

Breaking the curse of dimensionality in nonparametric testing

Pascal Lavergne

Simon Fraser University

Valentin Patilea

CREST-ENSAI

November 2005

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 approach 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.

We thank participants of ESRC 2005 (Bristol), CESG 2005 (Vancouver), in particular Yanqin Fan, and of the econometrics seminar at the University of Toulouse, for helpful comments. Special thanks to Jean-Marie Rolin for a very helpful suggestion.

Address correspondence to: Pascal Lavergne, Dept. of Economics, Simon Fraser University, 8888 University Drive, Burnaby BC, V5A 1S6 CANADA.

Emails: pascal.lavergne@sfu.ca 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 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 (2005).

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 by Zhu and Li (1998), who put forward a similar idea that we develop here for testing a linear

parametric regression, but did not study the related test.

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 approach that could apply to many nonparametric testing problems. To show the potential benefits of our approach, we apply it 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 *almost* as if the dimension of the regressors was 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 approach. As shown in Section 2, many testing problems consider a null hypothesis of the form

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

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 for reducing the dimensionality of the problem is to use single linear indices $X'\beta$ as a conditioning variable instead of X and to then to look at $\mathbb{E}[U(\theta_0)|X'\beta]$ for all directions defined by β of norm one 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 the case. Hence looking at the index that makes an estimate of $\mathbb{E}[U(\theta_0)|X'\beta]$ furthest away from zero yields a consistent test, and hopefully a powerful one. Under the null hypothesis however, the latter quantity is zero for every direction β . Hence, we introduce a criterion that favors one of the direction under H_0 and then yields a simple behavior. Our dimension-reduction approach 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 (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. 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. Against general directional alternatives of the form

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

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 , 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 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, where $\rho(Y, X, \theta) = Y^2 - \theta^2$. Delgado, Dominguez and Lavergne (2005) provide several more instances. 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 NONPARAMETRIC RESTRICTIONS. In this context, θ is infinite-dimensional. Some instances are as follows. 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), 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

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

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 (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

approach 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).

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 $\mathbb{E}|Z_j| < \infty$, $j = 1, \dots, c$.*

A) The following statements are equivalent:

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

$$\mathbb{E}(Z | X'\beta) = \mathbb{E}(Z) \quad \text{almost surely.} \quad (2.1)$$

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

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

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

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 σ -field generated by $X'\beta$ is the same as the σ -field generated by $X'\beta/\|\beta\|$. By Condition (2.1) and elementary properties of the conditional expectation, we obtain that for any β , including $\beta = 0$,

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

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

$$\mathbb{E}[\exp\{iX'\beta\}\mathbb{E}(Z^+ | X)] = \mathbb{E}[\exp\{iX'\beta\}\mathbb{E}(Z^- | X)] .$$

As distinct positive finite measures cannot have the same characteristic function, this implies that $\mathbb{E}(Z^+ | X) = \mathbb{E}(Z^- | X)$ and hence $\mathbb{E}(Z | X) = 0$ 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 $\mathbb{P}[\mathbb{E}(Z | X) = 0] < 1$,

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

where $A_1, \dots, A_q \subset \mathbb{R}$ contain 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 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 $\mathbb{E}|U_j(\theta)| < \infty$, $j = 1, \dots, c$, 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*

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

Lemma 2.1 can actually be deduced from Theorem 1 of Bierens (1982), who showed that $\mathbb{E}[Z | X] = 0$ 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 $\mathbb{E}[Z | X] = 0$ is equivalent to

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

whenever $\phi(\cdot)$ is analytic nonpolynomial. 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

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

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 $\mathbb{E}[Z\phi(X'\beta)\omega(X'\beta)]$ 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 $\hat{\theta}_n$ of θ_0 and denote by $U_i(\theta)$ the data-dependent vector function of θ for observation i . Let $\hat{\gamma}_i(X'_i\beta, \theta)$ be a consistent estimator of $\mathbb{E}(U(\theta)|X'_i\beta)\omega(X'_i\beta)$ and

$$Q_n(\theta, \beta) = \frac{1}{n} \sum_{i=1}^n U'_i(\theta) \hat{\gamma}_i(X'_i\beta; \theta) . \quad (2.3)$$

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

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

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

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

Our penalization scheme is related to Tikhonov's regularization in estimation, see Carasco, Florens and Renault (2006), but applied to a testing problem. 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 $\max_{\|\beta\|=1} Q_n(\hat{\theta}_n, \beta)$. However, the penalty should not perturb the behavior of the maximum when H_0 does not hold, hence we should have $\pi_n(t) \rightarrow 0$ fast enough for all t as n grows. A normalized version of $Q_n(\hat{\theta}_n, \hat{\beta}_n)$ is then taken as the test statistic. As will become apparent in the next section, the choice of the vector β_0 is theoretically irrelevant for consistency of the test. Though, the test should have greater power when $\mathbb{E}[U(\theta_0)|X'\beta_0]$ 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 (1.1) holds with $U(\theta) = Y - \mu(X; \theta)$. We thus apply our general approach to this testing problem using kernel estimators. To avoid handling denominators close to zero, 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 . Define

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

By Corollary 2.2, the regression model is then correctly specified iff $\max_{\|\beta\|=1} Q(\theta_0, \beta) = 0$. Assume $(Y_i, X_i)'$, $i = 1, \dots, n$ is a random sample from the distribution of $(Y, X)'$. The parameter θ_0 can be estimated in a variety of ways. For instance, $\hat{\theta}_n$ can be the nonlinear

least-squares (NLLS) estimator of θ solving

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

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

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

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 $Q_n(\hat{\theta}_n, \beta)$ 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 $Q_n(\hat{\theta}_n, \beta)$ for checking a linear regression model. However, their test is based on the maximum plus a term of the form $(1/n) \sum_{i=1}^n U_i(\hat{\theta}_n) \phi(\|X_i\|)$, 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

$$\hat{\beta}_n = \arg \max_{\|\beta\|=1} \{Q_n(\theta, \beta) - \alpha_n \mathbb{I}[\beta \neq \beta_0]\}, \quad (3.2)$$

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. Our choice for the penalty function corresponds to the one of Bierens (1990) and is made for simplicity. We will prove that $\hat{\beta}_n = \beta_0$ with probability tending to 1 under H_0 . Since $Q_n(\hat{\theta}_n, \hat{\beta}_n)$ behaves like $Q_n(\hat{\theta}_n, \beta_0)$, a test is easily constructed. With at hand a consistent estimator $\hat{v}_n^2(\beta)$ of the variance of $nh^{1/2}Q_n(\hat{\theta}_n, \beta)$, let

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

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 $(\varepsilon_1, X'_1)', \dots, (\varepsilon_n, X'_n)'$ are independent copies of the random vector $(\varepsilon, X)' \in \mathbb{R}^{q+1}$ with $\mathbb{E}(\varepsilon | X) = 0$ and $\mathbb{E}(\varepsilon^4) < \infty$.*

(b) *Let $\sigma^2(x) = \mathbb{E}(\varepsilon^2 | X = x)$. 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 $\mathbb{E}[\Phi_1^4(X)] < \infty$;

(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 $\mathbb{E}[\Phi_2^4(X)] < \infty$; and

(iii) 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 (3.1) as stated in Lemma 6.1 of Section 6. 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) $h \rightarrow 0$ and $(nh^2)^\alpha / \ln n \rightarrow \infty$ for some $\alpha \in (0, 1)$.

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

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

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

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

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

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

For instance, one can consider

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

where b_n 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 $\hat{v}_n^2(\beta)/v_n^2(\beta) = 1 + o_{\mathbb{P}}(1)$ 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

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

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 $\widehat{\beta}_n$ under H_0 .

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

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

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

Theorem 3.3 *Under Assumptions D, M, K, and (3.3), if $\alpha_n n h^{1/2+\gamma} \rightarrow C > 0$ 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 (3.3), $\widehat{v}_n^2(\beta_0) = v_n^2(\beta_0)(1 + o_p(1))$. 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

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

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

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 (2.2). As will be seen shortly, the study under fixed and directional alternatives is simpler.

3.4 Behavior under directional alternatives

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

$$\begin{aligned} T'_n \geq T_n &= \frac{nh^{1/2}Q_n(\widehat{\theta}_n, \widehat{\beta}_n)}{\widehat{v}_n(\beta_0)} \\ &= \frac{nh^{1/2}}{\widehat{v}_n(\beta_0)} \left[\max_{\|\beta\|=1} \left\{ Q_n(\widehat{\theta}_n, \beta) - \alpha_n \mathbb{I}(\beta \neq \beta_0) \right\} + \alpha_n \mathbb{I}(\widehat{\beta}_n \neq \beta_0) \right] \\ &\geq \frac{nh^{1/2}}{\widehat{v}_n(\beta_0)} \left[\max_{\|\beta\|=1} Q_n(\widehat{\theta}_n, \beta) - \alpha_n \right] \\ &\geq \frac{nh^{1/2}}{v_n(\beta_0)(1 + o_{\mathbb{P}}(1))} \left\{ Q_n(\widehat{\theta}_n, \beta) - \alpha_n \right\} \quad \text{for any } \beta. \end{aligned} \quad (3.4)$$

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 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

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

and the sequence of alternatives defined as

$$H_{1n} : \mathbb{E}[Y|X] = \mu(\cdot; \theta_0) + r_n \delta(X), \quad n \geq 1. \quad (3.6)$$

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} , $\hat{\theta}_n - \theta_0 = O_{\mathbb{P}}(n^{-1/2})$ 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 $r_n^2 n h^{1/2+\gamma} \rightarrow \infty$ 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, $r_n^2 n h^{q/2} \rightarrow \infty$ 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 can remain arbitrarily close to the power obtained in the unidimensional case.

Theorem 3.4 *Under Assumptions D, M, K, and (3.3), if $r_n^2 n h^{1/2} \rightarrow \infty$ and $\alpha_n/r_n^2 \rightarrow 0$, the tests based on T_n and T'_n are consistent against the sequence of alternatives H_{1n} with $\delta(X)$ satisfying (3.5).*

Proof. By Assumption D-(b), $v_n^2(\beta) \leq \bar{\sigma}^4 n^2 h \|W_\beta\|^2$, where W_β is the matrix with generic element $\mathbb{I}(i \neq j) K_h((X_i - X_j)' \beta_0) / (hn(n-1))$. Lemma 6.2 then ensures that $v_n^2(\beta_0)$ is bounded in probability from above for any β_0 . Under H_{1n} , $U_i(\hat{\theta}_n) = \mu(X_i; \theta_0) + r_n \delta(X_i) + \varepsilon_i - \mu(X_i; \hat{\theta}_n)$. Then by simple algebra, $Q_n(\hat{\theta}_n, \beta)$ can be decomposed for any β as

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

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

Since $v_n^2(\beta) \leq \bar{\sigma}^4 n^2 h \|W_\beta\|^2 = O_{\mathbb{P}}(1)$, $nh^{1/2}Q_{0n}(\beta) = O_{\mathbb{P}}(1)$ 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

$$Q_{3n}(\widehat{\theta}_n, \beta) = O_{\mathbb{P}}(r_n n^{-1/2}) \quad (3.7)$$

$$Q_{4n}(\beta) = O_{\mathbb{P}}(r_n n^{-1/2}) \quad (3.8)$$

$$Q_{5n}(\beta) = r_n^2 \mathbb{E} \left[\mathbb{E}^2[\delta(X)|X'\beta] f_\beta(X'\beta) \right] + o_{\mathbb{P}}(r_n^2) . \quad (3.9)$$

Collecting results and using $\alpha_n/r_n^2 \rightarrow 0$, 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 (3.4). ■

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 $Y_i^* = \mu(X_i, \widehat{\theta}_n) + \widehat{\sigma}(X_i)\omega_i$, $i = 1, \dots, n$. 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 $(1-\alpha)$ -th 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 bootstrap 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 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 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 $U[-\sqrt{3}, \sqrt{3}]$ to get unit variance. We sampled errors from a standard normal distribution and constructed the response variable as

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

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 $d = 0.1, 0.2, \dots, 0.7$. 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 1 in the other cases, and b varies in $\{0.5, 1, 1.5, \dots, 4\}$. To set α_n , we computed v_0 , the mean of $v_n(\hat{\beta}_n)$, which was found to vary little with β_0 . We then chose $\alpha_n = av_0n^{-1}h^{-0.51}$, with $v_0 = 0.65$ in our case, and we let a vary in a grid from 2 to 10.

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 when $a = 2, 7$ and 10 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 almost exactly similar whatever the value of α_n . 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, 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

$$Y_i = 0.3 d \left\{ \cosh \left(\frac{3X_{1i} + X_{2i}}{\sqrt{3}} \right) + \cosh \left(\frac{X_{2i} + 3X_{3i}}{\sqrt{3}} \right) - e \right\} + \varepsilon_i \quad i = 1, \dots, 50$$

with e a centering constant equal to $2 \sinh(1) \sinh(3)/3$. Other features of the experiments were unchanged, except that we now considered as a benchmark Zheng's test based on the two linear indices entering the regression function. Figures 5 to 8 sum up our 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 better performances than the benchmark test for different values of β_0 and h , as soon as α_n is not too large.

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 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 that aims at obtaining a simple behavior under the null hypothesis. 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 and behaves against directional 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. The uniform behavior of our test against smooth alternatives should be studied. Also, 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 these issues. Finally, future work should be devoted to applying our approach to other testing problems.

6 Technicalities

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

Lemma 6.1 *Under Assumptions D-(a) and M, $\|\hat{\theta}_n - \theta_0\| = O_{\mathbb{P}}(n^{-1/2})$ 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 $\hat{\theta}_n \xrightarrow{p} \theta_0$. Now, by definition of $\hat{\theta}_n$,

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

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

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

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

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 $\mathbb{I}(i \neq j) K_h((X_i - X_j)' \beta) / (h n(n-1))$. Under Assumptions D-(c) and K, $\text{Sp}(W_\beta) = O_{\mathbb{P}}(n^{-1})$ and $nh^{1/2} \|W_\beta\| \asymp_{\mathbb{P}} 1$ for any β .*

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

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

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

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

since $nh^2 \rightarrow \infty$. Moreover, using the proof of (3.9) with $\delta(X) \equiv 1$, deduce

$$\mathbb{E} \{ \mathbb{E}[h^{-1} K_h((X_i - X_j)' \beta) | X_i] \} \rightarrow \mathbb{E}[f_\beta(X'_i \beta)] \leq C$$

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

This gives the first result. For the second result,

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

follows by adapting the proof of (3.9) below with $\delta(X) \equiv 1$. The last quantity is bounded from above and below by Assumptions D-(c) and K-(a). ■

Proof of Lemma 3.1. i) Consider the degenerate U -process

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

defined by the functions \tilde{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 $\{\tilde{g} : \|\beta\| = 1, h > 0\}$ is Euclidean for an envelope with bounded fourth moment. By Sherman's (1994a) Main Corollary with $p = 1$ and Holder's inequality,

$$\mathbb{E} \sup_{\|\beta\|=1, h>0} |nU_n \tilde{g}| \leq \Lambda \left[\mathbb{E} \left\{ \sup_{\|\beta\|=1, h>0} (U_{2n} \tilde{g}^2)^\alpha \right\} \right]^{1/2} \leq \Lambda \mathbb{E}^{\alpha/2} \left\{ \sup_{\|\beta\|=1, h>0} U_{2n} \tilde{g}^2 \right\}, \quad (6.10)$$

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

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

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

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

with $C > 0$ independent of h and β . Inequality (6.10) and $nh^2 \rightarrow \infty$ then yield

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

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

(ii) Consider $V_n(\theta_0) = \{\theta \in \Theta : \|\theta - \theta_0\| \leq n^{-1/2}M\}$ a shrinking neighborhood of θ_0 . By Lemma 6.1, $\mathbb{P}[\hat{\theta}_n \in V_n(\theta_0)] \rightarrow 1$ whenever $M \rightarrow \infty$. Let $W = (\varepsilon, X)'$ and

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

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 $F(W_i, W_j) = |\varepsilon_i| \tilde{\Phi}(X_j)$ where $\tilde{\Phi}(\cdot) = C \sum_{i=1}^2 \Phi_i(\cdot)$, 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 $U_n \bar{g}_{\theta, h, \beta}$, with $\bar{g}_{\theta, h, \beta}(W_i, W_j) = g_{\theta, h, \beta}(W_i, W_j) - \mathbb{E}[g_{\theta, h, \beta}(W_i, W_j) | W_i]$. By Lemma 5 of Sherman (1994a), the family $\bar{g}_{\theta, h, \beta}$, $\theta \in \Theta$, $h \in (0, 1]$, $\|\beta\| = 1$, is Euclidean for a squared-integrable envelope. From the Main Corollary of Sherman with $p = 1$ and $k = 2$,

$$\mathbb{E} \left[\sup_{\theta \in V_n(\theta_0), h, \beta} |nU_n \bar{g}_{\theta, h, \beta}| \right] \leq \Lambda \left[\mathbb{E} \sup_{\theta \in V_n(\theta_0), h, \beta} \{U_{2n} \bar{g}_{\theta, h, \beta}^2\}^\alpha \right]^{1/2} \quad (6.11)$$

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

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

for some constants C, C' . Hence, from Inequality (6.11),

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

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

$$\sup_{\|\beta\|=1} |nh^{-1/2} U_n \bar{g}_{\theta, h, \beta}| = O_{\mathbb{P}} \left((nh^{1/\alpha})^{-\alpha/2} \right). \quad (6.12)$$

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

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

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

of Sherman with $p = k = 1$ yields

$$\mathbb{E} \left[\sup_{h, \beta} \left| n^{1/2} \mathbb{P}_n \tilde{g}_{1,s} \right| \right] \leq \Lambda \left[\mathbb{E} \sup_{h, \beta} \left\{ \mathbb{P}_{2n} \tilde{g}_{1,s}^2 \right\}^\alpha \right]^{1/2} \quad s = 1, \dots, p, \quad (6.13)$$

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

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

for some $C > 0$. From (6.13) and Chebyshev's inequality, $\sup_\beta \|n^{1/2} \mathbb{P}_n \tilde{g}_1\| = O_{\mathbb{P}}(h^{3\alpha/4})$, thus

$$\sup_{\theta \in V_n(\theta_0), \beta} \left| nh^{-1/2} (\theta - \theta_0)' \mathbb{P}_n \tilde{g}_1 \right| = O_{\mathbb{P}} \left(h^{(3\alpha/4) - 1/2} \right). \quad (6.14)$$

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

$$\sup_{\theta \in V_n(\theta_0), \beta} \left| nh^{-1/2} (\theta - \theta_0)' \mathbb{P}_n \tilde{g}_2 (\theta - \theta_0) \right| = O_{\mathbb{P}} \left(n^{-1/2} h^{(3\alpha/4) - 1/2} \right). \quad (6.15)$$

From Equations (6.12), (6.14), and (6.15) with $\alpha > 2/3$ and using $nh^2 \rightarrow \infty$,

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

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

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

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

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

Proof of (3.7). Since $|u' W_\beta v| \leq \|u\| \|v\| \text{Sp}(W_\beta)$, then for any β ,

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

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

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

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

Proof of (3.8). Denote by \mathbb{E}_n the conditional expectation given the X_i and let

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

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

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

using $\|(\delta(X_1), \dots, \delta(X_n)) W_\beta\| \leq \|(\delta(X_1), \dots, \delta(X_n))\| \text{Sp}(W_\beta)$, Lemma 6.2 and the weak law of large numbers. ■

Proof of (3.9). Consider $U_n = r_n^{-2} Q_{5n}(\beta)$. By straightforward computations,

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

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

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

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

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

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

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.
- 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.
- CHEN, P.D. (1991). L^1 -theory of approximation by ridge functions. *Chinese Sci. Bull.* **36**, 1761-1764.
- DELECROIX, M., HRISTACHE, M., AND PATILEA, V. (2005). On semiparametric M -estimation in single-index regression. *J. Statist. Plann. Inference*, forthcoming.
- 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. (2005). Consistent tests of conditional moment restrictions. *Annales d'Economie et Statistique*, forthcoming.
- 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., 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. 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., 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.
- 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.
- 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). Jackknife, 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. *Statist. Sinica* **14**, 1-39 (with discussion).
- 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: $\beta_0 = (1, 1, 1)/\sqrt{3}$

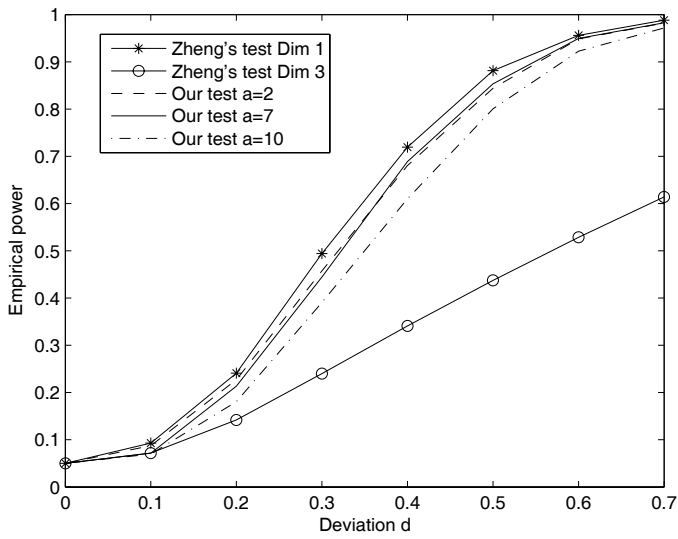


Figure 2: $\beta_0 = (1, 2, 3)/\sqrt{14}$

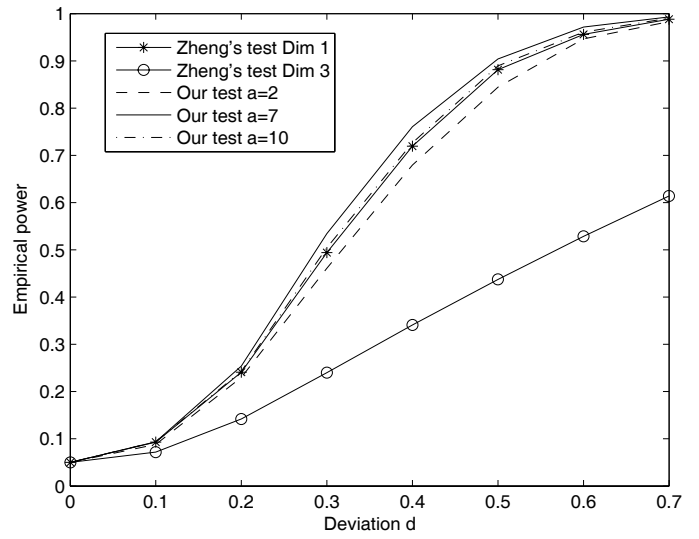


Figure 3: $\beta_0 = (0, 1, 0)$

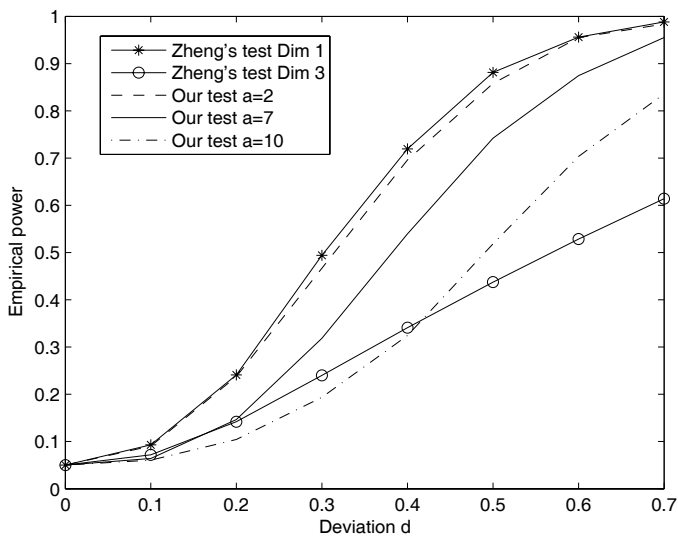


Figure 4: $\beta_0 = (1, 1, 1)/\sqrt{3}$ and $d = 0.4$

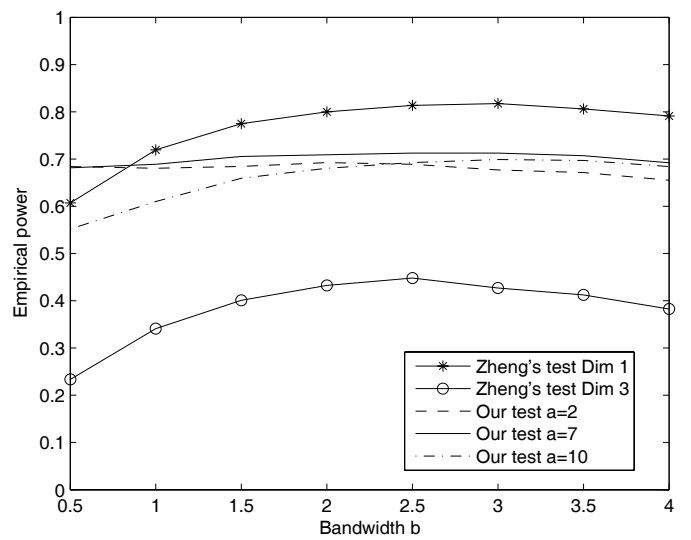


Figure 5: $\beta_0 = (1, 1, 1)/\sqrt{3}$

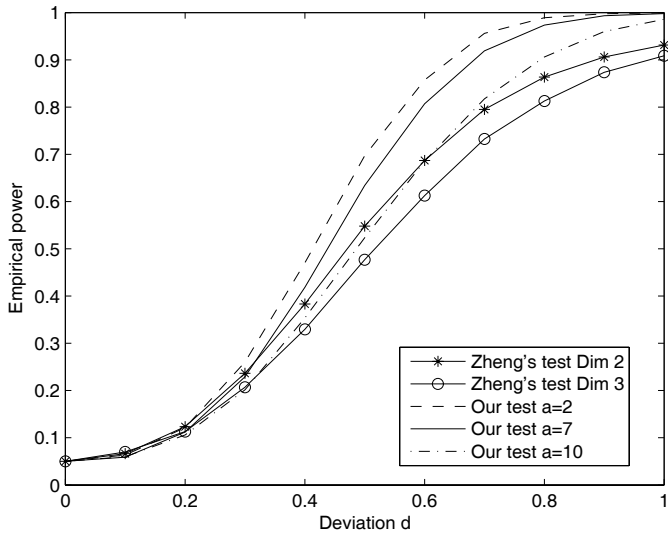


Figure 6: $\beta_0 = (3, 1, 0)/\sqrt{10}$

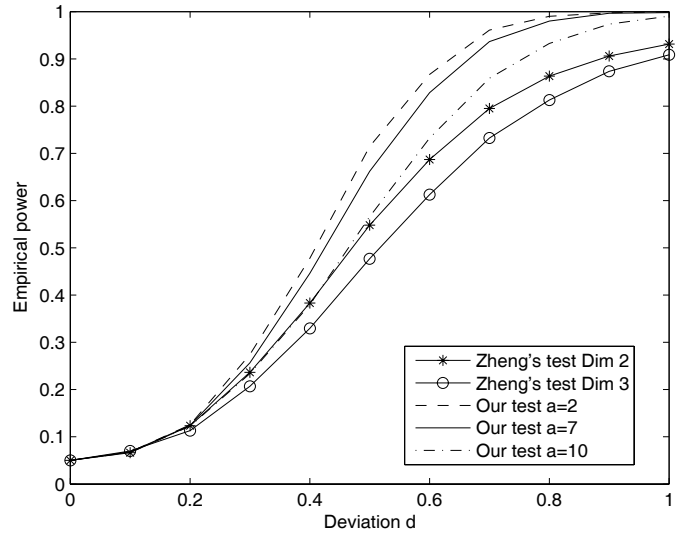


Figure 7: $\beta_0 = (0, 1, 0)$

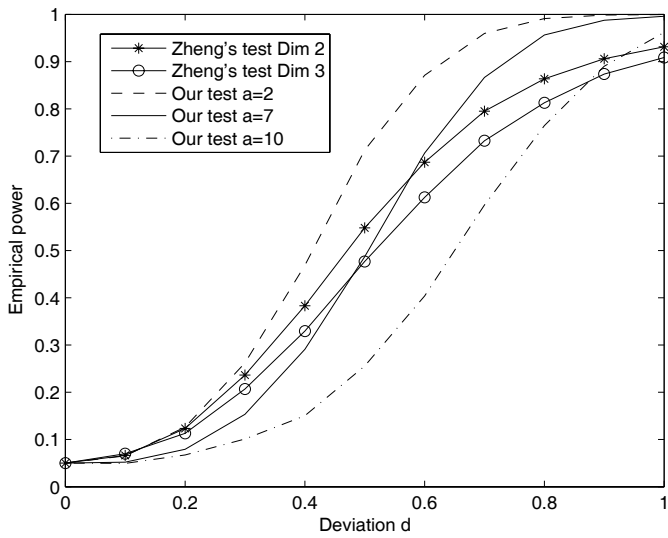


Figure 8: $\beta_0 = (1, 1, 1)/\sqrt{3}$ and $d = 0.5$

