

# Breaking the curse of dimensionality in nonparametric testing

Pascal Lavergne, Simon Fraser University

Valentin Patilea, CREST-ENSAI

## **Abstract**

For tests based on nonparametric methods, power crucially depends on the dimension of the conditioning variables, and specifically decreases with this dimension. This is known as the “curse of dimensionality”. We propose a new general approach to nonparametric testing in high dimensional settings and we show how to implement it when testing for a parametric regression. The resulting test behaves against directional local alternatives almost as if the dimension of the regressors was one. It is also almost optimal against classes of one-dimensional alternatives for a suitable choice of the smoothing parameter. The test performs well in small samples compared to several other tests.

**Keywords:** Curse of dimensionality, Testing, Nonparametric methods.

**JEL classification:** Primary C52; Secondary C14.

Address correspondence to: Pascal Lavergne, Dept. of Economics, Simon Fraser University, 8888 University Drive, Burnaby BC, V5A 1S6 CANADA. Tel: (1) 604 268 7161 Fax : (1) 604 291 5944 Emails: [pascal.lavergne@sfu.ca](mailto:pascal.lavergne@sfu.ca) [patilea@ensai.fr](mailto:patilea@ensai.fr)

# 1 Introduction

The “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 dimension increases, see Stone (1980). This issue is a prominent reason for the study of dimension-reduction models in nonparametric estimation. For instance, when a 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).

Many consistent specification tests of a (semi)parametric model contrast the latter model with a completely nonparametric one. As 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. 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 applications, where data often exhibit conditional volatility. After we wrote a first version of this paper, we discovered a former work on testing a linear parametric regression by Zhu and Li (1998), who put forward a idea similar to that we develop here, but did not study the related test. Escanciano (2006) is a closely related work that essentially uses the same premises to improve on Stute’s (1997) test.

The purpose of this paper is to propose a general approach for nonparametric testing

in high dimensional settings. Many testing problems consider a null hypothesis of the form

$$H_0 : \mathbb{E}[U(\theta_0)|X] = 0 \quad \text{almost surely (a.s.),} \quad (1)$$

where  $\theta_0$  is an unknown parameter to be estimated and  $X \in \mathbb{R}^q$ . We thus want to check whether a zero conditional moment restriction holds for almost any value of  $X$ . Our proposal is to use single linear indices  $X'\beta$  as a conditioning variable instead of  $X$ , and then to look for the direction  $\beta$  that makes  $\mathbb{E}[U(\theta_0)|X'\beta]$  furthest away from zero. Our dimension-reduction approach is thus the testing counterpart of single-index model in estimation, with the fundamental difference that the function under test does not need to depend on a single-index only. The key idea relies on an insight by Bierens (1982) as recently used by Escanciano (2006). However, both papers dealt with empirical-process based tests, while our work deal with smoothing-based tests.

Our first main finding, as stated in Section 2, is that our approach yields a test which is consistent against *any* nonparametric alternative. To show the potential benefits of our approach, we apply it to testing for a parametric regression in Section 3. Our second main finding is that the resulting test behaves against general nonparametric directional alternatives *almost* as if the dimension of the regressors was one. Hence there is a cost to dimensionality, but this cost is low and is paid only once, as it does not increase with the dimension of the regressors. Our third main finding is that for a suitable choice of the smoothing parameter, the test is almost optimal against classes of single-index alternatives. This is at odds with multidimensional tests, which have the same detection properties irrespective of the dimension of the alternatives. Finally, our simulation study in Section 4 shows that our test outperforms the standard “multidimensional” test of Zheng (1996) and the tests by Bierens (1982) and Escanciano (2006). Section 5 concludes and Section 6 gathers technical results and proofs.

## 2 Dimension reduction in nonparametric testing

### 2.1 Testing against nonparametric alternatives

In the null hypothesis (1), the unknown parameter  $\theta_0$  can be of finite or infinite dimension and is to 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.

Section 3 deals with testing for a parametric regression. In that case,  $U(\theta) = Y - \mu(X; \theta)$ , where  $\mu(\cdot; \cdot)$  belongs to a parametric family and  $\theta$  belongs to a subset of  $\mathbb{R}^d$ . Tests using smoothing methods have been proposed by Härdle and Mammen (1993), Hong and White (1995), and Zheng (1996), among others, see Hart (1997) for a review. For nonparametric directional alternatives of the form

$$\mathbb{E}[Y|X] = \mu(X; \theta_0) + r_n \delta(X), \quad (2)$$

$r_n$  should be of higher order than  $n^{-1/2}h^{-q/4}$  to obtain consistency. Uniform consistency against a class of alternatives of known smoothness  $s$  requires that the alternatives lie at distance  $n^{-2s/4s+q}$  from the null hypothesis, see Guerre and Lavergne (2002). When the smoothness index  $s$  is unknown, the so-called adaptive rate is less by a small factor, see Spokoiny (1996), Horowitz and Spokoiny (2001), and Guerre and Lavergne (2005). Other consistent tests of a parametric regression are based on transforms of the cumulative process of parametric residuals, see in particular Bierens (1982, 1990), Stute (1997), and Escanciano (2006). Theoretical results are mixed: while such tests are consistent against directional alternatives (2) whenever  $r_n^2 n$  diverges, they exhibit poor performances against sets of regular alternatives, see Guerre and Lavergne (2002).

Econometricians are often interested in conditional moment restrictions beyond the regression case, where  $U(\theta) = \rho(Y, X, \theta)$  with  $\rho(\cdot, \cdot, \cdot)$  a multivariate function known up to a finite-dimensional parameter  $\theta$ . When testing for homoscedasticity in a parametric regression model,  $\rho(Y, X, \theta, \sigma^2) = \{Y - \mu(X, \theta)\}^2 - \sigma^2$ . Delgado, Dominguez and Lavergne (2006) provide more examples. Stinchcombe and White (1998) and Whang (2001) study single conditional moment restrictions, Donald, Imbens and Newey (2003) and Delgado, Dominguez and Lavergne (2006) study multiple ones.

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  $\beta$  and an unknown univariate function  $m(\cdot)$ , see Fan and Li (1996), Stute and Zhu (2005), and Xia and al. (2005). When testing for the significance of some regressors  $Z$  in a nonparametric regression on  $X = (X_1, Z)$ ,  $U(\theta) = Y - \mathbb{E}(Y|X_1)$ , see Fan and Li (1996), Lavergne and Vuong (2000), Aït-Sahalia, Bickel and Stoker (2001), Delgado and Gonzalez-Manteiga (2001), and Lavergne (2001). Chen and Fan (1999) consider other types of nonparametric restrictions.

Finally, when testing for a parametric conditional distribution function,

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

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 (the random)  $X$ , but also by (the non-random)  $y$ . Such a pattern also appears when testing for conditional independence, see Delgado and Gonzalez-Manteiga (2001). Though we do not pursue this issue, our approach could be extended to these hypotheses by rewriting  $H_0$  as depending on  $X$  only through an integral over the domain of  $y$ , as done by 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  $\mathbb{E}\|Z\| < \infty$ .*

A)  $\mathbb{E}(Z | X) = \mathbb{E}(Z)$  a.s.  $\iff \mathbb{E}(Z | X'\beta) = \mathbb{E}(Z)$  a.s.  $\forall \beta \in \mathbb{R}^q : \|\beta\| = 1$ .

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

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

*is included in a finite union of circles and points on the unitary sphere, has Lebesgue measure zero on the sphere and is not dense.*

As noted by Escanciano (2006), Lemma 2.1-A can be deduced from Theorem 1 of Bierens (1982). It can also be found in Chen (1991). We provide here a short proof of this result.

**Proof.** A) Without loss of generality, take  $c = 1$  and  $\mathbb{E}(Z) = 0$ . The implication is immediate. To prove the reverse, 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 elementary properties of the conditional expectation, we obtain that for any  $\beta$ , including  $\beta = 0$ ,

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

From the unicity of Fourier transforms,  $\mathbb{E}(Z|X)f(X) = 0$  a.s., and then  $\mathbb{E}(Z|X) = 0$  a.s.

B) By Bierens and Ploberger (1997, Theorem 1),  $A = \{\beta \in \mathbb{R}^q : \mathbb{E}[\exp\{iX'\beta\}Z] = 0\}$  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 unitary sphere is thus a finite union of circles and points. ■

As Lemma 2.1 extends to any one-to-one transform of  $X$ , assuming a bounded  $X$  entails no loss of generality. Part B then ensures that when the conditional expectation  $\mathbb{E}(Z|X)$  is not constant, the search for a direction  $\beta$  such that  $\mathbb{E}(Z|X'\beta) \neq \mathbb{E}(Z)$  is not vain. Our result readily yields a new formulation of  $H_0$ .

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

$$\max_{\|\beta\|=1} \mathbb{E}[U'(\theta_0)\mathbb{E}(U(\theta_0)|X'\beta)\omega(X'\beta)] = 0 \quad \text{for some } \theta_0 \in \Theta. \quad (3)$$

As already mentioned, our approach is closely related to Bierens (1982), who showed that for bounded  $X$ ,  $\mathbb{E}[Z | X] = 0$  iff  $\mathbb{E}[Z \exp\{iX'\beta\}] = 0$  for all  $\beta$ . Stinchcombe and White extended Bierens' result showing that  $\mathbb{E}[Z | X] = 0$  is equivalent to  $\mathbb{E}[Z\phi(X'\beta)] = 0$  for

any  $\beta$ , whenever  $\phi(\cdot)$  is analytic non polynomial. Our approach is however different in a key aspect. Instead of working with a particular known  $\phi(\cdot)$  at the outset, we choose for each  $\beta$  a function of  $X'\beta$  maximizing  $\mathbb{E}^2 [Z\phi(X'\beta)]$ . This strategy enables better detection of departures from the null hypothesis. It is easily shown that the solution is proportional to  $\mathbb{E}(Z | X'\beta)$ . Now, looking for the *least favorable direction*  $\beta$  for the null hypothesis yields (3) 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  $\mathbb{E}^2 [Z\phi(X'\beta)\omega(X'\beta)]$  and  $\omega(\cdot)$  is not identically one.

### 2.3 A general dimension-reduction approach

Our goal is thus to estimate the quantity in (3). Assume we have at our disposal a consistent estimator  $\hat{\theta}_n$  of  $\theta_0$  and denote by  $U_i(\theta)$  the data-dependent vector function of  $\theta$  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

$$Q_n(\theta, \beta) = \frac{1}{n} \sum_{i=1}^n U'_i(\theta) \hat{\gamma}_i(X'_i\beta; \theta) , \quad (4)$$

which should converge uniformly in  $\beta$  to  $Q(\theta, \beta) = \mathbb{E}[U'(\theta)\mathbb{E}(U(\theta)|X'\beta)\omega(X'\beta)]$  under suitable conditions. When  $H_0$  does not hold, the maximum of  $Q_n(\hat{\theta}_n, \beta)$  over  $\beta$  stays away from zero almost surely and a test based on it is consistent. Under  $H_0$  however,  $Q_n(\hat{\theta}_n, \beta)$  converges to zero for any  $\beta$ . Hence, we introduce a penalized criterion and define

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

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

The penalty  $\pi_n(\cdot)$  is a nonnegative function that equals zero only at zero. Provided it is large enough with respect to  $\max_{\|\beta\|=1} Q_n(\hat{\theta}_n, \beta)$ , it forces the maximum to be attained at  $\beta_0$  under  $H_0$ . The critical value will then be the one of the test based on  $Q_n(\hat{\theta}_n, \beta_0)$ . Our criterion thus yields a pivotal distribution for the test statistic under  $H_0$ , which is of foremost interest in practice. While bootstrap can be used in small and moderate samples, one can avoid costly numerical simulations using asymptotic critical values in large samples. When  $H_0$  does not hold, the penalty should not perturb the behavior of the

maximum, hence  $\pi_n(t)$  should decrease towards zero fast enough for all  $t$  as  $n$  grows. The penalized criterion will thus select a direction different from  $\beta_0$  only when this direction gives more power to the test. The choice of the penalty is then crucial to control the level of the test and to ensure high power.

The direction  $\beta_0$  is theoretically irrelevant for consistency, but matters in practice. Since

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

and as the asymptotic critical values of the tests based on  $\widehat{\beta}_n$  and  $\beta_0$  are the same, our test is at least as powerful than the test tailored for alternatives depending on the single index  $X'\beta_0$ . These represent the favorite alternatives of the practitioner and thus allow to incorporate some a priori information. As shown by Janssen (2000), any omnibus test implicitly favors some type of alternatives. Our test makes this explicit by favoring alternatives depending on  $X'\beta_0$  or correlated with a function of  $X'\beta_0$ . If one suspect possible deviations from the model due to, say, the first component of  $X$ , this is easily accounted for in the test by choosing the first component of  $\beta_0$  large relative to the other components. We illustrate this characteristic feature in our simulation study.

### 3 Testing for a parametric regression

#### 3.1 The test

Let  $(Y, X)'$  be a random vector in  $\mathbb{R}^{1+q}$ . We consider the  $q$ -variate regression  $m(X) = \mathbb{E}(Y|X)$  and continuous  $X$ , as discrete regressors do not strictly speaking yield a “curse of dimensionality.” Let the parametric regression model be  $\{\mu(\cdot; \theta) : \theta \in \Theta\}$  with  $\Theta \subset \mathbb{R}^d$ . We apply our general approach to this testing problem using kernel estimators. To avoid handling denominators close to zero, we set the weight function  $\omega(\cdot)$  in (3) equal to the density of  $X'\beta$ , denoted by  $f_\beta(\cdot)$ , which is assumed to exist for any  $\beta$ . Let

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



By Corollary 2.2, the regression model is correctly specified iff  $\max_{\|\beta\|=1} Q(\theta_0, \beta) = 0$ . Let  $(Y_i, X_i)'$ ,  $i = 1, \dots, n$ , be a random sample from the distribution of  $(Y, X)'$ . The parameter  $\theta_0$  can be estimated for instance by nonlinear least-squares (NLLS), i.e.

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta} \sum_{i=1}^n (Y_i - \mu(X_i; \theta))^2, \quad (5)$$

with an appropriate convention in case of ties. In view of Equation (4), define

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

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  $\beta$ , 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  $\beta$  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  $\beta$  as

$$\hat{\beta}_n = \arg \max_{\|\beta\|=1} \left\{ nh^{1/2} Q_n(\hat{\theta}_n, \beta) - \alpha_n \mathbb{I}[\beta \neq \beta_0] \right\}, \quad (6)$$

where  $\beta_0$  is user-chosen and  $\alpha_n$ ,  $n \geq 1$ , is a sequence of positive real numbers decreasing to zero at an appropriate rate. Note that, with respect to Section 2.3, we have introduced the rate of convergence of  $Q_n(\theta, \beta)$  in our criterion, as it yields more transparent results. Our choice for the penalty function corresponds to the one of Bierens (1990) and is made for simplicity. We will prove that  $\hat{\beta}_n = \beta_0$  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  $\alpha$ -level test is given by  $\mathbb{I}(T_n \geq z_{1-\alpha})$ , where  $z_\alpha$  is the  $(1 - \alpha)$ -th quantile of the standard normal distribution. 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 take the minimum of the two variance estimators to improve the small sample power of our test and consider  $\mathbb{I}(T'_n \geq z_{1-\alpha})$ , where

$$T'_n = nh^{1/2} \frac{Q_n(\widehat{\theta}_n, \widehat{\beta}_n)}{\min\left(\widehat{v}_n(\beta_0), \widehat{v}_n(\widehat{\beta}_n)\right)}.$$

### 3.2 Assumptions

**Assumption D** (a) The random vectors  $(\varepsilon_1, X'_1)', \dots, (\varepsilon_n, X'_n)'$  are independent draws from the random vector  $(\varepsilon, X)'$  from  $\mathbb{R}^{1+q}$  with  $\mathbb{E}(\varepsilon | X) = 0$  and  $\mathbb{E}|\varepsilon|^{11} < \infty$ .

(b)  $\exists \underline{\sigma}^2$  and  $\bar{\sigma}^2$  such that  $0 < \underline{\sigma}^2 \leq \sigma^2(x) := \mathbb{E}(\varepsilon^2 | X = x) \leq \bar{\sigma}^2 < \infty \forall x$ .

(c) For any  $\beta$  of norm one,  $X'\beta$  admits a density  $f_\beta(\cdot)$  that is bounded uniformly in  $\beta$ .

We now 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  $\mathbb{E}[\Phi_2^2(X)] < \infty$ ; and

(iii)  $\forall \varepsilon > 0$ , there is a  $\eta > 0$  such that  $\mathbb{E} \sup_{\|\theta_1 - \theta_2\| \leq \eta} \|\dot{\mu}(X; \theta_1, \theta_2) - \dot{\mu}(X; \theta_2, \theta_2)\| < \varepsilon$ .

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

A large range of parametric models satisfies Assumption M. Together with our assumptions on the design, it ensures the  $\sqrt{n}$ -consistency of the NLLS estimator (5) as stated in Lemma 6.1. We make the following assumptions on the kernel and bandwidth.

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

We need to estimate the asymptotic conditional variance of  $nh^{1/2}Q_n(\widehat{\theta}_n, \beta)$ , which writes

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

With at hand a nonparametric estimator of the conditional variance such that

$$\sup_{1 \leq i \leq n} \left| \frac{\widehat{\sigma}^2(X_i)}{\sigma^2(X_i)} - 1 \right| = o_{\mathbb{P}}(1) , \quad (7)$$

$$\widehat{v}_n^2(\beta) = \frac{2}{n(n-1)} \sum_{j \neq i} \widehat{\sigma}^2(X_i) \widehat{\sigma}^2(X_j) h^{-1} K_h^2((X_i - X_j)' \beta) = v_n^2(\beta) (1 + o_{\mathbb{P}}(1)) . \quad (8)$$

Different estimators can be used. For instance, one can consider

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

where  $b$  is a bandwidth parameter chosen independently of  $h$ . Guerre and Lavergne (2005) provide some primitive conditions for (7). Given our focus, we shall proceed under (8).

### 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  $\beta$  under  $H_0$ . It has the following decomposition

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

**Lemma 3.1** *Under Assumptions D and K (i)  $\sup_{\|\beta\|=1} |Q_{0n}(\beta)| = O_{\mathbb{P}}(n^{-1}h^{-1/2} \ln n)$  under  $H_0$ . (ii)  $\sup_{\|\beta\|=1} |Q_n(\widehat{\theta}_n, \beta) - Q_{0n}(\beta)| = o_{\mathbb{P}}(n^{-1}h^{-1/2})$  if M holds and  $\|\widehat{\theta}_n - \theta_0\| = O_{\mathbb{P}}(n^{-1/2})$ .*

The proof is given in Section 5. We now describe the behavior of  $\widehat{\beta}_n$  under  $H_0$ .

**Lemma 3.2** *Under Assumptions D, M, and K, for  $\alpha_n \geq 0$  such that  $\alpha_n/\ln n \rightarrow \infty$ ,  $\mathbb{P}(\widehat{\beta}_n = \beta_0) \rightarrow 1$  under  $H_0$ .*

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

**Theorem 3.3** *Under Assumptions D, M, K, (8), and  $\alpha_n/\ln n \rightarrow \infty$ , the tests based on  $T_n$  or  $T'_n$  have asymptotic level  $\alpha$  given the design.*

**Proof.** From Lemma 3.2,  $\mathbb{P} \left[ Q_n(\widehat{\theta}_n, \widehat{\beta}_n) = Q_n(\widehat{\theta}_n, \beta_0) \right]$  and  $\mathbb{P} \left[ \widehat{v}_n^2(\widehat{\beta}_n) = \widehat{v}_n^2(\beta_0) \right]$  both converge to one. By Condition (8),  $\widehat{v}_n^2(\beta_0) = v_n^2(\beta_0)(1 + o_p(1))$ . 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

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

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

### 3.4 Consistency

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

$$\begin{aligned} T'_n \geq T_n &= \frac{nh^{1/2}Q_n(\widehat{\theta}_n, \widehat{\beta}_n)}{\widehat{v}_n(\beta_0)} \\ &= \frac{1}{\widehat{v}_n(\beta_0)} \left[ \max_{\|\beta\|=1} \left\{ nh^{1/2}Q_n(\widehat{\theta}_n, \beta) - \alpha_n \mathbb{I}(\beta \neq \beta_0) \right\} + \alpha_n \mathbb{I}(\widehat{\beta}_n \neq \beta_0) \right] \\ &\geq \frac{1}{\widehat{v}_n(\beta_0)} \left[ \max_{\|\beta\|=1} nh^{1/2}Q_n(\widehat{\theta}_n, \beta) - \alpha_n \right] \geq \frac{nh^{1/2}Q_n(\widehat{\theta}_n, \beta) - \alpha_n}{v_n(\beta_0)(1 + o_{\mathbb{P}}(1))} \quad \forall \beta. \quad (9) \end{aligned}$$

Hence, the test based on  $T_n$  (or  $T'_n$ ) is consistent if the last minorant stays away from zero with probability going to one for some  $\beta$ . When the parametric model is misspecified, our test is consistent under the assumptions of Theorem 3.3 provided  $\widehat{\theta}_n$  converges to some pseudo-true value  $\theta^*$ , since there exists at least one  $\beta$  for which  $Q(\theta^*, \beta) > 0$ .

### 3.4.1 Behavior against nonparametric directional alternatives

Consider a real-valued function  $\delta(X)$  such that

$$\mathbb{E}[\delta(X)\dot{\mu}(X; \theta_0)] = 0 \quad \text{and} \quad 0 < \mathbb{E}[\delta^4(X)] < \infty, \quad (10)$$

and the sequence of alternatives defined as

$$H_{1n} : m_n(X) = \mu(X; \theta_0) + r_n \delta(X), \quad n \geq 1. \quad (11)$$

We show below that such directional alternatives can be detected as soon as  $r_n^2 n h^{1/2} / \alpha_n$  tends to infinity. The conditions of Theorem 3.2 yield that  $\alpha_n$  is of the form  $a_n \ln n$  with  $a_n$  diverging at an arbitrary slow rate. Hence to obtain consistency against (11), we should have  $r_n^2 n h^{1/2} / (a_n \ln n) \rightarrow \infty$  where  $h$  applies to the univariate variable defined by a single linear index in  $X$ . By comparison, when one uses a standard multidimensional smooth test,  $r_n^2 n h^{q/2} \rightarrow \infty$  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 remains close to the power obtained in the unidimensional case. Note that we do not impose any smoothness restriction on  $\delta(\cdot)$ , as is frequent in this kind of analysis, see e.g. Zheng (1996).

**Theorem 3.4** *Under Assumptions D, M, K, (8), and  $r_n^2 n h^{1/2} / \alpha_n \rightarrow \infty$ , the tests based on  $T_n$  and  $T'_n$  are consistent given the design against the sequence of alternatives  $H_{1n}$  with  $\delta(X)$  satisfying (10).*

### 3.4.2 Behavior against classes of low-dimensional alternatives

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

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

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

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

$$\forall n, \quad \mathbb{E}[\delta_n(X'\beta)\dot{\mu}(X; \theta_0)] = 0 \quad \text{and} \quad 0 < \mathbb{E}[\delta_n^4(X'\beta)] < C < \infty, \quad (12)$$

and the sequence of alternatives defined as

$$H'_{1n} : m_n(X) = \mu(X; \theta_0) + \delta_n(X'\beta), \quad n \geq 1. \quad (13)$$

For such alternatives, a specific inference problem arises. As a high-dimensional design is projected onto a lower dimension space, the density of the low dimension vector generally goes to zero at the endpoints of its support. For instance, even if the density of  $X$  is bounded away from zero on  $(0, 1)^p$ , the density of  $X'\beta$  behaves as  $x^{q-1}$  around zero if  $\beta$  is not parallel to a side. This issue has been investigated for nonparametric curve estimation by Hall *et al.* (1997). In nonparametric testing, it is impossible to detect any alternative that concentrates in a low density area, see Guerre and Lavergne (2002, Prop. 2). This explains why these authors, as well as Horowitz and Spokoiny (2001), assume a design whose density is bounded away from zero. In our setting, such an assumption is clearly not generally relevant, and explains our following assumptions.

**Assumption N** (a)  $X$  is bounded. (b)  $K(\cdot)$  has a nonnegative Fourier transform and  $\int |u|K(u) du < \infty$ . (c) For any  $\beta$ , either (i)  $\exists C > 0$  and  $a > 0$  such that  $|f_\beta(t) - f_\beta(t')| \leq C|t - t'|^a$  for any  $t, t'$  and  $\exists c_0 > 0$  and an integer  $k_0$  such that for each  $\beta$  and  $0 < c \leq c_0$ ,  $A_{\beta,c} = \{u : f_\beta(u) \geq c\}$  is a union of at most  $k_0$  intervals on the real line, or (ii)  $f_\beta(\cdot)$  is bounded away from zero.

**Theorem 3.5** Let  $\kappa_n \rightarrow \infty$  and consider the class of alternatives (13) with unknown  $\theta_0$  and  $\beta$ ,  $\delta_n(\cdot) \in C(L, s)$  for some  $s > 3/4$  and  $L > 0$ ,  $\mathbb{E}\delta_n^2 = o(1)$ ,

$$\left[ n^{-1} \sum_{i=1}^n \delta_n^2(X'_i\beta) \right]^{1/2} \geq \kappa_n(1 + o_{\mathbb{P}}(1)) \left( \frac{\alpha_n}{n} \right)^{\frac{2s}{4s+1}}. \quad (14)$$

In case of Assumption N-(c)(i), let  $a_n \rightarrow 0$  such that  $a_n\sqrt{nh} \rightarrow \infty$ ,  $a_n h^{-a} \rightarrow \infty$ , and

$$\left[ n^{-1} \sum_{i=1}^n \delta_n^2(X_i' \beta) \mathbb{I}(X_i' \beta \notin A_{\beta, a_n}) \right] = o_{\mathbb{P}}(1) \left[ n^{-1} \sum_{i=1}^n \delta_n^2(X_i' \beta) \right]. \quad (15)$$

In case of Assumption N-(c)(ii), let  $a_n = 1$ . Under Assumptions D, M, K, N, and (8), if  $h$  is of order  $(\alpha_n/n)^{\frac{2}{4s+1}}$ , the tests based on  $T_n$  and  $T'_n$  are consistent given the design if  $\kappa_n^2 a_n \rightarrow \infty$ .

To justify Condition (15), note that outside  $A_{\beta, a_n}$  the density is  $O(a_n)$  by definition, and the volume of the complement of  $A_{\beta, a_n}$  is typically an  $o(1)$  when  $X$  is bounded. What our condition excludes is thus alternatives that concentrate on low density areas. Aside this technicality, the rate obtained in Theorem 3.5 is almost optimal, and differs from the optimal rate only because  $\alpha_n$  enters in its formula. This is a low price to pay that is independent of the dimension of  $X$ . Note that we have to impose  $s > 3/4$  because of our Assumption K-(b).

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

## 4 Implementation

### 4.1 Bootstrap critical values

We use the smooth conditional moments bootstrap introduced by Gozalo (1997). It consists in drawing  $n$  i.i.d. random variables  $\omega_i$  independent from the original sample with  $\mathbb{E}\omega_i = 0$ ,  $\mathbb{E}\omega_i^2 = 1$ , and  $\mathbb{E}\omega_i^4 < \infty$ , to obtain bootstrap observations of  $Y_i$  as  $Y_i^* = \mu(X_i, \hat{\theta}_n) + \hat{\sigma}(X_i)\omega_i$ ,  $i = 1, \dots, n$ . A bootstrap test statistic is built from the bootstrap sample as was the original test statistic. When this scheme is repeated many times,

the bootstrap critical value  $z_{1-\alpha,n}^*$  at level  $\alpha$  is the empirical  $(1-\alpha)$ -th quantile of the bootstrapped test statistics. This critical value is then compared to the initial test statistic. The validity of this procedure follows from our results and its proof is thus omitted.

**Theorem 4.1** *Under the assumptions of Theorem 3.3 and (7), the bootstrap critical values yield a test based on  $T_n$  or  $T'_n$  with asymptotic level  $\alpha$  given the design.*

## 4.2 Practical choices in the procedure

For the construction of our basic statistics  $Q_n(\hat{\theta}_n, \beta)$ , choices of a kernel and a bandwidth are required. The kernel is not expected to have much influence, though the bandwidth choice may be important. Note that the same bandwidth is used for all directions  $X'\beta$ . Since our fundamental lemma applies to any one-to-one transform of  $X$ , one can transform  $X$  so that its covariance matrix is identity or its support is  $(0, 1)^p$ . A data-driven choice of the bandwidth in the spirit of Guerre and Lavergne (2005) should be investigated, but this is outside the scope of our paper.

Maximizing our penalized criterion can be done on a fine enough grid on the hypersphere. Lemma 2.1-B ensures that under the alternative any direction yields a consistent test but a set of measure zero. The specific choices for our procedure are  $\beta_0$  and  $\alpha_n$ . As explained before,  $\beta_0$  corresponds to a class of favored alternatives to be determined in each particular case depending on a priori information. We investigate in the next section how its choice affects the small sample performances of our test. The choice of  $\alpha_n$  is crucial and reflects the weight given to the favored alternatives. It should be of an slightly higher order than  $nh^{1/2}Q_n(\hat{\theta}_n, \beta_0)$  times  $\ln n$  under the null hypothesis. In practice, one can always simulate this statistic under  $H_0$  and obtain an estimate of its standard deviation to calibrate  $\alpha_n$ . We implemented this device in our simulations with success.



### 4.3 Simulation study

Our first aim was to study the sensitivity of our test to the penalty  $\alpha_n$ , the direction defined by  $\beta_0$  and the bandwidth  $h$ . Our second aim was to compare the small sample power of our test to the multivariate test of Zheng (1996) and Li and Wang (1998), to the Integrated Conditional Moment (ICM) test of Bierens (1982), and to the test recently proposed by Escanciano (2006). We chose the version of the ICM test studied by Escanciano (2006), which is based on

$$n \int_{\mathbb{R}^q} \left| \frac{1}{n} \sum_{i=1}^n U_i(\theta) \exp(iX_i'\beta) \right|^2 \phi(\beta) d\beta = \frac{1}{n} \sum_{i,j} U_i(\theta) U_j(\theta) \exp\left(-\frac{\|X_i - X_j\|^2}{2}\right),$$

where  $\phi(\beta)$  is the standard normal density on  $\mathbb{R}^q$ , see Bierens (1982, p. 111). Though the asymptotic theory developed by Bierens and Ploberger (1997) applies for a measure with compact support, it makes no practical difference to consider the standard normal or a version of it truncated at an arbitrarily large value. Dominguez (2004) shows that the wild bootstrap is valid and preserves admissibility of the test, so we used this method to obtain critical values. Escanciano's test is based on the statistic

$$\frac{1}{n^2} \sum_{i,j} U_i(\theta) U_j(\theta) \left( \frac{1}{n} \sum_k \int_{\{\beta \in \mathbb{R}^q: \|\beta\|=1\}} \mathbb{I}(X_i'\beta \leq X_k'\beta) \mathbb{I}(X_j'\beta \leq X_k'\beta) d\beta \right).$$

Computation of the statistic was performed using Escanciano's (2006) analytic results, see his Appendix B, and the wild bootstrap was used to obtain critical values.

We considered as the null hypothesis a linear model

$$H_0 : \mathbb{E}(Y|X) = \theta_0 + \sum_{j=1}^3 \theta_j X_j.$$

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  $[-\sqrt{3}, \sqrt{3}]$  to get unit variance. We sampled independent  $N(0, 1)$  errors and constructed the response variable as

$$Y_i = 0.5X_1 - X_2 + 0.5X_3 + d \cos\left(\pi X' \beta^* / \sqrt{3}\right) + \varepsilon_i \quad i = 1, \dots, 50,$$

with varying  $d$  and different vectors  $\beta^*$ .

For the nonparametric-based tests, 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 compute the test statistics, we used the normal kernel and we selected the bandwidth as  $h = bn^{-2/(8+q)}$ , with  $q = 3$  in Case (ii) and  $q = 1$  in the other cases, and  $b$  varies in  $\{0.5, 1, 1.5, \dots, 4\}$ . The errors’ conditional variance was estimated by a kernel estimator with normal kernel and bandwidth  $2n^{-1/6}$ . For our test, optimization was carried out on a grid of 2000 points uniformly sampled on the three-dimensional hypersphere of unit radius. To set  $\alpha_n$ , we computed  $v_0$ , the mean of  $v_n(\beta_0)$ , which was found to vary little with  $\beta_0$ . We then chose  $\alpha_n = a \ln(50)v_0n^{-1}h^{-1/2}$ , with  $v_0 = 0.3$  in our case, and we let  $a = 1, 2, 4$ . For each considered test, we used 199 bootstrap samples for each replication to compute the critical value.

*Insert Figure 1 here*

We first set  $\beta^* = (1, 2, 3)'/\sqrt{14}$  and  $\beta_0 = (1, 1, 1)'/\sqrt{3}$ , a natural choice if one does not favor any regressor at the outset. We then drew the power curves of the different tests based on 399 samples for each point of the grid  $d = 0.2, 0.4, \dots, 1.2$  for the smooth tests and 1000 samples for the other tests. Figure 1 compares the power curves of the tests when the bandwidth constant  $b$  is set to 1.5. Clearly there is a 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. Our test outperforms Zheng’s test in dimension 3, as well as Bierens’ and Escanciano’s ones. The power curves become closer as  $\alpha_n$  increases, as expected. Still the gain in power with respect to Zheng’s test is large even for  $a = 4$ . To get a more precise idea, for  $d = 0.6$  the power is 30.2% for Escanciano’s test, around 37% for Bierens’ and Zheng’s test, and 51.6% for our test. It is not surprising that the ICM test has good performances against a cosine alternative, since the test is based on the correlation between the parametric residuals and trigonometric functions. Figure 2 corresponds to the case where  $b = 1$ . The same patterns appear, but the level of our test is not well calibrated. We thus recommend that a large number of bootstrap samples should be used in practice.

*Insert Figure 2 here*

We then set  $\beta^* = (0, 0, 1)$  and  $\beta_0 = \beta^*$ . This corresponds to the situation where we favor alternatives depending upon one particular variable and the deviation depends on this variable only. The power curves when  $b = 1.5$  are drawn in Figure 3. Irrespective of the value of  $\alpha_n$ , our test performs almost as well as the the infeasible test.

*Insert Figure 3 here*

Additional simulations results reported in Lavergne and Patilea (2006) show that our test exhibit good performances against other forms of alternatives even if they do not depend on a single-index only.

## 5 Conclusion

Our general approach to testing conditional moment restrictions in high dimensions relies on the equivalence of testing  $\mathbb{E}(Z|X) = 0$  or  $\mathbb{E}(Z\mathbb{E}(Z|X'\beta)) = 0$  for all  $\beta$  of norm 1. In practice, an index is selected by maximizing an estimator of the previous quantity minus a penalty function. Our approach potentially applies to many testing problems as explained in Section 2.1. We have studied testing for a parametric regression function. The test has known asymptotic critical values. It behaves against directional alternatives and against a class of low-dimensional regular alternatives almost as if the dimension of  $X$  was one. Our simulations results illustrate the good behavior of the test in small samples.

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

## 6 Technicalities

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

The proof is based on a uniform law of large numbers, see e.g. Pakes and Pollard (1989) and standard arguments, and is omitted, see Lavergne and Patilea (2006) for details.

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

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

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

$$\begin{aligned} \mathbb{P} \left( \sup_{\beta} |U_n \tilde{g}| \geq \frac{th^{1/2} \ln n}{n-1} \right) &= \mathbb{P} \left( \sup_{\beta} \left| \frac{1}{n} \sum_{j \neq i} \frac{\eta_i^M}{M} \frac{\eta_j^M}{M} K_h((X_i - X_j)' \beta) \right| \geq \frac{th^{1/2} \ln n}{M^2} \right) \\ &\leq C_1 C_2 \exp \left\{ -C_3 \left( \frac{th^{1/2} \ln n}{M^2 \sigma_M} \right) \right\} \quad \text{for any } t > 0, \end{aligned} \quad (16)$$

$$\text{provided} \quad n\sigma_M^2 \geq \frac{th^{1/2} \ln n}{M^2 \sigma_M} \geq C_4 [C_5 + \max(\ln C_2 / \ln n, 0)]^{3/2} \ln \frac{2}{\sigma_M} \quad (17)$$

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

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

From Assumption D and the Lebesgue dominated convergence theorem, there is a constant  $C > 0$  independent of  $n$  such that  $C^{-1} \leq \sigma_M^2 M^4 / h \leq C$ . Take  $M^4 = nh(\ln n)^{-(1+\delta)}$  with  $\delta > 0$  arbitrarily small. Hence  $\sigma_M^2$  is of order  $n^{-1} \ln^{1+\delta} n$  and for any  $t$

$$n\sigma_M^2 \geq \frac{nh}{CM^4} = C^{-1} \ln^{1+\delta} n \geq \frac{th^{1/2} \ln n}{M^2 \sigma_M} \quad (18)$$

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

$$\frac{th^{1/2} \ln n}{M^2 \sigma_M} \geq C^{-1/2} t \ln n \geq C' \ln n \quad (19)$$

for any sufficiently large  $t$ . Since  $(\ln n)^{-1} \ln(2/\sigma_M)$  tends to a nonnegative constant as  $n$  goes to  $\infty$ , Equations (18) and (19) show that (17) is satisfied for our  $M$ , with  $n$  and  $t$  large enough. Hence (16) yields  $U_n \tilde{g} = O_P(n^{-1} h^{1/2} \ln n)$ . Now, it remains to study the tails of  $\varepsilon_i$ , that is we have to derive the orders of the remainder terms

$$2R_{1n} + R_{2n} = \frac{2}{n(n-1)} \sum_{j \neq i} \eta_i^M \xi_j K_h((X_i - X_j)' \beta) + \frac{1}{n(n-1)} \sum_{j \neq i} \xi_i \xi_j K_h((X_i - X_j)' \beta)$$

where  $\xi_i = \varepsilon_i - \eta_i^M = \varepsilon_i \mathbb{I}(|\varepsilon_i| > M) - \mathbb{E}[\varepsilon_i \mathbb{I}(|\varepsilon_i| > M) | X_i]$ .

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

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

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

(ii) For  $V_{n,M'}(\theta_0) = \{\theta \in \Theta : \|\theta - \theta_0\| \leq M'/n^{1/2}\}$ ,  $\liminf_{M' \rightarrow \infty} \lim_n \mathbb{P}[\hat{\theta}_n \in V_{n,M'}(\theta_0)] = 1$  by Lemma 6.1. 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  $\mathbb{E}[g_{\theta,h,\beta}(W_i, W_j) | W_j] = 0$ . 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  $hQ_{1n}(\theta, \beta)$  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. By the Main Corollary of Sherman (1994a) with  $p = 1$  and  $k = 2$ ,

$$\mathbb{E} \left[ \sup_{\theta \in V_n(\theta_0), h, \beta} |n U_n \bar{g}_{\theta,h,\beta}| \right] \leq \Lambda \left[ \mathbb{E} \sup_{\theta \in V_n(\theta_0), h, \beta} \{U_{2n} \bar{g}_{\theta,h,\beta}^2\}^{\bar{\alpha}} \right]^{1/2} \quad (20)$$

where  $\Lambda$  is a universal constant and any  $\bar{\alpha}$  such that  $0 < \bar{\alpha} < 1$ . 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  $C$  and  $C'$ . Hence from Inequality (20) and Chebyshev's inequality,

$$\sup_{\|\beta\|=1} |nh^{-1/2}U_n\bar{g}_{\theta_n, h, \beta}| = O_{\mathbb{P}}\left((nh^{1/\bar{\alpha}})^{-\bar{\alpha}/2}\right). \quad (21)$$

Next, let  $\mathbb{P}_n\tilde{g}$  denote the empirical process in Hoeffding's decomposition of  $hQ_{1n}(\theta, \beta)$ , where

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

$\tilde{g}_1(W_i) = \mathbb{E}[\dot{\mu}(X_j; \theta_0) K_h((X_i - X_j)' \beta) \mid X_i] \varepsilon_i$  and  $\tilde{g}_2(W_i) = \mathbb{E}[\ddot{\mu}(X_j; \theta_0, \theta_0) K_h((X_i - X_j)' \beta) \mid X_i]$ .

Reasoning as above and using Assumption M-(a), it can be shown that

$$\sup_{\theta \in V_n(\theta_0), \beta} \left| nh^{-1/2}(\theta - \theta_0)' \mathbb{P}_n \tilde{g}_1 \right| = O_{\mathbb{P}}\left(h^{(3\bar{\alpha}/4)-1/2}\right) \quad (22)$$

$$\text{and } \sup_{\theta \in V_n(\theta_0), \beta} \left| nh^{-1/2}(\theta - \theta_0)' \mathbb{P}_n \tilde{g}_2(\theta - \theta_0) \right| = O_{\mathbb{P}}\left(n^{-1/2}h^{(3\bar{\alpha}/4)-1/2}\right). \quad (23)$$

From Equations (21), (22), and (23) with  $\bar{\alpha} > 2/3$ ,

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

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

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

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

**Lemma 6.2** *Let  $W_{\beta}$  be the matrix with generic element  $\mathbb{I}(i \neq j) K_h((X_i - X_j)' \beta) / (hn(n-1))$ . 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  $\beta$ .*

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

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

From Assumption K, 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,  $\mathbb{E} [h^{-1} K_h((X_i - X_j)' \beta) | X_i] \leq C$  uniformly in  $i$  by Assumptions D-(c), K-(a), and a change of variables. This gives the first result. Next,

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

follows like in the proof of (26) below with  $\delta(X) \equiv 1$  and  $K(\cdot)$  replaced by  $K^2(\cdot)/\int K^2(u) du$ . The last quantity is bounded from above and below by Assumptions D-(c) and K-(a). ■

**Proof of Theorems 3.4 and 3.5.** By Assumption D-(b),  $v_n^2(\beta) \leq \bar{\sigma}^4 n^2 h \|W_\beta\|^2$ , and by Lemma 6.2,  $v_n^2(\beta)$  and  $nh^{1/2} Q_{0n}(\beta)$  are bounded in probability for any  $\beta$ . Under  $H_{1n}$ ,  $U_i(\hat{\theta}_n) = \mu(X_i; \theta_0) + \delta_n(X_i) + \varepsilon_i - \mu(X_i; \hat{\theta}_n)$ , with  $\delta_n(\cdot) = r_n \delta(\cdot)$  for Theorem 3.4. By simple algebra,  $Q_n(\hat{\theta}_n, \beta)$  writes for any  $\beta$  as

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

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

Lemma 3.1-(ii) deals with  $Q_{1n}(\hat{\theta}_n, \beta)$  and  $Q_{2n}(\hat{\theta}_n, \beta)$ .

For Theorem 3.4, it is shown below that for any  $\beta$

$$Q_{3n}(\widehat{\theta}_n, \beta) = O_{\mathbb{P}}(r_n n^{-1/2}) \quad (24)$$

$$Q_{4n}(\beta) = O_{\mathbb{P}}(r_n n^{-1/2}) \quad (25)$$

$$Q_{5n}(\beta) = r_n^2 \mathbb{E} \left[ \mathbb{E}^2[\delta(X)|X'\beta] f_{\beta}(X'\beta) \right] + o_{\mathbb{P}}(r_n^2). \quad (26)$$

Collecting results, it follows that for any  $\beta$

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

Choose  $\beta$  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 (9).

For Theorem 3.5, it is shown below that for any  $\beta$

$$Q_{3n}(\widehat{\theta}_n, \beta) = O_{\mathbb{P}}(n^{-1/2}) \left[ n^{-1} \sum_{i=1}^n \delta_n^2(X'_i \beta) \right]^{1/2} \quad (27)$$

$$Q_{4n}(\beta) = O_{\mathbb{P}}(n^{-1/2}) \left[ n^{-1} \sum_{i=1}^n \delta_n^2(X'_i \beta) \right]^{1/2} \quad (28)$$

$$Q_{5n}(\beta) \geq C(1 + o_{\mathbb{P}}(1)) \left[ \sqrt{a_n} \left( n^{-1} \sum_{i=1}^n \delta_n^2(X'_i \beta) \right)^{1/2} - h^s \right]^2, \quad (29)$$

where (29) holds with probability going to one. Collecting results, it follows that

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

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

**Proof of (24) and (27).** Since  $|u'W_{\beta}v| \leq \|u\| \|v\| \text{Sp}(W_{\beta})$ , then for any  $\beta$ ,

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

By Lemma 6.2,  $n \text{Sp}(W_{\beta}) = O_{\mathbb{P}}(1)$ . By Assumption M,  $[\mu(X_i; \widehat{\theta}_n) - \mu(X_i; \theta_0)]^2 \leq \tilde{\Phi}^2(X_i) \|\widehat{\theta}_n - \theta_0\|^2$  for some  $\tilde{\Phi}(\cdot)$  with  $\mathbb{E}[\tilde{\Phi}^4(X)] < \infty$  and from Lemma 6.1,  $\sum_{i=1}^n [\mu(X_i; \widehat{\theta}_n) - \mu(X_i; \theta_0)]^2 = O_{\mathbb{P}}(1)$ . Hence the initial quantity is bounded by  $O_{\mathbb{P}}(n^{-1/2}) \left[ n^{-1} \sum_{i=1}^n \delta_n^2(X_i) \right]^{1/2}$ . ■



**Proof of (25) and (28).** Denote by  $\mathbb{E}_n$  the conditional expectation given the  $X_i$  and let

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

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

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

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

**Proof of (26).** Consider  $U_n = r_n^{-2} Q_{5n}(\beta)$ . By straightforward computations,

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

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

$$\mathbb{E}(U_n) = \frac{1}{2\pi} \int |\mathbb{E}[\mathbb{E}[\delta(X)|X'\beta] \exp(itX'\beta)]|^2 \widehat{K}(ht) dt .$$

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

**Proof of (29).** We treat the case of Assumption N(c)(i), the proof for case (ii) follows the same lines but is much simpler since the density of  $f_\beta(\cdot)$  is bounded away from zero.

Let  $\overline{W}_\beta$  be the matrix with generic element

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

Let  $\delta_n = (\delta_n(X'_1 \beta), \dots, \delta_n(X'_n \beta))'$ , and define  $\delta_n \mathbb{I}_{A_{\beta, a_n}}$  and  $\delta_n \mathbb{I}_{A_{\beta, a_n}^c}$  similarly, where  $A^c$  stands for the complementary set of  $A$ . We have

$$\begin{aligned} \delta'_n (W_\beta - \overline{W}_\beta) \delta_n &= \delta'_n (W_\beta - \overline{W}_\beta) \delta_n \mathbb{I}_{A_{\beta, a_n}^c} + \delta'_n \mathbb{I}_{A_{\beta, a_n}^c} (W_\beta - \overline{W}_\beta) \delta_n \mathbb{I}_{A_{\beta, a_n}} \\ \text{and} \quad |\delta'_n (W_\beta - \overline{W}_\beta) \delta_n| &\leq 2 \text{Sp}(W_\beta - \overline{W}_\beta) \|\delta_n \mathbb{I}_{A_{\beta, a_n}^c}\| \|\delta_n\| . \end{aligned}$$

As in Lemma 6.2,  $n\text{Sp}(W_\beta - \overline{W}_\beta)$  has the same order as

$$\max_i \mathbb{E} \left[ h^{-1} K_h \left( (X_i - X_j)' \beta \right) \left( 1 - \mathbb{I}(X_i' \beta, X_j' \beta \in A_{\beta, a_n}) \right) | X_i \right].$$

Since  $1 - \mathbb{I}(x, y \in A_{\beta, a_n}) \leq \mathbb{I}(y \in A_{\beta, a_n}^c) + \mathbb{I}(x \in A_{\beta, a_n}^c)$ ,

$$\begin{aligned} & \mathbb{E} \left[ h^{-1} K_h \left( (X_i - X_j)' \beta \right) \left( 1 - \mathbb{I}(X_i' \beta, X_j' \beta \in A_{\beta, a_n}) \right) | X_i \right] \\ &= \int K(u) f_\beta(X_i' \beta - hu) \left( 1 - \mathbb{I}(X_i' \beta, X_i' \beta - hu \in A_{\beta, a_n}) \right) du \\ &\leq 2a_n \int K(u) du + C \int |hu|^a K(u) du = O(a_n) \end{aligned}$$

uniformly in  $i$ , using Assumptions N-(b), (c)(i), and  $a_n h^{-a} \rightarrow \infty$ . Hence, from Condition (15),

$$|\delta'_n(W_\beta - \overline{W}_\beta) \delta_n| \leq o_{\mathbb{P}}(a_n) \left[ n^{-1} \sum_{i=1}^n \delta_n^2(X_i' \beta) \right].$$

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

$$\begin{aligned} u' \overline{P}_\beta u &= \frac{1}{n(n-1)} \sum_{i, j} u_i u_j \mathbb{I}(X_i' \beta \in A_{\beta, a_n}) \mathbb{I}(X_j' \beta \in A_{\beta, a_n}) h^{-1} K_h \left( (X_i - X_j)' \beta \right) \\ &= \frac{n}{2\pi(n-1)h} \int \left| n^{-1} \sum_j u_j \exp \left( \frac{it X_j' \beta}{h} \right) \mathbb{I}(X_j' \beta \in A_{\beta, a_n}) \right|^2 \widehat{K}(t) dt \end{aligned} \quad (30)$$

is nonnegative and vanishes iff  $u_j \mathbb{I}(X_j' \beta \in A_{\beta, a_n}) = 0$  for all  $j$ . For  $\pi(\cdot) \in \Pi_{s, h}$ , let  $\pi$  denote the column vector  $(\pi(X_1' \beta), \dots, \pi(X_n' \beta))'$  and let

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

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

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

Then

$$\Lambda_n^2 \geq \inf_{b, J_k \subset A_{\beta, a_n}} \frac{\pi'_{b,k} \bar{P}_\beta \pi_{b,k}}{n^{-1} \sum_{i=1}^n \pi_{b,k}^2(X'_i \beta) \mathbb{I}(X'_i \beta \in A_{\beta, a_n})}, \quad (31)$$

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

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

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

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

Using (30) and  $t_k \in A_{\beta, a_n}$  implies that with probability going to one

$$\frac{1}{a_n} \frac{\pi'_{b,k} \bar{P}_\beta \pi_{b,k}}{n^{-1} \sum_{i=1}^n \pi_{b,k}^2(X'_i \beta) \mathbb{I}(X'_i \beta \in A_{\beta, a_n})} \geq \frac{\int |\hat{\pi}_b(t)|^2 \hat{K}(t) dt}{\int_{(-1/2, 1/2)} \pi_b^2(u) du}, \quad (32)$$

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

$$\begin{aligned} (\delta'_n \bar{P}_\beta \delta_n)^{1/2} &\geq (\pi'_n \bar{P}_\beta \pi_n)^{1/2} - ((\delta_n - \pi_n)' \bar{P}_\beta (\delta_n - \pi_n))^{1/2} \\ &\geq \Lambda_n n^{-1/2} \|\pi_n \mathbb{I}_{A_{\beta, a_n}}\| - \text{Sp}^{1/2}(\bar{P}_\beta) \|\delta_n - \pi_n\| \\ &\geq \Lambda_n n^{-1/2} \|\delta_n \mathbb{I}_{A_{\beta, a_n}}\| - \left( \Lambda_n + n^{1/2} \text{Sp}^{1/2}(\bar{P}_\beta) \right) Ch^s. \end{aligned} \quad (33)$$

From Conditions (14), (15), and  $\text{Sp}^{1/2}(\bar{P}_\beta) \leq \text{Sp}^{1/2}(\bar{W}_\beta) + \text{Sp}^{1/2}(\bar{W}_\beta - \bar{P}_\beta) = O_{\mathbb{P}}(n^{-1/2})$ , the above quantity is positive with probability going to one. Since  $\Lambda_n^{-2} n \text{Sp}(\bar{W}_h - \bar{P}_h) = o_{\mathbb{P}}(1)$ , deduce that with probability going to one

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

## Acknowledgement

The first author gratefully acknowledge support from NSERC under project 328474-06. We thank participants of ESRC 2005 (Bristol), CESG 2005 (Vancouver), ESF exploratory workshop on specification testing 2005 (Santander), seminars at the University of Toulouse and UBC, as well as the editor and a referee for various comments. Special thanks to Jean-Marie Rolin for a very helpful suggestion.

## References

- Aït-Sahalia, Y., Bickel, P., Stoker, T., 2001. Goodness-of-fit tests for kernel regression with an application to option implied volatilities. *Journal of 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. *Journal of Econometrics* 20, 105–134.
- Bierens, H.J., 1990. A consistent conditional moment test of functional form. *Econometrica* 58, 1443–1458.
- Bierens, H.J., Ploberger W., 1997. Asymptotic theory of integrated conditional moment tests. *Econometrica* 65, 1129–1151.
- Chen, X., Fan, Y., 1999. Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series. *Journal of Econometrics* 91, 373–401.
- Chen, P.D., 1991.  $L^1$ -theory of approximation by ridge functions. *Chinese Science Bulletin* 36, 1761–1764.
- Delgado, M.A., González-Manteiga, W., 2001. Significance testing in nonparametric regression based on the bootstrap. *Annals of Statistics* 29, 1469–1507.
- Delgado, M.A., Dominguez, M.A., Lavergne, P., 2006. Consistent tests of conditional moment restrictions. *Annales d’Economie et Statistique* 81, 33–67.
- Dominguez, M.A., 2004. On the power of bootstrapped specification tests. *Econometric Reviews* 23, 215–228.
- Donald, S.G., Imbens, G.W., Newey, W.K., 2003. Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics* 117, 55–93.
- Escanciano, J.C., 2006. A consistent diagnostic test for regression models using projections.

Econometric Theory 22, 1030–1051.

Fan, Y., 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. *Journal of Econometrics* 81, 357–393.

Gozalo, P.L., Linton, O.B., 2001. Testing additivity in generalized nonparametric regression models with estimated parameters. *Journal of Econometrics* 104, 1–48.

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

Guerre, E., Lavergne, P., 2005. Data-driven rate-optimal specification testing in regression models. *Annals of Statistics* 33, 840–870.

Hall, P., Marron, J.S., Neumann, M.H., Titterton, D.M., 1997. Curve estimation when the design density is low. *Annals of Statistics* 25, 756–770.

Hall, P., Yatchew, A., 2005. Unified approach for testing functional hypotheses in semiparametric contexts. *Journal of Econometrics* 127, 225–252.

Härdle, W., Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21, 1296–1947.

Härdle, W., Stoker, T., 1989. Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association* 84, 986–995.

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

Hong, Y., 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. *Journal of Econometrics* 58, 71–120.

Janssen, A., 2000. Global power functions of goodness of fit tests. *Annals of Statistics* 28, 239–253.

Lavergne, P., 2001. An equality test across nonparametric regressions. *Journal of Econometrics* 103, 307–344.

Lavergne, P., Patilea, V., 2006. Breaking the curse of dimensionality in nonparametric testing.

Working paper, Simon Fraser University. <http://www.sfu.ca/~pascall>

Lavergne, P., Vuong, Q., 2000. Nonparametric significance testing. *Econometric Theory* 16, 576–601.

Li, Q., Wang, S., 1998. A simple consistent bootstrap test for a parametric regression function. *Journal of Econometrics* 87, 145–165.

Major, P., 2006. An estimate on the supremum of a nice class of stochastic integrals and  $U$ -statistics. *Probability Theory and Related Fields* 134, 489–537.

Nolan, D., Pollard, D., 1987.  $U$ -processes : Rates of convergence. *Annals of Statistics* 15, 780–799.

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

Powell, J.L., Stock, J.H., 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. *Annals of Mathematical Statistics* 24, 220–238.

Rudin, W., 1987. Real and complex analysis. McGraw-Hill International Editions: Mathematics Series.

Sherman, R.P., 1994a. Maximal inequalities for degenerate  $U$ -processes with applications to optimization estimators. *Annals Statistics*. 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. *Annals of Statistics* 24, 2477–2498.

Stinchcombe, M.B., 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. *Annals of Statistics* 8, 1348–1360.

Stute, W., 1997. Nonparametric models checks for regression. *Annals of Statistics* 25, 613–641.

Stute, W., Zhu, L.X., 2005. Nonparametric checks for single-index models. *Annals of Statistics*

33, 1048–1083.

Whang, Y.J., (001. Consistent specification testing for conditional moment restrictions. *Economics Letters* 71, 299–306.

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

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

Zhu, L.X., Li R., 1998. Dimension-reduction type test for linearity of a stochastic model. *Acta Mathematicae Applicatae Sinica* 14, 165-175.

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

Figure 1:  $\beta_0 = (1, 1, 1)/\sqrt{3}$  and  $b = 1.5$

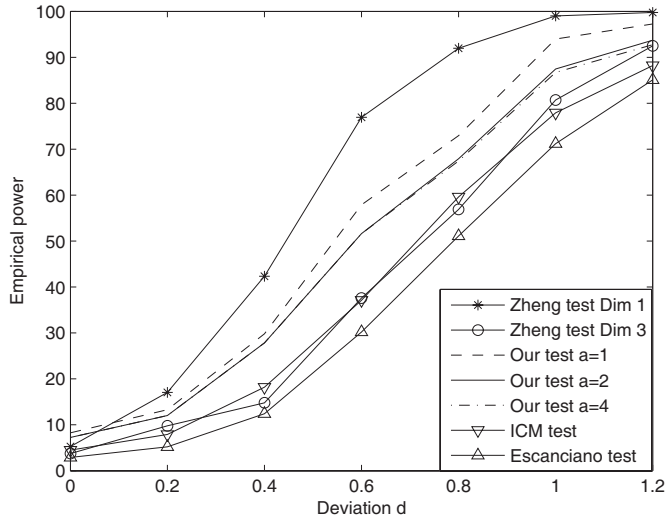


Figure 2:  $\beta_0 = (1, 1, 1)/\sqrt{3}$  and  $b = 1$

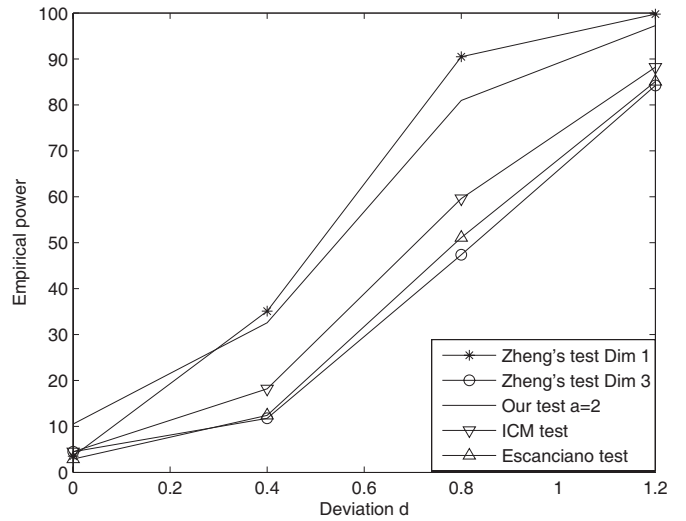


Figure 3:  $\beta_0 = (0, 0, 1)$

