One for all and all for one: Regression checks with many regressors

Pascal Lavergne
Simon Fraser University

Valentin Patilea
CREST-ENSAI

April 2007

Abstract

We develop a novel approach to consistent checks of parametric regression models when many regressors are present. The principle is to replace the nonparametric alternative by a class of semiparametric alternatives, namely single-index models, that is rich enough to allow detection of any nonparametric alternative. We propose an omnibus test based on the kernel method that performs against a sequence of directional local nonparametric alternatives as if there was one regressor only, whatever the number of regressors. This test can also be viewed as a smooth version of the integrated conditional moment (ICM) test of Bierens. For these reasons, we label our test the smooth ICM test. Moreover, qualitative information can be easily incorporated in the procedure to further improve its power. In an extensive simulation study, we provide evidence that our test is little sensitive to the smoothing parameter and performs better than several known lack-of-fit tests in multidimensional settings.

Keywords: Dimensionality, Hypothesis testing, Nonparametric methods.

AMS classification: Primary 62G10 ; Secondary 62G08.

The first author gratefully acknowledge support from the NSERC under project 328474-06. We thank seminar participants at Penn State University and Yale University for helpful comments.

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

Parametric forms are frequently used in regression models to estimate the association between a response variable and predictors. Checking the adequacy of a parametric regression function is then useful in many applications, whether in econometrics or in other applied fields. Popular graphical displays of residuals against fitted values or covariates can fail to detect an inadequate model when many covariates are present. Hence, since the end of the eighties, many regression checks have been developed. With few exceptions, notably Bierens (1982, 1990) and Stute, Gonzalez Manteiga and Presedo Quindimil (1998), most rely on some smoothing method, such as kernels, splines, local polynomials, or orthogonal series, from the earlier work of Cox and al. (1988), Azzalini, Bowman and Härdle (1989), Eubank and Spiegleman (1990), Hart and Wehrly (1992), Eubank and Hart (1993), to the more recent papers by Dette (1999), Aerts, Claeskens and Hart (1999), Spokoiny (2001), Baraud, Huet and Laurent (2003). The nice monograph by Hart (1997) reviews this statistical literature, but almost exclusively deals with the one predictor case. Among the authors who explicitly studied the many regressors case, Härdle and Mammen (1993) used an L^2 distance between the parametric regression and the nonparametric one; Zheng (1996), Aerts, Claeskens and Hart (1999), and Guerre and Lavergne (2005) used a score approach; Fan, Zhang and Zhang (2001) adopted a likelihood-ratio approach. The ability of these omnibus tests to detect deviations from the parametric model quickly vanes when there is more than a couple of regressors. Indeed, since the nonparametric estimators suffer from the "curse of dimensionality" as shown by Stone (1980), so too do the related tests, see e.g. Guerre and Lavergne (2002). Hence, their usefulness is questionable for many applications, in particular in econometrics where the number of covariates can be large. To circumvent this issue, one can aim at testing the parametric regression against some non-saturated semiparametric alternatives. Fan, Zhang and Zhang (2001) studied varying coefficients linear models. Aerts, Claeskens and Hart (2000) and Guerre and Lavergne (2005) proposed tests tailored for additive alternatives. Hart (1997, Section 9.3) considered alternatives of the form $m(t(X))$, where $m(\cdot)$ is nonparametric and $t(X)$ is the vector of the first principal components of the covariance matrix X; he noted that there is however no guarantee that lack-of-fit will manifest itself along principal components. Fan and Huang (2001) similarly relied on scores from principal components analysis. The alternative dimension-reduction test of Zhu (2003) assumes *independence* of the parametric residuals with the regressors. All these proposals thus rely on some auxiliary assumptions that allow to restrict the alternative model, but they yield tests that are not omnibus.

Our goal is to device a powerful regression check that researchers could confidently apply in the presence of many regressors without imposing restrictions on the form of the alternative. In this aim, we develop a novel approach that improves on known regression checks based on smoothing methods. The approach is related to a previous proposal by Zhu and Li (1998) that we discuss further on. It can also be viewed as a further elaboration of the integrated conditional moment (ICM) test proposed by Bierens (1982), and for this reason we label our test the smooth ICM test. Moreoever, our approach allows to incorporate a priori qualitative information the procedure to improve its power. Our theoretical results show that the smooth ICM test is consistent against any alternative, yet it is not affected by the dimension of the regressors, since it behaves as if there was only one regressor. In practice, we found that the test is more powerful than known lack-of-fit tests in multidimensional settings. Specifically, it outperforms not only the kernel-based test of Zheng (1996), but also the ICM test by Bierens (1982) and the projection-based test recently proposed by Escanciano (2006). Moreover, it is little sensitive to the smoothing parameter choice.

Acknowledging that testing directly against saturated alternatives yield low power, our key principle is to replace the nonparametric alternative by a *class* of a semiparametric alternatives that is rich enough to allow detection of any nonparametric alternative, thus reducing the dimension of the problem while preserving consistency. Specifically, we look at the class of single-index regression models. Formally, let $(Y_1, X_1')', \ldots (Y_n, X_n')'$ be independent observations from a population $(Y, X')' \in \mathbb{R}^{1+q}$, where X is a continuous random vector. We want to check whether the regression function $\mathbb{E}(Y|X)$ belongs to a parametric family $\{\mu(\cdot, \theta) : \theta \in \Theta\}$, for instance of linear or logistic functions. Our null hypothesis then writes

$$H_0 \ : \mathbb{E}\left[Y - \mu(X, \theta_0)|X\right] = 0 \quad \text{for some } \theta_0 \ . \tag{1.1}$$

As we face the "curse of dimensionality" in estimating the above conditional expectation, the resulting estimate will be imprecise in small and moderate samples, and the related test will lack power. Our approach consists in estimating conditional expectations given a linear index $X'\beta$ for any β and thus to replace *one* conditional expectation given *all* the regressors by *all* conditional expectations given *one* single linear index only. The advantage is that each expectation can be estimated accurately for a reasonable sample size since it depends on a single linear index only. The apparent drawback is that we have to estimate many conditional expectations. However, this cumbersome task can be avoided by combining expectations into a single integral and estimating this integral at once. We show indeed below that H_0 is equivalent to

$$\int_{\mathbb{S}^q} \mathbb{E}\left[\mathbb{E}^2\left(Y - \mu(X, \theta_0)|X'\beta\right) f_\beta(X'\beta)\right]\, d\beta = 0 \quad \text{for some } \theta_0\ , \tag{1.2}$$

where $\mathbb{S}^q$ is the hypersphere $\{\beta \in \mathbb{R}^q : \|\beta\| = 1\}$ and $f_\beta(\cdot)$ is the density of the linear index $X'\beta$. Our approach thus reduces the dimension of the problem without any knowledge about the form of the alternatives. The resulting test is truly omnibus and the rate of convergence of the test statistic under H_0 equals the rate one would obtain in the one-dimensional case. Moreover, it behaves against local directional alternatives as if there was one regressor only. We also show that when the regressors are bounded, it is sufficient to consider the above integral on a subset of the hypersphere with nonempty interior. This readily allows to incorporate some qualitative information in the procedure. For instance, if it is known that the marginal effects of two regressors X_1 and X_2 always have the same sign, one can choose B as the domain where the corresponding β_1 and β_2 are positive.

The rest of the paper is organized as follows. Section 2 explains the principle on which our approach relies. In Section 3, we propose a test statistic based on the kernel method, detail its practical computation, and study its asymptotic behavior under the null hypothesis. We also justify the validity of a bootstrap method to obtain critical values for samples of small or moderate size. In Section 4, we study the test under a sequence of directional alternatives and report the results of an extensive simulation study that compare our approach to different tests previously proposed in the literature. We also provide some evidence that explains why the smooth ICM test is powerful. The technical proofs are gathered in the Appendix.

2 The principle

The following lemma is the crux of our approach. It provides a direct justification for considering *all* conditional expectations given *one* single linear index for testing H_0. Part (ii) shows that when X is bounded, it is even sufficient to consider *infinitely many* of these conditional expectations. Note that X is bounded can be assumed without loss of generality, since we can always find a one-to-one transformation that maps X in a bounded set and retains all conditioning information, see e.g. Bierens (1982).

Lemma 2.1 *Let $\mathbb{S}^q = \{\beta \in \mathbb{R}^q : \|\beta\| = 1\}$ be the hypersphere with radius one. Consider random vectors $Z \in \mathbb{R}$ with $\mathbb{E}(Z^2) < \infty$ and $X \in \mathbb{R}^q$ with bounded density $f(\cdot)$. Let $f_\beta(\cdot)$ be the density of $X'\beta$ and assume that for some C, $|f_\beta(\cdot)| \leq C$ for any $\beta \in \mathbb{S}^q$.*
(i) $\mathbb{E}(Z \mid X) = 0$ is equivalent to

$$\int_{\mathbb{S}^q} \mathbb{E}\left[\mathbb{E}^2\left(Z|X'\beta\right) f_\beta(X'\beta)\right] \, d\beta = 0 \, . \tag{2.3}$$

(ii) If X is bounded, then $\mathbb{E}(Z \mid X) = 0$ is equivalent to

$$\int_{B} \mathbb{E}\left[\mathbb{E}^2\left(Z|X'\beta\right) f_\beta(X'\beta)\right] \, d\beta = 0 \tag{2.4}$$

for any $B \subset \mathbb{S}^q$ with nonempty interior.

Lemma 2.1 can be deduced from Bierens (1982, Theorem 1), but since it is the key of our approach, we provide here a simple proof and we comment it thereafter. **Proof.** (i) The implication is straightforward. By elementary properties of the conditional expectation, for any $\beta \in \mathbb{S}^q$ and any $t \in \mathbb{R}$,

$$\psi_\beta(t) := \mathbb{E}\left[\exp\{itX'\beta\}\mathbb{E}(Z \mid X'\beta)\right] = \mathbb{E}\left[\exp\{itX'\beta\}\mathbb{E}(Z \mid X)\right] \, , \tag{2.5}$$

where $i = \sqrt{-1}$. For any $\beta \in \mathbb{S}^q$, $\mathbb{E}\left(Z|X'\beta\right) f_\beta(X'\beta) \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, and Parseval's formula yields, see e.g. Rudin (1987),

$$\int_{\mathbb{R}} |\psi_\beta(t)|^2 \, dt = 2\pi \; \mathbb{E}\left[\mathbb{E}^2\left(Z|X'\beta\right) f_\beta(X'\beta)\right]$$

$$\text{and} \qquad \int_{\mathbb{S}^q}\int_{\mathbb{R}} |\psi_\beta(t)|^2 \, dt \, d\beta = 2\pi \int_{\mathbb{S}^q} \mathbb{E}^2\left[\mathbb{E}\left(Z|X'\beta\right) f_\beta(X'\beta)\right] \, d\beta \, .$$

If this integral equals zero, this implies $\psi_\beta(t) = 0$ for all β and all t. By the unicity of the Fourier transform, $\mathbb{E}(Z \mid X) = 0$.

(ii) Clearly, $\mathbb{E}(Z \mid X) = 0$ implies (2.4). The equivalence for $B = \mathbb{S}^q$ follows from Part (i). Since

$$2\pi \int_{B} \mathbb{E}\left[\mathbb{E}^2\left(Z|X'\beta\right) f_\beta(X'\beta)\right] \, d\beta = \int_{B}\int_{\mathbb{R}} |\psi_\beta(t)|^2 \, dt \, d\beta \, ,$$

(2.4) implies $\psi_\beta(t) = 0$ for all $\beta \in B$ and t. Since X is bounded, this yields $\mathbb{E}(Z|X) = 0$ by Theorem 1 of Bierens (1982). ∎

The proof clearly shows how (2.3) naturally appears from Fourier analysis. It is also useful to see that, because of the symmetry of the Fourier transform, our lemma holds not only for the hypersphere $\mathbb{S}^q$, but for any half-hypersphere. By half-hypersphere, we mean any subset H of $\mathbb{S}^q$ such that (i) $H \cup H^- = \mathbb{S}^q$, where $H^- = \{\beta^- : \beta^- = -\beta, \beta \in H\}$ and (ii) $H \cap H^-$ has Lebesgue measure zero. Hence, the assumption of a bounded X is necessary for Part (ii) only if B is not included in a half-hypersphere.

As a consequence of Lemma 2.1, our null hypothesis (1.1) can be written as (1.2). Moreover, when X is bounded, we can incorporate some qualitative information by considering only a restricted subset B. Our approach is related to the ICM test of Bierens (1982) and Bierens and Ploberger (1997), which is based on the fact that for X bounded, $\mathbb{E}(Z|X) = 0$ iff

$$\int_{\mathbb{R}^q} \left|\mathbb{E}\left[Z\psi\left(X'u\right)\right]\right|^2 \, d\mu(u) = 0 \ ,$$

for some probability measure $\mu(\cdot)$ and a well-chosen function $\psi(\cdot)$, such as $\exp(\cdot)$. As is clear from Lemma 2.1, our approach is similar, but instead of choosing a particular $\psi\left(\cdot\right)$ at the outset, we select for each β the function that maximizes the above quantity. Formally, we look for the L^2-function that maximizes $\mathbb{E}\left[Z\psi\left(X'\beta\right) f_\beta(X'\beta)\right]$. However, we do have to normalize the function $\psi(\cdot)$ to obtain a finite solution. Under our assumptions, a convenient normalization is given by

$$\mathbb{E}\left[\psi^2\left(X'\beta\right) f_\beta(X'\beta)\right] = \mathbb{E}\left[\mathbb{E}^2(Z \mid X'\beta) f_\beta(X'\beta)\right] \ . \tag{2.6}$$

Hence $\psi(\cdot)$ minimizes $\mathbb{E}\left[\left(Z - \psi\left(X'\beta\right)\right)^2 f_\beta(X'\beta)\right]$ under this constraint, and the solution is clearly $\mathbb{E}(Z \mid X'\beta)$.

3 The smooth ICM test

3.1 The test statistic

Let $(Y_i, X_i')'$, $i = 1, \ldots n$, be a random sample from $(Y, X')' \in \mathbb{R}^{1+q}$. The vector X is assumed to be continuously distributed, since regressors with fixed discrete support have no theoretical influence on the asymptotic power of a regression check. The model to be checked writes

$$Y = \mu(X, \theta_0) + \varepsilon, \quad \mathbb{E}\left(\varepsilon|X\right) = 0 \ .$$

An estimated candidate $\widehat{\theta}_n$ for the parameter θ_0 can be obtained by least-squares. The parametric residuals are then $\widehat{U}_i = Y_i - \mu(X_i, \widehat{\theta}_n), i = 1, \ldots n$. We use the kernel method to estimate (2.4), as it yields a very tractable statistic. We could certainly accommodate for other nonparametric methods, such as splines, local polynomials, or orthogonal series, but we do not pursue this issue here. We first define

$$\frac{1}{n(n-1)}\sum_{j\neq i}\widehat{U}_i\widehat{U}_j\frac{1}{h}K_h\left((X_i-X_j)'\beta\right)\ , \tag{3.7}$$

as an estimator of $\mathbb{E}\left[\mathbb{E}^2\left(Y-\mu(X,\theta_0)|X'\beta\right)f_\beta(X'\beta)\right]$. Here $K_h\left(\cdot\right)=K\left(\cdot/h\right)$, where $K(\cdot)$ is an univariate symmetric density and h a bandwidth. This statistic is the one studied by Zheng (1996) and Li and Wang (1998) applied to the index $X'\beta$ and has an asymptotic centered normal distribution with rate $nh^{1/2}$ under H_0. As noted by Dette (1999), Zheng's statistic is comparable to Härdle and Mammen's one (1993) with weight function equal to the squared density, which is exactly what is needed here. The quantity in (2.4) is thus estimated by

$$I_n=I_n(B)=\frac{1}{n(n-1)}\sum_{j\neq i}\widehat{U}_i\widehat{U}_j\frac{1}{h}\int_B K_h\left((X_i-X_j)'\beta\right)d\beta\ .$$

Zhu and Li (1998) first proposed to use an unweighed integral of expectations conditional upon single linear indices, yielding a statistic close to, but different than, I_n for checking a linear regression model. However, they do not study the related test. Instead, their test is based on their integral statistic plus a term of the form $(1/n)\sum_{i=1}^n\widehat{U}_i\phi\left(\|X_i\|\right)$, where $\phi(\cdot)$ is the standard normal univariate density (or any other known function). Hence, they combine a test statistic based on nonparametric methods with a directional test statistic. The asymptotic behavior of their test statistic under H_0 is completely driven by the second one. By contrast, we directly base our test on the integral statistic I_n. Let v_n^2 be the variance of $nh^{1/2}I_n$ under H_0, which is positive and finite as shown below. With at hand a consistent estimator $\widehat{v}_n^2$, an asymptotic α-level test is given by

$$\text{Reject } H_0 \quad \text{if } nh^{1/2}I_n\geq z_{1-\alpha}\widehat{v}_n\ ,$$

where $z_{1-\alpha}$ is the $(1-\alpha)$-th quantile of the standard normal distribution. The conditional variance of $nh^{1/2}I_n$ writes

$$v_n^2=\frac{2}{n(n-1)}\sum_{j\neq i}\sigma^2(X_i)\sigma^2(X_j)h^{-1}\mathbb{E}_B^2\left[K_h\left((X_i-X_j)'\beta\right)\right]\ ,$$

where $\mathbb{E}_B\left[g(\beta)\right]:=\int_B g(\beta)\,d\beta$ for any function $g(\cdot)$ of β. In general, the conditional variance $\sigma^2\left(\cdot\right)$ is unknown, but with at hand a nonparametric estimator such that

$$\sup_{1\leq i\leq n}\left|\frac{\widehat{\sigma}^2(X_i)}{\sigma^2(X_i)}-1\right|=o_{\mathbb{P}}(1)\ , \tag{3.8}$$

v_n^2 can be consistently estimated by

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

Many nonparametric estimators could be used. For instance, one can consider

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

where b is a bandwidth parameter converging to zero as the sample size increases, which can be selected independently of h. Guerre and Lavergne (2005) provide some primitive conditions for (3.8). It is then straightforward to show that $\widehat{v}_n^2 / v_n^2 = 1 + o_{\mathbb{P}}(1)$ under H_0. Given our focus, we will proceed assuming this condition holds.

The use of a nonparametric estimator of the error's variance does not affect the test at a first order. A simpler alternative is to plug estimated parametric residuals in the expression of v_n^2 in place of the unknown variance components, which gives

$$\widehat{v}_n^2 = \frac{2}{n(n-1)} \sum_{j \neq i} \widehat{U}_i^2 \widehat{U}_j^2 h^{-1} \mathbb{E}_B^2 \left[K_h \left((X_i - X_j)'\beta \right) \right] \ .$$

This alternative estimator is consistent for v_n^2 under H_0, but overestimates it when the parametric model is incorrect, and thus likely yields some loss in power for the test. For this reason, we do not recommend its use in practice. Nevertheless, our asymptotic results allows for its use.

3.2 Practical considerations

A first practical issue relates to the fact that the same bandwidth is used for all directions $X'\beta$. Hence it is desirable to transform the regressors to make different linear combinations comparable. An easy way is to center and rescale the matrix of observations on X so that it has mean zero and variance identity. Alternatively, as suggested by Bierens (1982) for the ICM test, one can map each regressor onto $(0, 1)$.

Implementation of our test requires integration on the (half) hypersphere or a subset of it. To approximate the integral in practice (up to a constant), it is sufficient to draw a large number of points randomly distributed on the (half) hypersphere, to evaluate the function under the integral for each draw and to compute the average. A draw can be easily performed by sampling independent $z_i, i = 1, \ldots q$, distributed as $N(0, 1)$ and to define β as the vector $z/\|z\|$. By the radial symmetry of the normal distribution, this gives points uniformly distributed on the hypersphere. In some cases, it may be possible to derive the analytic form of the integral. From the previous arguments, we have that

$$\int_{\mathbb{S}^q} K(u'\beta)\, d\beta = \int_{\mathbb{R}^q} K(\frac{u'z}{\|z\|})\phi(z)dz$$

where $\phi(\cdot)$ is the q-variate standard normal density. By a suitable change of variables, this equals

$$\int_{\mathbb{R}^q} K(\|u\|\frac{z_1}{\|z\|})\phi(z)dz \ ,$$

and thus depends only depends on $\|u\|$. However, deriving the analytic formula of this function can be quite tedious, even with symbolic computation engines, while numerical approximation is quite fast and easy. Matlab codes to implement the test are available from the authors upon request.

3.3 Behavior under the null hypothesis

To avoid technicalities, the parametric regression is taken to be linear in variables. However, we do not restrict the data to exhibit normality or homoscedasticity. Our results extend to a general parametric regression, see for instance Lavergne and Patilea (2006) for necessary assumptions. We first state our general assumptions on the data-generating process, the kernel and smoothing parameter.

Assumption D *(a) The random vectors* $(\varepsilon_1, X_1')', \ldots, (\varepsilon_n, X_n')'$ *are independent copies of the random vector* $(\varepsilon, X')' \in \mathbb{R}^{1+q}$*, where* $\mathbb{E}(\varepsilon \mid X) = 0$ *and* $\mathbb{E}(\varepsilon^4) < \infty$*.*
(b) Let $\sigma^2(x) = \mathbb{E}(\varepsilon^2 \mid X = x)$*. There exist constants* $\underline{\sigma}^2$ *and* $\overline{\sigma}^2$ *such that for any* x $0 < \underline{\sigma}^2 \leq \sigma^2(x) \leq \overline{\sigma}^2 < \infty$*.*
(c) X *is continuous with bounded density* $f(\cdot)$*, and the density* $f_\beta(\cdot)$ *of* $X'\beta$ *is such that for some* C*,* $|f_\beta(\cdot)| \leq C$ *for any* $\beta \in B$*. If* B *is not included in a half-hypershpere,* X *is assumed to be bounded.*
(d) Let $Z = [Z_i, i = 1, \ldots n] = [(1, X_i'), i = 1, \ldots n]$ *be the design matrix. There exists a positive definite matrix* A *such that* $n^{-1}Z'Z \xrightarrow{p} A$*.* $\theta \in \Theta$*, a compact of* $\mathbb{R}^{1+q}$*.*

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

Assumptions D(c) comes from our Lemma 2.1 and rules out multicollinearity among the regressors. For a bounded X, a bounded density for X implies that $f_\beta(\cdot)$ is bounded uniformly in $\beta \in \mathbb{S}^q$. The assumptions on the kernel $K(\cdot)$ are satisfied by most kernels used in practice. The restrictions on the bandwidth are compatible with optimal choices for regression checks, see Guerre and Lavergne (2002). The following theorem states the asymptotic validity of the smooth ICM test.

Theorem 3.1 *Under Assumptions D and K and if $\widehat{v}_n^2/v_n^2 = 1 + o_{\mathbb{P}}(1)$ under H_0, the test based on I_n has asymptotic level α conditionally on the X_i.*

While the test can be implemented using asymptotic critical values for large samples, the asymptotic approximation is likely not accurate for small or moderate samples, as is the case for most regression checks. The wild bootstrap, initially proposed by Wu (1986), is thus often used to compute small sample critical values, see e.g. Härdle and Mammen (1993) and Stute and al. (1998). 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 $Y_i^* = \mu(X_i, \widehat{\theta}_n) + \widehat{\sigma}(X_i)\omega_i$, $i = 1, \ldots n$. A bootstrap test statistic is built from the bootstrap sample as the original test statistic was. When this scheme is repeated many times, the bootstrap critical value $z_{1-\alpha,n}^*$ at level α is the empirical $(1-\alpha)$-th quantile of the bootstrapped test statistic. This critical value is then compared to the initial test statistic. The following theorem can be shown following the lines of Theorem 3.1's proof.

Theorem 3.2 *Under the assumptions of Theorem 3.1 and Condition (2.7), the bootstrap critical value yields a test based on I_n with asymptotic level α conditionally on the X_i.*

4 Power analysis

4.1 Power under local alternatives

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

$$\mathbb{E}[(1, X')\delta(X)] = \mathbf{0} \quad \text{and} \quad 0 < \mathbb{E}[\delta^4(X)] < \infty \,. \tag{4.9}$$

The first condition ensures that $\delta(\cdot)$ is orthogonal to any linear combination of the regressors. We do not impose smoothness restrictions on the function $\delta(\cdot)$ as is frequent in this kind of analysis. We consider the sequence of local directional alternatives

$$H_{1n}: \; \mathbb{E}\left[Y|X\right] = (1, X')\,\theta_0 + r_n\delta(X), \quad n \geq 1 \,. \tag{4.10}$$

Such directional alternatives can be detected if $r_n^2 nh^{1/2} \to \infty$, where h applies to the univariate variable defined by a single linear index in X. By comparison, when one uses a regression check based on a standard "multidimensional" nonparametric estimator, $r_n^2 nh^{q/2} \to \infty$ is needed for consistency. Hence, from the theoretical point of view, the asymptotic power of our test against directional alternatives is not affected by the dimension of the regressors.

Theorem 4.1 *Under Assumptions D and K, if $\widehat{v}_n^2/v_n^2 = O_{\mathbb{P}}(1)$ and $r_n^2 nh^{1/2} \to \infty$, the test based on I_n is consistent conditionally on the X_i against the sequence of alternatives H_{1n} with $\delta(X)$ satisfying (4.9).*

4.2 Small sample power

Our simulation study had two main objectives. First, we wanted to determine the sensitivity of the smooth ICM test to the smoothing parameter h. Second, we wanted to compare its small sample power to the test of Zheng (1996) and Li and Wang (1998) on the one hand, and to the tests of Bierens (1982) and Escanciano (2006) on the other hand.

Let us first present briefly the different tests we considered. Zheng's test is based on the statistic (3.7) where $h^{-1}K_h\left((X_i - X_j)'\beta\right)$ is replaced by $h^{-q}K_h\left(X_i - X_j\right)$ with a multivariate kernel. When properly normalized, this statistic has an asymptotic standard normal distribution. Li and Wang (1998) investigated application of the wild bootstrap to this test. We used the smooth conditional moment bootstrap, which is also valid as can be shown from standard arguments. The ICM test is based on the statistic

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

where $\phi(\beta)$ is the standard normal density on $\mathbb{R}^q$, see Bierens (1982, p. 111). Escanciano (2006) also used this form of the ICM test for comparison. The asymptotic theory developed by Bierens and Ploberger (1997) applies only if the measure used in integration has compact support, so that the normal distribution should be truncated at some possibly very large values. For all practical matters however, this does not make any substantial difference. The ICM statistic thus resembles ours, with a kernel depending only on the norm $\|X_i - X_j\|$ but a fixed bandwidth. Dominguez (2004) shows that the wild bootstrap is valid and preserves admissibility of the test, consequently we used this method to obtain critical values. Finally, Escanciano's test is based on the statistic

$$\frac{1}{n^2}\sum_{i,j} U_i(\theta)U_j(\theta)\left(\frac{1}{n}\sum_{k}\int_{\mathbb{S}^q}\mathbb{I}(X_i'\beta \leq X_k'\beta)\mathbb{I}(X_j'\beta \leq X_k'\beta)\, d\beta\right) ,$$

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

We consider X with dimension four and the null hypothesis

$$H_0 \ : \ \mathbb{E}(Y|X) = (1, X)'\theta_0 \quad \text{for some } \theta_0 \ .$$

We generated samples of 100 observations from independent uniformly distributed variables for each component of X. The support was chosen as $U\left[-\sqrt{3}, \sqrt{3}\right]$ to get unit variance. We sampled errors from a standard normal distribution and we constructed the response variable as

$$Y_i = (1, X_i)'\theta_0 + d\ \delta(X_i'\beta_0) + \varepsilon_i \quad i = 1, \ldots 100 \ ,$$

with $\theta_0 = (0.5, 0.5, 0.5, -1.5)$, and different d and $\delta(\cdot)$. For each experiment, the number of replications is 5000 under the null hypothesis and 1000 under each alternative. The number of bootstrap samples is 199 for each replication and the level is 5%. We considered the following nonparametric tests: (i) Zheng's test when the index $X'\beta_0$ is considered as the only regressor, labeled as "Zheng's test Dim 1" in our figures; (ii) Zheng's test when all four regressors are taken into account, labeled as "Zheng's test Dim 4;" (iii) the smooth ICM test where numerical integration is performed on a grid of 5000 points on an half-hypersphere; (iv) the smooth ICM test where integration is performed on the subset B of the hypersphere for which the first three components of β are positive. This last situation corresponds to using the qualitative information that the influences of the first three components of X on Y are of the same sign. To compute the test statistics, we used a normal kernel and we selected the bandwidth as $h = b\,n^{-2/(8+q)}$, with $q = 4$ in Case (ii) and 1 in the other cases, and b varies in $\{0.5, 1, \ldots, 3\}$. The errors' conditional variance was estimated by a kernel estimator with normal kernel and bandwidth $2n^{-1/6}$. We applied this estimator to the parametric residuals, since it yielded a better behavior for Zheng's multidimensional test.

In our first set of simulations, $\delta(X) = 0.1 \times (X'\beta_0/\sqrt{3})^2$, where $\beta_0 = (1, 2, 3, -2)/\sqrt{18}$. Figure 1 illustrates that residual plots may not be informative on whether the model is misspecified when many regressors are present. Partial residuals are defined as $Y -$

$\sum_{j\neq k}\widehat{\theta}_j Z_j$, see Cook (1993) and the references therein, and the data were generated with $d=6$. Figure 2 compares the power of Zheng's tests and our test on the sphere when $d=6$ and the bandwidth constant b varies. The key insight is that the performance of Zheng's tests is quite variable depending on the bandwidth, while our test appears to be little sensitive to this parameter. Figure 3 compares the power curves of the different tests for varying d and $b=1$. Empirical levels are well approximated by the bootstrap for all smooth tests. Bierens' and Escanciano's tests are under rejecting, with respective levels 3.8 and 2.52. In terms of power, there is a large loss in power for Zheng's test when going from dimension one to four. In practice however, the test based on the unknown single linear index is infeasible. The second striking fact is that the smooth ICM test largely outperforms Zheng's test in dimension 4, as well as the other tests. The performance of the smooth ICM test is close to the one of the infeasible test. When incorporating some qualitative information, they practically cannot be distinguished.

In our second set of simulations, we considered the hyperbolic sine alternative $\delta(X)=\sinh(X'\beta_0/\sqrt{3})$. This alternative is particularly difficult to detect, because it resembles very much a linear function. Other features of the experiments are unchanged. Figure 4 illustrates the performances of the different smooth tests for $d=6$ when the bandwidth varies. As can be seen, Zheng's tests are very sensitive to the bandwidth, while the smooth ICM test's power is almost stable for a bandwidth constant varying from 0.5 to 2. When the bandwidth further increases, the power decreases to the power of the ICM test. Figure 5 is the analog of Figure 3 for hyperbolic sine alternatives. While our test is not as powerful as Zheng's infeasible test, it outperforms all the considered competing tests. The smooth ICM test on B has now power close to the one of the infeasible test. Finally, for this alternative, Zheng's test does a better job than the Bierens' test, which itself outperforms Escanciano's test.

In a third step, we considered the sine alternative $\delta(X)=0.1\times\sin(\pi X'\beta_0/\sqrt{3})$. This alternative is favorable to Bierens' test, which is based on the correlation between residuals and trigonometric functions. Figure 6 compares the power curves of the different tests. As expected, Bierens' test performs better than Escanciano's test, but surprisingly Zheng's test does better, and the smooth ICM test outperforms them all. Restricting integration on B further improves its power.

To show that our conclusions are not tied to single-index alternatives, we considered the two-indexes alternative $\delta(X)=\sinh(X'\beta_1/\sqrt{3})+\sinh(X'\beta_2/\sqrt{3})$, where $\beta_1=(0,2,1,-1)/\sqrt{6}$ and $\beta_2=(1,0,2,-1)/\sqrt{6}$. As a benchmark, we took Zheng's test based on the two linear indices entering the regression function, labeled as "Zheng's test Dim 2." Figure 7 shows the previous qualitative results still hold.

4.3 What makes our test powerful?

To understand why our test outperforms the ICM test, recall that this procedure, as well as the one by Escanciano, estimates a quantity of the form

$$\int_{\mathbb{R}^q} \left|\mathbb{E}\left[Z\psi\left(X'u\right)\right]\right|^2 \, d\mu(u) = 0 \, .$$

From a theoretical viewpoint, they are consistent against sequences of local alternatives of the form (4.10) whenever $r_n^2 n \to \infty$, because the standard deviation of their statistic goes to zero at rate n. Now, instead of working with a particular known $\psi(\cdot)$ at the outset, we estimate the function of $X'\beta$ that maximizes weighted correlation with Z. Since this function needs to be nonparametrically estimated, our test statistic has a larger variance. However, it is also expected to have a higher mean under any alternative. Since the power of the test depends of both mean and variance, our test can have higher power than its competitors.

Ideally, one would like to derive the analytical power function of each test to compare them. This is however quite intricate and would involve some asymptotic approximations that may not be accurate in small samples. Instead, we report in Figure 8 the densities of the statistics under the null hypothesis and the quadratic alternative with $d = 10$. One can clearly see from the upper part of Figure 8 that, as expected, the smooth ICM statistic has a lower mean than Zheng's statistic but a much more concentrated distribution. It also shows that, as already noted in previous simulation studies, these statistics are biased in small samples, which calls for the use of resampling methods. The lower part of Figure 8 shows that Escanciano's statistic is less variable and has a lower mean than the ICM statistic. When comparing the smooth ICM and ICM statistics, we note that their mean respectively increase by 0.67 and 0.50 when d goes from zero to 10, see Table 1. This confirms that the smooth ICM test uses a one-dimensional function that maximizes the difference in mean between the null and alternative hypotheses and then yields better power performances.

We also consider in Figure 9 the case of a sine alternative with $d = 10$ because it yields another interesting insight. Remember that the form of the ICM test we used is directed against such alternatives. Indeed, we found that the mean of the ICM statistic increases by 0.67, while the mean of the smooth ICM statistic increases only by 0.49. However, the latter statistic is much more concentrated, with a standard deviation of 0.22, to be compared to 0.35 for the ICM statistic. Hence in that case, the higher power of the smooth ICM test seems to come from its lower dispersion. Indeed, even the variance of the ICM statistic decreases at a faster rate with the sample size, this does not imply that in practice its variance will be smaller.

Our analysis thus sheds light on two main facts. First, when many regressors are present, the smooth ICM test should be powerful in most cases because it is based on nonparametric estimation and thus maximizes the difference in behavior under the null and alternative hypothesis. Second, even for alternatives against which the ICM test is directed, the smooth ICM test can do a better job in small samples because it has less dispersion. It seems that in small samples with many regressors, our test is able to balance central tendency and dispersion so as to yield a powerful test. Clearly, such findings are not accounted for by standard asymptotic results.

Appendix

For any function $g(\cdot) \in L^1(\mathbb{R}^q) \cap L^2(\mathbb{R}^q)$, its Fourier and inverse Fourier transforms are respectively defined as $\widehat{g}(t) = (2\pi)^{-q/2} \int_{\mathbb{R}^q} \exp(it'x) g(x)\, dx$ and $(2\pi)^{-q/2} \int_{\mathbb{R}}^q \exp(-it'x)\widehat{g}(t)\, dt$. In what follows, C denotes a positive constant that may vary from line to line. We first show two lemmas that are useful for proving our main results.

Lemma 4.2 *Let $\delta(\cdot)$ be any non-zero function of X on the support of X and $h \to 0$. Under Assumptions D(c) and K(a), (i) If $\mathbb{E}\delta^2(X) < \infty$, $\mathbb{E}\left\{\delta(X_1)\delta(X_2)h^{-1}\mathbb{E}_B\left[K_h\left((X_1 - X_2)'\beta\right)\right]\right\}$ has a strictly positive finite limit. (ii) If $\mathbb{E}\delta^4(X) < \infty$ and $nh \to \infty$, then $U_n - \mathbb{E}(U_n) = o_{\mathbb{P}}(1)$ where*

$$U_n = \frac{1}{n(n-1)} \sum_{j \neq i} \delta(X_i)\delta(X_j)h^{-1}\mathbb{E}_B\left[K_h\left((X_i - X_j)'\beta\right)\right].$$

Proof. (i) Denoting by $\widehat{K}(\cdot)$ the Fourier transform of $K(\cdot)$,

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

As $\left|\widehat{K}(\cdot)\right| \leq \widehat{K}(0) = (2\pi)^{-1/2}$, Lebesgue's dominated convergence yields the limit

$$(2\pi)^{q-1/2} \int_{\mathbb{R}} \int_B \left|\widehat{\delta f}(t\beta)\right|^2 \, d\beta \, dt \, ,$$

provided it is finite. But the above quantity is bounded by

$$\int_{\mathbb{R}} \int_B \left|\widehat{\delta f}(t\beta)\right|^2 \, d\beta \, dt < \infty \, .$$

Finally, the limit is shown to be strictly positive as in the proof of Lemma 2.1.

$$\text{(ii)} \qquad \begin{aligned} \operatorname{Var}(U_n) &\leq \tfrac{C}{n} \operatorname{Var}\left[\delta(X_1)\delta(X_2)h^{-1}\mathbb{E}_B K_h\left((X_1 - X_2)'\beta\right)\right] \\ &\leq \tfrac{C}{nh} \mathbb{E}\left[\delta^2(X_1)\delta^2(X_2)h^{-1}\mathbb{E}_B K_h\left((X_1 - X_2)'\beta\right)\right] \, , \end{aligned}$$

and the above expectation converges to a finite limit from Part (i). ■
Let W be the matrix with generic element $\mathbb{E}_B\left[K_h\left((X_i - X_j)'\beta\right)\right] \mathbb{I}\left(i \neq j\right) / \left(hn(n-1)\right)$ and define its spectral radius as $\operatorname{Sp}(W) = \sup_{u \neq 0} \|Wu\| / \|u\|$.

Lemma 4.3 *Under Assumptions D(c) and K, (i)* $\operatorname{Sp}(W) = O_{\mathbb{P}}(n^{-1})$ *and (ii)* $n^2 h \|W\|^2$ *has a strictly positive limit, where* $\|W\|$ *denotes the Euclidean matrix norm.*

Proof. (i) For any $u \in \mathbb{R}^n$,

$$\begin{aligned} \|Wu\|^2 &= \sum_{i=1}^n \left(\sum_{j=1, j \neq i}^n w_{ij} \, u_j \right)^2 \leq \sum_{i=1}^n \left(\sum_{j=1, j \neq i}^n w_{ij} \right) \sum_{j=1, j \neq i}^n w_{ij} \, u_j^2 \\ &\leq \|u\|^2 \left[\max_{1 \leq i \leq n} \left(\sum_{j=1, j \neq i}^n w_{ij} \right) \right]^2 \, . \end{aligned}$$

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

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

since $nh^2 \to \infty$ by Assumption K(b). Now

$$\mathbb{E}\left[h^{-1}\mathbb{E}_B K_h\left((X_i - X_j)'\beta\right)|X_i\right] = \int_B \int_{\mathbb{R}} K(u) f_\beta(X_i'\beta - hu)\,du\,d\beta$$

is bounded uniformly in i by Assumptions D(c) and K(a).

(ii) Write $n^2 h \|W\|^2 = \frac{1}{(n-1)^2}\sum_{i\neq j} h^{-1}\mathbb{E}_B^2 K_h\left((X_i - X_j)'\beta\right)$. Hoeffding's (1963) inequality for U-statistics yields for any $\alpha \in (0,1)$

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

by Assumption K(b). We have

$$\begin{aligned}
&\mathbb{E}\left[h^{-1}\mathbb{E}_B^2 K_h\left((X_i - X_j)'\beta\right)\right] \\
&= \mathbb{E}\left[h^{-1}\int_B K_h\left((X_i - X_j)'\beta\right)\,d\beta \int_B K_h\left((X_i - X_j)'\alpha\right)\,d\alpha\right] \\
&= (2\pi)^{q-1} h \int_{\mathbb{R}}\int_{\mathbb{R}}\int_B\int_B \widehat{K}(ht)\widehat{K}(hu)\left|\widehat{f}\left(t\beta + u\alpha\right)\right|^2 dt\,du\,d\beta\,d\alpha\ .
\end{aligned}$$

By Assumption K(a),

$$\begin{aligned}
&h\int_{\mathbb{R}}\int_{\mathbb{R}}\int_B\int_B |\widehat{K}(hu)|\left|\widehat{f}\left(t\beta + u\alpha\right)\right|^2 dt\,du\,d\beta\,d\alpha \\
&= \int_{\mathbb{R}\times B}\left|\widehat{f}\left(t\beta\right)\right|^2 dt\,d\beta \int_B d\alpha \int_{\mathbb{R}} |\widehat{K}(u)|\,du < \infty\ .
\end{aligned} \tag{4.11}$$

Now

$$\begin{aligned}
&\left|h\int_{\mathbb{R}}\int_{\mathbb{R}}\int_B\int_B \left(\widehat{K}(ht) - \widehat{K}(0)\right)\widehat{K}(hu)\left|\widehat{f}\left(t\beta + u\alpha\right)\right|^2 dt\,du\,d\beta\,d\alpha\right| \\
&\leq C \sup_{|ht|\leq M}\left|\widehat{K}(ht) - \widehat{K}(0)\right| \\
&\quad + 2(2\pi)^{-1/2} h \int_{|t|\geq M/h}\int_{\mathbb{R}}\int_B\int_B |\widehat{K}(hu)|\left|\widehat{f}\left(t\beta + u\alpha\right)\right|^2 dt\,du\,d\beta\,d\alpha\ .
\end{aligned}$$

From the uniform continuity of $\widehat{K}(\cdot)$ and Equation (4.11), the right-hand side can be rendered arbitrarily small by choosing M small enough then letting h tend to zero. Therefore $\mathbb{E}\left[h^{-1}\mathbb{E}_B^2 K_h\left((X_i - X_j)'\beta\right)\right]$ tends to

$$(2\pi)^{q-1}\widehat{K}(0)\int_{\mathbb{R}\times B}\left|\widehat{f}\left(t\beta\right)\right|^2 dt\,d\beta\int_B d\alpha\int_{\mathbb{R}}\widehat{K}(u)\,du = (2\pi)^{q-1}K(0)\int_B d\alpha\int_{\mathbb{R}\times B}\left|\widehat{f}\left(t\beta\right)\right|^2 dt\,d\beta\ ,$$

using Assumption K(a). ■

Proof of Theorem 3.1. Let $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$. We have

$$I_n = I_{0n} - 2I_{1n} + I_{2n} = \varepsilon' W \varepsilon - 2(\widehat{\theta}_n - \theta_0)' Z' W \varepsilon + (\widehat{\theta}_n - \theta_0)' Z' W Z (\widehat{\theta}_n - \theta_0)\,,$$

Under Assumption D, $\widehat{\theta}_n - \theta_0 = O_{\mathbb{P}}(n^{-1/2})$. Hence $I_{2n} \leq \mathrm{Sp}(W) \| Z(\widehat{\theta}_n - \theta_0) \|^2 = O_{\mathbb{P}}(n^{-1})$ by Lemma 4.3(i). Let $\mathbb{E}_n$ denote the conditional expectation given the X_i, Z_k be any column of Z, $k = 1, \ldots d+1$, and $\overline{Z}_k = Z_k' W$. Then Marcinkiewicz-Zygmund's and Minkowski's inequalities imply that there is some C independent of n such that

$$\begin{aligned}
\mathbb{E}_n \left| Z_k' W \varepsilon \right| \;&\leq\; C \left\{ \mathbb{E}_n^2 \left| \sum_{i=1}^n \overline{Z}_{ki}^2 \varepsilon_i^2 \right|^{1/2} \right\}^{1/2} \leq C \left\{ \sum_{i=1}^n \overline{Z}_{ki}^2 \mathbb{E}_n^2 |\varepsilon_i| \right\}^{1/2} \\
&\leq\; C \| Z_k' W \| \leq C \mathrm{Sp}(W) \| Z_k \| = O_{\mathbb{P}}(n^{-1/2})\,.
\end{aligned}$$

Hence $I_{1n} = O_{\mathbb{P}}(n^{-1})$. Now from Lemma 2(i) by Guerre and Lavergne (2005), $nh^{1/2} I_{0n}/v_n$ converges to a standard normal conditionally on the X_i if $\|W\|^{-1} \mathrm{Sp}(W) = o_{\mathbb{P}}(1)$. Lemma 4.3 allows to conclude. ■

Proof of Theorem 4.1. Under H_{1n}, $U_i(\widehat{\theta}_n) = \varepsilon_i - Z_i(\widehat{\theta}_n - \theta_0) + r_n \delta(X_i)$. Letting $\delta = [\delta(X_1), \ldots \delta(X_n)]'$, I_n can be decomposed as $I_{0n} - 2I_{1n} + I_{2n} - 2I_{3n} - 2I_{4n} + I_{5n}$, where $I_{3n} = r_n \delta' W Z (\widehat{\theta}_n - \theta_0)$, $I_{4n} = r_n \delta' W \varepsilon$, and $I_{5n} = r_n^2 \delta' W \delta$. By Assumption D(c) and Lemma 4.3(ii), $v_n^2 \leq \overline{\sigma}^4 n^2 h \|W\|^2 = O_{\mathbb{P}}(1)$. Hence $nh^{1/2} I_{0n} = O_{\mathbb{P}}(1)$. Because under our assumptions, $\widehat{\theta}_n - \theta_0 = O_{\mathbb{P}}(n^{-1/2})$, I_{1n} and I_{2n} are both $O_{\mathbb{P}}(n^{-1})$ as in Theorem 3.1's proof. Since $|u' W v| \leq \|u\| \|v\| \mathrm{Sp}(W)$, $r_n^{-1} I_{3n} \leq \|\delta\| \|Z(\widehat{\theta}_n - \theta_0)\| \mathrm{Sp}(W) = O_{\mathbb{P}}(n^{-1/2})$. Also $I_{4n} = O_{\mathbb{P}}(r_n n^{-1/2})$ by the same arguments used for dealing with I_{1n}. Lemma 4.2(ii) yields $I_{5n} = r_n^2 C + o_{\mathbb{P}}(r_n^2)$ with $C > 0$. Collecting results, it follows that $nh^{1/2} I_n = nh^{1/2} r_n^2 C + o_{\mathbb{P}}(r_n^2 n h^{1/2})$. Deduce from $\widehat{v}_n^2 / v_n^2 = O_{\mathbb{P}}(1)$ and $r_n^2 n h^{1/2} \to \infty$ that $nh^{1/2} I_n / \widehat{v}_n$ diverges in probability. ■

REFERENCES

Aerts, M., G. Claeskens and J.D. Hart (1999). Testing the fit of a parametric function. *J. Amer. Statist. Assoc.* **94**, 869–879.

Aerts, M., G. Claeskens and J.D. Hart (2000). Testing lack of fit in multiple regression. *Biometrika* **87 (2)**, 405–4242.

Azzalini, A., A.W. Bowman and W. Härdle (1989). On the use of nonparametric regression for model checking. *Biometrika* **76 (1)**, 1–11.

BARAUD, Y., S. HUET AND B. LAURENT (2003). Adaptive tests of linear hypotheses by model selection. *Ann. Statist.* **31 (1)**, 225–251.

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 (6)**, 1443-1458.

BIERENS, H.J. AND PLOBERGER W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica* **65**, 1129-1151.

COOK, R.D. (1993). Exploring partial residual plots. *Technometrics* **35 (4)**, 351–362.

COX, D., E. KOH, G. WAHBA AND B.S. YANDELL (1988). Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models. *Ann. Statist.* **16 (1)**, 113–119.

DETTE, H. (1999). A consistent test for the functional form of a regression based on a difference of variance estimators. *Ann. Statist.* **27 (3)**, 1012–1040.

DOMINGUEZ, M.A. (2004). On the power of boootstrapped specification tests. *Econometric Rev.* **23 (3)**, 215–228.

EUBANK, R.L. AND C.H. SPIEGELMAN (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques. *J. Amer. Statist. Assoc.* **85 (410)**, 387–392.

EUBANK, R.L. AND J.D. HART (1993). Commonality of cusum, von Neumann and smoothing-based goodness-of-fit tests. *Biometrika* **80 (1)**, 89–98.

ESCANCIANO, J.C. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory* **22 (6)** 1030–1051.

FAN, J. AND L.S. HUANG (2001). Goodness-of-fit tests for parametric regression models. *J. Amer. Statist. Assoc.* **96 (454)**, 640–652.

FAN, J., C. ZHANG AND J. ZHANG (2001). Generalized lihelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* **29 (1)** ,153–193.

GOZALO, P.L. (1997). Nonparametric bootstrap analysis with applications to demographic effects in demand functions. *J. Econometrics* **81 (2)**, 357-393.

GUERRE, E., AND LAVERGNE, P. (2002). Optimal minimax rates for nonparametric specification testing in regression models. *Econometric Theory* **18 (5)**, 1139-1171.

GUERRE, E., AND P. LAVERGNE (2005). Data-driven rate-optimal specification testing in regression models. *Ann. Statist.* **33 (2)**, 840-870.

HÄRDLE, W., AND E. MAMMEN (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21 (4)**, 1296-1947.

Hart, J.D., and T.E. Wehrly (1992). Kernel regression when the boundary region is large, with an application to testing the adequacy of polynomial models. *J. Amer. Statist. Assoc.* **87 (420)**, 1018–1024.

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

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58 (301)**, 13–30.

Lavergne, Q., and V. Patilea (2006). Breaking the curse of dimensionality in nonparametric testing. Working paper.

Li, Q., and S. Wang (1998). A simple consistent bootstrap test for a parametric regression function. *J. Econometrics* **87 (1)**, 145–165.

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

Spokoiny, V. (2001). Data-driven testing the fit of linear models. *Math. Methods Statist.* **10 (4)**, 465–497.

Stone, C.J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8 (6)**, 1348–1360.

Stute, W., W. Gonzalez-Manteiga and M. Presedo (1998). Bootstrap approximations in model checks for regression. *J. Amer. Statist. Assoc.* **93 (441)**, 141–149.

Wu, C.F.J. (1986). Jacknife, bootstrap and other resampling methods in regression analysis (with discussion). *Ann. Statist.* **14 (4)**, 1261–1350.

Zheng, J.X. (1996). A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics* **75 (2)**, 263–289.

Zhu, L.X., and R. Li (1998). Dimension-reduction type test for linearity of a stochastic model. *Acta Math. Appli. Sinica* **14 (2)**, 165–175.

Zhu, L.X. (2003). Model checking of dimension-reduction type for regression. *Statist. Sinica* **13 (2)**, 283–296.

Figure 1: Quadratic alternative: Residuals plots

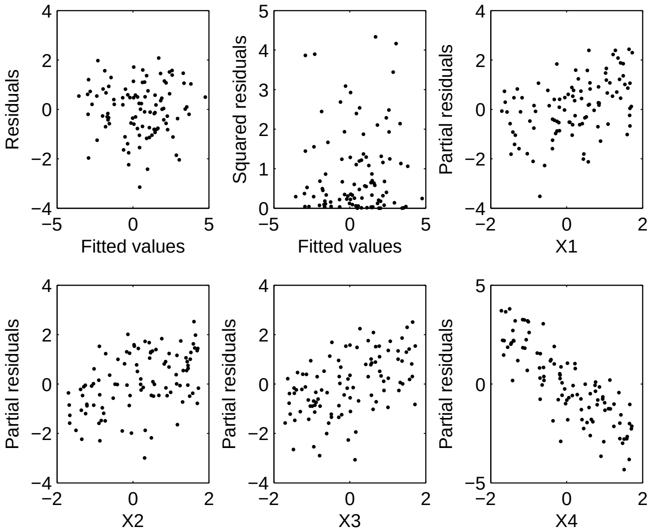

Figure 2: Quadratic alternative — varying bandwidth

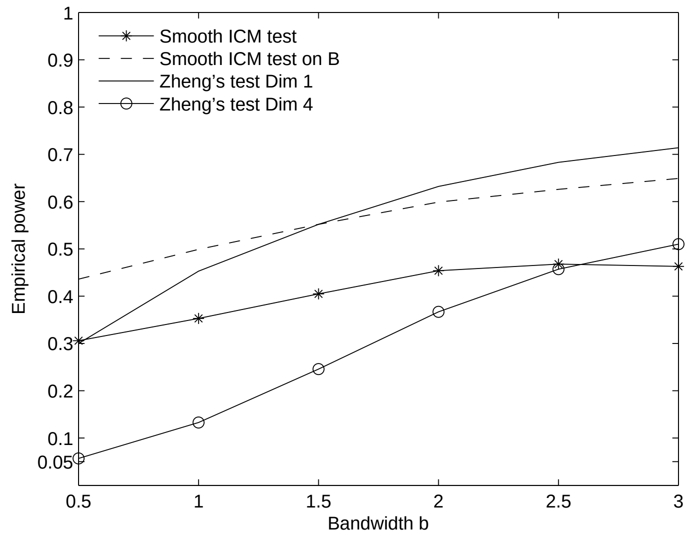

Figure 3: Quadratic alternative

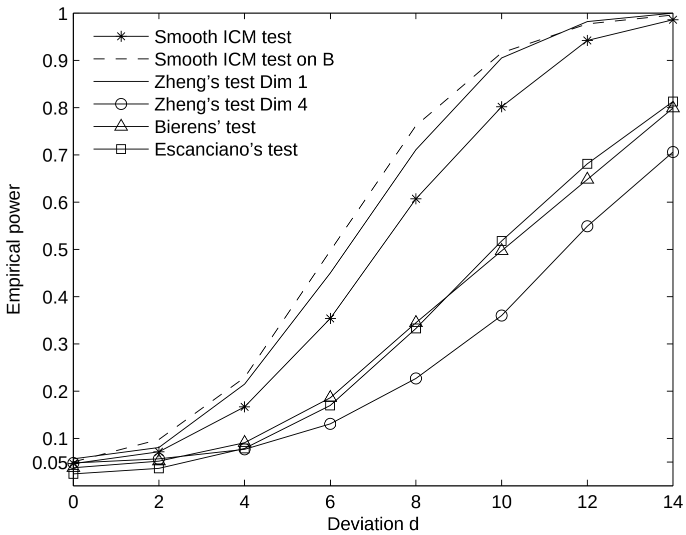

Figure 4: Sinh alternative — varying bandwidth

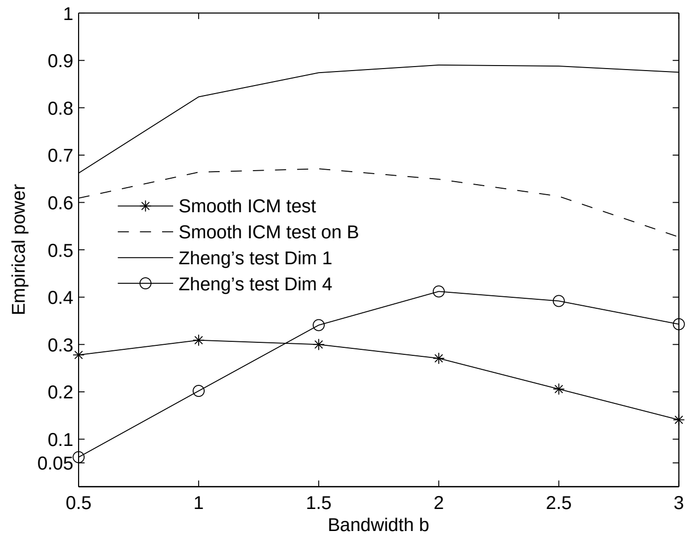

Figure 5: Sinh alternative

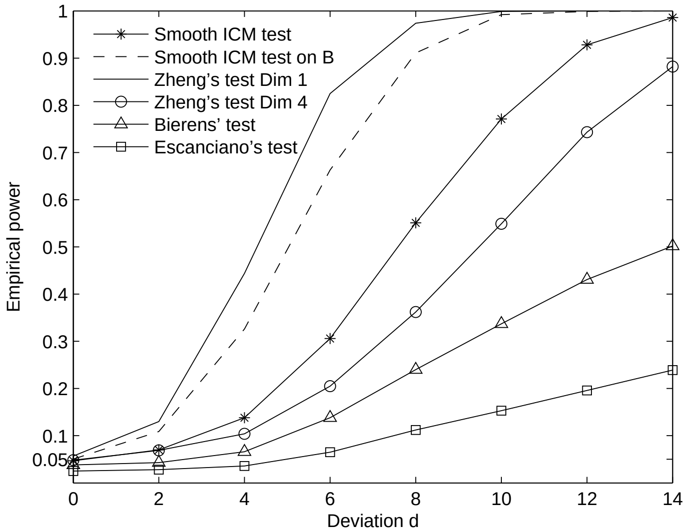

Figure 6: Sine alternative

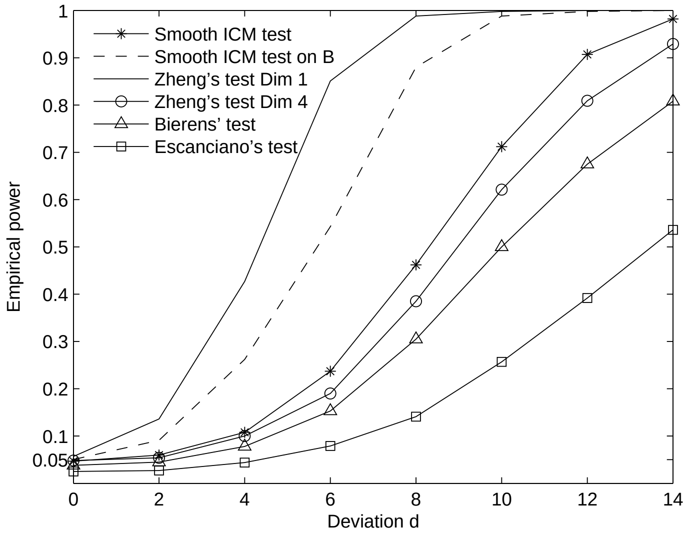

Figure 7: Two-indexes alternative

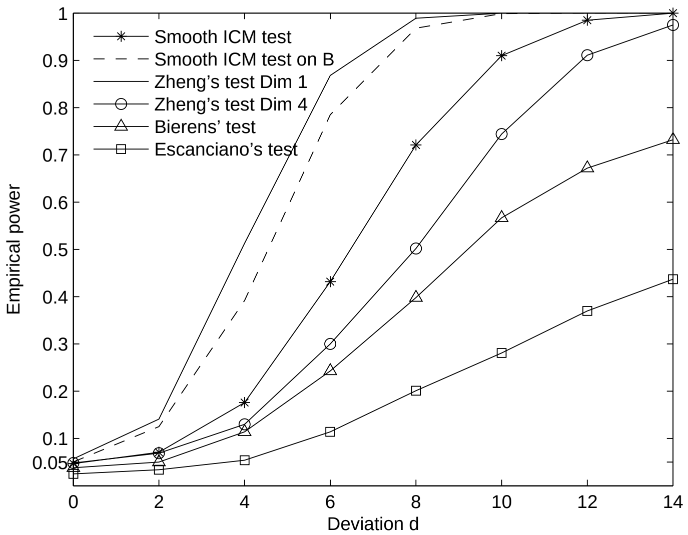

Figure 8: Densities for quadratic alternative

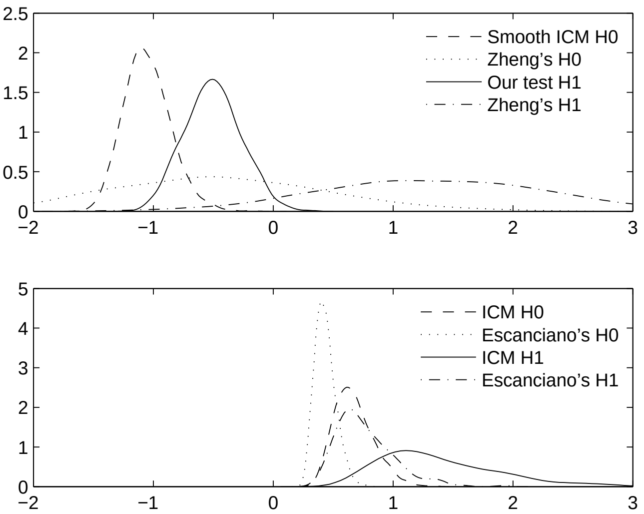

Table 1:
Descriptive statistics

	Under H_0			Quadratic alternative			Sine alternative		
	Integral	ICM	Esc.	Integral	ICM	Esc.	Integral	ICM	Esc.
Mean	-1.0507	0.6741	0.4340	-0.3768	1.1785	0.7194	-0.5615	1.3403	0.7390
S. d.	0.1967	0.1649	0.0875	0.3484	0.3642	0.1986	0.2158	0.3500	0.1636
Min	-1.6317	0.2993	0.2021	-1.4041	0.3737	0.3063	-1.1207	0.5250	0.3695
Median	-1.0648	0.6542	0.4248	-0.3747	1.1409	0.6968	-0.5636	1.3166	0.7276
Max	0.0014	1.4150	0.8289	0.9261	3.1427	1.6997	0.4706	2.7068	1.4374

Figure 10: Densities for sine alternative