Breaking the curse of dimensionality in nonparametric testing

Pascal Lavergne
University of Toulouse (GREMAQ and INRA)

Valentin Patilea
CREST-ENSAI

June 30th, 2005

Incomplete version

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 general method to circumvent this problem 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.

Keywords: Curse of dimensionality, Testing, Nonparametric methods.

JEL classification: Primary C52; Secondary C14.

Emails: Pascal.Lavergne@univ-tlse1.fr patilea@ensai.fr

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 dimensionality-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 (2005).

The curse of dimensionality also appears when using smoothing methods for testing, since the variables dimension adversely affects the power of the test. 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. As illustrated by our simulations in Section 4, this is not only of theoretical interest but also of practical relevance.

The purpose of this paper is to propose a cure for the curse of dimensionality in nonparametric testing. In Section 2, we propose a general method that could apply to many nonparametric testing problems. To show the potential benefits of our method, we apply our method to testing for the parametric form of a regression and derive its properties in Section 3. Our main finding is that the resulting test behaves against directional alternatives as if the dimension of the regressors was *almost* one. Hence there is a cost to dimensionality, but this cost can be set arbitrarily low and is paid only once, as it does not increase with the dimension of the regressors. Our simulation study in Section 4 confirms these theoretical findings.

Before entering into details, let us give a flavor of our general method. As shown in

Section 2, many testing problems consider a null hypothesis of the form

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

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 any value of the variables X. Our proposal for reducing the dimensionality of the problem is to use single linear indices $X'\beta$ as conditioning variables instead of X and to then to look at $\mathbb{E}\left[U(\theta_0)|X'\beta\right]$ for all directions defined by β in $\mathbb{R}^q$. This clearly helps in breaking the curse of dimensionality, but is it sufficient to obtain a consistent test? Our fundamental lemma in Section 2 shows that it is indeed equivalent to search among single linear indices $X'\beta$ or among all possible values of X. Then a consistent test can be built by looking for an index that makes $\mathbb{E}\left[U(\theta_0)|X'\beta\right]$ furthest away from zero. This should yield high power for the test. However, this method cannot be used directly as under the null hypothesis, every index $X'\beta$ should yield a zero function. Hence, we introduce a penalized criterion that favors one of the index and then yields a simple behavior under H_0. Our dimensionality-reduction technique is thus the testing counterpart of single-index model in estimation, but with the fundamental difference that the function under test does not need to depend on a single-index only.

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

$$H_0: \ \mathbb{E}\left[U(\theta_0)|X\right] = 0 \quad \text{a. s. for some } \theta_0 \in \Theta \,, \tag{2.1}$$

with $X \in \mathbb{R}^q$. 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. A large literature has been devoted to checking the functional form of a regression function. 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 also Hart (1997) for a review. When looking at the behavior of tests against alternatives of the form

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

one usually finds that r_n should be of higher order than $n^{-1/2}h^{-q/4}$ for consistency of the test. When looking at the uniform consistency of the test against a class of alternatives of known smoothness s, it is found that the alternatives should be at distance $n^{-2s/4s+q}$ for consistency, 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 for instance 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 have trivial uniform power 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 example is testing for homoscedasticity, where $\rho(Y, X, \theta) = Y^2 - \theta^2$. Delgado, Dominguez and Lavergne (2005) provide several more examples. Stinchcombe and White (1998), Koul and Stute (1999), and Whang (2001) study single conditional moment restrictions, Donald, Imbens and Newey (2003) and Delgado, Dominguez and Lavergne (2005) study multiple ones.

Example 3: Testing restrictions in semiparametric and nonparametric models. Constructing a test for such restrictions is usually more technically involved, as the unknown parameter θ is infinite-dimensional. When testing for additivity, $U(\theta) = Y - \sum_{l=1}^{q} m_l(X_l)$, where the unknown functions $m_l(X_l)$ are properly normalized, see Gozalo and Linton (2001). Another example is testing for the significance of some regressors X_1 in a nonparametric regression on $X = (X_1, Z)$, i.e. $U(\theta) = Y - \mathbb{E}[Y|X_1]$, see Fan and Li (1996), Lavergne and Vuong (2000), Aït-Sahalia, Bickel and Stoker (2001), Delgado and Gonzalez-Manteiga (2001), Lavergne (2001). Chen and Fan (1999) consider other types of restrictions, such as portfolio efficiency or rational expectation behavior.

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

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

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 (random) X, but also by (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 method could be generalized to these more general hypotheses, for instance by rewriting H_0 as a conditional moment restriction upon X only through an integral over the domain of y, see Hall and Yatchew (2005) on this method.

2.2 The fundamental lemma

Our method 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 *Consider random vectors $Z \in \mathbb{R}^c$ and $X \in \mathbb{R}^q$. Assume that $\mathbb{E}(\|Z\|) < \infty$, where $\|\cdot\|$ denotes the Euclidean norm.*

A) The following statements are equivalent:

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

$$\mathbb{E}(Z \mid X'\beta) = \mathbb{E}(Z) \quad \textit{almost surely.} \tag{2.2}$$

(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 contours 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 (2.2) 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 $A = \{\beta \in \mathbb{R}^q : \mathbb{E}\left[\exp\{iX'\beta\}Z\right] = 0\}$ 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 $\mathbb{P}\left[\mathbb{E}(Z \mid X) = 0\right] < 1$,

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

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

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 $\theta \in \Theta$ and $X \in \mathbb{R}^q$. Assume that $\mathbb{E}(\|U(\theta)\|) < \infty$ for all θ. Then H_0 in (2.1) is equivalent to*

$$\max_{\|\beta\|=1} \mathbb{E}\left[U'(\theta_0)\mathbb{E}\left(U(\theta_0)|X'\beta\right)\omega(X'\beta)\right] = 0 \quad \text{for some } \theta_0 \in \Theta \tag{2.3}$$

for a function $\omega(\cdot)$ such that for any β, $\omega(X'\beta) > 0$ on the support of $\mathbb{E}\left(U(\theta_0)|X'\beta\right)$.

Part B of Lemma 2.1 is a version of Theorem 1 of Bierens (1990), who showed that $\mathbb{E}\left[Z \mid X\right] = 0$ is equivalent to $\mathbb{E}\left[Z \exp\{X'\beta\}\right] = 0$ for all β, see also Bierens (1982). Stinchcombe and White extended Bierens' result showing that $\mathbb{E}\left[Z \mid X\right] = 0$ is equivalent to $\mathbb{E}\left[Z\phi\left(X'\beta\right)\right] = 0$ for all β whenever $\phi\left(\cdot\right)$ is analytic nonpolynomial. These authors then write the null hypothesis as

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

Our approach is closely related to theirs, but different in a key aspect. Instead of working with a particular known $\phi\left(\cdot\right)$ 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 $\mathbb{E}(Z \mid X'\beta = \cdot)$, 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 (2.3) with $\omega\left(\cdot\right) \equiv 1$. This is in the spirit of the well-known *union-intersection principle* in the classical multivariate analysis (cf. Roy, 1953). A similar reasonning 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 method

In view of the previous corollary, our goal is to estimate the quantity in (2.3). For simplicity, we consider that U is unidimensional. Assume we have at our disposal a consistent estimator $\widehat{\theta}_n$ of θ_0 and denote by $U_i(\theta)$ the data-dependent function of θ. Let $\widehat{\gamma}_i(X_i'\beta, \theta)$ be a consistent estimator of $\mathbb{E}\left(U(\theta)|X_i'\beta\right)\omega(X_i'\beta)$. Define

$$Q_n\left(\theta, \beta\right) = \frac{1}{n}\sum_{i=1}^{n} U_i\left(\theta\right)\widehat{\gamma}_i\left(X_i'\beta; \theta\right) \ . \tag{2.4}$$

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

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

Thus a consistent test could be based on the maximum of $Q_n(\widehat{\theta}_n, \beta)$ over β. When the restriction given by H_0 does not hold, this maximum should stay away from zero almost surely and a test based on it should be consistent. Under H_0 however, $Q_n(\widehat{\theta}_n, \beta)$ converges to zero for any β. This means that the theoretical behavior of a test based on this maximum would be difficult to study. More importantly, considering $\max_{\|\beta\|=1} Q_n(\widehat{\theta}_n, \beta)$ would yield high critical values and then low power. To avoid this, we introduce a penalized criterion and define

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

Here β_0 is a fixed vector chosen by the practitioner. The penalty $\pi_n(\cdot)$ is a nonnegative function that equals zero only at zero. This forces the maximum to be attained at β_0 under H_0, provided the penalty is large enough with respect to $\max_{\|\beta\|=1} Q_n(\widehat{\theta}_n, \beta)$. However, the penalty should not perturb the behavior of the maximum when H_0 does not hold, hence we should have $\pi_n(t) \to 0$ fast enough for all t as n grows. A normalized version of $Q_n(\widehat{\theta}_n, \widehat{\beta}_n)$ is then taken as the test statistic.

Our penalization method is related to Tikhonov's regularization in estimation, see Carrasco, Florens and Renault (2006), but applied to a testing problem. As will become apparent in the next section, the choice of the vector β_0 is irrelevant for the general power of the test. Though, the test would 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. By contrast, the choice of the penalty is crucial to control the level of the test and to ensure high power.

3 Testing for a parametric regression

3.1 The test

Consider a random vector $(Y, X')' \in \mathbb{R}^{1+q}$. 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." Consider the parametric regression model $\{\mu(\cdot;\theta) : \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^d$. The model is correctly specified if and only if H_0 as defined in (2.1) holds with $U(\theta) = Y - \mu(X;\theta)$. We thus apply our general method to this testing problem using kernel estimators. To avoid handling denominators close to zero, set the weight function $\omega(\cdot)$ in (2.3) equal to the density of $X'\beta$, denoted by $f_\beta(\cdot)$, which is assumed to exist for any β. Define

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

By Corollary 2.2, the regression model is then correctly specified if and only if

$$\max_{\|\beta\|=1} Q(\theta_0,\beta) = 0. \tag{3.1}$$

Assume $(Y_i, X_i')'$, $i = 1,\ldots,n$ is 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

$$\widehat{\theta}_n = \arg\min_{\theta\in\Theta} \sum_{i=1}^{n} \left(Y_i - \mu(X_i;\theta)\right)^2 \ , \tag{3.2}$$

with an appropriate convention in case of ties. In view of Equation (2.4), 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.

We choose β as

$$\widehat{\beta}_n = \arg\max_{\|\beta\|=1} \{Q_n(\theta, \beta) - \alpha_n \mathbb{I}[\beta \neq \beta_0]\} \ , \tag{3.3}$$

where β_0 is a fixed vector, and α_n, $n \geq 1$, is a sequence of positive real numbers decreasing to zero at an appropriate rate. 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. As $Q_n(\widehat{\theta}_n, \widehat{\beta}_n)$ behaves as $Q_n(\widehat{\theta}_n, \beta_0)$, 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

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

An asymptotic α-level test is given by $\mathbb{I}(T_n \geq z_\alpha)$, where z_α is the upper α-percentile 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

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

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 $(\varepsilon, X')' \in \mathbb{R}^{q+1}$ with $\mathbb{E}(\varepsilon \mid X) = 0$ and $\mathbb{E}(\varepsilon^4) < \infty$.*

(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 $\left[\sum_{kl} a_{kl}^2\right]^{1/2}$.

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

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, the latter ensures the $\sqrt{n}-$consistency of the NLLS estimator (3.2) as stated in Lemma 6.1 of Section 6.

We make the following asumptions 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 $\left(nh^2\right)^\alpha / \ln n \to \infty$ 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

$$\sup_{1\leq i\leq n} \left| \frac{\widehat{\sigma}^2(X_i)}{\sigma^2(X_i)} - 1 \right| = o_{\mathbb{P}}(1) \ , \tag{3.4}$$

$v_n^2(\beta)$ 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

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

where b_n is a bandwidth parameter chosen independently of h. Guerre and Lavergne (2005) provide some primitive conditions such that (3.4) holds. 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 then proceed under (3.4).

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

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

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

(i) $\sup_{\|\beta\|=1} |Q_{0n}(\beta)| = O_{\mathbb{P}}(n^{-1} h^{-(1/2+\gamma)})$ *for any* $\gamma > 0$,

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

The proof is given in Section 6.

We now describe the behavior of $\widehat{\beta}_n$ defined in (3.3) under the null hypothesis.

Lemma 3.2 *Let Assumptions D, M, and K. Consider a positive sequence* α_n *such that* $\alpha_n n h^{1/2+\gamma} \to C > 0$ *for some* $\gamma > 0$. *Under* H_0, $\mathbb{P}(\widehat{\beta}_n = \beta_0) \to 1$.

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

The properties of our test under the null hypothesis can then be stated.

Theorem 3.3 *Under Assumptions D, M, K, and (3.4), if* $\alpha_n n h^{1/2+\gamma} \to C > 0$ *for some* $\gamma > 0$, *then the tests based on* T_n *or* T_n' *have asymptotic level* α.

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 (3.4), $\widehat{v}_n^2(\beta_0) = v_n^2(\beta_0)(1 + o_p(1))$. 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

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

and $\mathrm{Sp}(W_\beta)$ is the spectral radius of the matrix W_β. Lemma 6.2 allows to conclude. ■

Some technical comments. 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 of Sherman (1994a) we use here. It is an open question whether the self-normalized 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 (2.3). As will be seen shortly, the study under the alternative is simpler.

3.4 Behavior under directional alternatives

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

$$\begin{aligned}
T_n' \geq T_n &= \frac{nh^{1/2}Q_n(\widehat{\theta}_n, \widehat{\beta}_n)}{\widehat{v}_n(\beta_0)} \\
&\geq \frac{nh^{1/2}}{\widehat{v}_n(\beta_0)} \left[\max_{\|\beta\|=1} \left\{ Q_n(\widehat{\theta}_n, \beta) - \alpha_n \mathbb{I}\,(\beta \neq \beta_0) \right\} + \alpha_n \mathbb{I}\left(\widehat{\beta}_n \neq \beta_0 \right) \right] \\
&\geq \frac{nh^{1/2}}{\widehat{v}_n(\beta_0)} \left[\max_{\|\beta\|=1} Q_n(\widehat{\theta}_n, \beta) - \alpha_n \right] \\
&\geq \frac{nh^{1/2}}{v_n(\beta_0)(1 + o_{\mathbb{P}}(1))} \left\{ Q_n(\widehat{\theta}_n, \beta) - \alpha_n \right\} \quad \text{for any } \beta\ .
\end{aligned} \tag{3.5}$$

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 fro some β_1. It is easily seen that our test is consistent under the assumptions of Theorem 3.3 provided $\widehat{\theta}_n$ converges to some pseudo-true value θ^*. Indeed, when the model is misspecified, there exists at least one β for which $Q(\theta^*, \beta) > 0$.

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

$$\mathbb{E}[\delta(X)\dot{\mu}(X; \theta_0)] = 0 \quad \text{and} \quad 0 < \mathbb{E}[\delta^4(X)] < \infty\ , \tag{3.6}$$

and the sequence of alternatives defined as

$$H_{1n}: \; m_n(\cdot) = \mu(\cdot;\theta_0) + r_n\delta(\cdot), \quad n \geq 1 \,. \tag{3.7}$$

Under H_{1n}, $\widehat{\theta}_n - \theta_0 = O_{\mathbb{P}}(n^{-1/2})$ as proved by Lemma 6.1 in Section 6. We show below that such directional alternatives can be detected if α_n/r_n^2 tends to zero. Given the conditions of Theorem 3.2, this means that $r_n^2 nh^{1/2+\gamma} \to \infty$ for some small $\gamma > 0$, 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, $r_n^2 nh^{q/2} \to \infty$ 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 arbitrarily close to the power obtained in the unidimensional case.

Theorem 3.4 *Under Assumptions D, M, K, and (3.4), if $r_n^2 nh^{1/2} \to \infty$ and $\alpha_n/r_n^2 \to 0$, the tests based on T_n and T_n' are consistent against the sequence of alternatives H_{1n} with $\delta(X)$ satisfying (3.6).*

Proof. By Assumption D-(b), $v_n^2(\beta) \leq \overline{\sigma}^4 n^2 h \|W_\beta\|^2$, where W_β is the matrix with generic element $\mathbb{I}(i \neq j)\, K_h\left((X_i - X_j)'\beta_0\right)/(hn(n-1))$. Lemma 6.2 then ensures that $v_n^2(\beta)$ is bounded in probability from above for any β. Under H_{1n}, $U_i(\widehat{\theta}_n) = \mu(X_i;\theta_0) + r_n\delta(X_i) + \varepsilon_i - \mu(X_i;\widehat{\theta}_n)$. Then by simple algebra, $Q_n(\widehat{\theta}_n,\beta)$ can be decomposed for any β as

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

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

Since $v_n^2(\beta) \leq \overline{\sigma}^4 n^2 h \|W_\beta\|^2 = O_{\mathbb{P}}(1)$, $nh^{1/2} Q_{0n}(\beta) = O_{\mathbb{P}}(1)$ for any β. Lemma 3.1-(ii) deals with $Q_{1n}(\widehat{\theta}_n,\beta)$ and $Q_{2n}(\widehat{\theta}_n,\beta)$. It is shown in Section 6 that for any β

$$Q_{3n}(\widehat{\theta}_n,\beta) = O_{\mathbb{P}}(r_n n^{-1/2}) \tag{3.8}$$

$$Q_{4n}(\beta) = O_{\mathbb{P}}(r_n n^{-1/2}) \tag{3.9}$$

$$Q_{5n}(\beta) = r_n^2 \mathbb{E}\left[\mathbb{E}^2[\delta(X)|X'\beta] f_\beta(X'\beta)\right] + o_{\mathbb{P}}(r_n^2) \,. \tag{3.10}$$

Collecting results, it follows that for any β

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

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 (3.5). ■

4 Small sample 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 $Y_i^* = \mu(X_i,\widehat{\theta}_n)+\widehat{\sigma}_n(X_i)\omega_i, i = 1,\ldots,n$. Bootstrap test statistics are 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 $1-\alpha$ quantile of the bootstrapped test statistic. This critical value is then compared to the initial test statistic.

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

The proof follows easily from our previous results and is thus omitted.

4.2 Simulation study

Our focus is first to determine the sensitivity of our test to the penalty parameter α_n and second to compare the small sample power of our test to the multivariate test of Zheng (1996) and Li and Wang (1996). For 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 is chosen as $U\left[-\sqrt{3},\sqrt{3}\right]$ to get unit variance. We then sampled errors from a standard normal distribution and constructed the response variable sample as

$$Y_i = 3b\cosh\left(\frac{X_{1i}}{\sqrt{3}}\right) - 3\sinh(1) + \varepsilon_i \quad i = 1,\ldots 50\ .$$

By sampling under the null hypothesis, i.e. $b = 0$, and computing the tests statistics, we obtained small sample critical values that allows to calibrate the level of the different tests at 5%. This is equivalent to bootstrapping since there is no estimated parameter under H_0. We then drew the power curves of the different tests when b varies. Specifically, we considered (i) Zheng's test when only X_1 is considered (ii) Zheng's test when all three regressors are taken into account (iii) our test based on T_n (iv) our test based on T'_n.

For each $b = 0, 0.2, \ldots, 1.4$, we run 1000 replications. To compute our test statistics, we used a quartic kernel with support $[-1, 1]$. For the bandwidth, we chose $h = cn^{-2/(4s+q)}$ where q is either 3 in Case (ii) or 1 for all other cases, and c varies in $0.5, 1, 1.5$. For our tests, β_0 is set to $(1,1,1)'/\sqrt{3}$, a natural choice if one does not want to favor any regressor at the outset. The parameter α is set to ??? Last, to speed up computations, we assume that the errors' variance is known to be one for all the tests.

To be completed

5 Conclusion

We have proposed a general approach to circumvent the curse of dimensionality in testing moment restrictions conditional upon a multivariate random variable X. Lemma 2.1 is the key of our method. 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 that aims at obtaining a simple behavior under the null hypothesis.

We have applied our method to testing for a parametric regression function. The test has known asymptotic critical values and behaves against directional alternatives almost as if the dimension of X was one. Our simulations results confirm the good power of the test. The penalty function could have been specified differently and it is an open question whether one is always preferable. From our key result, other testing procedures could be constructed such as an integrated conditional moment test in the spirit of Bierens (1982). We are currently investigating such an alternative test.

Our method applies to many testing problems as explained in Section 2.1. Testing for general moment restrictions depending upon unknown finite dimensional parameters as in Delgado, Dominguez and Lavergne (2005) should not raise any technical difficulties per se. However, designing a test for semi or nonparametric restrictions following our approach appears in some cases more intricate from the theoretical viewpoint. This will be the topic of future work.

6 Proofs

6.1 Preliminary results

In the proofs, C is a positive constant that may vary from line to line.

Lemma 6.1 *Under Assumptions D–(a) and M,* $\|\widehat{\theta}_n - \theta_0\| = O_{\mathbb{P}}(n^{-1/2})$ *under* H_0 *and* H_{1n}*.*

Proof. We will proceed under H_{1n}, and the result under H_0 follows taking $r_n = 0$. By a uniform law of large numbers for Euclidean families, see e.g. Pakes and Pollard (1989, Lemma 2.8),

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

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 $\widehat{\theta}_n$,

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

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

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

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

■

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

Lemma 6.2 *Let W_β be the matrix with generic element $\mathbb{I}\left(i \neq j\right) K_h\left((X_i - X_j)'\beta\right)/(h\,n(n-1))$. 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, $\mathrm{Sp}(W_\beta) = \sup_{u\neq 0} \|W_\beta u\|/\|u\|$ and for any $u \in \mathbb{R}^n$,

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

Hence $n\mathrm{Sp}(W_\beta) \leq \max_{1\leq i\leq n} \sum_{j\neq i} \frac{1}{(n-1)h} K_h\left((X_i - X_j)'\beta\right)$. For all j and β, $|K_h\left((x - X_j)'\beta\right)| \leq C$ and $\mathrm{Var}\left[K_h\left((x - X_j)'\beta\right)\right] \leq C$. Thus the Bernstein inequality yields for any $t > 0$

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

since $nh^2 \to \infty$. Moreover, using the proof of (3.10) with $\delta(X) \equiv 1$, deduce

$$\mathbb{E}\left\{\mathbb{E}\left[h^{-1}K_h\left((X_i - X_j)'\beta\right)|X_i\right]\right\} \to \mathbb{E}\left[f_\beta(X_i'\beta)\right] \leq C$$

for some C independent of β. By Markov inequality, $\mathbb{E}\left[h^{-1}K_h\left((X_i - X_j)'\beta\right)|X_i\right] = O_{\mathbb{P}}(1)$. This gives the first result.

For the second result,

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

follows by adapting the proof of (3.10) below with $\delta(X)\equiv 1$. The last quantity is bounded from above and below by Assumptions D-(c) and K-(a). ■

6.2 Proof of Lemma 3.1

i) Consider the degenerate U-process

$$U_n\widetilde{g} = \frac{1}{n(n-1)}\sum_{j\neq i}\varepsilon_i\varepsilon_j K_h\left(X_i'\beta - X_j'\beta\right)$$

defined by the functions $\widetilde{g}$ indexed by h and β with $\|\beta\| = 1$. By Assumption D, 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 an envelope with bounded fourth moment. By Sherman's (1994a) Main Corollary with $p = 1$ and Holder's inequality,

$$\mathbb{E}\sup_{\|\beta\|=1,h>0}|nU_n\widetilde{g}| \leq \Lambda\left[\mathbb{E}\left\{\sup_{\|\beta\|=1,h>0}\left(U_{2n}\widetilde{g}^2\right)^{\alpha}\right\}\right]^{1/2} \leq \Lambda\mathbb{E}^{\alpha/2}\left\{\sup_{\|\beta\|=1,h>0}U_{2n}\widetilde{g}^2\right\}, \qquad (6.11)$$

where Λ is a universal constant and $0 < \alpha < 1$. Apply Hoeffding's decomposition for $U_{2n}\widetilde{g}^2$ and Corollary 4(i) of Sherman (1994a) to deduce that

$$\begin{aligned}\mathbb{E}\left\{\sup_{\|\beta\|=1,h>0}U_{2n}\widetilde{g}^2\right\} &\leq \mathbb{E}\left\{\sup_{\|\beta\|=1,h>0}\left|U_{2n}\widetilde{g}^2 - \mathbb{E}\left(\widetilde{g}^2\right)\right|\right\} + \sup_{\|\beta\|=1,h>0}\mathbb{E}\left(\widetilde{g}^2\right)\\ &\leq O(n^{-1/2}) + \sup_{\|\beta\|=1,h>0}\mathbb{E}\left(\widetilde{g}^2\right).\end{aligned}$$

Using the boundedness of $\sigma^2(\cdot)$ and $f_\beta(\cdot)$,

$$\mathbb{E}\left(\widetilde{g}^2\right) \leq \overline{\sigma}^4\mathbb{E}\left\{K_h^2\left(X_i'\beta - X_j'\beta\right)\right\} \leq \overline{\sigma}^4\int_{\mathbb{R}}\left\{\int_{\mathbb{R}}K^2(t)f_\beta(v+th)\,dt\right\}f_\beta(v)dv \leq Ch$$

with $C > 0$ independent of h and β. Inequality (6.11) and $nh^2\to\infty$ then yield

$$\mathbb{E}\sup_{\|\beta\|=1}\left|nh^{(1/2+\gamma)}Q_{0n}(\beta)\right| = \mathbb{E}\sup_{\|\beta\|=1}\left|nh^{(-1/2+\gamma)}U_n\widetilde{g}\right| = O(h^{\gamma+(\alpha-1)/2}).$$

Choose $\alpha \geq 1 - 2\gamma$ to obtain the result.

(ii) Consider $V_n(\theta_0) = \left\{\theta \in \Theta : \|\theta - \theta_0\| \leq n^{-1/2}M\right\}$ a shrinking neighborhood of θ_0. By Lemma 6.1, $\mathbb{P}\left[\widehat{\theta}_n \in V_n(\theta_0)\right] \to 1$ whenever $M \to \infty$. 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 $F(W_i, W_j) = |\varepsilon_i|\, \widetilde{\Phi}(X_j)$ 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 $k = 2$,

$$\mathbb{E}\left[\sup_{\theta\in V_n(\theta_0),h,\beta} \left|nU_n\overline{g}_{\theta,h,\beta}\right|\right] \leq \Lambda\left[\mathbb{E}\sup_{\theta\in V_n(\theta_0),h,\beta}\left\{U_{2n}\overline{g}^2_{\theta,h,\beta}\right\}^{\alpha}\right]^{1/2} \tag{6.12}$$

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

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

for some constants C, C'. Deduce that for any $\theta \in V_n(\theta_0)$, h, and β,

$$U_{2n}\overline{g}^2_{\theta,h,\beta} \leq C\varepsilon_i^2\left\{\widetilde{\Phi}\left(X_j\right)+1\right\}^2 \|\theta-\theta_0\|^2\,\varepsilon_i^2\left\{\widetilde{\Phi}\left(X_j\right)+1\right\}^2 \leq \frac{CM^2}{n}$$

for some C independent of θ, h, and β . By Inequality (6.12),

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

for some $C > 0$. Hence by Chebyshev's inequality

$$\sup_{\|\beta\|=1} |nh^{-1/2}U_n\overline{g}_{\theta_n,h,\beta}| = O_{\mathbb{P}}\left((nh^{1/\alpha})^{-\alpha/2}\right). \tag{6.13}$$

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

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

Let $\widetilde{g}_{1,s}(\cdot)$, $1 \le s \le d$, 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

$$\mathbb{E}\left[\sup_{h,\beta}\left|n^{1/2}\mathbb{P}_n\widetilde{g}_{1,s}\right|\right] \le \Lambda\left[\mathbb{E}\sup_{h,\beta}\left\{\mathbb{P}_{2n}\widetilde{g}_{1,s}^2\right\}^{\alpha}\right]^{1/2} \qquad s = 1, \ldots, p, \tag{6.14}$$

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

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

for some $C > 0$. From (6.14) and Chebyshev's inequality, $\sup_\beta\left\|n^{1/2}\mathbb{P}_n\widetilde{g}_1\right\| = O_{\mathbb{P}}(h^{3\alpha/4})$, thus

$$\sup_{\theta\in V_n(\theta_0),\beta}\left|nh^{-1/2}(\theta - \theta_0)'\mathbb{P}_n\widetilde{g}_1\right| = O_{\mathbb{P}}\left(h^{(3\alpha/4)-1/2)}\right). \tag{6.15}$$

Similar arguments apply to each of the components $\widetilde{g}_{2,kl}$, $1 \le k, l \le d$, of the square matrix $\widetilde{g}_2$, so that

$$\sup_{\theta\in V_n(\theta_0),\beta}\left|nh^{-1/2}(\theta - \theta_0)'\mathbb{P}_n\widetilde{g}_2(\theta - \theta_0)\right| = O_{\mathbb{P}}\left(n^{-1/2}h^{(3\alpha/4)-1/2}\right). \tag{6.16}$$

From Equations (6.13), (6.15), and (6.16) with $\alpha > 2/3$ and using $nh^2 \to \infty$,

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

For $Q_{2n}(\widehat{\theta}_n, \beta)$, 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 $\theta \in V_n(\theta_0)$,

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

6.3 Proof of (3.8)–(3.10)

Proof of (3.8). Since $|u'W_\beta v| \le \|u\|\|v\|\mathrm{Sp}(W_\beta)$, then for any β,

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

by Lemma 6.2 and the weak law of large numbers. Now, under Assumption M,

$$
\left(\mu(X_i;\widehat{\theta}_n)-\mu(X_i;\theta_0)\right)^2 \le \tilde{\Phi}^2(X_i)\|\widehat{\theta}_n-\theta_0\|^2
$$

for some $\tilde{\Phi}(\cdot)$ with bounded fourth moment and $\frac{1}{n}\sum_{i=1}^{n}\left(\mu(X_i;\widehat{\theta}_n)-\mu(X_i;\theta_0)\right)^2 = O_{\mathbb{P}}(n^{-1})$.

Proof of (3.9). Denote by $\mathbb{E}_n$ the conditional expectation given the X_i and let

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

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

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

using $\|\left(\delta(X_1),\dots,\delta(X_n)\right)W_\beta\| \le \|\left(\delta(X_1),\dots,\delta(X_n)\right)\|\,\mathrm{Sp}(W_\beta)$, Lemma 6.2 and the weak law of large numbers.

Proof of (3.10). Consider $U_n = r_n^{-2}Q_{5n}(\beta)$. By straightforward computations,

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

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

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

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

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(5)**, 1129-1151.

Carrasco, M., Florens J.P. and Renault E.(2006). Linear inverse problems in structural econometrics: estimation based on spectral decomposition and regularization. *Handbook of Econometrics vol. 6* J. Heckman and E. Leamer eds. North Holland.

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

Delecroix, M., Hristache, M., and Patilea, V. (2005). On semiparametric $M-$estimation in single-index regression. *J. Statist. Plann. Inference*, in press.

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

Delgado, M.A., M.A. Dominguez and P. Lavergne (2005). Consistent tests of conditional moment restrictions. *Annales d'Economie et Statistique*, to appear.

Donald, S.G., G.W. Imbens and W.K. Newey (2003). Empirical likelhood 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 applicatins 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**, No. 2, 840-870.

HALL, P., YATCHEW, A. (2005). Unified approach for testing functional hypotheses in semiparametric contexts. *J. Econometrics*, **127/2**, 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. Statist. Amer. 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., 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.

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.

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

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.

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

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.

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

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

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.