting H_0 as depending on X only through an integral over the domain of y, see Hall and Yatchew (2005).

2.2 The fundamental lemma

Our approach relies on the following lemma, which shows that for checking constancy of a conditional expectation, it is equivalent to consider expectations conditional on X and expectations conditional on single linear indices of X.

Lemma 2.1 *Let $X \in \mathbb{R}^q$ and $Z \in \mathbb{R}^c$ be random vectors, with $\mathbb{E}\|Z_j\| < \infty$, $j = 1, \ldots c$.*

A) The following statements are equivalent:

(i) for any (non random) $\beta \in \mathbb{R}^q$ with $\|\beta\| = 1$,

$$\mathbb{E}(Z \mid X'\beta) = \mathbb{E}(Z) \quad \textit{almost surely.} \tag{2.1}$$

(ii) $\mathbb{E}(Z \mid X) = \mathbb{E}(Z)$ almost surely.

B) If X is bounded and $\mathbb{P}\left[\mathbb{E}(Z \mid X) = \mathbb{E}(Z)\right] < 1$, then the set

$$S = \{\beta \in \mathbb{R}^q : \|\beta\| = 1, \ \mathbb{E}(Z \mid X'\beta) = \mathbb{E}(Z) \ \textit{almost surely} \ \}$$

is included in a finite union of circles and points on the sphere $\{\beta \in \mathbb{R}^q : \|\beta\| = 1\}$. In particular, S has Lebesgue measure on the sphere equal to zero and it is not dense on the sphere.

Proof. A) That (ii) implies (i) is immediate. To prove that (i) implies (ii), it suffices to consider the case $c = 1$ and $\mathbb{E}(Z) = 0$. Note that for any $\beta \neq 0$, the $\sigma-$field generated by $X'\beta$ is the same as the $\sigma-$field generated by $X'\beta/\|\beta\|$. By Condition (2.1) and elementary properties of the conditional expectation, we obtain that for any β, including $\beta = 0$,

$$0 = \mathbb{E}\left[\exp\{iX'\beta\}\mathbb{E}(Z \mid X'\beta)\right] = \mathbb{E}\left[\exp\{iX'\beta\}Z\right] = \mathbb{E}\left[\exp\{iX'\beta\}\mathbb{E}(Z \mid X)\right] ,$$

where $i = \sqrt{-1}$. Write $Z = Z^+ - Z^-$ where Z^+ and Z^- are the positive and negative parts of Z, and deduce that for any β, $\mathbb{E}\left[\exp\{iX'\beta\}\mathbb{E}(Z^+ \mid X)\right] = \mathbb{E}\left[\exp\{iX'\beta\}\mathbb{E}(Z^- \mid X)\right]$. As distinct positive finite measures cannot have the same characteristic function, this implies that $\mathbb{E}(Z^+ \mid X) = \mathbb{E}(Z^- \mid X)$ and hence $\mathbb{E}(Z \mid X) = 0$ almost surely.

B) Without loss of generality, take $c = 1$ and $\mathbb{E}(Z) = 0$. By Theorem 1 of Bierens and Ploberger (1997), the set $A = \{\beta \in \mathbb{R}^q : \mathbb{E}[\exp\{iX'\beta\}Z] = 0\}$ has Lebesgue measure zero and is not dense in $\mathbb{R}^q$. Since $S \subset A$, the same conclusion holds for S. A careful inspection of the proofs of Lemma 1 of Bierens (1990) and Theorems 1 and 2 in Bierens (1982) actually shows that when $\mathbb{P}[\mathbb{E}(Z \mid X) = 0] < 1$,

$$A \subset B = \{A_1 \times \mathbb{R}^{q-1}\} \cup \{\mathbb{R}\times A_2 \times \mathbb{R}^{q-2}\} \cup \ldots \cup \{\mathbb{R}^{q-1} \times A_q\}$$

where $A_1, ..., A_q \subset \mathbb{R}$ contain only isolated points. The intersection of B with the set of vectors $\|\beta\| = 1$ is thus a finite union of circles and points. $\blacksquare$

Our lemma easily extends when X is replaced by any one-to-one transformation of X, and thus assuming that X is bounded entails no lost of generality. Part B then ensures that when the conditional expectation $\mathbb{E}(Z|X)$ is not constant, the search for a direction β such that $\mathbb{E}(Z|X'\beta) \neq \mathbb{E}(Z)$ is not vain. Note further that our result still holds when $X'\beta$ is replaced by $g(X'\beta)$ where $g(\cdot)$ is one-to-one. Our fundamental lemma readily yields a new formulation of the null hypothesis of interest.

Corollary 2.2 *Consider random vectors $U(\theta) \in \mathbb{R}^c$ depending on a parameter $\theta \in \Theta$, such that $\mathbb{E}\|U_j(\theta)\| < \infty$, $j = 1, \ldots c$, for all θ, and $X \in \mathbb{R}^q$. Then for any function $\omega(\cdot)$ such that for any β, $\omega(X'\beta) > 0$ on the support of $\mathbb{E}(U(\theta_0)|X'\beta)$, (1.1) is equivalent to*

$$\max_{\|\beta\|=1} \mathbb{E}[U'(\theta_0)\mathbb{E}(U(\theta_0)|X'\beta)\,\omega(X'\beta)] = 0 \quad \text{for some } \theta_0 \in \Theta\,. \tag{2.2}$$

Lemma 2.1 is closely related to Theorem 1 of Bierens (1982), who showed that for bounded X, $\mathbb{E}[Z \mid X] = 0$ is equivalent to $\mathbb{E}[Z\exp\{iX'\beta\}] = 0$ for all β. Part A can also be found in Chen (1991). Stinchcombe and White extended Bierens' result showing that $\mathbb{E}[Z \mid X] = 0$ is equivalent to

$$\mathbb{E}[Z\phi(X'\beta)] = 0 \quad \text{for any } \beta\ ,$$

whenever $\phi(\cdot)$ is analytic non polynomial. Our approach is closely related to theirs, but different in a key aspect. Instead of working with a particular known $\phi(\cdot)$ at the outset, we choose for each β the function of $X'\beta$ maximizing squared correlation with Z. Intuitively, this strategy should enable better detection of departures from the null hypothesis. It is easily shown that the solution is $\mathbb{E}(Z \mid X'\beta)$, so that our null hypothesis writes

$$\mathbb{E}\left[Z\mathbb{E}\left(Z|X'\beta\right)\right] = 0 \quad \text{for any } \beta\ .$$

Now, looking for the *least favorable direction* β for the null hypothesis yields (2.2) with $\omega(\cdot) \equiv 1$. This is in the spirit of the well-known *union-intersection principle* in classical multivariate analysis, cf. Roy (1953). A similar reasoning applies if one maximizes the square of $\mathbb{E}\left[Z\phi\left(X'\beta\right)\omega(X'\beta)\right]$ and $\omega(\cdot)$ is not identically one.

2.3 A general dimension-reduction approach

In view of the previous corollary, our goal is to estimate the quantity in (2.2). Assume we have at our disposal a consistent estimator $\widehat{\theta}_n$ of θ_0 and denote by $U_i(\theta)$ the data-dependent vector function of θ for observation i. Let $\widehat{\gamma}_i(X_i'\beta, \theta)$ be a consistent estimator of $\mathbb{E}\left(U(\theta)|X_i'\beta\right)\omega(X_i'\beta)$ and

$$Q_n\left(\theta, \beta\right) = \frac{1}{n}\sum_{i=1}^{n} U_i'\left(\theta\right)\widehat{\gamma}_i\left(X_i'\beta; \theta\right)\ . \tag{2.3}$$

Under suitable conditions, $Q_n(\widehat{\theta}_n, \beta)$ should converge uniformly in β to

$$Q(\theta, \beta) = \mathbb{E}\left[U'(\theta)\gamma(X'\beta; \theta)\right] = \mathbb{E}\left[U'(\theta)\mathbb{E}\left(U(\theta)|X'\beta\right)\omega(X'\beta)\right]\ .$$

When H_0 does not hold, the maximum of $Q_n(\widehat{\theta}_n, \beta)$ over β stays away from zero almost surely and a test based on it is consistent. Under H_0 however, $Q_n(\widehat{\theta}_n, \beta)$ converges to zero for any β. Hence, we introduce a penalized criterion and define

$$\widehat{\beta}_n = \arg\max_{\|\beta\|=1}\left\{Q_n(\widehat{\theta}_n, \beta) - \pi_n(\|\beta - \beta_0\|)\right\}\ .$$

A normalized version of $Q_n(\widehat{\theta}_n, \widehat{\beta}_n)$ is then taken as the test statistic.

The penalty $\pi_n(\cdot)$ is a nonnegative function that equals zero only at zero. Provided it is large enough with respect to $\max_{\|\beta\|=1} Q_n(\widehat{\theta}_n, \beta)$, it forces the maximum to be attained at β_0 under H_0. The critical value will then be the one of the test based on $Q_n(\widehat{\theta}_n, \beta_0)$. Here β_0 is a fixed vector representing the favorite alternative of the practitioner. When H_0 does not hold, the penalty should not perturb the behavior of the maximum, hence $\pi_n(t)$ should decrease towards zero fast enough for all t as n grows. The penalized criterion will thus select a direction different from β_0 only when this direction gives more power to the test. As will become apparent in the next section, the choice of the penalty is crucial to control the level of the test and to ensure high power. By contrast, the choice of β_0 is theoretically irrelevant for consistency of the test. Though, the test should have greater power when $\mathbb{E}\left[U(\theta_0)|X'\beta_0\right]$ is different from zero, and in practice the choice of β_0 may have some influence. We view this small sample phenomenon more as an advantage than as a drawback, since it introduces some flexibility. For instance, when testing for a linear regression model, if one suspect possible nonlinearities in, say, the first component of X, this is easily accounted for in the procedure. Alternatively, one may want to use, say, the least-squares estimator itself in the preferred index, in the spirit of Ramsey (1969).

Our penalized criterion yields two main properties. First, it allows to obtain a pivotal distribution under H_0. This is of foremost interest in practice. In small and moderate samples, bootstrapping techniques are asymptotically valid, as we will show below. In large samples, one can avoid costly numerical simulations by using asymptotic critical values. Second, the test we obtain is always as powerful as the test based on the favorite direction β_0. Indeed, as $\pi(0) = 0$,

$$Q_n(\widehat{\theta}_n, \widehat{\beta}_n) \geq \max_{\|\beta\|=1} \left\{ Q_n(\widehat{\theta}_n, \beta) - \pi_n(\|\beta - \beta_0\|) \right\} \geq Q_n(\widehat{\theta}_n, \beta_0) \,.$$

Since the critical value of our test is the one of the test based on $Q_n(\widehat{\theta}_n, \beta_0)$, we obtain a test that is consistent against nonparametric alternatives and at the same time is always more powerful than a test tailored for alternatives of the form $\varphi(X'\beta_0)$.

An alternative way to proceed would be to base a test on $\max_{\|\beta\|=1} Q_n(\widehat{\theta}_n, \beta)$. The trouble here is that any β is solution of (2.2), and thus the maximizer is not identified in the limit problem, though the objective function itself might be well-behaved, as in the case of set identification studied by Chernozhukov, Hong and Tamer (2004). This is however unclear in our case, as $Q_n(\widehat{\theta}_n, \beta)$ is not a smooth process with respect to β. Even if this technical difficulty were overcame, the resulting test would have higher critical values than our test, and its power would be affected.

3 Testing for a parametric regression

3.1 The test

Let $(Y, X')'$ be a random vector in $\mathbb{R}^{1+q}$. We consider the q-variate regression $m(X) = \mathbb{E}(Y|X)$ and continuous X, as discrete regressors do not strictly speaking yield a "curse of dimensionality." Let the parametric regression model be $\{\mu(\cdot;\theta) : \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^d$. The model is correctly specified iff H_0 as defined in (1.1) holds with $U(\theta) = Y - \mu(X;\theta)$. We apply our general approach to this testing problem using kernel estimators. To avoid handling denominators close to zero, we set the weight function $\omega(\cdot)$ in (2.2) equal to the density of $X'\beta$, denoted by $f_\beta(\cdot)$, which is assumed to exist for any β. Let

$$Q(\theta, \beta) = \mathbb{E}\{U(\theta)\mathbb{E}[U(\theta) \mid X'\beta] f_\beta(X'\beta)\} = \mathbb{E}\{\mathbb{E}^2[U(\theta) \mid X'\beta] f_\beta(X'\beta)\}.$$

By Corollary 2.2, the regression model is correctly specified iff $\max_{\|\beta\|=1} Q(\theta_0, \beta) = 0$. Let $(Y_i, X_i')'$, $i = 1, \ldots n$ be a random sample from the distribution of $(Y, X')'$. The parameter θ_0 can be estimated in a variety of ways. For instance, $\widehat{\theta}_n$ can be the nonlinear least-squares (NLLS) estimator of θ solving

$$\widehat{\theta}_n = \arg\min_{\theta\in\Theta} \sum_{i=1}^{n} \left(Y_i - \mu(X_i;\theta)\right)^2 \ , \tag{3.1}$$

with an appropriate convention in case of ties. In view of Equation (2.3), define the estimator of $Q(\theta, \beta)$ as

$$Q_n(\theta, \beta) = \frac{1}{n(n-1)} \sum_{j\neq i} U_i(\theta) \, U_j(\theta) \, \frac{1}{h} K_h\left((X_i - X_j)'\beta\right)$$

where $U_i(\theta) = Y_i - \mu(X_i;\theta)$ and $K_h(\cdot) = K(\cdot/h)$, where $K(\cdot)$ is a kernel and h a bandwidth. For a fixed β, the estimator $Q_n(\widehat{\theta}_n, \beta)$ is the statistic studied by Li and Wang (1998) and Zheng (1996) applied to the index $X'\beta$, and has an asymptotic centered normal distribution with rate $nh^{1/2}$ under H_0.

Zhu and Li (1998) first proposed to use the maximum over β of a statistic close to $Q_n(\widehat{\theta}_n, \beta)$ for checking a linear regression model. However, their test is based on the maximum plus a term of the form $(1/n)\sum_{i=1}^n U_i(\widehat{\theta}_n)\phi(\|X_i\|)$, where $\phi(\cdot)$ is the standard normal univariate density (or any other known function). Hence, they combine a consistent test based on nonparametric methods with an inconsistent M-type test, so that the asymptotic behavior under H_0 is completely driven by the M-test statistic. Instead, we apply our penalization method and we choose β as

$$\widehat{\beta}_n = \arg\max_{\|\beta\|=1} \left\{ nh^{1/2} Q_n(\widehat{\theta}_n, \beta) - \alpha_n \mathbb{I}\left[\beta \neq \beta_0\right] \right\} , \tag{3.2}$$

where β_0 is user-chosen and α_n, $n \geq 1$, is a sequence of positive real numbers decreasing to zero at an appropriate rate. Note that, with respect to Section 2.3, we have introduced the rate of convergence of $Q_n(\theta, \beta)$ in our criterion, as it yields more transparent results. Our choice for the penalty function corresponds to the one of Bierens (1990) and is made for simplicity. We will prove that $\widehat{\beta}_n = \beta_0$ with probability tending to 1 under H_0. Since $Q_n(\widehat{\theta}_n, \widehat{\beta}_n)$ behaves like $Q_n(\widehat{\theta}_n, \ \beta_0)$, a test is easily constructed. With at hand a consistent estimator $\widehat{v}_n^2(\beta)$ of the variance of $nh^{1/2} Q_n(\widehat{\theta}_n, \beta)$, let

$$T_n = nh^{1/2} \frac{Q_n(\widehat{\theta}_n, \widehat{\beta}_n)}{\widehat{v}_n(\beta_0)} \ .$$

An asymptotic α-level test is given by $\mathbb{I}(T_n \geq z_\alpha)$, where z_α is the $(1-\alpha)$-th quantile of the standard normal distribution. Moreover, as both $\widehat{v}_n^2(\widehat{\beta}_n)$ and $\widehat{v}_n^2(\beta_0)$ estimate the variance of $Q_n(\widehat{\theta}_n, \widehat{\beta}_n)$ under H_0, we can also consider $\mathbb{I}(T_n' \geq z_\alpha)$, where

$$T_n' = nh^{1/2} \frac{Q_n(\widehat{\theta}_n, \widehat{\beta}_n)}{\min\left(\widehat{v}_n(\beta_0), \widehat{v}_n(\widehat{\beta}_n)\right)} \ .$$

The purpose of taking the minimum of the two variance estimators is to improve the small sample power of our test.

3.2 Assumptions

We consider the following assumptions on the data-generating process.

Assumption D *(a) The random vectors* $(\varepsilon_1, X_1')', \ldots, (\varepsilon_n, X_n')'$ *are independent copies of the random vector* $(\varepsilon, X')' \in \mathbb{R}^{1+q}$ *with* $\mathbb{E}(\varepsilon \mid X) = 0$ *and* $\mathbb{E}|\varepsilon|^{11} < \infty$.
(b) Let $\sigma^2(x) = \mathbb{E}(\varepsilon^2 \mid X = x)$. *There exist constants* $\underline{\sigma}^2$ *and* $\overline{\sigma}^2$ *such that for any* x $0 < \underline{\sigma}^2 \leq \sigma^2(x) \leq \overline{\sigma}^2 < \infty$.
(c) For any β *of norm one,* $X'\beta$ *admits a density* $f_\beta(\cdot)$ *that is bounded uniformly in* β.

Next, we introduce assumptions on the regression model. For any matrix A of generic element a_{kl}, let $\|A\|$ denote the matrix norm $[\sum_{kl} a_{kl}^2]^{1/2}$.

Assumption M *(a) Let* $\Theta \subset \mathbb{R}^d$ *be a compact set. For any* $\theta_1, \theta_2 \in \Theta$,

$$\mu(\cdot;\theta_1) - \mu(\cdot;\theta_2) = (\theta_1 - \theta_2)' \,\dot{\mu}(\cdot;\theta_2) + (\theta_1 - \theta_2)' \,\ddot{\mu}(\cdot;\theta_1,\theta_2)(\theta_1 - \theta_2) \ ,$$

where (i) $\dot{\mu}(\cdot;\theta)$ *is such that* $\sup_{\theta\in\Theta} \|\dot{\mu}(X;\theta)\| \leq \Phi_1(X)$ *with* $\mathbb{E}\left[\Phi_1^4(X)\right] < \infty$;
(ii) $\ddot{\mu}(\cdot;\theta_1,\theta_2)$ *is such that* $\sup_{\theta_1,\theta_2\in\Theta} \|\ddot{\mu}(X;\theta_1,\theta_2)\| \leq \Phi_2(X)$ *with* $\mathbb{E}\left[\Phi_2^2(X)\right] < \infty$; *and*
(iii) $\forall \varepsilon > 0$, *there is a* $\eta > 0$ *such that* $\mathbb{E}\sup_{\|\theta_1-\theta_2\|\leq\eta} \|\ddot{\mu}(X;\theta_1,\theta_2) - \ddot{\mu}(X;\theta_2,\theta_2)\| < \varepsilon$.

(b) (Identification condition) There exists a real valued function $\Phi_3(\cdot)$ *that is not almost surely zero such that for any* $\theta \in \Theta$ *and* X, $|\mu(X;\theta) - \mu(X;\theta_0)| \geq \Phi_3(X)\,\|\theta - \theta_0\|$.

A large range of parametric models satisfies Assumption M. Together with our assumptions on the design, it ensures the $\sqrt{n}$-consistency of the NLLS estimator (3.1) as stated in Lemma 6.1. We make the following assumptions on the kernel and bandwidth.

Assumption K *(a) The kernel* $K(\cdot)$ *is a bounded symmetric density of bounded variation. (b)* $h \to 0$ *and* $(nh^2)^\alpha / \ln n \to \infty$ *for some* $\alpha \in (0,1)$.

Last, we need some assumptions to estimate the asymptotic variance of $nh^{1/2}Q_n(\widehat{\theta}_n, \beta)$, which writes, conditionally upon the X_i,

$$v_n^2(\beta) = \frac{2}{n(n-1)} \sum_{j\neq i} \sigma^2(X_i)\sigma^2(X_j) h^{-1} K_h^2\left((X_i - X_j)'\beta\right) \ .$$

In general, the conditional variance $\sigma^2(\cdot)$ is unknown. However, with at hand a nonparametric estimator of the conditional variance such that

$$\sup_{1\leq i\leq n}\left|\frac{\widehat{\sigma}^2(X_i)}{\sigma^2(X_i)}-1\right|=o_{\mathbb{P}}(1)\ , \tag{3.3}$$

$v_n^2(\beta)$ can be consistently estimated by

$$\widehat{v}_n^2(\beta)=\frac{2}{n(n-1)}\sum_{j\neq i}\widehat{\sigma}^2(X_i)\widehat{\sigma}^2(X_j)h^{-1}K_h^2\left((X_i-X_j)'\beta\right)\ .$$

For instance, one can consider

$$\widehat{\sigma}^2(x)=\frac{\sum_{i=1}^n Y_i^2\mathbb{I}\left\{\|x-X_i\|\leq b\right\}}{\sum_{i=1}^n \mathbb{I}\left\{\|x-X_i\|\leq b\right\}}-\left(\frac{\sum_{i=1}^n Y_i\mathbb{I}\left\{\|x-X_i\|\leq b\right\}}{\sum_{i=1}^n \mathbb{I}\left\{\|x-X_i\|\leq b\right\}}\right)^2\ ,$$

where b is a bandwidth parameter chosen independently of h. Guerre and Lavergne (2005) provide some primitive conditions for (3.3). Then it is straightforward to show that $\widehat{v}_n^2(\beta)/v_n^2(\beta)=1+o_{\mathbb{P}}(1)$ for any β. Given our focus, we shall proceed under (3.3).

3.3 Behavior under the null hypothesis

Our first task is to study the behavior of the process $Q_n(\widehat{\theta}_n,\beta)$ as indexed by β under H_0. It has the following decomposition

$$\begin{aligned}Q_n(\widehat{\theta}_n,\beta)&=Q_{0n}(\beta)-2Q_{1n}(\widehat{\theta}_n,\beta)+Q_{2n}(\widehat{\theta}_n,\beta)=\frac{1}{n(n-1)}\sum_{j\neq i}\varepsilon_i\varepsilon_j\frac{1}{h}K_h\left((X_i-X_j)'\beta\right)\\&-\frac{2}{n(n-1)}\sum_{j\neq i}\varepsilon_i\left\{\mu(X_j;\widehat{\theta}_n)-\mu(X_j;\theta_0)\right\}\frac{1}{h}K_h\left((X_i-X_j)'\beta\right)\\&+\frac{1}{n(n-1)}\sum_{j\neq i}\left\{\mu(X_i;\widehat{\theta}_n)-\mu(X_i;\theta_0)\right\}\left\{\mu(X_j;\widehat{\theta}_n)-\mu(X_j;\theta_0)\right\}\frac{1}{h}K_h\left((X_i-X_j)'\beta\right)\ .\end{aligned}$$

Lemma 3.1 *Let Assumptions D and K hold. Then*

(i) $\sup_{\|\beta\|=1}|Q_{0n}(\beta)|=O_{\mathbb{P}}(n^{-1}h^{-1/2}\ln n)$ *under* H_0.

(ii) *Under Assumption M,* $\sup_{\|\beta\|=1}|2Q_{1n}(\widehat{\theta}_n,\beta)-Q_{2n}(\widehat{\theta}_n,\beta)|=o_{\mathbb{P}}(n^{-1}h^{-1/2})$ *if* $\|\widehat{\theta}_n-\theta_0\|=O_{\mathbb{P}}(n^{-1/2})$.

The proof is given in Section 6. We now describe the behavior of $\widehat{\beta}_n$ under H_0.

Lemma 3.2 *Let Assumptions D, M, and K hold. Consider a positive sequence α_n such that $\alpha_n/\ln n \to \infty$. Under H_0, $\mathbb{P}(\widehat{\beta}_n = \beta_0) \to 1$.*

Proof. By definition, for all $n \geq 1$, $nh^{1/2}Q_n(\widehat{\theta}_n, \beta_0) \leq nh^{1/2}Q_n(\widehat{\theta}_n, \widehat{\beta}_n) - \alpha_n \mathbb{I}(\widehat{\beta}_n \neq \beta_0)$. This implies that $0 \leq \mathbb{I}(\widehat{\beta}_n \neq \beta_0) \leq nh^{1/2}\alpha_n^{-1}\left\{Q_n(\widehat{\theta}_n, \widehat{\beta}_n) - Q_n(\widehat{\theta}_n, \beta_0)\right\}$. From Lemma 6.1, $\|\widehat{\theta}_n - \theta_0\| = O_{\mathbb{P}}(n^{-1/2})$ under H_0 and then from Lemma 3.1, $Q_n(\widehat{\theta}_n,\widehat{\beta}_n) - Q_n(\widehat{\theta}_n,\beta_0) = O_{\mathbb{P}}(n^{-1}h^{-1/2}\ln n)$. Then $\alpha_n/\ln n \to \infty$ yields $\mathbb{I}(\widehat{\beta}_n \neq \beta_0) = o_{\mathbb{P}}(1)$. Use the boundedness of $\mathbb{I}(\cdot)$ to conclude that $\mathbb{P}(\widehat{\beta}_n \neq \beta_0) = \mathbb{E}\left[\mathbb{I}(\widehat{\beta}_n \neq \beta_0)\right] \to 0$. ■

The asymptotic behavior of our tests under the null hypothesis can now be stated.

Theorem 3.3 *Under Assumptions D, M, K, and (3.3), if $\alpha_n/\ln n \to \infty$, then the tests based on T_n or T_n' have asymptotic level α given the design.*

Proof. From Lemma 3.2, $\mathbb{P}\left[Q_n(\widehat{\theta}_n, \widehat{\beta}_n) = Q_n(\widehat{\theta}_n, \beta_0)\right]$ and $\mathbb{P}\left[\widehat{v}_n^2(\widehat{\beta}_n) = \widehat{v}_n^2(\beta_0)\right]$ both converge to one. By Condition (3.3), $\widehat{v}_n^2(\beta_0) = v_n^2(\beta_0)(1 + o_p(1))$. From Lemmas 6.1 and 3.1, $nh^{1/2}Q_n(\widehat{\theta}_n, \beta_0) = nh^{1/2}Q_{0n}(\beta_0) + o_p(1)$. From Lemma 2-(i) by Guerre and Lavergne (2005), $nh^{1/2}Q_{0n}(\beta_0)/v_n(\beta_0)$ converges to a standard normal conditionally upon the X_i if

$$\frac{\mathrm{Sp}\,(W_{\beta_0})}{\|W_{\beta_0}\|} \xrightarrow{p} 0\,, \quad \text{where } W_{\beta_0} = \left[\mathbb{I}\,(i \neq j)\,K_h\left((X_i - X_j)'\beta_0\right)/(h\,n(n-1)), i, j = 1, \ldots n\right]$$

and $\mathrm{Sp}(W_\beta)$ is the spectral radius of the matrix W_β. Lemma 6.2 allows to conclude. ■
Lemma 3.1 is the theoretical key that drives our results. The quantity $Q_{0n}(\beta)$ is in probability of order $(nh^{1/2})^{-1}$ for any β, however the supremum over β is not regular enough to be shown of the same order, at least from the results we use here. It is an open question whether the process $Q_{0n}(\beta)/v_n(\beta)$ has a more regular behavior. The study of $Q_{1n}(\widehat{\theta}_n, \beta)$ raises a similar problem. Namely, for fixed β, standard empirical processes methods show that $Q_{1n}(\widehat{\theta}_n, \beta) = O_{\mathbb{P}}(n^{-1})$, see e.g. Guerre and Lavergne (2005), but the same does not seem to hold uniformly over β. All the trouble here comes from that any β is solution of (2.2). As will be seen shortly, the study under the alternative is simpler.

3.4 Consistency

A simple inequality is at the heart of the consistency of our test. Indeed, we have

$$\begin{aligned}
T_n' \geq T_n &= \frac{nh^{1/2} Q_n(\widehat{\theta}_n, \widehat{\beta}_n)}{\widehat{v}_n(\beta_0)} \\
&= \frac{1}{\widehat{v}_n(\beta_0)} \left[\max_{\|\beta\|=1} \left\{ nh^{1/2} Q_n(\widehat{\theta}_n, \beta) - \alpha_n \mathbb{I}(\beta \neq \beta_0) \right\} + \alpha_n \mathbb{I}(\widehat{\beta}_n \neq \beta_0) \right] \\
&\geq \frac{1}{\widehat{v}_n(\beta_0)} \left[\max_{\|\beta\|=1} nh^{1/2} Q_n(\widehat{\theta}_n, \beta) - \alpha_n \right] \\
&\geq \frac{nh^{1/2} Q_n(\widehat{\theta}_n, \beta) - \alpha_n}{v_n(\beta_0)(1 + o_{\mathbb{P}}(1))} \quad \text{for any } \beta \; . \qquad (3.4)
\end{aligned}$$

Hence, the test based on T_n (or T_n') is consistent if the last minorant stays away from zero with probability tending to one for some β. It is easily seen that when the parametric model is misspecified, our test is consistent under the assumptions of Theorem 3.3 provided $\widehat{\theta}_n$ converges to some pseudo-true value θ^*. Indeed, there exists at least one β for which $Q(\theta^*, \beta) > 0$. We also note as a particular feature of our penalization approach that our test has better power than the test based on the favored direction β_0. This follows since

$$T_n' \geq T_n \geq \frac{nh^{1/2} Q_n(\widehat{\theta}_n, \beta_0)}{\widehat{v}_n(\beta_0)} \; ,$$

and *all these statistics have the same limiting distribution under* H_0.

3.4.1 Behavior against nonparametric directional alternatives

Let us now investigate the ability of our test to detect nonparametric directional departures from the null hypothesis. Consider a real-valued function $\delta(X)$ such that

$$\mathbb{E}[\delta(X)\dot{\mu}(X;\theta_0)] = 0 \quad \text{and} \quad 0 < \mathbb{E}[\delta^4(X)] < \infty \; , \qquad (3.5)$$

and the sequence of alternatives defined as

$$H_{1n} : \; m_n(X) = \mu(X;\theta_0) + r_n \delta(X), \quad n \geq 1 \; . \qquad (3.6)$$

Note that there is no smoothness restriction on the function $\delta(\cdot)$ as is frequent in this kind of analysis, see e.g. Zheng (1996). Under H_{1n}, $\widehat{\theta}_n - \theta_0 = O_{\mathbb{P}}(n^{-1/2})$ as shown by Lemma

6.1 in Section 6. We show below that such directional alternatives can be detected as soon as $r_n^2 n h^{1/2}/\alpha_n$ tends to infinity. The conditions of Theorem 3.2 yield that α_n is of the form $a_n \ln n$ with a_n diverging to infinity at an arbitrary slow rate. Hence to obtain consistency against (3.6), we should have $r_n^2 n h^{1/2}/\left(a_n \ln n\right) \to \infty$, where h applies to the univariate variable defined by a single linear index in X. By comparison, when one uses a standard multidimensional smooth test, $r_n^2 n h^{q/2} \to \infty$ is needed for consistency. In other words, from the theoretical point of view, our test *does not suffer from the curse of dimensionality* against directional alternatives, that is, whatever the number of regressors, the power remains close to the power obtained in the unidimensional case. The proof of the following result is deferred to Section 6.

Theorem 3.4 *Under Assumptions D, M, K, and (3.3), if $r_n^2 n h^{1/2}/\alpha_n \to \infty$, the tests based on T_n and T_n' are consistent given the design against the sequence of alternatives H_{1n} with $\delta(X)$ satisfying (3.5).*

3.4.2 Behavior against classes of low-dimensional alternatives

For nonparametric multidimensional tests, Guerre and Lavergne (2002) showed that a suitable choice of the smoothing parameter yields an optimal test against nonparametric alternatives of known smoothness s. Specifically, the smoothing parameter h should balance the bias in estimating the L^2-norm of the regression function, which is of order h^{2s}, with the variance of the basic statistic given by $nh^{q/2}$. Unfortunately, such an optimal test will have the same power properties against classes of low-dimensional alternatives. Here we show that our tests are almost optimal against classes of one-dimensional alternatives.

Define a class of regular functions as follows. For any real s, let $\lfloor s \rfloor$ be the lower integer part of s, i.e. $\lfloor s \rfloor < s \leq \lfloor s \rfloor + 1$. Define the Hölder class $C(L,s)$ as

$$C(L,s) = \{m(\cdot); |m(x) - m(y)| \leq L|x-y|^s \text{ for all } x, y\ \} \text{ for } s \in (0,1],$$
$$C(L,s) = \{m(\cdot);\ \text{the } \lfloor s \rfloor\text{-th derivative of } m(\cdot) \text{ are in } C(L, s - \lfloor s \rfloor)\ \} \text{ for } s > 1\ .$$

Consider a sequence of functions $\delta_n(X'\beta)$ of a single-index such that

$$\forall n, \qquad \mathbb{E}[\delta_n(X'\beta)\dot{\mu}(X;\theta_0)] = 0 \quad \text{and} \quad 0 < \mathbb{E}[\delta_n^4(X'\beta)] < C < \infty\ , \qquad (3.7)$$

and the sequence of alternatives defined as

$$H_{1n}' : \; m_n(X) = \mu(X;\theta_0) + \delta_n(X'\beta), \quad n \geq 1 \, . \tag{3.8}$$

For such classes of alternatives, a specific inference problem arises. Indeed, as a high-dimensional design is projected onto a lower dimension space, the density of the low dimension vector decreases to zero at the endpoints of its support, even when the original design density is bounded away from zero. This issue has been investigated for nonparametric curve estimation by Hall *et al.* (1997). In nonparametric testing, it is hard, and often impossible, to detect any alternative that concentrates in a low density area, see Proposition 2 in Guerre and Lavergne (2002). This explains why the latter, as well as Horowitz and Spokoiny (2001), assume a design whose density is bounded away from zero. In our setting, such an assumption is clearly irrelevant, and explains the new assumption that follows as well as our Condition (3.10).

Assumption N *(a) X is bounded.*

(b) For any β (i) there exist constants $C > 0$ and $a > 0$ such that $|f_\beta(t) - f_\beta(t')| \leq C|t - t'|^a$ for any t, t' and (ii) there exist a constant $c_0 > 0$ and an integer k_0 such that for each β and $0 < c \leq c_0$ the set $A_{\beta,c} = \{u : f_\beta(u) \geq c\}$ is a union of at most k_0 intervals on the real line.

(c) $K(\cdot)$ has a nonnegative Fourier transform.

We are now ready to state our next result, whose proof is deferred to Section 6.

Theorem 3.5 *Let $a_n \to 0$ and consider the class of alternatives (3.8) with unknown θ_0 and β, $\delta_n(\cdot) \in C(L, s)$ for some $s > 3/4$ and $L > 0$, $\mathbb{E}\delta_n^2 = o(1)$,*

$$\left[n^{-1} \sum_{i=1}^{n} \delta_n^2(X_i'\beta) \right]^{1/2} \geq \kappa_n (1 + o_{\mathbb{P}}(1)) \left(\frac{\alpha_n}{n} \right)^{\frac{2s}{4s+1}} \, , \tag{3.9}$$

$$\text{and} \qquad \left[n^{-1} \sum_{i=1}^{n} \delta_n^2(X_i'\beta) \mathbb{I}(X_i'\beta \notin A_{\beta,a_n}) \right] = o_{\mathbb{P}}(a_n) \left[n^{-1} \sum_{i=1}^{n} \delta_n^2(X_i'\beta) \right] \, . \tag{3.10}$$

Under Assumptions D, M, K, N, and (3.3), if h of order $(\alpha_n/n)^{\frac{2}{4s+1}}$, $a_n\sqrt{n}h \to \infty$ and $a_n h^{-a} \to \infty$, the tests based on T_n and T_n' are consistent given the design whenever $\kappa_n^2 a_n$ diverges.

To justify Condition (3.10), note that outside A_{β,a_n} the density is $O(a_n)$ by definition, and the volume of the complement of A_{β,a_n} is typically an $o(1)$ when the density of X is bounded away from zero: when the domain of X is an ellipsoid, it is an $O(a_n^{2/(q-1)})$, while when the domain of X is rectangular, it is an $O(a_n^{1/(q-1)})$ when β is not parallel to a side of the domain, see Hall *et al.* (1997). What our condition excludes is thus alternatives that concentrate on low density areas. Aside this technicality, the rate obtained in Theorem 3.5 is almost optimal, and differs from the optimal rate only because α_n enters in its formula. This is a low price to pay that is independent of the dimension of X. Note that we have to impose $s > 3/4$ because of our Assumption K-(b).

If we pursued further the investigation of our test against low-dimensional alternatives, we could show using similar arguments that our test is not optimal against m-dimensional alternatives, but can be more powerful than a multidimensional test. Specifically the rate would be close to $n^{-(2s+1-m)/(4s+1)}$. This can be better than the usual rate for a multidimensional test, that is $n^{-2s/(4s+p)}$. For instance, for $s = 2$, our test is more powerful against classes of double-indices alternatives whenever $p > 4$.

4 Practical implementation

4.1 Bootstrap critical values

The wild bootstrap, initially proposed by Wu (1986), is often used in smooth tests to compute small sample critical values, see e.g. Härdle and Mammen (1993). Here we use a generalization of this method, the smooth conditional moments bootstrap introduced by Gozalo (1997). It consists in drawing n i.i.d. random variables ω_i independent from the original sample with $\mathbb{E}\omega_i = 0$, $\mathbb{E}\omega_i^2 = 1$, and $\mathbb{E}\omega_i^4 < \infty$, and to generate bootstrap observations of Y_i as $Y_i^* = \mu(X_i, \widehat{\theta}_n) + \widehat{\sigma}(X_i)\omega_i$, $i = 1, \dots n$. A bootstrap test statistic is built from the bootstrap sample as was the original test statistic. When this scheme is repeated many times, the bootstrap critical value $z^*_{\alpha,n}$ at level α is the empirical $(1-\alpha)$-th quantile of the bootstrapped test statistic. This critical value is then compared to the initial test statistic. The validity of this procedure follows easily from our previous results and the proof is thus omitted.

Theorem 4.1 *Under the assumptions of Theorem 3.3, the bootstrap critical values yield a test based on T_n or T'_n with asymptotic level α given the design.*

4.2 Practical choices in the procedure

Different choices should be made in implementing the procedure. For the construction of our basic statistics $Q_n(\widehat{\theta}_n, \beta)$, choices of a kernel and a bandwidth are required. As usual, the kernel is not expected to have much influence. The bandwidth choice in contrast is important. A data-driven choice of the bandwidth in the spirit of Guerre and Lavergne (2005) should be investigated further, but this is outside the scope of our paper. However, one striking result of our simulation study reported below is that our test is little sensitive to the bandwidth. We should also comment here on the fact that the same bandwidth is used for all directions. As commented on in Section 2, our fundamental lemma also applies to a one-to-one transformation of the X. For instance, one can transform X so that its covariance matrix is identity or its support is $(0,1)^p$. Alternatively, one can map all the linear indices $X'\beta$ through a bijective function into, say, $(0,1)$.

To compute our test statistic, we have to maximize our penalized criterion. In practice, this can be done on a fine enough grid on the hypersphere. Lemma 2.1-B ensures that under the alternative any direction yields a consistent test, but a set of measure zero when X is bounded, which is not restrictive as already noted. The specific choices for our procedure are β_0 and α_n. As explained before, β_0 corresponds to a favored alternative to be determined in each particular case depending on a priori information. We investigate in the next section how its choice can affect the small sample performances of our test. The choice of α_n is more crucial and reflects the weight given to the favored alternative.

In view of our results, α_n should be of an order slightly higher than $nh^{1/2}Q_n(\widehat{\theta}_n, \beta_0)$ times $\ln n$ under the null hypothesis. In practice, one can simulate this statistic to obtain an estimate of its standard deviation and inflate it by some small factor. This device was implemented in our simulation study and appeared to work well.

4.3 Simulation study

Our focus was first to compare the small sample power of our test to the multivariate test of Zheng (1996) and Li and Wang (1998) and second to determine the sensitivity of our test to the penalty α_n, the direction defined by β_0 and the smoothing parameter h. For the sake of simplicity, we considered the null hypothesis

$$H_0 \ : \mathbb{E}(Y|X) = 0 \ .$$

We generated samples of 50 observations from independent uniformly distributed variables X_1, X_2, X_3. The support of each variable was chosen as $\left[-\sqrt{3}, \sqrt{3}\right]$ to get unit variance. We sampled errors from a standard normal distribution and constructed the response variable as

$$Y_i = \frac{d}{5}\left\{\cosh\left(\frac{X_{1i} + 2X_{2i} + 3X_{3i}}{\sqrt{3}}\right) - e\right\} + \varepsilon_i \quad i = 1, \ldots 50$$

with e a centering constant equal to $\sinh(1)\sinh(2)\sinh(3)/6$.

We considered (i) Zheng's test when the index $(X_1 + 2X_2 + 3X_3)/\sqrt{14}$ is considered as the only regressor; (ii) Zheng's test when all three regressors are taken into account; (iii) our test based on T_n' (results for T_n differed little and are not reported). To speed up computations, we assumed that the errors' variance was known for all the tests. The optimization was carried out on a grid of 5000 points sampled from the uniform distribution on the three-dimensional hypersphere of unit radius. From 5000 samples generated under the null hypothesis, i.e. with $d = 0$, we computed the tests statistics and obtained small sample critical values that allows to calibrate the level of the different tests at 5%. This is equivalent to bootstrapping since no parameter is estimated under H_0. We then drew the power curves of the different tests based on 2000 samples for each point of the grid $d = 0.1, 0.2, \ldots 0.7$. To compute the test statistics, we used a biweight kernel with support $[-1, 1]$ and we selected the bandwidth as $h = bn^{-2/(8+q)}$, with $q = 3$ in Case (ii) and $q = 1$ in the other cases, and b varies in $\{0.5, 1, 1.5, \ldots, 4\}$. To set α_n, we computed v_0, the mean of $v_n(\beta_0)$, which was found to vary little with β_0. We then chose $\alpha_n = av_0n^{-1}h^{-1/2}$, with $v_0 = 0.65$ in our case, and we let a vary. Specifically, since $\ln(50) \approx 1.70$, we chose $a = 2, 7$ and 10. Such a wide variation allows to study the sensitivity of our test to α_n, and should not be viewed as a recommendation.

We first set β_0 to $(1, 1, 1)'/\sqrt{3}$, a natural choice if one does not favor any regressor at the outset. Figure 1 compares the power curves of Zheng's tests and of our test for the different values of α and the bandwidth constant b is set to one. The first striking fact is the large loss in power for Zheng's test when going from dimension one to three. In practice however, the test based on the unknown single linear index is infeasible. The second striking fact is that our test largely outperforms Zheng's test in dimension 3. The curve power of our dimension-reduction test is very close to the one of the infeasible test for $a = 2$ and goes away from it as α_n increases, as expected. Still the gain in power with respect to Zheng's test is large for $a = 10$.

We then considered two polar cases. In Figure 2, β_0 was chosen as the true unknown index. In that case, the power curves of the infeasible test and of our test are very similar whatever the value of α_n. This confirms our theoretical finding that our test is as powerful as the test based on the index $X'\beta_0$. Figure 3 depicts the less perfect case where β_0 is set to $(0, 1, 0)$, that is we "favor" alternatives depending upon X_2 *only*. When α_n is small, our test performs well, but its power decreases when α_n increases. For the largest considered penalty, our test is beaten by Zheng's test for small alternatives, as expected, but the reverse holds for larger alternatives.

In Figure 4 we drew the power of the tests as a function of the bandwidth constant b for $d = 0.4$ to illustrate that our main findings were very little dependent of the chosen smoothing parameter. For a small bandwidth, our test can even outperform the infeasible test. Moreover, the power of our test is very stable while the performances of Zheng's test is much more variable in either dimension.

In a second set of simulations, the true regression depended on two linear indices, and we generated samples as

$$Y_i = \frac{3}{10}\, d \left\{ \cosh\left(\frac{3X_{1i} + X_{2i}}{\sqrt{3}}\right) + \cosh\left(\frac{X_{2i} + 3X_{3i}}{\sqrt{3}}\right) - e \right\} + \varepsilon_i \quad i = 1, \ldots 50$$

with e a centering constant equal to $2\sinh(1)\sinh(3)/3$. This setup was chosen to evaluate the behavior of our test against an alternative of intermediate dimension, and we considered as a benchmark Zheng's test based on the two linear indices entering the regression function. Given our previous findings, we discarded the value $a = 10$ as giving too much weight to the favored direction. Other features of the experiments were unchanged, Figures 5 to 8 summarize findings for different values for α_n, β_0 and h. Compared to our first set of experiments, the benchmark test has power close to Zheng's test in dimension 3 and our test has equal or better performances than the benchmark test for different values of β_0, α_n, and h.

5 Conclusion

We have proposed a general approach to testing conditional moment restrictions in high dimension. Lemma 2.1 is the key to our approach. It shows that for testing $\mathbb{E}(Z|X) = 0$, it is sufficient to test whether $\mathbb{E}(Z\mathbb{E}(Z|X'\beta)) = 0$ for all β of norm 1. In practice, an index is selected by maximizing an estimator of the previous quantity minus a penalty function. Our approach applies to many testing problems as explained in Section 2.1. We have applied it to testing for a parametric regression function. The test has known asymptotic critical values. It behaves against directional alternatives and against a class of low-dimensional regular alternatives almost as if the dimension of X was one. Our simulations results confirm the good power of the test.

Much work remains to be done to elaborate further on our general approach. From our fundamental lemma, other testing procedures could be constructed such as an integrated conditional moment test in the spirit of Bierens (1982). An automatic bandwidth choice should be proposed, in the line of Horowitz and Spokoiny (2001) or Guerre and Lavergne (2005). We are currently investigating these issues. Finally, future work should be devoted to applying our approach to other testing problems as outlined in this paper.

6 Technicalities

In the following, C and C' are positive constants that may vary from line to line.

Lemma 6.1 *Under Assumptions D–(a) and M and for a sequence of alternatives* $m_n(X) = \mu(X;\theta_0) + \delta_n(X)$ *with* $\mathbb{E}[\delta_n(X)\dot{\mu}(X;\theta_0)] = 0$, $\mathbb{E}\delta_n^2 = o(1)$, *and* $0 < \mathbb{E}[\delta_n^4(X)] < C < \infty$, $\|\widehat{\theta}_n - \theta_0\| = O_{\mathbb{P}}(n^{-1/2})$.

Proof. By a uniform law of large numbers for Euclidean families, see e.g. Pakes and Pollard (1989, Lemma 2.8),

$$\sup_{\theta}\left|\frac{1}{n}\sum_{i=1}^{n}\left\{\varepsilon_i + \mu(X_i;\theta_0) + \delta_n(X_i) - \mu(X_i;\theta)\right\}^2 - \mathbb{E}\left\{\varepsilon + \mu(X;\theta_0) - \mu(X;\theta)\right\}^2\right| = o_{\mathbb{P}}(1).$$

The Euclidean property is ensured by Assumption M–(a). As θ_0 is identifiable from Assumption M–(b), deduce that $\widehat{\theta}_n \xrightarrow{p} \theta_0$. Now, by definition of $\widehat{\theta}_n$,

$$\begin{aligned}
0 &\leq \frac{1}{n}\sum_{i=1}^{n}\left[\varepsilon_i + \delta_n(X_i)\right]^2 - \frac{1}{n}\sum_{i=1}^{n}\left[\varepsilon_i + \delta_n(X_i) - \left\{\mu(X_i;\widehat{\theta}_n) - \mu(X_i;\theta_0)\right\}\right]^2 \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left\{\mu(X_i;\widehat{\theta}_n) - \mu(X_i;\theta_0)\right\}^2 + \frac{2}{n}\sum_{i=1}^{n}\left\{\varepsilon_i + \delta_n(X_i)\right\}\left\{\mu(X_i;\widehat{\theta}_n) - \mu(X_i;\theta_0)\right\} \\
&\leq -\|\widehat{\theta}_n - \theta_0\|^2\left\{\frac{1}{n}\sum_{i=1}^{n}\Phi_3^2(X_i)\right\} + (\widehat{\theta}_n - \theta_0)'\left\{\frac{2}{n}\sum_{i=1}^{n}\left[\varepsilon_i + \delta_n(X_i)\right]\dot{\mu}(X_i;\theta_0)\right\} \\
&\quad + (\widehat{\theta}_n - \theta_0)'\left\{\frac{1}{n}\sum_{i=1}^{n}\left[\varepsilon_i + \delta_n(X_i)\right]\ddot{\mu}(X_i;\widehat{\theta}_n,\theta_0)\right\}(\widehat{\theta}_n - \theta_0) \\
&= : -A_n\|\widehat{\theta}_n - \theta_0\|^2 + (\widehat{\theta}_n - \theta_0)'B_n + (\widehat{\theta}_n - \theta_0)'C_n(\widehat{\theta}_n - \theta_0)\,.
\end{aligned}$$

Now $A_n - A = O_{\mathbb{P}}(n^{-1/2})$, where $A = \mathbb{E}\left[\Phi_3^2(X)\right] > 0$ and $\|B_n\| = O_{\mathbb{P}}(n^{-1/2})$. On the event $E_n = \{A_n \geq 3A/4\} \cap \{\mathrm{Sp}(C_n) \leq A/4\}$, where Sp denotes spectral radius, we then have

$$A\|\widehat{\theta}_n - \theta_0\|^2 - 2\,\|B_n\|\,\|\widehat{\theta}_n - \theta_0\| \leq 0\,,$$

that is $\|\widehat{\theta}_n - \theta_0\| \le 2A^{-1}\|B_n\|$. As $\widehat{\theta}_n \xrightarrow{p} \theta_0$ and by Assumption M–(a), $\mathbb{P}(E_n) \to 1$ and thus $\|\widehat{\theta}_n - \theta_0\| = O_{\mathbb{P}}(n^{-1/2})$. ■

Proof of Lemma 3.1. (i) Let $M = M_n$ depend on n in a way that will be specified later, define $\eta_i^M = \varepsilon_i \mathbb{I}(|\varepsilon_i| \le M) - \mathbb{E}(\varepsilon_i \mathbb{I}(|\varepsilon_i| \le M) \mid X_i)$ and consider the degenerate U-process

$$U_n\widetilde{g} = \frac{1}{n(n-1)}\sum_{j\neq i} \eta_i^M \eta_j^M K_h\left((X_i - X_j)'\beta\right) = \frac{1}{n(n-1)}\sum_{j\neq i} \widetilde{g}((\eta_i^M, X_i), (\eta_j^M, X_j); h, \beta)$$

defined by the functions $\widetilde{g}$ indexed by h and β with $\|\beta\| = 1$. By Assumption D and K–(a), Lemma 22(ii) of Nolan and Pollard (1987) and Lemma 2.14(ii) of Pakes and Pollard (1989), the family $\{\widetilde{g} : \|\beta\| = 1, h > 0\}$ is Euclidean for a constant envelope. By Theorem 2 of Major (2006) and its corollary (we may assume without loss of generality that $0 \le K(\cdot) \le 1$),

$$\begin{aligned}\mathbb{P}\left(\sup_\beta |U_n\widetilde{g}| \ge \frac{th^{1/2}\ln n}{n-1}\right) &= \mathbb{P}\left(\sup_\beta \left|\frac{1}{n}\sum_{j\neq i} \frac{\eta_i^M}{M}\frac{\eta_j^M}{M} K_h\left((X_i - X_j)'\beta\right)\right| \ge \frac{th^{1/2}\ln n}{M^2}\right) \\ &\le C_1C_2\exp\left\{-C_3\left(\frac{th^{1/2}\ln n}{M^2\sigma_M}\right)\right\} \qquad \text{for any } t > 0\ , \end{aligned} \tag{6.1}$$

$$\text{provided} \qquad n\sigma_M^2 \ge \frac{th^{1/2}\ln n}{M^2\sigma_M} \ge C_4\left[C_5 + \max\left(\ln C_2/\ln n, 0\right)\right]^{3/2} \ln\frac{2}{\sigma_M} \tag{6.2}$$

where $C_1, \ldots C_5 > 0$ are some constants independent on n, h and M and

$$\sigma_M^2 = \mathbb{E}\left[\left(\frac{\eta_i^M}{M}\right)^2\left(\frac{\eta_j^M}{M}\right)^2 K_h^2\left(X_i'\beta - X_j'\beta\right)\right].$$

Note that there exists some constant (independent of n) $C > 0$ such that $C^{-1} \le \sigma_M^2 M^4/h \le C$. In view of these bounds and of (6.2), take $M^4 = nh(\ln n)^{-(1+\delta)}$ with $\delta > 0$ arbitrarily small. This yields σ_M^2 of order $O(n^{-1}(\ln n)^{1+\delta})$ and then for any t

$$n\sigma_M^2 \ge \frac{nh}{CM^4} = C^{-1}\ln^{1+\delta} n \ge \frac{th^{1/2}\ln n}{M^2\sigma_M} \tag{6.3}$$

provided that n is large enough. On the other hand, for any constant $C' > 0$

$$\frac{th^{1/2}\ln n}{M^2\sigma_M} \ge C^{-1/2}t\ln n \ge C'\ln n \tag{6.4}$$

if t is sufficiently large. Since $(\ln n)^{-1}\ln(2/\sigma_M)$ tends to a positive constant as $n \uparrow \infty$, the equations (6.3) and (6.4) show that (6.2) is satisfied for this choice of M, if $t \uparrow \infty$ and $t\ln^{-\delta} n \downarrow 0$. Hence (6.1) yields $U_n\widetilde{g} = O_P\left(n^{-1}h^{1/2}\ln n\right)$. Now, it remains to study the tails of ε_i, that is we have to derive the orders of the remainder terms

$$2R_{1n} + R_{2n} = \tfrac{2}{n(n-1)}\sum_{j\neq i}\eta_i^M\xi_j K_h\left((X_i - X_j)'\beta\right) + \tfrac{1}{n(n-1)}\sum_{j\neq i}\xi_i\xi_j K_h\left((X_i - X_j)'\beta\right)$$
$$\text{where} \qquad \xi_i = \varepsilon_i - \eta_i^M = \varepsilon_i\mathbb{I}\left(|\varepsilon_i| > M\right) - \mathbb{E}\left[\varepsilon_i\mathbb{I}\left(|\varepsilon_i| > M\right) \mid X_i\right] .$$

First, $\mathbb{E}\left[\sup_\beta |R_{1n}|\right] \leq C\mathbb{E}\left(\left|\eta_i^M\right| |\xi_j|\right) \leq 2C\mathbb{E}\left(|\varepsilon_i|\right)\mathbb{E}\left(|\xi_j|\right) \leq C'\mathbb{E}\left(|\xi_j|\right)$, and thus by Hölder's and Chebyshev's inequalities

$$\mathbb{E}\left(|\xi_i|\right) \leq 2\mathbb{E}\left[|\varepsilon_i|\,\mathbb{I}\left(|\varepsilon_i| > M\right)\right] \leq 2\mathbb{E}^{1/11}\left[|\varepsilon_i|^{11}\right]\mathbb{P}^{10/11}\left[|\varepsilon_i| > M\right] \leq 2\mathbb{E}\left[|\varepsilon_i|^{11}\right] M^{-10}.$$

By Assumption K-(b) and our choice of M, $M^{-10} = o\left(n^{-1}h^{1/2}\ln n\right)$. Also it is clear that $\sup_\beta |R_{2n}|$ is of smaller order than $\sup_\beta |R_{1n}|$.

(ii) Let $V_n(\theta_0) = \left\{\theta \in \Theta : \|\theta - \theta_0\| \leq M/n^{1/2}\right\}$, with M not necessarily the same as in (i). By Lemma 6.1, $\liminf_{M\to\infty}\lim_{n\to\infty}\mathbb{P}\left[\widehat{\theta}_n \in V_n(\theta_0)\right] = 1$. Let $W = (\varepsilon, X')'$ and

$$g_{\theta,h,\beta}(W_i, W_j) = \varepsilon_i\left\{\mu(X_j;\theta) - \mu(X_j;\theta_0)\right\}K_h\left((X_i - X_j)'\beta\right) ,$$

which is such that $\mathbb{E}[g_{\theta,h,\beta}(W_i, W_j) \mid W_j] = 0$. From our assumptions, the class of functions $g_{\theta,h,\beta}(\cdot,\cdot)$, $\theta \in \Theta$, $h \in (0,1]$, $\|\beta\| = 1$, is Euclidean for a squared-integrable envelope $F\left(W_i, W_j\right) = |\varepsilon_i|\,\widetilde{\Phi}(X_j)$ where $\widetilde{\Phi}(\cdot) = C\sum_{i=1}^2\Phi_i(\cdot)$, for some suitable constant C, cf. Nolan and Pollard (1987, Lemma 22(ii)) and Pakes and Pollard (1989, Lemma 2.13 and Lemma 2.14 (ii)). Apply Hoeffding's decomposition to the U-process $hQ_{1n}(\theta,\beta)$ and consider the second order degenerate U-process in this decomposition $U_n\overline{g}_{\theta,h,\beta}$, with $\overline{g}_{\theta,h,\beta}(W_i, W_j) = g_{\theta,h,\beta}(W_i, W_j) - \mathbb{E}[g_{\theta,h,\beta}(W_i, W_j) \mid W_i]$. By Lemma 5 of Sherman (1994a), the family $\overline{g}_{\theta,h,\beta}$, $\theta \in \Theta$, $h \in (0,1]$, $\|\beta\| = 1$, is Euclidean for a squared-integrable envelope. From the Main Corollary of Sherman with $p = 1$ and $k = 2$,

$$\mathbb{E}\left[\sup_{\theta\in V_n(\theta_0),h,\beta}\left|nU_n\overline{g}_{\theta,h,\beta}\right|\right] \leq \Lambda\left[\mathbb{E}\sup_{\theta\in V_n(\theta_0),h,\beta}\left\{U_{2n}\overline{g}^2_{\theta,h,\beta}\right\}^{\overline{\alpha}}\right]^{1/2} \qquad (6.5)$$

where Λ is a universal constant and $0 < \overline{\alpha} < 1$. We have

$$\left|\overline{g}_{\theta,h,\beta}(W_i, W_j)\right| \leq C\left\|\theta-\theta_0\right\||\varepsilon_i|\left\{\widetilde{\Phi}\left(X_j\right) + \mathbb{E}[\widetilde{\Phi}\left(X_j\right) \mid W_i]\right\} \leq C'\left\|\theta-\theta_0\right\||\varepsilon_i|\{\widetilde{\Phi}\left(X_j\right)+1\}$$

for some constants C and C'. Hence from Inequality (6.5),

$$\mathbb{E} \sup_{\theta\in V_n(\theta_0),h,\beta} \left|nU_n\overline{g}_{\theta,h,\beta}\right| \leq \left(\frac{CM^2}{n}\right)^{\overline{\alpha}/2},$$

for some $C > 0$, and by Chebyshev's inequality,

$$\sup_{\|\beta\|=1} |nh^{-1/2}U_n\overline{g}_{\theta_n,h,\beta}| = O_{\mathbb{P}}\left((nh^{1/\overline{\alpha}})^{-\overline{\alpha}/2}\right). \tag{6.6}$$

We now study the U-process of order 1 in Hoeffding's decomposition of $hQ_{1n}(\theta,\beta)$. Let $\mathbb{P}_n\widetilde{g}$ denote this empirical process, where

$$\begin{aligned}
\widetilde{g}(W_i) &= \widetilde{g}_{\theta,h,\beta}(W_i) = \mathbb{E}[g_{\theta,h,\beta}(W_i,W_j) \mid W_i] \\
&= \varepsilon_i \mathbb{E}\left[\{\mu(X_j;\theta) - \mu(X_j;\theta_0)\} K_h\left((X_i - X_j)'\beta\right) \mid X_i\right] \\
&= (\theta-\theta_0)'\mathbb{E}\left[\dot{\mu}(X_j;\theta_0)K_h\left((X_i - X_j)'\beta\right) \mid X_i\right]\varepsilon_i \\
&\quad + (\theta-\theta_0)'\mathbb{E}\left[\ddot{\mu}(X_j;\theta,\theta_0)K_h\left((X_i - X_j)'\beta\right) \mid X_i\right](\theta-\theta_0)\varepsilon_i \\
&= : (\theta-\theta_0)'\widetilde{g}_1(W_i) + (\theta-\theta_0)'\widetilde{g}_2(W_i)(\theta-\theta_0)\ .
\end{aligned}$$

Let $\widetilde{g}_{1,s}(\cdot)$, $1 \leq s \leq d$, denote the components of $\widetilde{g}_1(\cdot)$. For each s, by our assumptions, Lemma 22(ii) of Nolan and Pollard (1987) and Lemma 5 of Sherman (1994), the family of functions $\widetilde{g}_{1,s}(\cdot)$, indexed by h and β is Euclidean for a squared integrable envelope. The Main Corollary of Sherman with $p = k = 1$ yields

$$\mathbb{E}\sup_{h,\beta}\left|n^{1/2}\mathbb{P}_n\widetilde{g}_{1,s}\right| \leq \Lambda\left[\mathbb{E}\sup_{h,\beta}\left\{\mathbb{P}_{2n}\widetilde{g}_{1,s}^2\right\}^{\overline{\alpha}}\right]^{1/2} \qquad s = 1,\ldots,p, \tag{6.7}$$

where Λ is a universal constant and $0 < \overline{\alpha} < 1$. If $\dot{\mu}_s(\cdot;\cdot)$ denotes the sth component of $\dot{\mu}(\cdot;\cdot)$,

$$\begin{aligned}
\left|\widetilde{g}_{1,s}\left(W_i\right)\right| &= \left|\mathbb{E}\left[\dot{\mu}_s(X_j;\theta_0)K_h\left((X_i{-}X_j)'\beta\right) \mid X_i\right]\varepsilon_i\right| \leq \mathbb{E}\left[\Phi_1(X_j)K_h\left((X_i{-}X_j)'\beta\right) \mid X_i\right]|\varepsilon_i| \\
&\leq \mathbb{E}^{1/4}\left[\Phi_1^4\left(X\right)\right]\mathbb{E}^{3/4}\left[K_h^{4/3}\left((X_i - X_j)'\beta\right) \mid X_i\right]|\varepsilon_i| \leq Ch^{3/4}\left|\varepsilon_i\right|
\end{aligned}$$

for some $C > 0$. From (6.7) and Chebyshev's inequality, $\sup_\beta \left\|n^{1/2}\mathbb{P}_n\widetilde{g}_1\right\| = O_{\mathbb{P}}(h^{3\overline{\alpha}/4})$, thus

$$\sup_{\theta\in V_n(\theta_0),\beta}\left|nh^{-1/2}(\theta-\theta_0)'\mathbb{P}_n\widetilde{g}_1\right| = O_{\mathbb{P}}\left(h^{(3\overline{\alpha}/4)-1/2}\right). \tag{6.8}$$

Use Assumption M(a)–(i) and apply similar arguments to each of the components of the square matrix $\widetilde{g}_2$ to show that

$$\sup_{\theta\in V_n(\theta_0),\beta}\left|nh^{-1/2}(\theta-\theta_0)'\mathbb{P}_n\widetilde{g}_2(\theta-\theta_0)\right| = O_{\mathbb{P}}\left(n^{-1/2}h^{(3\overline{\alpha}/4)-1/2}\right)\ . \tag{6.9}$$

From Equations (6.6), (6.8), and (6.9) with $\overline{\alpha} > 2/3$,

$$\sup_{\|\beta\|=1} |nh^{1/2}Q_{1n}(\widehat{\theta}_n, \beta)| = o_{\mathbb{P}}(1).$$

For $Q_{2n}(\theta, \beta)$, use the expansion of $\mu(\cdot;\theta)$ and similar arguments to show that

$$\sup_{\theta\in V_n(\theta_0), \|\beta\|=1} nh^{1/2}\left[Q_{2n}(\theta,\beta) - \mathbb{E}Q_{2n}(\theta,\beta)\right] = o_{\mathbb{P}}(1).$$

Last, for $\theta \in V_n(\theta_0)$,

$$\begin{aligned} |\mathbb{E}Q_{2n}(\theta,\beta)| &= \left|\mathbb{E}\left[\{\mu(X_i;\theta) - \mu(X_i;\theta_0)\}\{\mu(X_j;\theta) - \mu(X_j;\theta_0)\} h^{-1}K_h\left((X_i - X_j)'\beta\right)\right]\right| \\ &\le \|\theta - \theta_0\|^2 \mathbb{E}^{1/2}\left[\tilde{\Phi}^4(X)\right]\mathbb{E}^{3/4}\left[h^{-4/3}K_h^{4/3}\left((X_i - X_j)'\beta\right)\right] \\ &= O_{\mathbb{P}}(n^{-1}h^{-1/4}) = o_{\mathbb{P}}(n^{-1}h^{-1/2})\,. \quad \blacksquare \end{aligned}$$

For real random variables, $A_n \asymp_{\mathbb{P}} B_n$ means that $\mathbb{P}(C^{-1} \le A_n/B_n \le C)$ goes to 1 when n grows.

Lemma 6.2 *Let W_β be the matrix with generic element $\mathbb{I}(i \ne j)\, K_h\left((X_i - X_j)'\beta\right)/(h\, n(n-1))$. Under Assumptions D–(c) and K, $\mathrm{Sp}(W_\beta) = O_{\mathbb{P}}(n^{-1})$ and $nh^{1/2}\|W_\beta\| \asymp_{\mathbb{P}} 1$ for any β.*

Proof. By definition, $\mathrm{Sp}(W_\beta) = \sup_{u\ne 0}\|W_\beta u\|/\|u\|$ and for any $u \in \mathbb{R}^n$,

$$\begin{aligned} \|W_\beta u\|^2 &= \sum_{i=1}^n \left(\sum_{j=1, j\ne i}^n \frac{K_h\left((X_i - X_j)'\beta\right)}{h\, n(n-1)} u_j\right)^2 \\ &\le \sum_{i=1}^n \left(\sum_{j=1, j\ne i}^n \frac{K_h\left((X_i - X_j)'\beta\right)}{h\, n(n-1)}\right) \sum_{j=1, j\ne i}^n \frac{K_h\left((X_i - X_j)'\beta\right)}{h\, n(n-1)} u_j^2 \\ &\le \|u\|^2 \left[\max_{1\le i\le n}\left(\sum_{j=1, j\ne i}^n \frac{K_h\left((X_i - X_j)'\beta\right)}{h\, n(n-1)}\right)\right]^2 . \end{aligned}$$

Hence $n\mathrm{Sp}(W_\beta) \le \max_{1\le i\le n}\sum_{j\ne i}\frac{1}{(n-1)h}K_h\left((X_i - X_j)'\beta\right)$. For all j and β, $|K_h\left((x - X_j)'\beta\right)| \le$

C and $\mathrm{Var}\left[K_h\left((x-X_j)'\beta\right)\right] \le C$. Thus the Bernstein inequality yields for any $t>0$

$$
\begin{aligned}
&\mathbb{P}\left[\max_{1\le i\le n}\left(\frac{(nh^2)^\alpha}{\ln n}\right)^{1/2}\left|\sum_{j\neq i}\frac{1}{(n-1)}h^{-1}K_h\left((X_i-X_j)'\beta\right)-\mathbb{E}\left[h^{-1}K_h\left((X_i-X_j)'\beta\right)|X_i\right]\right|\ge t\right]\\
\le\ & \sum_{1\le i\le n}\mathbb{E}\left[\mathbb{P}\left[\left|\frac{1}{(n-1)}\sum_{j\neq i}K_h\left((X_i-X_j)'\beta\right)\right.\right.\right.\\
&\qquad\qquad\left.\left.\left.-\mathbb{E}\left[K_h\left((X_i-X_j)'\beta\right)|X_i\right]\right|\ge th\left(\frac{\ln n}{(nh^2)^\alpha}\right)^{1/2}\ |\ X_i\right]\right]\\
\le\ & 2n\exp\left(-\frac{t^2}{2}\frac{(nh^2)(\ln n)}{C((nh^2)^\alpha+th(nh^2)^{\alpha/2}(\ln n)^{1/2})}\right)\le 2\exp\left[\ln(n)-\frac{t^2}{C'}(\ln n)(nh^2)^{1-\alpha}\right]\to 0\ ,
\end{aligned}
$$

since $nh^2\to\infty$. Moreover, by Assumptions D-(c) and K-(a),

$$
\mathbb{E}\left[h^{-1}K_h\left((X_i-X_j)'\beta\right)|X_i\right]=\int K(u)f_\beta(X_i'\beta+hu)\,du\le C
$$

uniformly in i. This gives the first result. For the second result,

$$
n^2h\|W_\beta\|^2=\frac{1}{(n-1)^2}\sum_{i\neq j}\frac{1}{h}K_h^2\left((X_i-X_j)'\beta\right)\xrightarrow{p}\mathbb{E}\left[f_\beta(X'\beta)\right]\int K^2(u)\,du
$$

follows like in the proof of (6.12) below with $\delta(X)\equiv 1$ and $K(\cdot)$ replaced by $K^2(\cdot)/\int K^2(u)du$. The last quantity is bounded from above and below by Assumptions D-(c) and K-(a). ■

Proof of Theorems (3.4) and (3.5). By Assumption D-(b), $v_n^2(\beta)\le\overline{\sigma}^4n^2h\|W_\beta\|^2$, where W_β is the matrix with generic element $\mathbb{I}\left(i\neq j\right)K_h\left((X_i-X_j)'\beta\right)/(hn(n-1))$. Lemma 6.2 then ensures that $v_n^2(\beta)$ is bounded in probability from above for any β. Under H_{1n}, $U_i(\widehat{\theta}_n)=\mu(X_i;\theta_0)+r_n\delta(X_i)+\varepsilon_i-\mu(X_i;\widehat{\theta}_n)$. By simple algebra, $Q_n(\widehat{\theta}_n,\beta)$ writes for any β as

$$
Q_{0n}(\beta)-2Q_{1n}(\widehat{\theta}_n,\beta)+Q_{2n}(\widehat{\theta}_n,\beta)-2Q_{3n}(\widehat{\theta}_n,\beta)+2Q_{4n}(\beta)+Q_{5n}(\beta)\ ,
$$

$$
\begin{aligned}
\text{where}\quad Q_{3n}(\widehat{\theta}_n,\beta)&=\frac{r_n}{n(n-1)}\sum_{j\neq i}\delta(X_i)\left\{\mu(X_j;\widehat{\theta}_n)-\mu(X_j;\theta_0)\right\}\frac{1}{h}K_h\left((X_i-X_j)'\beta\right)\ ,\\
Q_{4n}(\beta)&=\frac{r_n}{n(n-1)}\sum_{j\neq i}\varepsilon_i\delta(X_j)\frac{1}{h}K_h\left((X_i-X_j)'\beta\right)\ ,\\
Q_{5n}(\beta)&=\frac{r_n^2}{n(n-1)}\sum_{j\neq i}\delta(X_i)\delta(X_j)\frac{1}{h}K_h\left((X_i-X_j)'\beta\right)\ .
\end{aligned}
$$

Since $v_n^2(\beta) = O_{\mathbb{P}}(1)$, $nh^{1/2}Q_{0n}(\beta) = O_{\mathbb{P}}(1)$ for any β. Lemma 3.1-(ii) deals with $Q_{1n}(\widehat{\theta}_n, \beta)$ and $Q_{2n}(\widehat{\theta}_n, \beta)$.

For Theorem 3.4, it is shown below that for any β

$$\begin{aligned} Q_{3n}(\widehat{\theta}_n, \beta) &= O_{\mathbb{P}}(r_n n^{-1/2}) && (6.10)\\ Q_{4n}(\beta) &= O_{\mathbb{P}}(r_n n^{-1/2}) && (6.11)\\ Q_{5n}(\beta) &= r_n^2 \mathbb{E}\left[\mathbb{E}^2[\delta(X)|X'\beta]f_\beta(X'\beta)\right] + o_{\mathbb{P}}(r_n^2)\,. && (6.12) \end{aligned}$$

Collecting results, it follows that for any β

$$\frac{nh^{1/2}Q_n(\widehat{\theta}_n, \beta) - \alpha_n}{v_n(\beta_0)(1 + o_{\mathbb{P}}(1))} \geq Cnh^{1/2}r_n^2 \left[\mathbb{E}\left[\mathbb{E}^2[\delta(X)|X'\beta]f_\beta(X'\beta)\right] + o(1)\right]\,.$$

Choose β such that $\mathbb{E}\left[\mathbb{E}^2[\delta(X)|X'\beta]f_\beta(X'\beta)\right] > 0$, which is possible from Lemma 2.1. The conclusion then follows from Inequality (3.4).

For Theorem 3.5, it is shown below that for any β

$$\begin{aligned} Q_{3n}(\widehat{\theta}_n, \beta) &= O_{\mathbb{P}}(n^{-1/2})\left[n^{-1}\sum_{i=1}^n \delta_n^2(X_i'\beta)\right]^{1/2} && (6.13)\\ Q_{4n}(\beta) &= O_{\mathbb{P}}(n^{-1/2})\left[n^{-1}\sum_{i=1}^n \delta_n^2(X_i'\beta)\right]^{1/2} && (6.14)\\ Q_{5n}(\beta) &\geq C\left(1 + o_{\mathbb{P}}(1)\right)\left[\sqrt{a_n}\left(n^{-1}\sum_{i=1}^n \delta_n^2(X_i'\beta)\right)^{1/2} - h^s\right]^2\,, && (6.15) \end{aligned}$$

where (6.15) holds with probability going to one. Collecting results, it follows that

$$\frac{nh^{1/2}Q_n(\widehat{\theta}_n, \beta) - \alpha_n}{v_n(\beta_0)(1 + o_{\mathbb{P}}(1))} \geq C\left(1 + o_{\mathbb{P}}(1)\right) nh^{1/2+2s}a_n\kappa_n^2 - \alpha_n = C\left(1 + o_{\mathbb{P}}(1)\right)\alpha_n a_n \kappa_n^2 - \alpha_n$$

diverges with probability going to one as $a_n\kappa_n^2$ diverges. ■

Proof of (6.10) and (6.13). Since $|u'W_\beta v| \le \|u\|\|v\|\mathrm{Sp}(W_\beta)$, then for any β,

$$\begin{aligned}
&\left|\frac{1}{n(n-1)}\sum_{j\neq i}\delta_n(X_i)\left\{\mu(X_j;\widehat{\theta}_n)-\mu(X_j;\theta_0)\right\}\frac{1}{h}K_h\left((X_i-X_j)'\beta\right)\right| \\
&\le\; n\left[n^{-1}\sum_{i=1}^n\delta^2(X_i)\right]^{1/2}\left[n^{-1}\sum_{i=1}^n\Big(\mu(X_i;\widehat{\theta}_n)-\mu(X_i;\theta_0)\Big)^2\right]^{1/2}\mathrm{Sp}(W_\beta) \\
&\le\; O_{\mathbb{P}}(1)\left[n^{-1}\sum_{i=1}^n\delta_n^2(X_i)\right]^{1/2}\left[n^{-1}\sum_{i=1}^n\Big(\mu(X_i;\widehat{\theta}_n)-\mu(X_i;\theta_0)\Big)^2\right]^{1/2},
\end{aligned}$$

by Lemma 6.2. Now by Assumption M, $\Big(\mu(X_i;\widehat{\theta}_n)-\mu(X_i;\theta_0)\Big)^2 \le \tilde{\Phi}^2(X_i)\|\widehat{\theta}_n-\theta_0\|^2$ for some $\tilde{\Phi}(\cdot)$ with bounded fourth moment and from Lemma 6.1 $n^{-1}\sum_{i=1}^n\Big(\mu(X_i;\widehat{\theta}_n)-\mu(X_i;\theta_0)\Big)^2 = O_{\mathbb{P}}(n^{-1})$. Hence the initial quantity is bounded by $O_{\mathbb{P}}(n^{-1/2})\left[n^{-1}\sum_{i=1}^n\delta_n^2(X_i)\right]^{1/2}$. ■

Proof of (6.11) and (6.14). Denote by $\mathbb{E}_n$ the conditional expectation given the X_i and let

$$\overline{\delta}_n(X_i)=\frac{1}{n(n-1)}\sum_{j=1,j\neq i}^n\delta_n(X_j)\frac{1}{h}K_h\left((X_i-X_j)'\beta\right).$$

Then Marcinkiewicz-Zygmund's and Jensen's inequalities implies that for any β, there is some constant C independent of n such that

$$\begin{aligned}
\mathbb{E}_n\left|\sum_{i=1}^n\varepsilon_i\overline{\delta}_n(X_i)\right| &\le C\mathbb{E}_n\left|\sum_{i=1}^n\varepsilon_i^2\overline{\delta}_n^2(X_i)\right|^{1/2}\le C\left\{\sum_{i=1}^n\overline{\delta}_n^2(X_i)\mathbb{E}_n(\varepsilon_i^2)\right\}^{1/2} \\
\le C\left\{\sum_{i=1}^n\overline{\delta}_n^2(X_i)\right\}^{1/2} &\le Cn^{1/2}\left\{n^{-1}\sum_{i=1}^n\delta_n^2(X_i)\right\}^{1/2}\mathrm{Sp}(W_\beta)=O_{\mathbb{P}}(n^{-1/2})\left\{n^{-1}\sum_{i=1}^n\delta_n^2(X_i)\right\}^{1/2},
\end{aligned}$$

using Lemma 6.2 and the weak law of large numbers. ■

Proof of (6.12). Consider $U_n=r_n^{-2}Q_{5n}(\beta)$. By straightforward computations,

$$\begin{aligned}
\mathrm{Var}\left(U_n\right) &\le \frac{C}{n}\mathrm{Var}\left[\delta(X_1)\delta(X_2)h^{-1}K_h\left((X_i-X_j)'\beta\right)\right] \\
&\le \frac{C}{n}\mathbb{E}\left[\delta^4(X)\right]\mathbb{E}^{1/2}\left[h^{-4}K_h^4\left((X_i-X_j)'\beta\right)\right]=O(n^{-1}h^{-3/2})=o(1).
\end{aligned}$$

Now, denoting by $\widehat{K}(\cdot)$ the Fourier transform of $K(\cdot)$,

$$\begin{aligned}\mathbb{E}\left(U_n\right) &= \mathbb{E}\left[\delta(X_1)\delta(X_2)h^{-1}K_h\left((X_1-X_2)'\beta\right)\right]\\ &= \frac{1}{2\pi}\mathbb{E}\left\{\delta(X_1)\delta(X_2)h^{-1}\int \exp\left(-it(X_1-X_2)'\beta/h\right)\widehat{K}(t)\,dt\right\}\\ &= \frac{1}{2\pi}\int \left|\mathbb{E}\left[\mathbb{E}[\delta(X)|X'\beta]\exp(itX'\beta)\right]\right|^2 \widehat{K}(ht)\,dt\,.\end{aligned}$$

As $\mathbb{E}\left[\delta(X)|X'\beta\right] f_\beta(X'\beta) \in L^1(\mathbb{R})\cap L^2(\mathbb{R})$, we obtain by the Plancherel theorem that

$$\frac{1}{2\pi}\int \left|\mathbb{E}\left[\mathbb{E}[\delta(X)|X'\beta]\exp(itX'\beta)\right]\right|^2\,dt = \mathbb{E}\left[\mathbb{E}^2[\delta(X)|X'\beta]f_\beta(X'\beta)\right]\ ,$$

see Rudin (1987). Since $\left|\widehat{K}(\cdot)\right| \leq 1$ and $\widehat{K}(0)=1$, the Lebesgue dominated convergence theorem yields $\mathbb{E}\left(U_n\right) \to \mathbb{E}\left[\mathbb{E}^2[\delta(X)|X'\beta]f_\beta(X'\beta)\right]$. ■

Proof of (6.15). Let $\overline{W}_\beta$ be the matrix with generic element

$$\mathbb{I}(X_i'\beta, X_j'\beta \in A_{\beta,a_n})\mathbb{I}\left(i\neq j\right)K_h\left((X_i-X_j)'\beta\right)/(hn(n-1))\,.$$

Let $\delta_n = (\delta_n(X_1'\beta),\ldots\delta_n(X_n'\beta))'$, and define $\delta_n\mathbb{I}_{A_{\beta,a_n}}$ and $\delta_n\mathbb{I}_{A^c_{\beta,a_n}}$ similarly, where A^c stands for the complementary set of A. From Condition 3.10,

$$|\delta_n'(W_\beta-\overline{W}_\beta)\delta_n| \leq \mathrm{Sp}(W_\beta)\left[2\|\delta_n\mathbb{I}_{A_{\beta,a_n}}\|\,\|\delta_n\mathbb{I}_{A^c_{\beta,a_n}}\| + \|\delta_n\mathbb{I}_{A^c_{\beta,a_n}}\|^2\right] = o_{\mathbb{P}}(a_n)\left[n^{-1}\sum_{i=1}^n \delta_n^2(X_i'\beta)\right].$$

Without loss of generality, assume that the bounded support of $X'\beta$ is $[0,1]$ and that $\overline{k}=h^{-1}$ is an integer number strictly larger than 1. For $k=0,\ldots\overline{k}-1$, let $t_k=t_k(h)=(k+1/2)h$, and $J_k=J_k(h)=t_k(h)+h[1/2,1/2)$, so that $\cup_{k=0,\ldots\overline{k}-1}J_k=[0,1)$. We can assume that $t_k\in A_{\beta,a_n}$ iff $J_k\subset A_{\beta,a_n}$. If not, by Assumption N-(b) we need to redefine at most $4k_0$ sets J_k as $J_k=t_k(h)+h(-a,b)$ with $0\leq a,b\leq 1$. Let $\Pi_{s,h}$ be the set of piecewise polynomial functions over the J_k with degree smaller than or equal to $\lfloor s\rfloor$. Then the definition of $C(L,s)$ and a standard piecewise Taylor expansion around the t_k show that $\sup_x|\delta_n(x'\beta)-\pi_n(x'\beta)|\leq Ch^s$ for some $\pi_n(\cdot)\in\Pi_{s,h}$. Denote by $\overline{P}_\beta$ the matrix with the same elements than $\overline{W}_\beta$ but with generic element $\mathbb{I}(X_i'\beta\in A_{\beta,a_n})K(0)/(n(n-1)h)$ on the diagonal. Then $\mathrm{Sp}(\overline{W}_\beta-\overline{P}_\beta)=O_{\mathbb{P}}(n^{-2}h^{-1})$

and

$$
\begin{aligned}
u'\overline{P}_\beta u &= \frac{1}{n(n-1)} \sum_{i,j} u_i u_j \, \mathbb{I}\left(X_i'\beta \in A_{\beta,a_n}\right) \mathbb{I}\left(X_j'\beta \in A_{\beta,a_n}\right) h^{-1} K_h\left((X_i - X_j)'\beta\right) \\
&= \frac{n}{2\pi(n-1)h} \int \left| n^{-1} \sum_i u_i \exp\left(\frac{it\, X_i'\beta}{h}\right) \mathbb{I}\left(X_i'\beta \in A_{\beta,a_n}\right)\right|^2 \widehat{K}(t)\, dt \qquad (6.16)
\end{aligned}
$$

is nonnegative and vanishes iff $u_i \mathbb{I}\left(X_i'\beta \in A_{\beta,a_n}\right) = 0$ for all i. For $\pi(\cdot) \in \Pi_{s,h}$, let π denote the column vector $(\pi(X_1'\beta), ..., \pi(X_n'\beta))'$ and let

$$
\Lambda_n^2 = \inf_{\pi(\cdot)\in\Pi_{s,h}} \frac{\pi'\overline{P}_\beta \pi}{n^{-1}\sum_{i=1}^n \ \pi^2(X_i\beta)\mathbb{I}(\ X_i\beta \in A_{\beta,a_n})} \ ,
$$

with the convention $0/0 = 1$. Consider the polynomial functions $\pi_{b,k}(\cdot)$ of degree $\lfloor s \rfloor$ with support $J_k(h)$, i.e.

$$
\pi_{b,k}(t) = \pi_b\left(\frac{t-t_k}{h}\right) = \sum_{0\le q\le \lfloor s\rfloor} b_q \left(\frac{t-t_k}{h}\right)^q \mathbb{I}\left(t \in J_k\right) \ .
$$

Then

$$
\Lambda_n^2 \ge \inf_{b, J_k \subset A_{\beta,a_n}} \frac{\pi_{b,k}'\overline{P}_\beta \pi_{b,k}}{n^{-1}\sum_{i=1}^n \pi_{b,k}^2(X_i'\beta)\mathbb{I}(X_i'\beta \in A_{\beta,a_n})} \ , \qquad (6.17)
$$

and we restrict to b in the unit hypersphere by homogeneity. The Main Corollary of Sherman (1994a) implies that uniformly in b and k such that $J_k \subset A_{\beta,a_n}$

$$
\begin{aligned}
&\frac{1}{nh}\sum_{i=1}^n \pi_b^2\left(\frac{X_i'\beta - t_k}{h}\right)\mathbb{I}(X_i'\beta \in A_{\beta,a_n}) = \frac{1}{h}\mathbb{E}\Big[\pi_b^2\left(\frac{X'\beta - t_k}{h}\right)\mathbb{I}(X'\beta \in A_{\beta,a_n})\Big] + O_{\mathbb{P}}\left(\frac{1}{\sqrt{nh}}\right) \\
&= \int_{(-1/2,1/2)} \pi_b^2(u) f_\beta(t_k + hu)\, du + o_{\mathbb{P}}(a_n) = f_\beta(t_k)\int_{(-1/2,1/2)} \pi_b^2(u)\, du + o_{\mathbb{P}}(a_n) \ ,
\end{aligned}
$$

from Assumption N and $a_n^{-1}h^a \to 0$. Similarly, uniformly in b, t and k such that $J_k \subset A_{\beta,a_n}$

$$
\begin{aligned}
&\frac{1}{\sqrt{2\pi}nh}\sum_{i=1}^n \pi_b\left(\frac{X_i'\beta - t_k}{h}\right)\exp\left(it\frac{X_i'\beta - t_k}{h}\right)\mathbb{I}(X_i'\beta \in A_{\beta,a_n}) \\
&= \frac{1}{\sqrt{2\pi}h}\mathbb{E}\left[\pi_b\left(\frac{X'\beta - \ t_k}{h}\right)\exp\left(it\frac{X'\beta - \ t_k}{h}\right)\ \mathbb{I}(X'\beta \in A_{\beta,a_n})\right] + O_{\mathbb{P}}\left(\frac{1}{\sqrt{nh}}\right) \\
&= \frac{1}{\sqrt{2\pi}}\int_{(-1/2,1/2)} \pi_b^2(u)\exp(itu) f_\beta(t_k + hu)\, du + o_{\mathbb{P}}(a_n) = f_\beta(\ t_k)\widehat{\pi}_b(t) + o_{\mathbb{P}}(a_n) \ .
\end{aligned}
$$

Using (6.16) and $t_k \in A_{\beta,a_n}$ implies that with probability going to one

$$
\frac{1}{a_n}\frac{\pi_{b,k}'\overline{P}_\beta \pi_{b,k}}{n^{-1}\sum_{i=1}^n \pi_{b,k}^2(\ X_i\beta)\mathbb{I}(\ X_i\beta \in A_{\beta,a_n})} \ge \frac{\int \left|\widehat{\pi}_b(t)\right|^2 \widehat{K}(t)\, dt}{\int_{(-1/2,1/2)} \pi_b^2(u)\, du} \ , \qquad (6.18)
$$

which stays away from zero for b in the unit sphere. Hence $\Lambda_n^2 \geq C\, a_n$ for some $C > 0$ with probability going to one. Now, from the triangular inequality,

$$\begin{aligned}
\left(\delta_n' \overline{P}_\beta \delta_n\right)^{1/2} &\geq \left(\pi_n' \overline{P}_\beta \pi_n\right)^{1/2} - \left((\delta_n - \pi_n)' \, \overline{P}_\beta \, (\delta_n - \pi_n)\right)^{1/2} \\
&\geq \Lambda_n n^{-1/2} \|\pi_n \mathbb{I}_{A_{\beta,a_n}}\| - \mathrm{Sp}^{1/2}(\overline{P}_\beta)\, \|\delta_n - \pi_n\| \\
&\geq \Lambda_n n^{-1/2} \left\|\delta_n \mathbb{I}_{A_{\beta,a_n}}\right\| - \left(\Lambda_n + n^{1/2}\mathrm{Sp}^{1/2}(\overline{P}_\beta)\right) C h^s \, . \qquad (6.19)
\end{aligned}$$

From Conditions (3.9), (3.10) and the fact that $\mathrm{Sp}^{1/2}(\overline{P}_\beta) \leq \mathrm{Sp}^{1/2}(\overline{W}_\beta) + \mathrm{Sp}^{1/2}(\overline{W}_\beta - \overline{P}_\beta) = O_{\mathbb{P}}(n^{-1/2})$ the above quantity is positive with probability going to one. Since $\Lambda_n^{-2} n \mathrm{Sp}(\overline{W}_h - \overline{P}_h) = o_{\mathbb{P}}(1)$, deduce that with probability going to one

$$\begin{aligned}
\delta_n' \overline{W}_\beta \delta_n &\geq \left[\, \Lambda_n n^{-1/2} \left\|\delta_n \mathbb{I}_{A_{\beta,a_n}}\right\| - \left(\Lambda_n + n^{1/2}\mathrm{Sp}^{1/2}(\overline{P}_\beta)\right) C h^s \right]^2 - \mathrm{Sp}(\overline{W}_h - \overline{P}_h) \left\|\delta_n \mathbb{I}_{A_{\beta,a_n}}\right\|^2 \\
&\geq \left[\left(\Lambda_n - n^{1/2}\mathrm{Sp}^{1/2}(\overline{W}_h - \overline{P}_h)\right) n^{-1/2} \left\|\delta_n \mathbb{I}_{A_{\beta,a_n}}\right\| - \left(\Lambda_n + n^{1/2}\mathrm{Sp}^{1/2}(\overline{P}_\beta)\right) C h^s \right]^2 \\
&\geq C\,(1 + o_{\mathbb{P}}(1)) \left[\sqrt{a_n} \left(n^{-1} \sum_{i=1}^n \delta_n^2(X_i'\beta)\right)^{1/2} - h^s \right]^2 . \quad \blacksquare
\end{aligned}$$

REFERENCES

Aït-Sahalia, Y., Bickel, P., and Stoker, T. (2001) Goodness-of-fit tests for kernel regression with an application to option implied volatilities. *J. Econometrics* **105**, 363-412.

Andrews, D.W.K. (1997). A conditional Kolmogorov test. *Econometrica* **65**, 1097-1128.

Bierens, H.J. (1982). Consistent model specification tests. *J. Econometrics* **20**, 105-134.

Bierens, H.J. (1990). A consistent conditional moment test of functional form. *Econometrica* **58**, 1443-1458.

Bierens, H.J. and Ploberger W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica* **65**, 1129-1151.

Chen, X., and Fan, Y. (1999). Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series. *J. Econometrics* **91**, 373-401.

Chen, P.D. (1991). L^1-theory of approximation by ridge functions. *Chinese Sci. Bull.* **36**, 1761-1764.

Chernozhukov, V., Hong, H., and Tamer, E. (2004). Inference on parameter sets in econometric models. Working paper, Massachusetts Institute of Technology.

Delecroix, M., Hristache, M., and Patilea, V. (2006). On semiparametric $M-$estimation in single-index regression. *J. Statist. Plann. Inference* **136**, 730-769.

Delgado, M.A. and González-Manteiga, W. (2001). Significance testing in nonparametric regression based on the bootstrap. *Ann. Statist.* **29**, 1469-1507.

Delgado, M.A., Dominguez, M.A., and Lavergne, P. (2006). Consistent tests of conditional moment restrictions. *Annales d'Economie et Statistique*, 81, 33-67.

Donald, S.G., Imbens, G.W. and Newey, W.K. (2003). Empirical likelihood estimation and consistent tests with conditional moment restrictions. *J. Econometrics* **117**, 55-93.

Fan, Y., and Li, Q. (1996). Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica* **64**, 865-890.

Gozalo, P.L. (1997). Nonparametric bootstrap analysis with applications to demographic effects in demand functions. *J. Econometrics* **81**, 357-393.

Gozalo, P.L., and Linton, O.B. (2001). Testing additivity in generalized nonparametric regression models with estimated parameters. *J. Econometrics* **104**, 1-48.

Guerre, E., and Lavergne, P. (2002). Optimal minimax rates for nonparametric specification testing in regression models. *Econometric Theory* **18**, 1139-1171.

Guerre, E., and Lavergne, P. (2005). Data-driven rate-optimal specification testing in regression models. *Ann. Statist.* **33**, 840-870.

Hall, P., Marron, J.S., Neumann, M.H., and Titterington, D.M. (1997). Curve estimation when the design density is low. *Ann. Statist.* **25**, 756-770.

Hall, P., and Yatchew, A. (2005). Unified approach for testing functional hypotheses in semiparametric contexts. *J. Econometrics* **127**, 225-252.

Härdle, W., and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21**, 1296-1947.

Härdle, W., and Stoker, T. (1989). Investigating smooth multiple regression by the method of average derivatives. *J. Amer. Statist. Assoc.* **84**, 986-995.

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

Hong, Y., and White, H. (1995). Consistent specification tests via nonparametric series regression. *Econometrica* **63**, 1133-1159.

Horowitz, J.L., and Spokoiny, V.G. (2001). An adaptive, rate-optimal test of a parametric model against a nonparametric alternative. *Econometrica* **69**, 599-631.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J. Econometrics* **58**, 71-120.

Lavergne, P. (2001). An equality test across nonparametric regressions. *J. Econometrics* **103**, 307-344.

Lavergne, P., and Vuong, Q. (2000). Nonparametric significance testing. *Econometric Theory* **16**, 576-601.

Li, Q., and Wang, S. (1998). A simple consistent bootstrap test for a parametric regression function. *J. Econometrics* **87**, 145-165.

Major, P. (2006). An estimate on the supremum of a nice class of stochastic integrals and U-statistics. *Probab. Theory Related Fields* **134(3)**, 489-537.

Nolan, D., and Pollard, D. (1987). $U-$processes : Rates of convergence. *Ann. Statist.* **15**, 780-799.

Pakes, A., and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica* **57**, 1027-1057.

Powell, J.L., Stock, J.H. and Stoker, T.M. (1989). Semiparametric estimation of index coefficients. *Econometrica* **57**, 1403-1430.

Ramsey, J.B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **31**, 511-534.

Roy, S.N. (1953). On a heuristic method of test construction and its use in multivariate analysis. *Ann. Math. Statist.* **24**, 220-238.

Rudin, W. (1987). *Real and complex analysis.* McGraw-Hill.

Sherman, R.P. (1994a). Maximal inequalities for degenerate $U-$processes with applications to optimization estimators. *Ann. Statist.* **22**, 439-459.

Sherman, R.P. (1994b). U-processes in the analysis of a generalized semiparametric regression estimator. *Econometric Theory* **10**, 372-395.

Spokoiny, V.G. (1996). Adaptive hypothesis testing using wavelets. *Ann. Statist.* **24**, 2477-2498.

Stinchcombe, M.B., and White, H.. (1998). Consistent specification testing with nuisance parameters present only on the alternative. *Econometric Theory* **14**, 295-324.

Stoker, T.M. (1986). Consistent estimation of scaled coefficients. *Econometrica* **54**, 1461-1481.

Stone, C.J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8**, 1348-1360.

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

Stute, W., and Zhu, L.X. (2005). Nonparametric checks for single-index models. *Ann. Statist.* **33**, 1048-1083.

Whang, Y.J. (2001). Consistent specification testing for conditional moment restrictions. *Economics Letters* **71**, 299–306.

Wu, C.F.J. (1986). Jacknife, bootstrap and other resampling methods in regression analysis (with discussion). *Ann. Statist.* **14**, 1261-1350.

Xia, Y., Li, W.K., Tong, H., and Zhang, D. (2004). A goodness-of-fit test for single-index models (with discussion). *Statist. Sinica* **14**, 1-39.

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

Zhu, L.X., and Li R. (1998). Dimension-reduction type test for linearity of a stochastic model. *Acta Math. Appli. Sinica* **14**, 165-175.

Zhu, L.X. (2003). Model checking of dimension-reduction type for regression. *Statist. Sinica* **3**, 283-296.

Figure 1: $\beta_0 = (1,1,1)/\sqrt{3}$

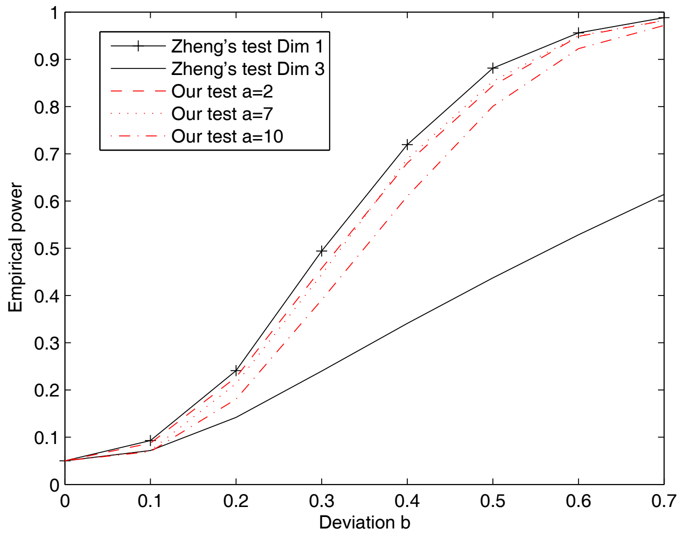

Figure 2: $\beta_0 = (1,2,3)/\sqrt{14}$

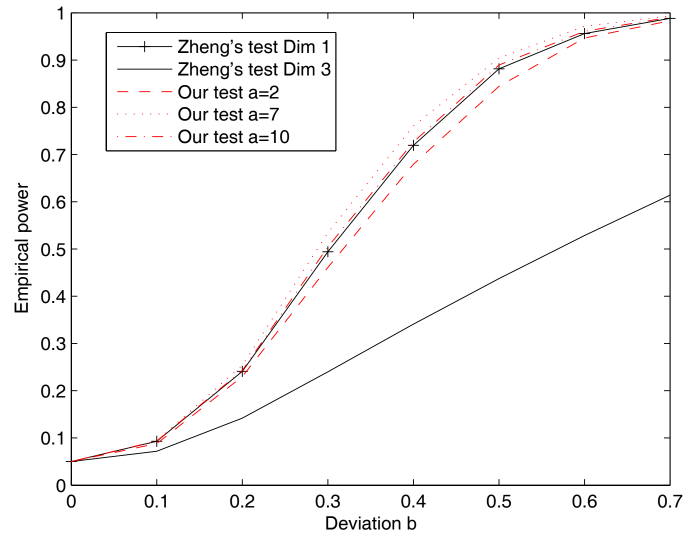

Figure 3: $\beta_0 = (0,1,0)$

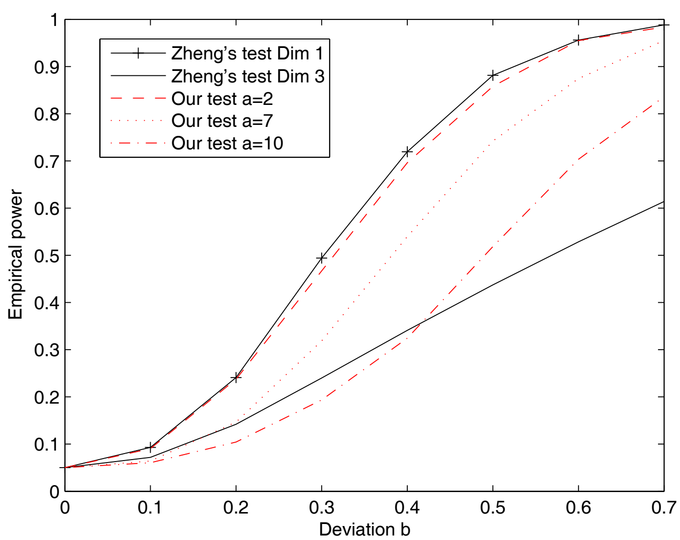

Figure 4: $\beta_0 = (1,1,1)/\sqrt{3}$ and $b = 0.4$

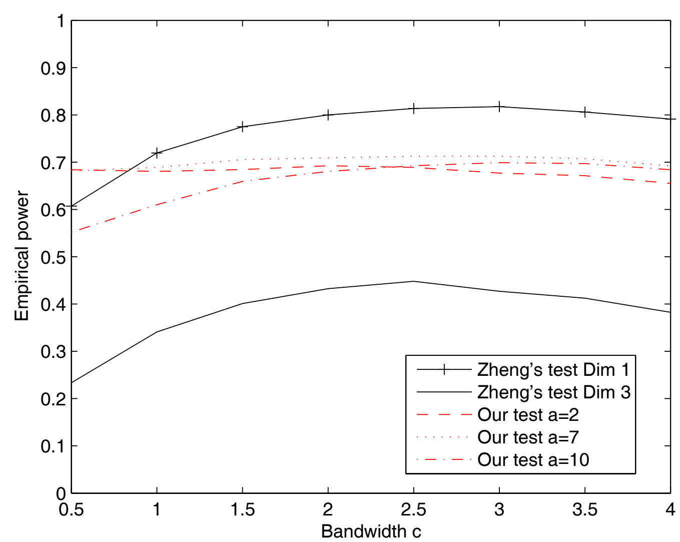

Figure 5: $\beta_0 = (1,1,1)/\sqrt{3}$

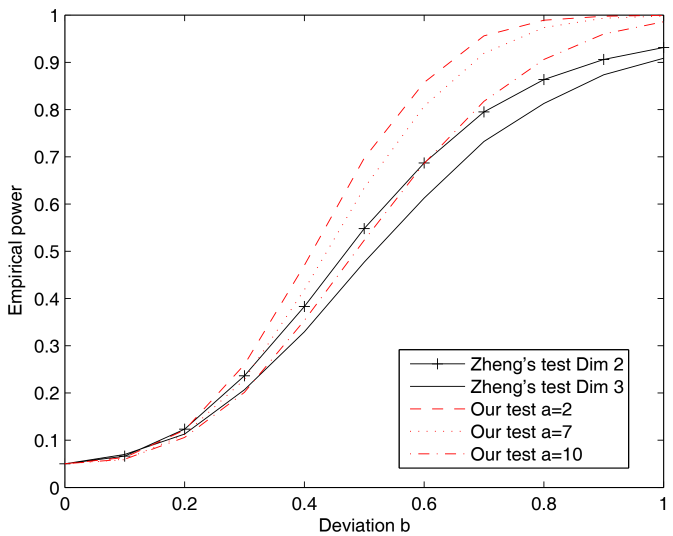

Figure 6: $\beta_0 = (3,1,0)/\sqrt{10}$

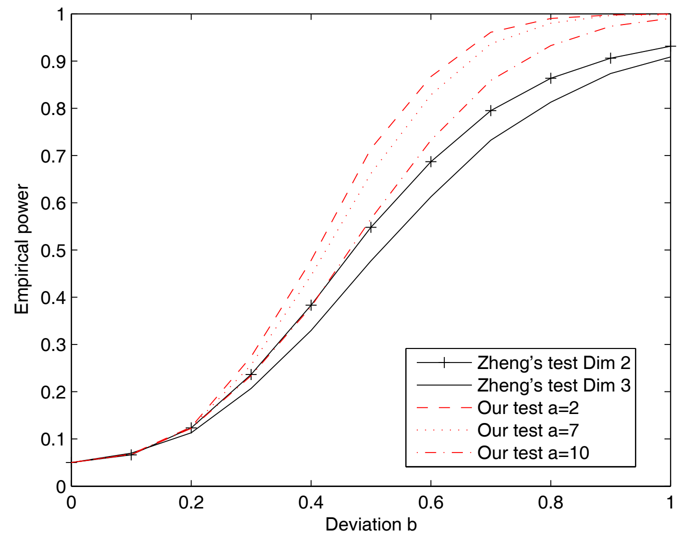

Figure 7: $\beta_0 = (0,1,0)$

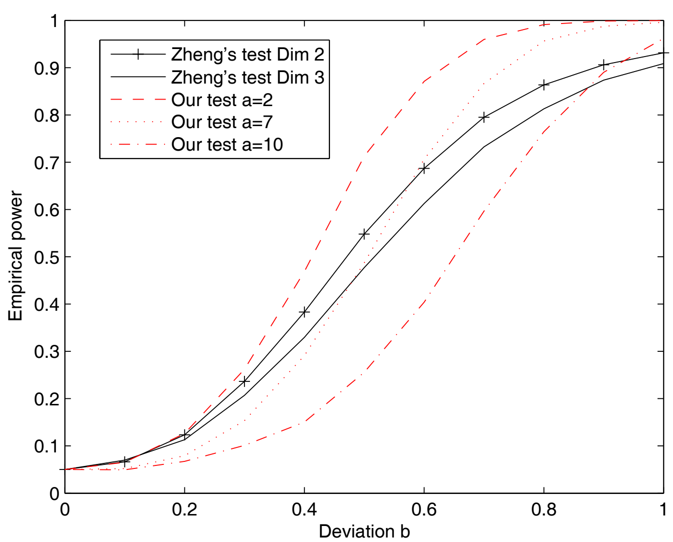

Figure 8: $\beta_0 = (1,1,1)/\sqrt{3}$ and $b = 0.5$